

AcCORD: um modelo assíncrono colaborativo para a reconciliação de dados

Dayse Silveira de Almeida

Orientadora: Prof^a. Dr^a. Cristina Dutra de Aguiar Ciferri

Instituto de Ciências Matemáticas e de Computação (ICMC)

Universidade de São Paulo (USP)



Roteiro

- Motivação
- Modelo AcCORD
- Repositório de operações
- Políticas de reconciliação
- Método de propagação de decisões de integração
- Resultados experimentais
- Conclusões

Roteiro

- **Motivação**
- Modelo AcCORD
- Repositório de operações
- Políticas de reconciliação
- Método de propagação de decisões de integração
- Resultados experimentais
- Conclusões

Motivação

- Reconciliação
 - Processo de prover uma visão consistente de dados provenientes de várias fontes de dados
- Problema
 - **Reconciliar** dados quando usuários **colaborativos** trabalham **assincronamente** sobre cópias distintas desses dados
- Usuários
 - Podem colaborar para compartilhar das suas decisões sobre a solução de inconsistências → **única visão consistente**
 - Podem discordar sobre quais são os valores corretos → **várias visões distintas**

Desafios

- Como compartilhar decisões de integração?
- Como possibilitar a geração de uma única visão consistente ou várias visões distintas?
- Como garantir decisões consistentes sobre o mesmo dado presente em diferentes fontes?

Modelo AcCORD

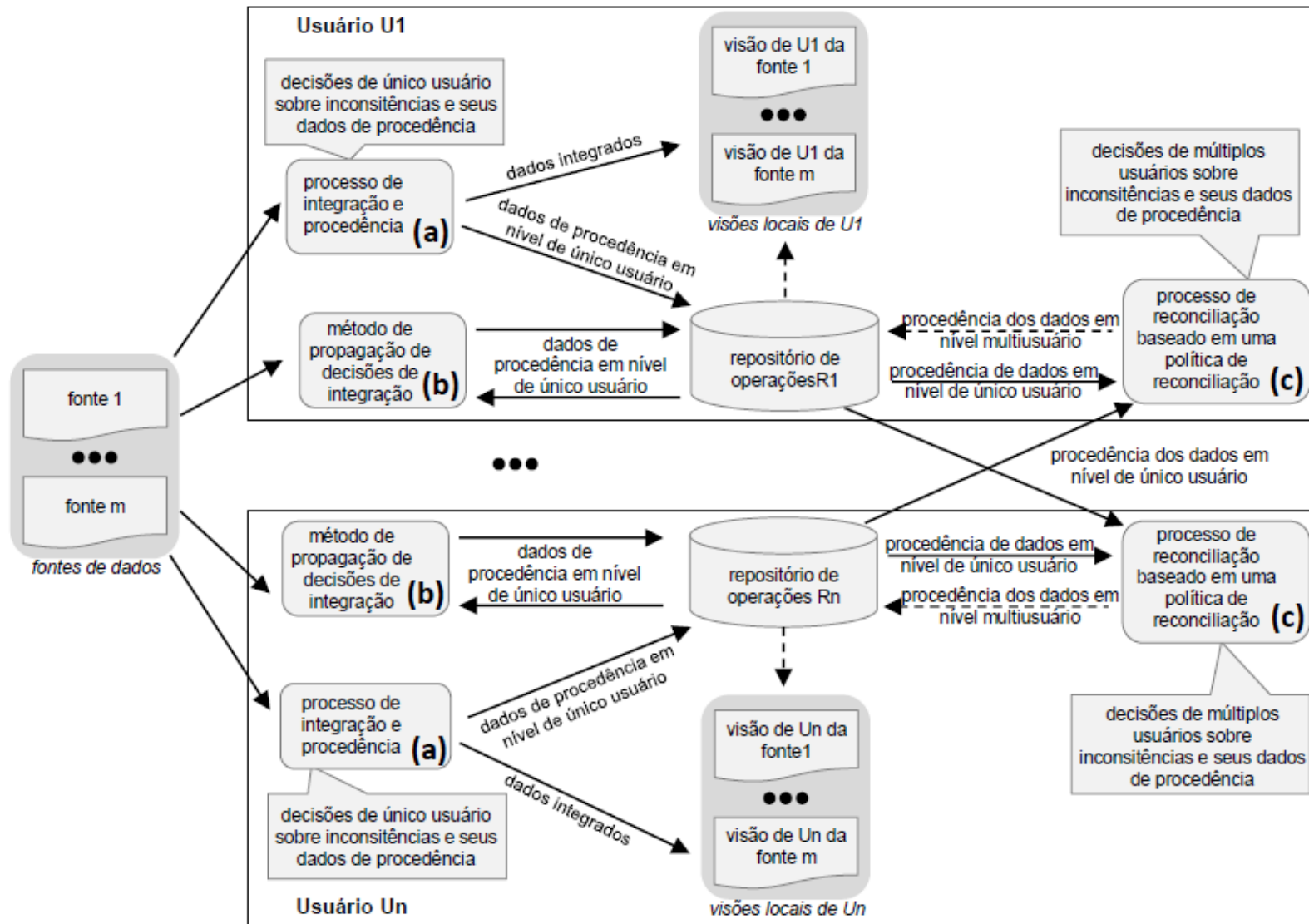
Asynchronous Collaborative data Reconciliation moDel

- Como compartilhar decisões de integração?
 - Armazenamento da procedência relacionada às decisões de integração em um **repositório de operações**
- Como possibilitar a geração de uma única visão consistente ou várias visões distintas?
 - Proposta de diferentes **políticas de reconciliação**
- Como garantir decisões consistentes sobre o mesmo dado presente em diferentes fontes?
 - Proposta do método de **propagação de decisões de integração**

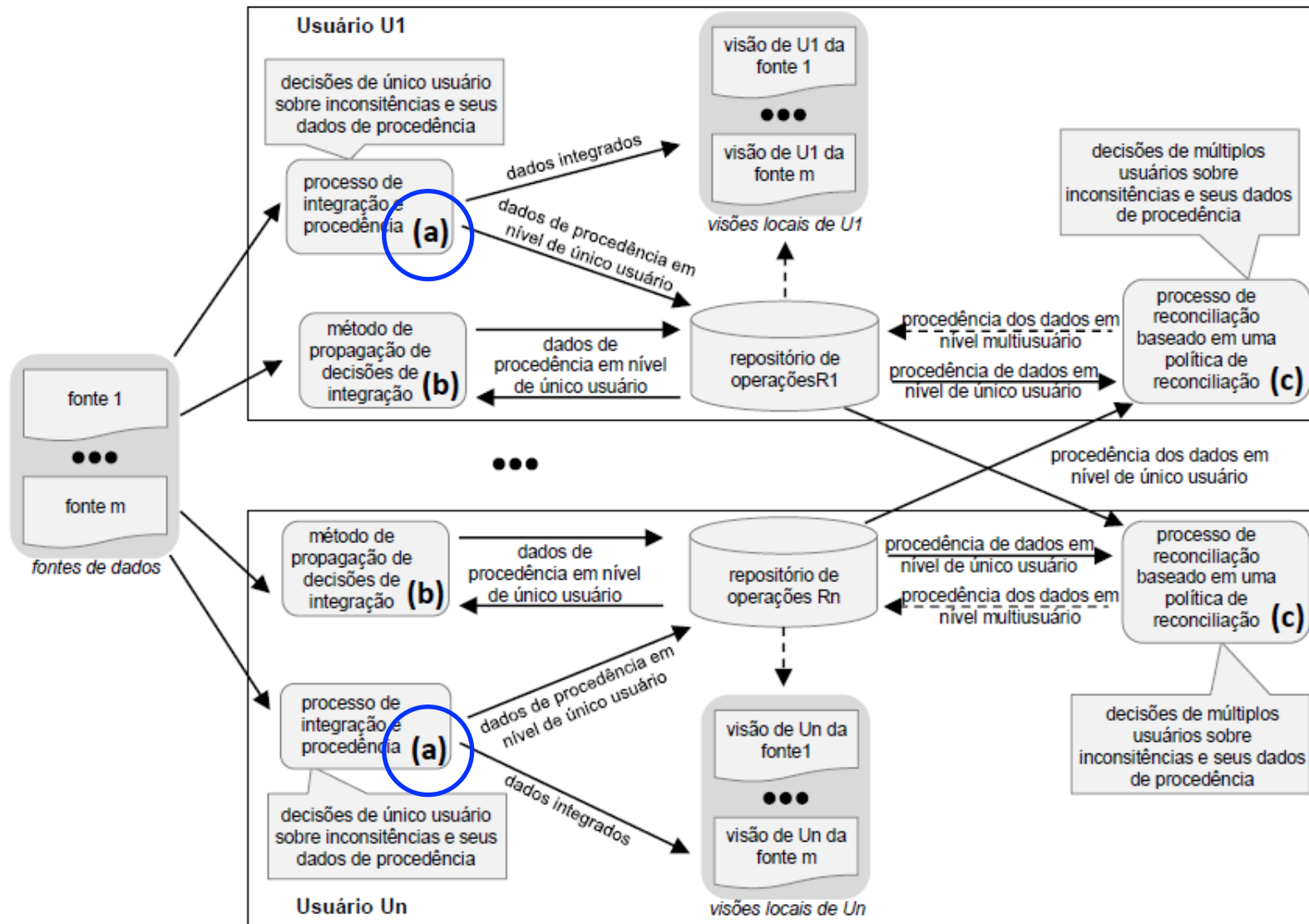
Roteiro

- Motivação
- **Modelo AcCORD**
- Repositório de operações
- Políticas de reconciliação
- Método de propagação de decisões de integração
- Resultados experimentais
- Conclusões

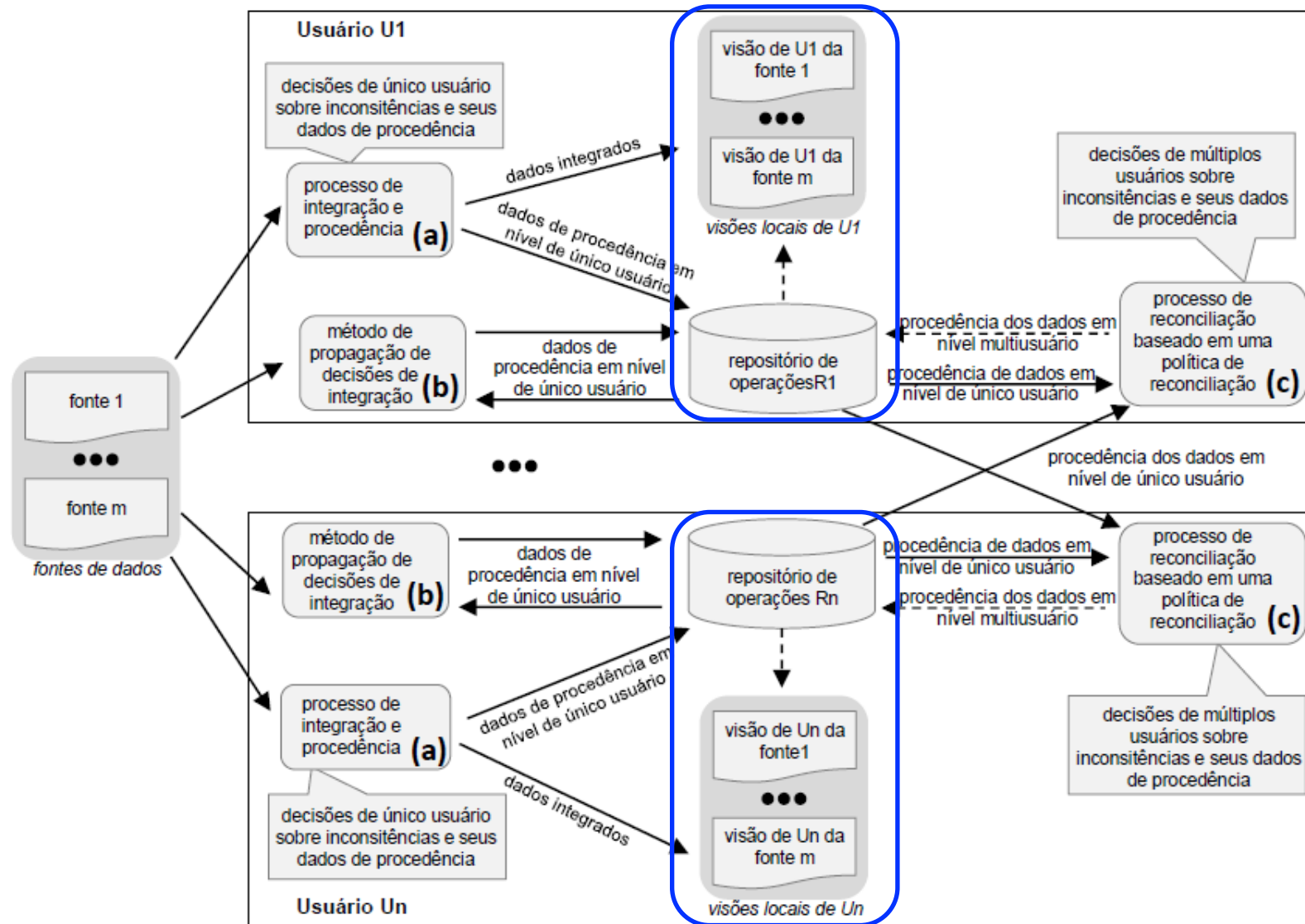
AcCORD



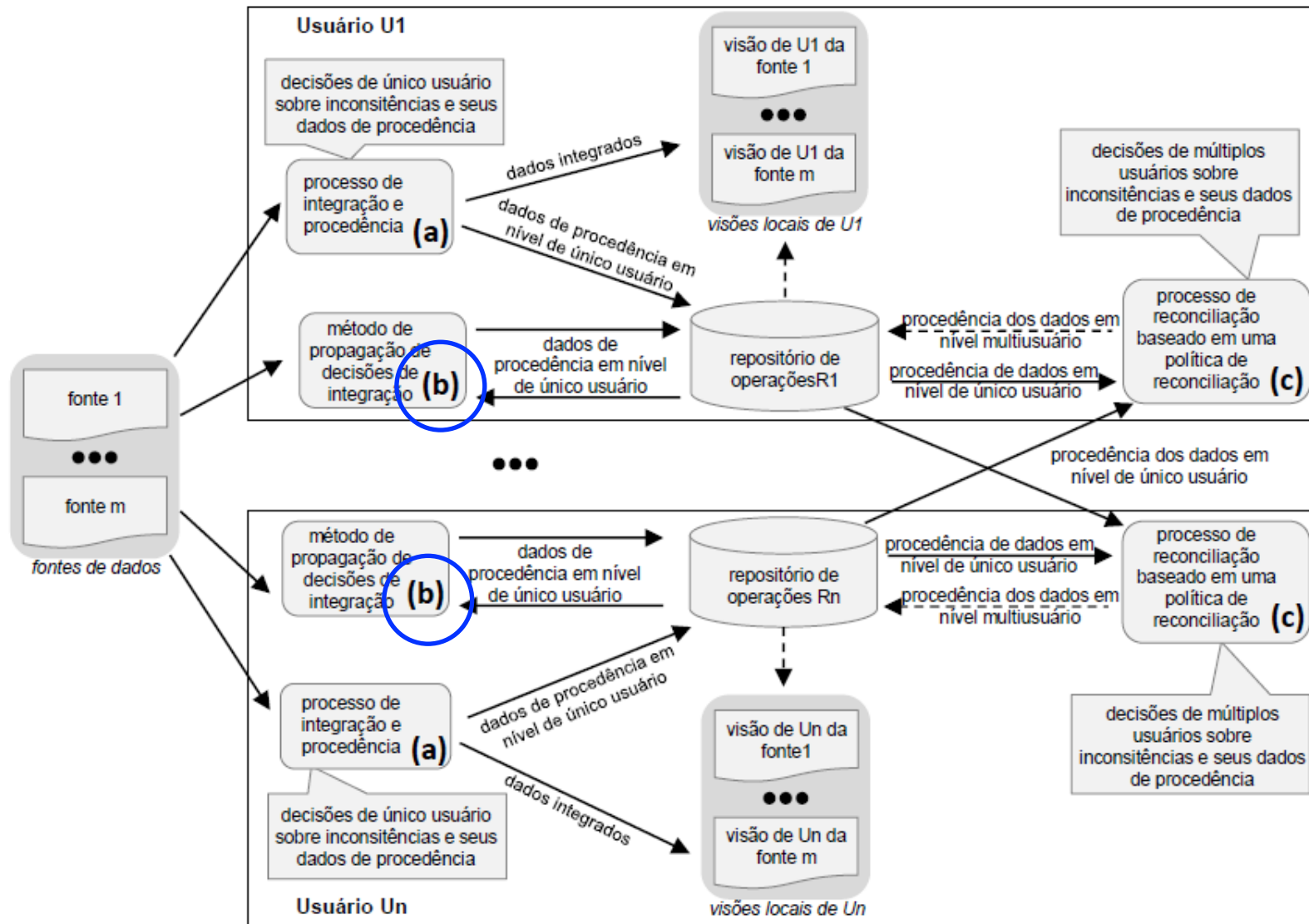
AcCORD



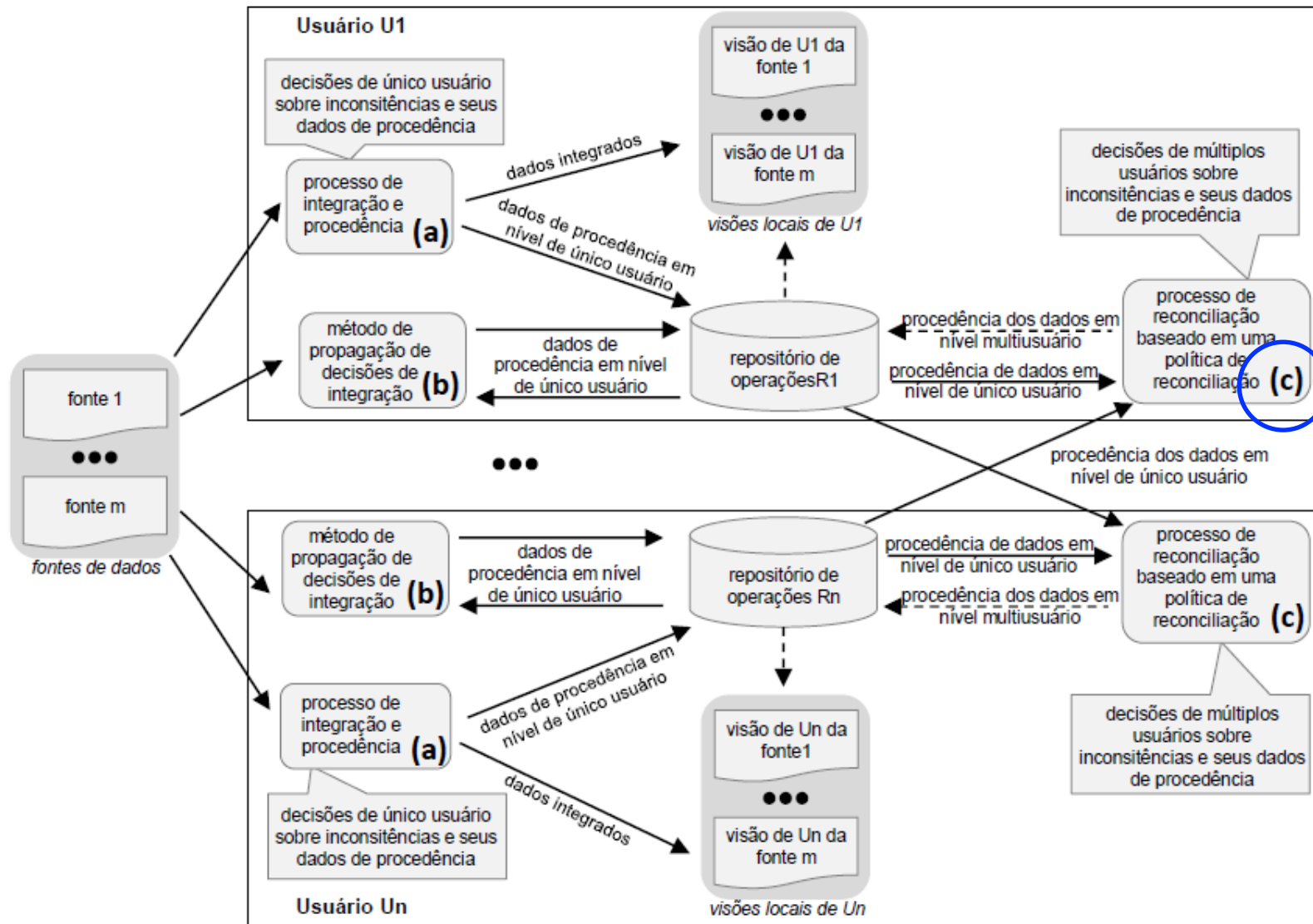
AcCORD



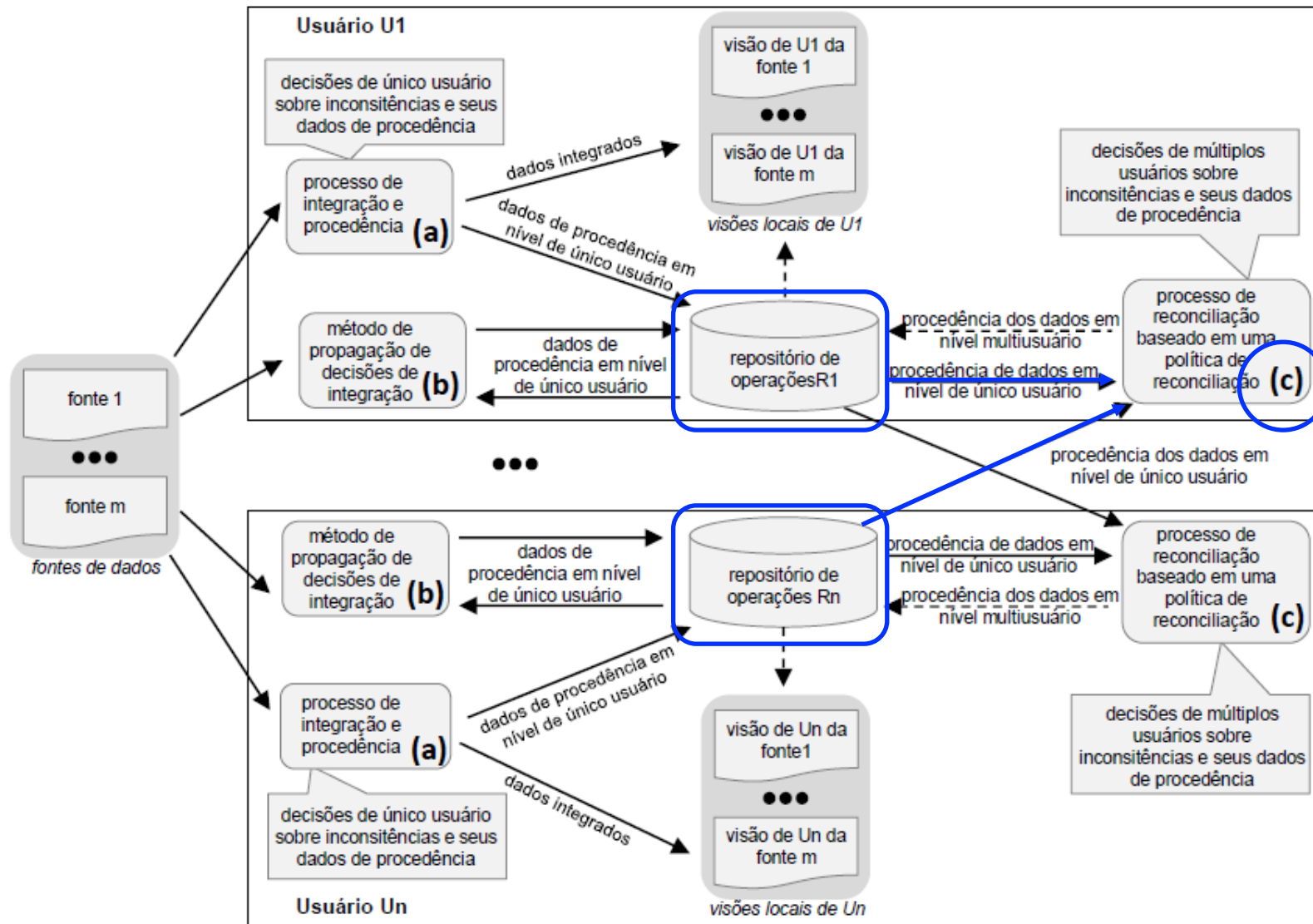
AcCORD



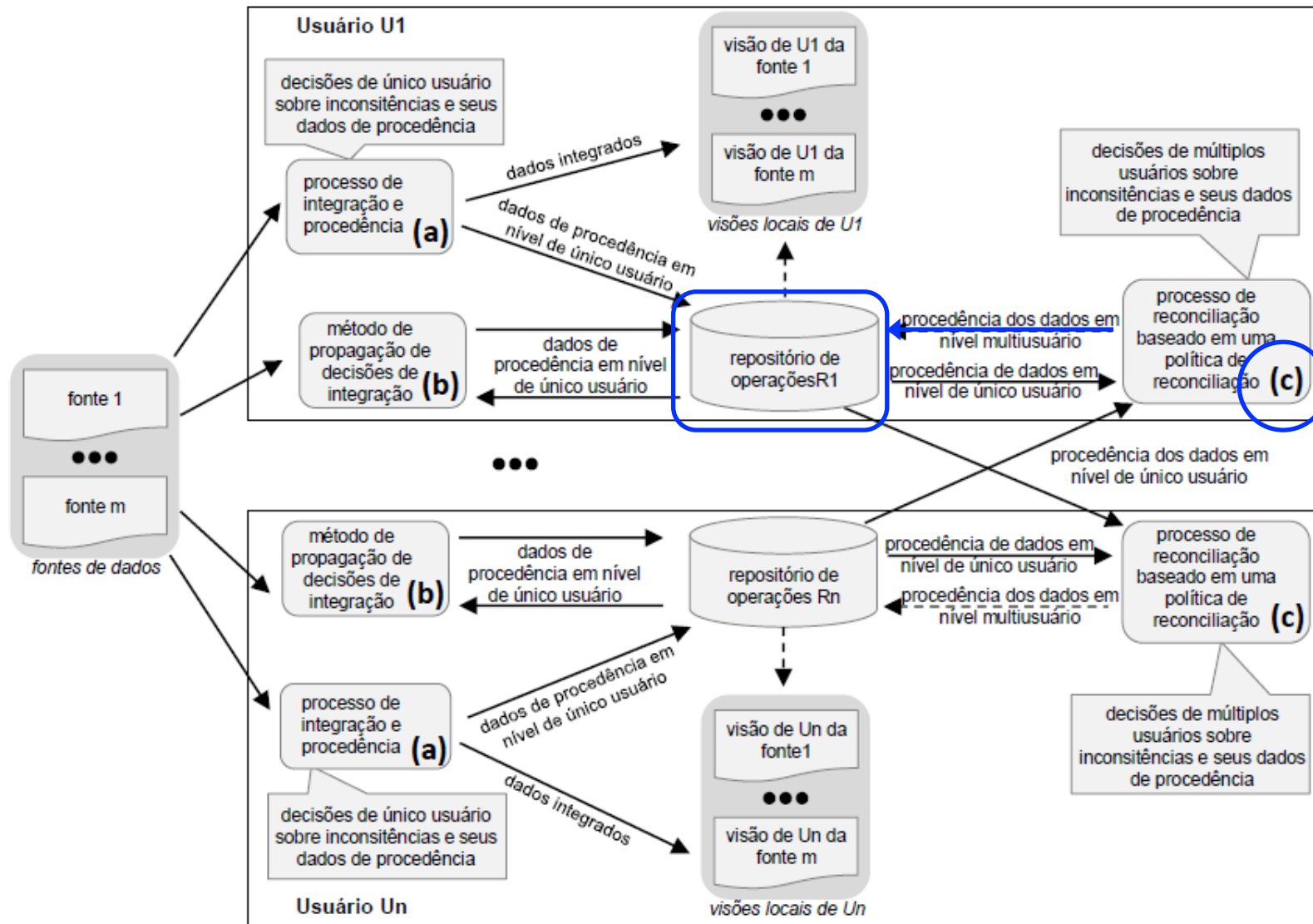
AcCORD



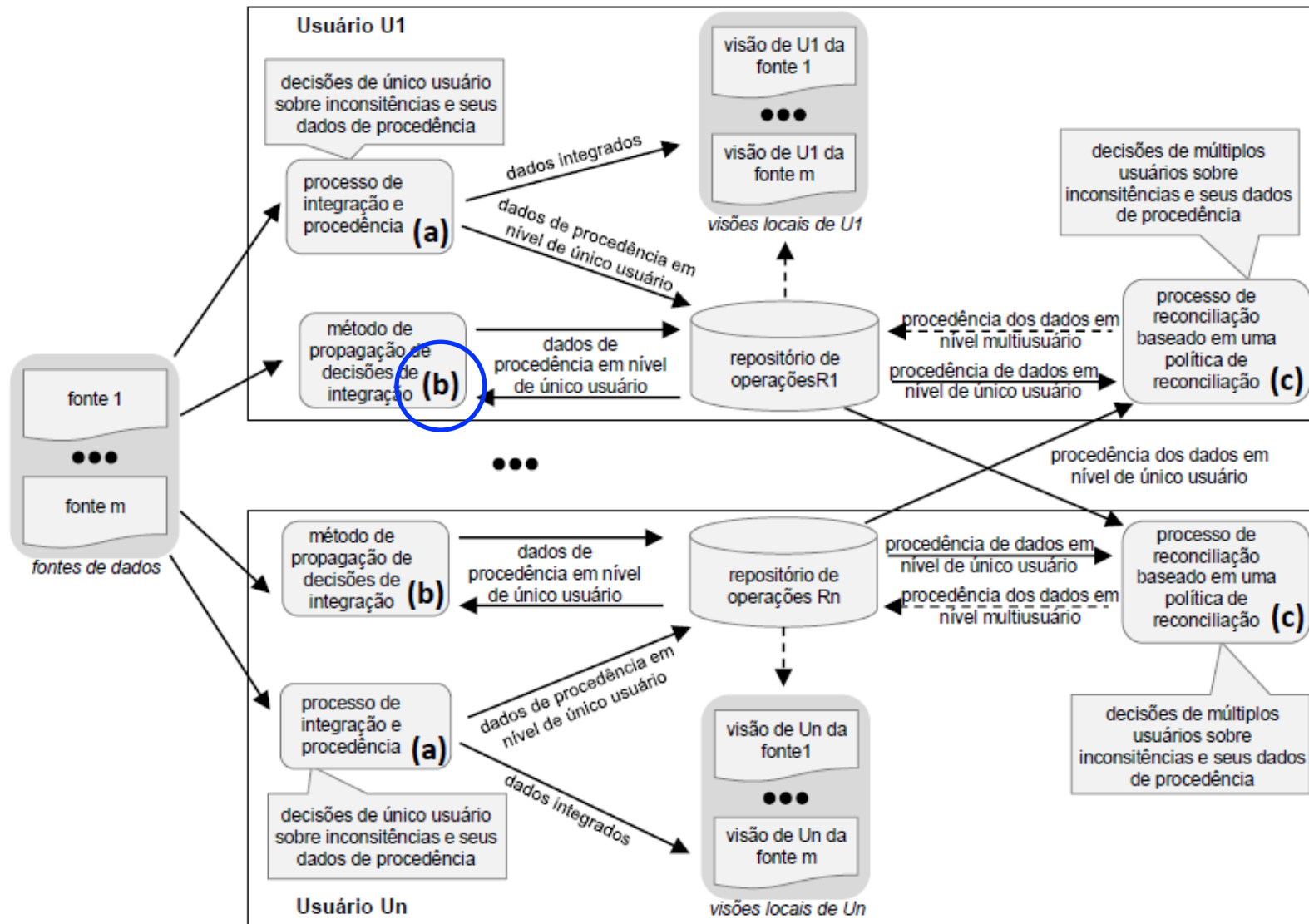
AcCORD



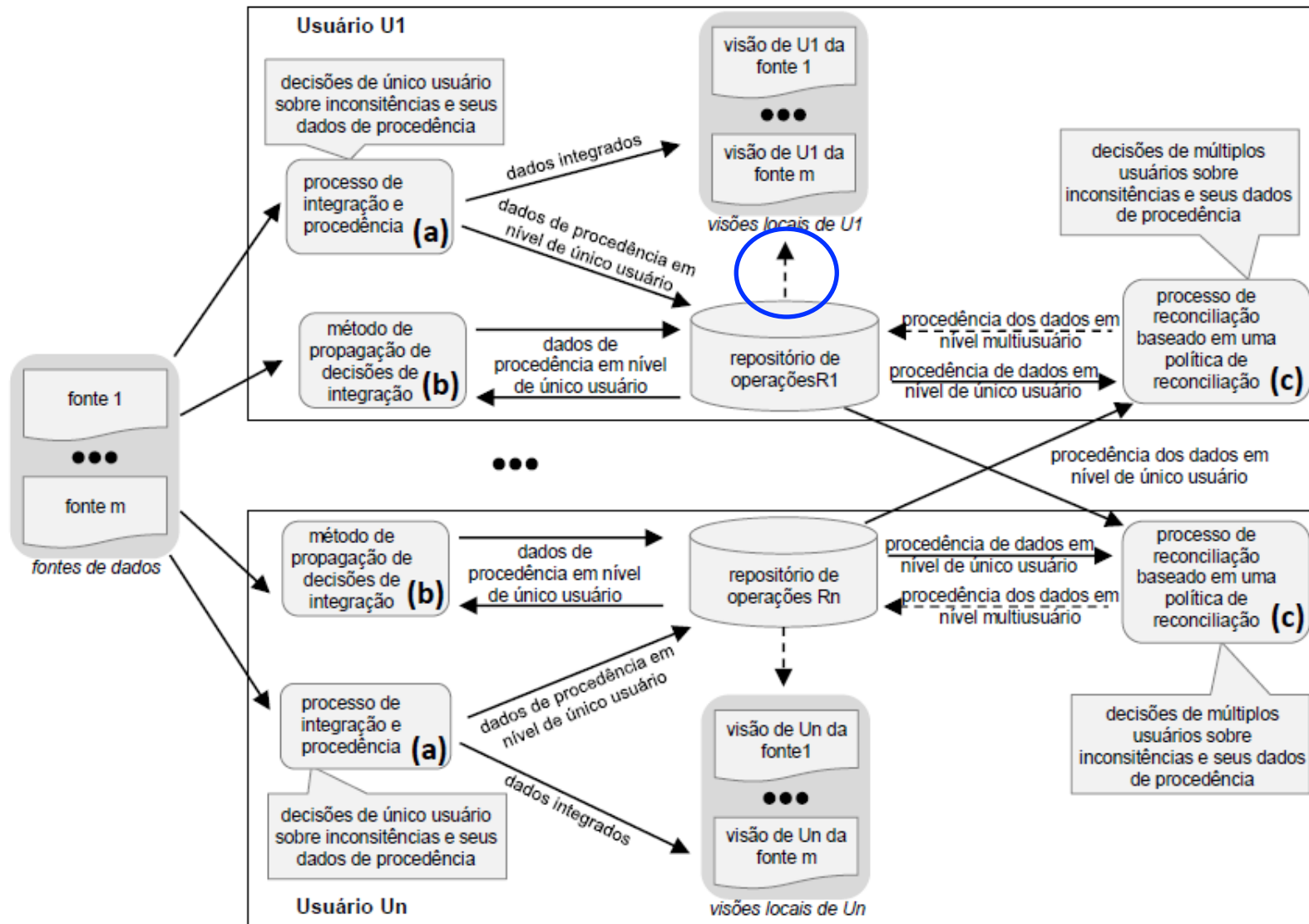
AcCORD



AcCORD



AcCORD



Roteiro

- Motivação
- Modelo AcCORD
- **Repositório de operações**
- Políticas de reconciliação
- Método de propagação de decisões de integração
- Resultados experimentais
- Conclusões

Repositório de operações

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

Repositório de operações

identificador único para usuário que tomou a decisão e criou a operação

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

Repositório de operações

usuários que confiaram em *op* em processos de reconciliação anteriores

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

Repositório de operações

operação que reflete uma decisão do usuário em um processo de integração

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

Repositório de operações

fonte que provê o valor correto de um atributo de um objeto

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

Repositório de operações

fonte na qual o valor do atributo foi atualizado por *op*

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

Repositório de operações

valor da chave que identifica um objeto

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

Repositório de operações

nome do atributo no qual *op* é realizada

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

Repositório de operações

valor do atributo na *origem*

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

Repositório de operações

valor do atributo no *destino* antes de ser sobrescrito por *valorOrigem*

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

Repositório de operações

marcador de tempo da execução de op

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

Dependência entre operações

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

$a(\text{usuário}, \text{destino}, \text{objeto}, \text{atributo}) = b(\text{usuário}, \text{origem}, \text{objeto}, \text{atributo}) \text{ e } a < b$

Operações conflitantes: tipo 1

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

$a(\text{destino}, \text{objeto}, \text{atributo}) = b(\text{destino}, \text{objeto}, \text{atributo})$

Operações conflitantes: tipo 2

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

$a(\text{destino}, \text{objeto}, \text{atributo}) = b(\text{origem}, \text{objeto}, \text{atributo})$

Operações conflitantes: tipo 3

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:29:01_03/18/2014
	3		cp	Bruno	Ana	Artigo[Título=AcCORD]	ano	2015	2013	08:29:17_03/18/2014
U2	1	U5	cp	Bruno	Daniel	Artigo[Título=AcCORD]	página final	24	15	13:41:47_03/18/2014
U3	1		ed	null	Daniel	Artigo[Título=AcCORD]	página final	22	15	12:24:55_03/18/2014
	2		cp	Daniel	Emanuel	Artigo[Título=AcCORD]	página final	20	1352	12:32:47_03/18/2014
U4	1	U6	cp	Daniel	Fábio	Artigo[Título=AcCORD]	página final	15	42	10:13:45_03/18/2014
	2	U6	cp	Daniel	Gabriel	Artigo[Título=AcCORD]	página final	15	25	10:14:21_03/18/2014
	3	U6	cp	Fábio	Hugo	Artigo[Título=AcCORD]	página final	15	32	10:15:42_03/18/2014
U5	1		cp	Bruno	Gabriel	Artigo[Título=AcCORD]	página final	24	25	09:45:57_03/18/2014
	2		cp	Gabriel	Ivan	Artigo[Título=AcCORD]	página final	24	72	09:50:02_03/18/2014
U6	1		cp	Bruno	Hugo	Artigo[Título=AcCORD]	página inicial	5	13	07:25:56_03/18/2014

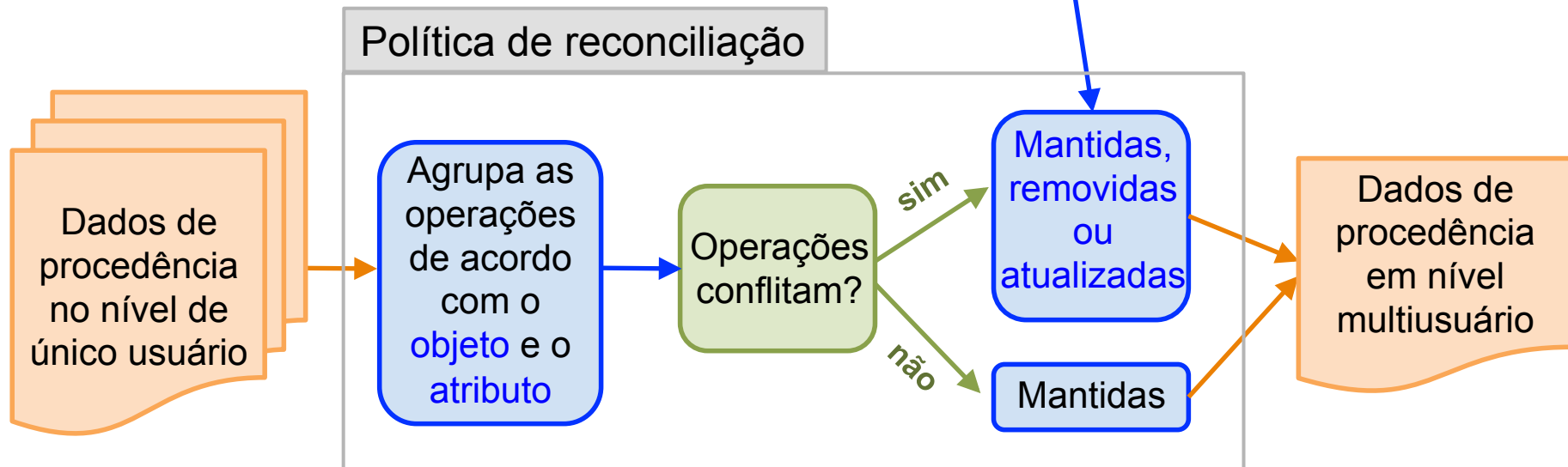
$a(\text{origem, objeto, atributo}) = b(\text{destino, objeto, atributo})$

Roteiro

- Motivação
- Modelo AcCORD
- Repositório de operações
- **Políticas de reconciliação**
- Método de propagação de decisões de integração
- Resultados experimentais
- Conclusões

Políticas de Reconciliação

políticas diferem uma da outra em relação à maneira como gerenciam as operações em conflito e suas operações dependentes



Política: Visão local

- **Prioriza operações conflitantes do usuário local**
 - Se quaisquer operações p e o conflitam e p é uma operação do usuário local
 - Mantém p e suas dependentes
 - Remove o e suas dependentes
- **Objetivo do trabalho colaborativo**
 - Gerar **várias visões distintas**

Política: Remove todos os conflitos

- Descarta operações conflitantes
 - Se quaisquer operações p e o conflitam
 - Remove p e suas dependentes
 - Remove o e suas dependentes
- Objetivo do trabalho colaborativo
 - Gerar uma **única visão consistente**

Política: *Timestamp*

- Prioriza a ordem temporal das operações conflitantes
 - Se quaisquer operações p e o conflitam e p tem o *timestamp* mais recente
 - Mantém p e suas dependentes
 - Remove o ou atualiza o com o valor de origem de p
 - Atualiza as operações dependentes de o com o valor de origem de p
- Objetivo do trabalho colaborativo
 - Gerar uma **única visão consistente**

Política: Votação

- Mantém as operações conflitantes que refletem a maioria das decisões
 - Se quaisquer operações p e o conflitam e
 - a) p pertence ao conjunto de operações vencedoras
 - Mantém p e suas dependentes
 - Remove o e suas dependentes ou atualiza o e suas dependentes com o valor de origem de p
 - b) não há operações vencedoras
 - remove o , p e todas as dependentes
- Objetivo do trabalho colaborativo
 - Gerar uma **única visão consistente**

Política: Confiança nas fontes

- Mantém as operações conflitantes cuja origem é a fonte mais confiável
 - Se quaisquer operações p e o conflitam e p tem como origem a fonte com maior taxa de confiança
 - mantém p e suas dependentes
 - remove o e suas dependentes ou atualiza o e suas dependentes com o valor de origem de p
- Objetivo do trabalho colaborativo
 - Gerar **várias visões distintas**

Taxa de confiança da fonte: determinada analisando-se as operações do repositório

Política: Confiança nos usuários

- Mantém as operações conflitantes que possuem a maior confiança
 - Se quaisquer operações p e o conflitam e p tem maior taxa de confiança nos usuários
 - mantém p e suas dependentes
 - remove o e suas dependentes ou atualiza o e suas dependentes com o valor de origem de p
- Objetivo do trabalho colaborativo
 - Gerar **várias visões distintas**

Taxa de confiança nos usuários: determinada analisando-se as operações do repositório, e usando como base a confiança no usuário que realizou a operação e nos usuários que confiaram naquela operação em outros processos

Roteiro

- Motivação
- Modelo AcCORD
- Repositório de operações
- Políticas de reconciliação
- **Método de propagação de decisões de integração**
- Resultados experimentais
- Conclusões

Método de Propagação de Decisões de Integração

- Objetivo
 - Propagar uma decisão de integração do usuário tomada para resolver o valor de um item de dado para cada fonte que contém aquele item de dado inconsistente
- Exemplo
 - Fontes: Ana, Bruno e Carlos

usuário	id	usuário conf	op	origem	destino	objeto	atributo	valor origem	valor destino	timestamp
U1	1		cp	Carlos	Bruno	Artigo[Título=AcCORD]	página final	234	24	08:28:43_03/18/2014
	2		cp	Bruno	Ana	Artigo[Título=AcCORD]	página final	234	20	08:28:43_03/18/2014

Roteiro

- Motivação
- Modelo AcCORD
- Repositório de operações
- Políticas de reconciliação
- Método de propagação de decisões de integração
- **Resultados experimentais**
- Conclusões

Resultados experimentais

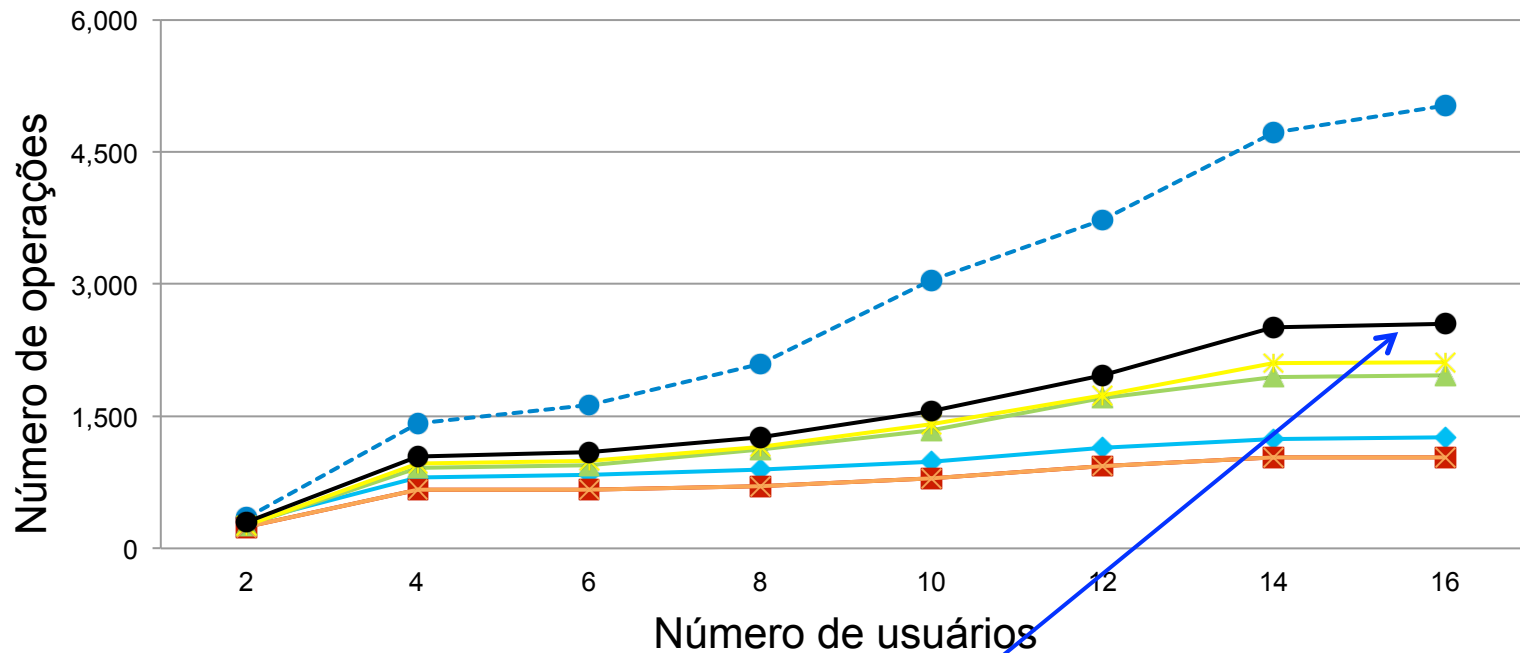
- Dados
 - Currículos de pesquisadores do ICMC/USP
- Implementação
 - C++ usando o Qt versão 5.5.1 e compilado com MingW versão 4.9.2
- Máquina
 - Processador Intel Core i5-4200U 1.6 GHz e 6 GB de memória principal
- Tipos de experimento
 - Investigar a eficiência e a eficácia do AcCORD
 - Investigar a opinião do usuário

Eficiência e eficácia

- Experimento 1: Menor volume de dados
 - Características
 - 4 currículos com grande numero de publicações em comum
 - 12 repositórios variando de 79 operações a 749 operações
 - Processo de reconciliação
 1. Primeiro processo
 2. 5 processos de reconciliação consecutivos
 3. 7 processos de reconciliação consecutivos
- Experimento 2: Maior volume de dados
 - Características
 - 120 currículos
 - 12 repositórios variando de 1.466 operações a 2.755 operações
 - Processo de reconciliação
 1. Primeiro processo
 2. 5 processos de reconciliação consecutivos

Resultados de desempenho
similares

Número de operações removidas (1/2)

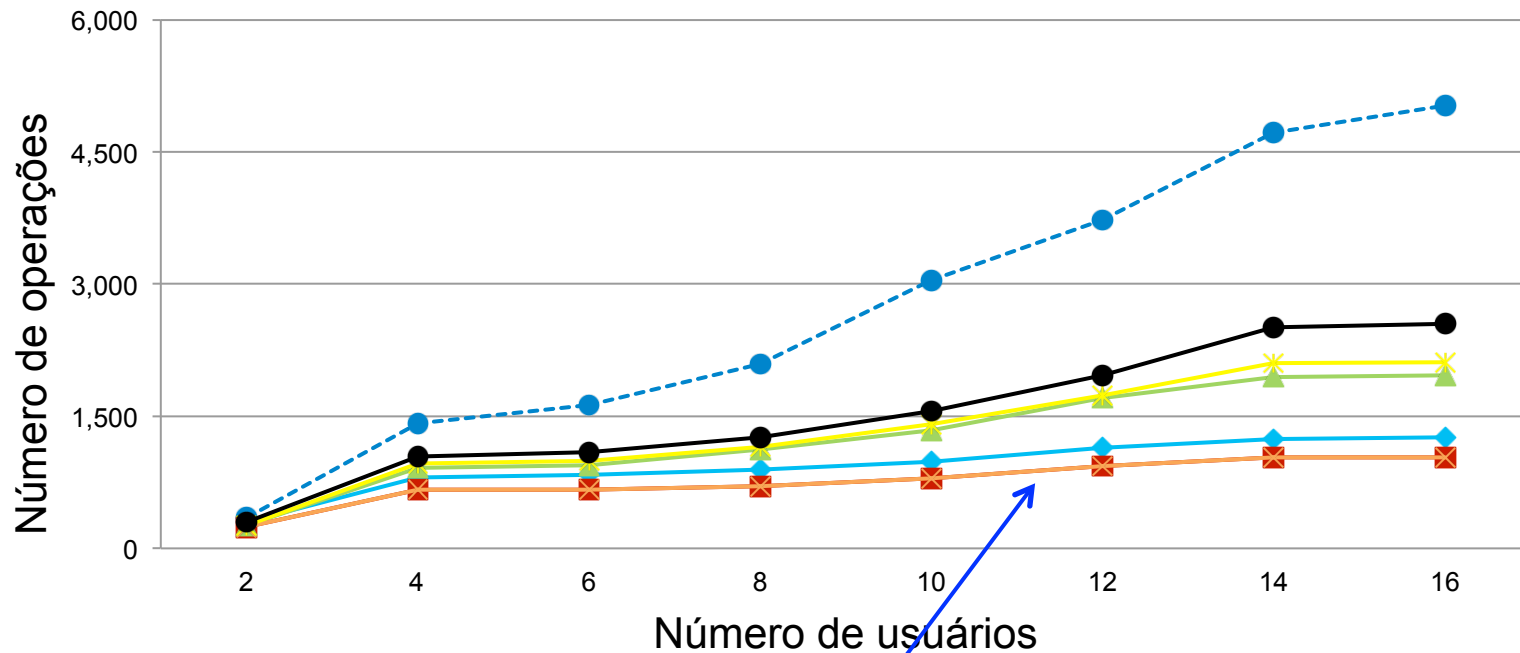


- Número total de operações
- ▲ Timestamp
- Confiança nos Usuários
- ◆ Visão Local
- × Votação
- Remove todos Conflitos
- * Confiança nas Fontes

A política baseada na confiança nos usuários removeu menos operações

- Causou a perda de um menor número de decisões de integração

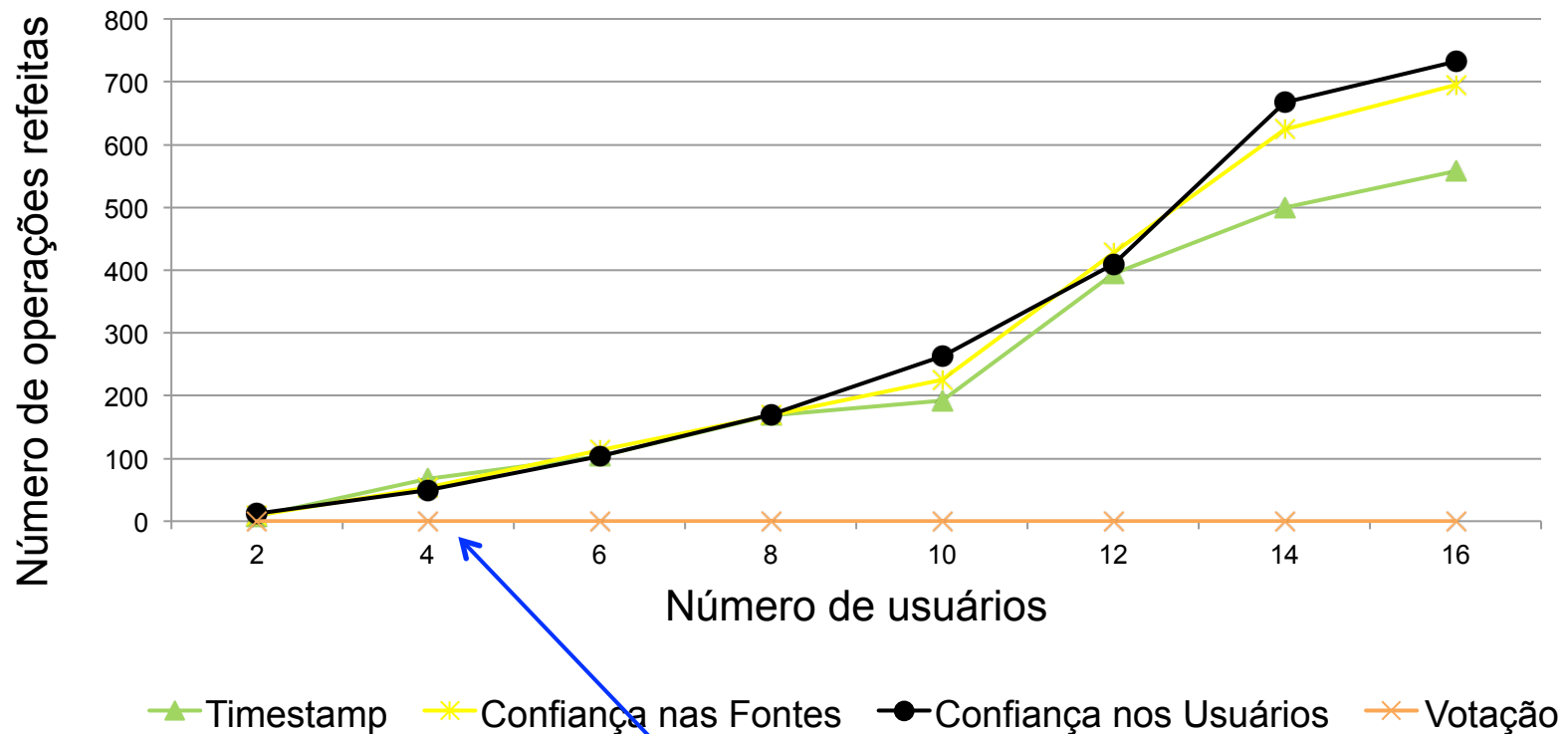
Número de operações removidas (2/2)



- Número total de operações
- ▲ Timestamp
- Confiança nos Usuários
- ◆ Visão Local
- × Votação
- Remove todos Conflitos
- * Confiança nas Fontes

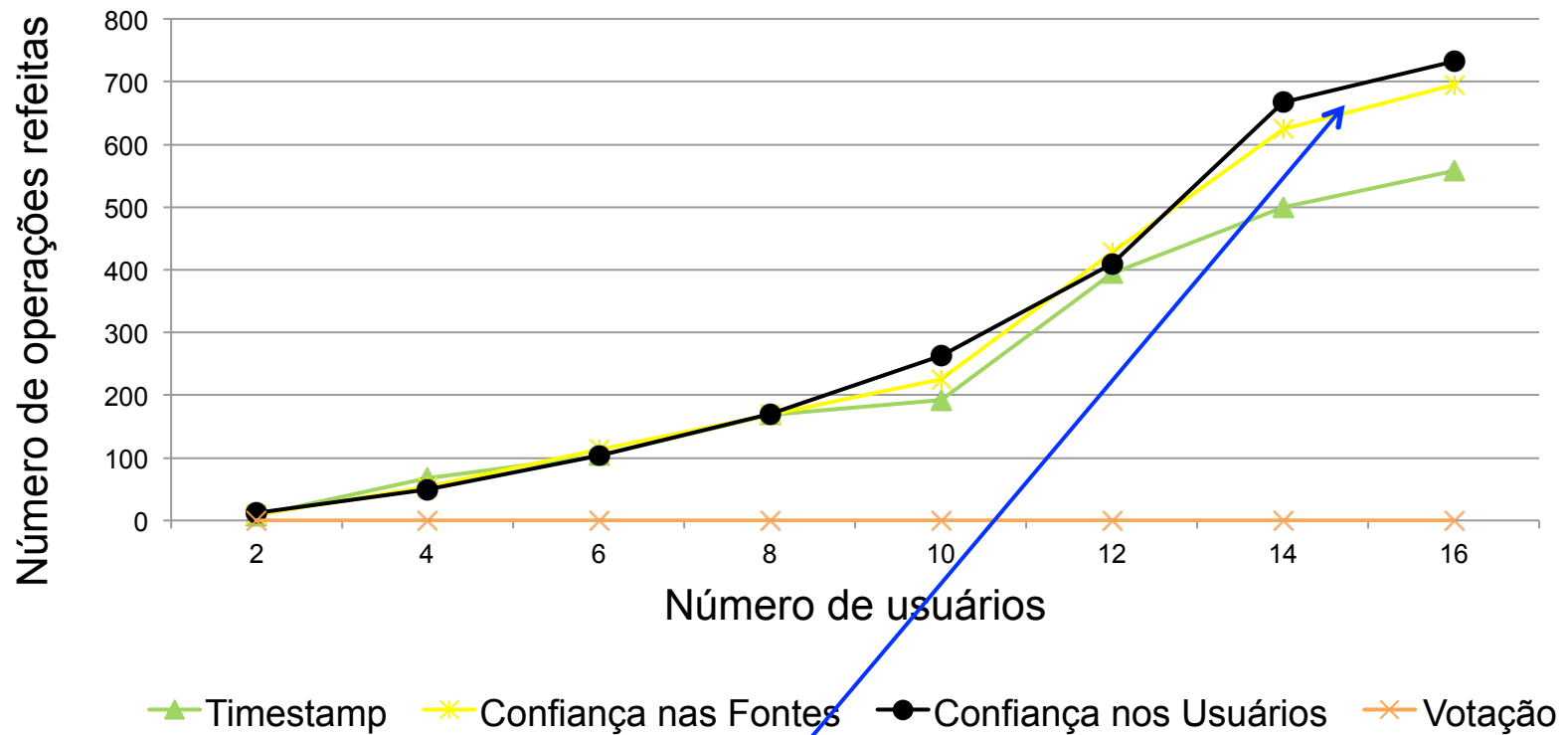
A política que remove todos os conflitos e a política baseada em votação removeram mais operações que as demais

Número de operações refeitas (1/3)



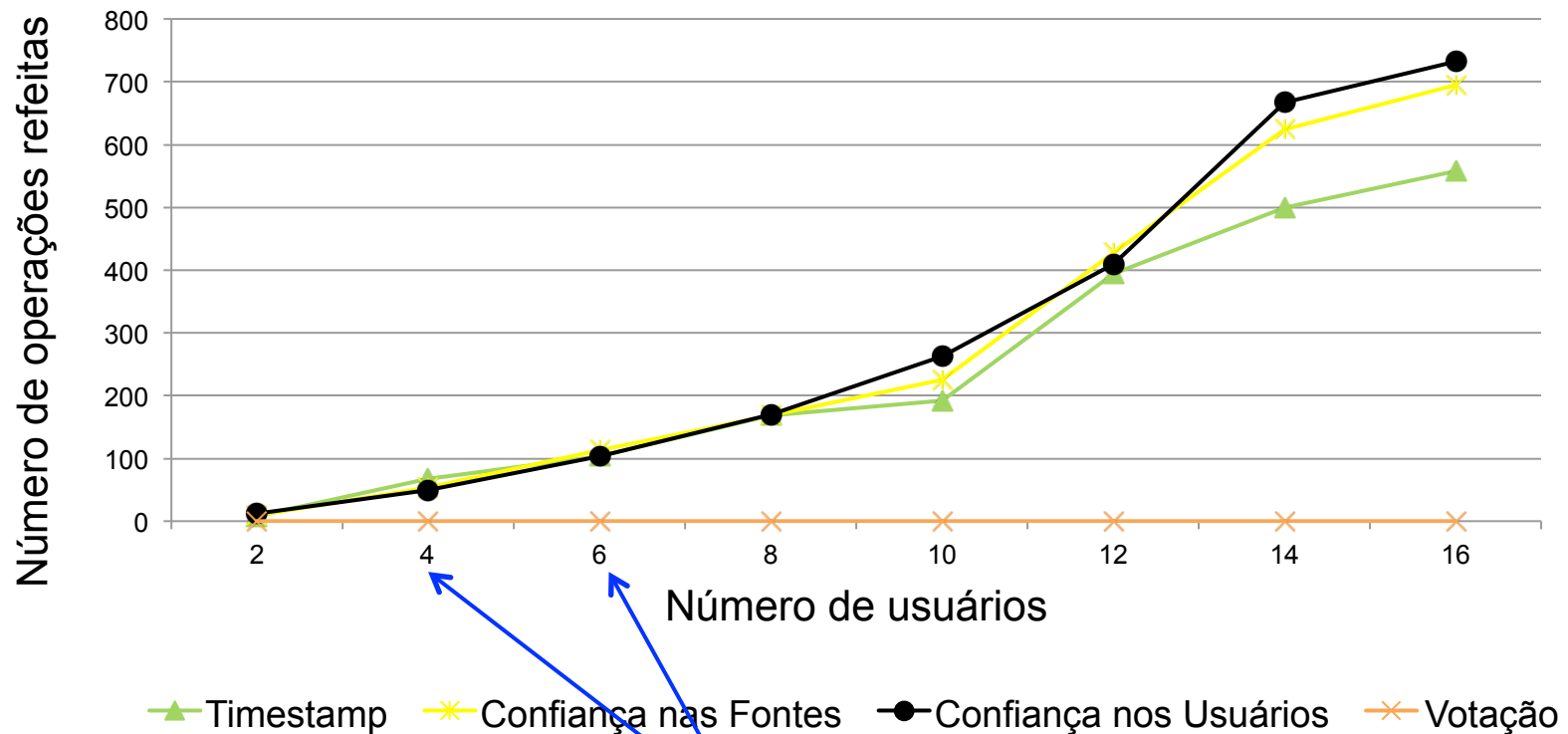
A política baseada em votação teve o mesmo comportamento da política que remove todos os conflitos e não refez qualquer operação

Número de operações refeitas (2/3)



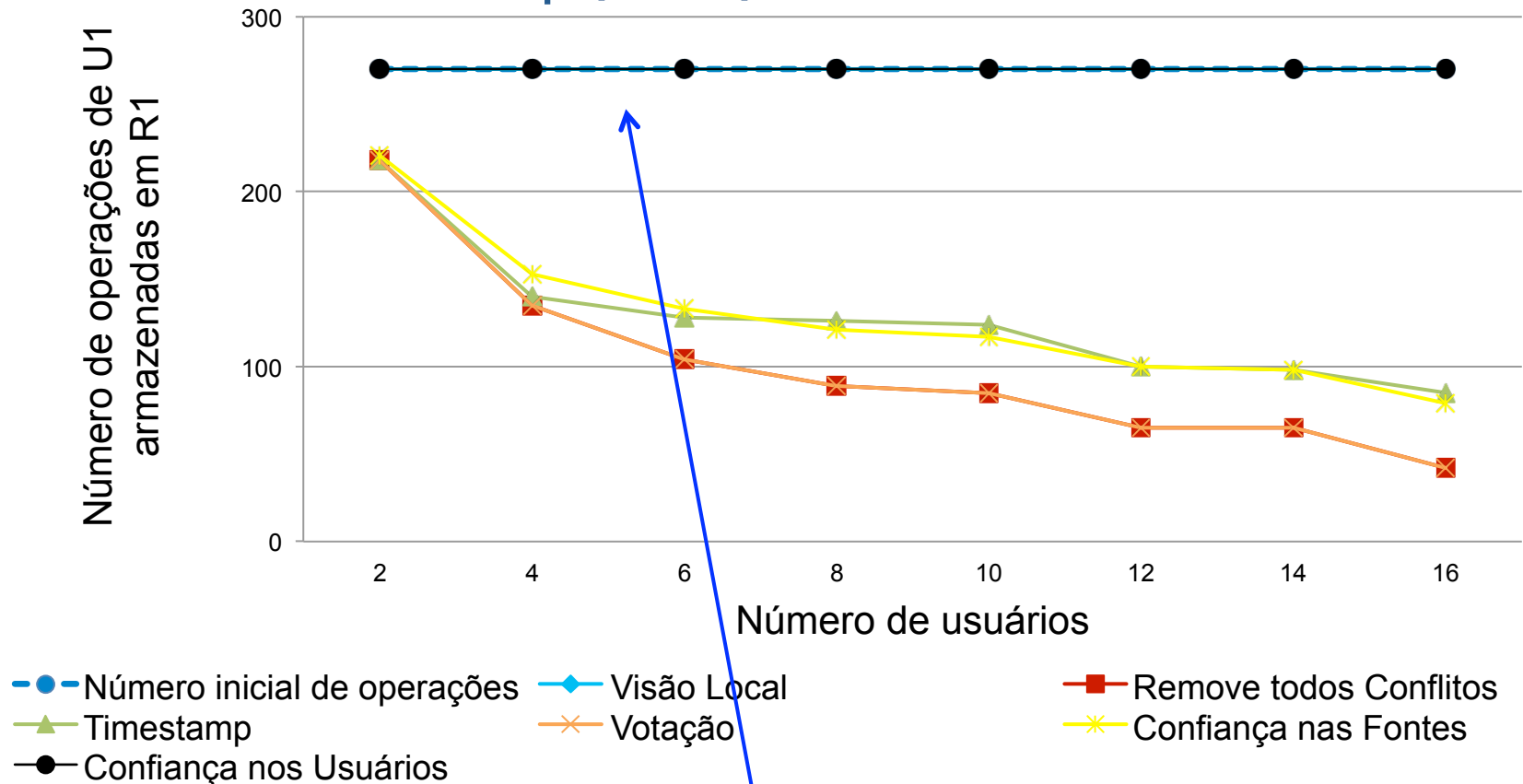
A política baseada na confiança nos usuários refez o maior número de operações na maioria dos casos

Número de operações refeitas (3/3)



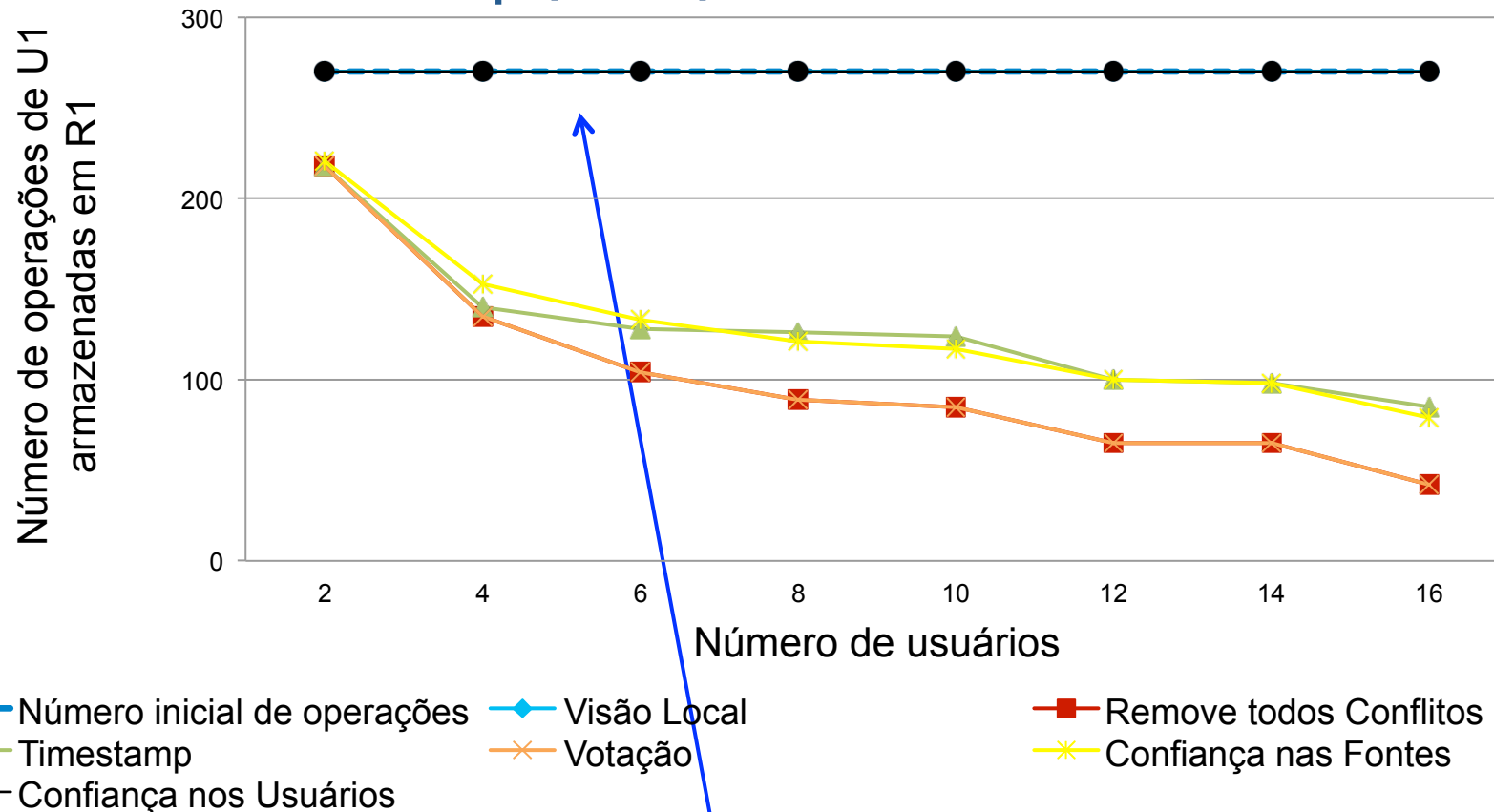
Apenas para 4 e 6 usuários que a *política baseada em timestamp* e a *política baseada na confiança nas fontes* obtiveram melhores resultados que a *política baseada na confiança nos usuários*

Número de operações removidas do repositório R_1 (1/4)



Política baseada na visão local não removeu qualquer operação

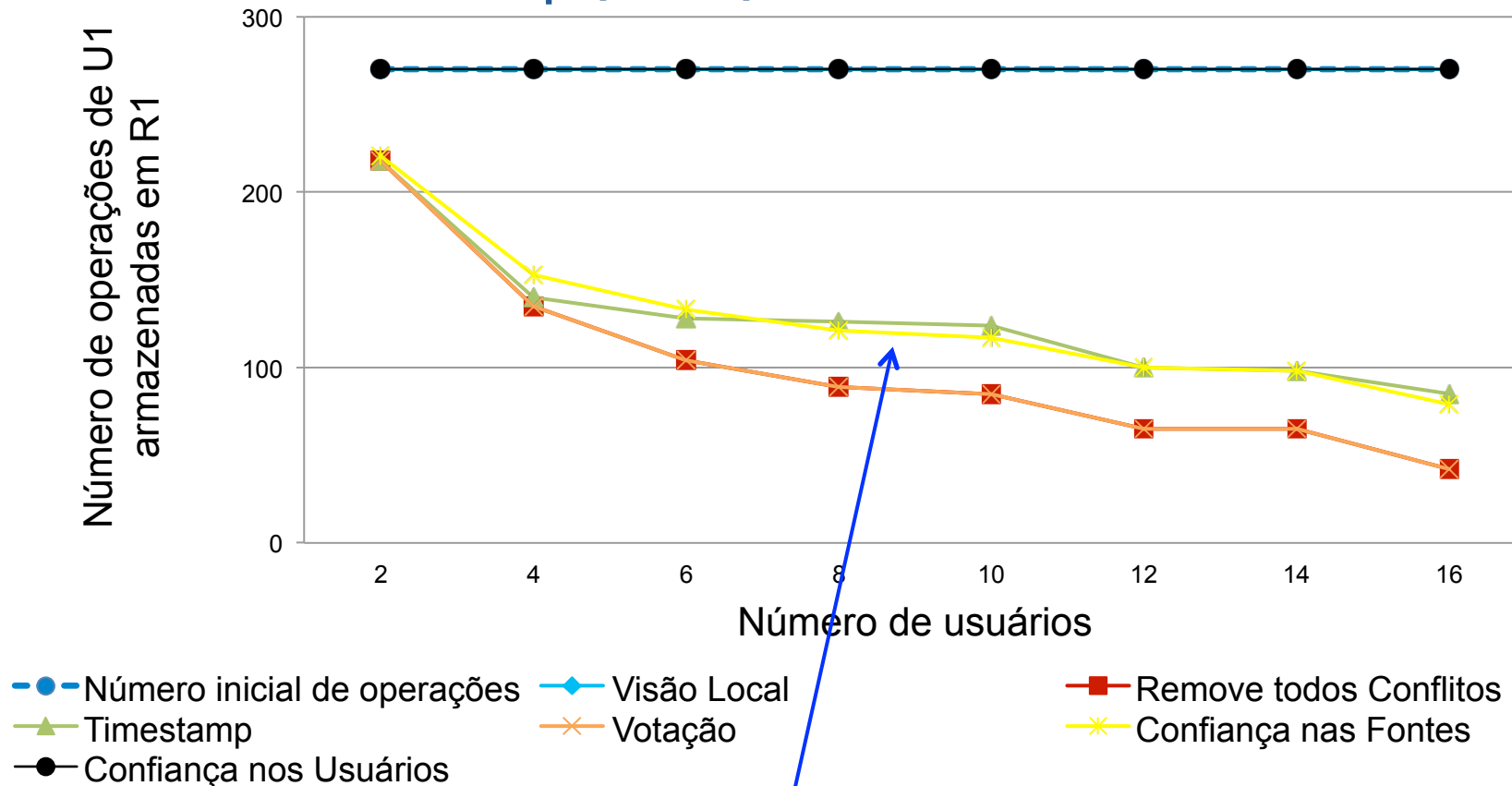
Número de operações removidas do repositório R_1 (2/4)



Como é o primeiro processo de reconciliação, a confiança nos usuários para a política baseada nessa confiança é determinada somente para o usuário local

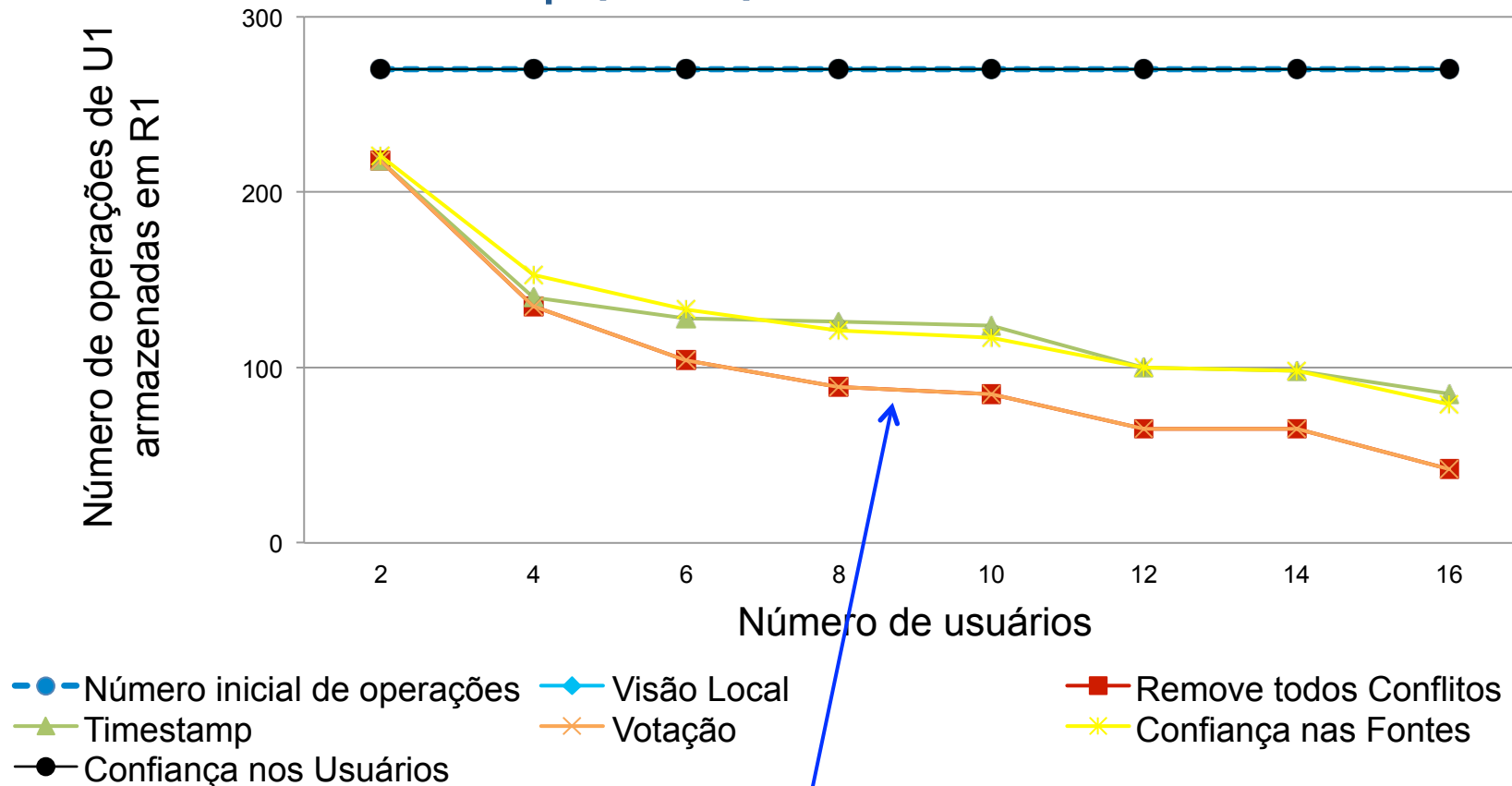
- Portanto, a maior confiança é a do próprio U1

Número de operações removidas do repositório R_1 (3/4)



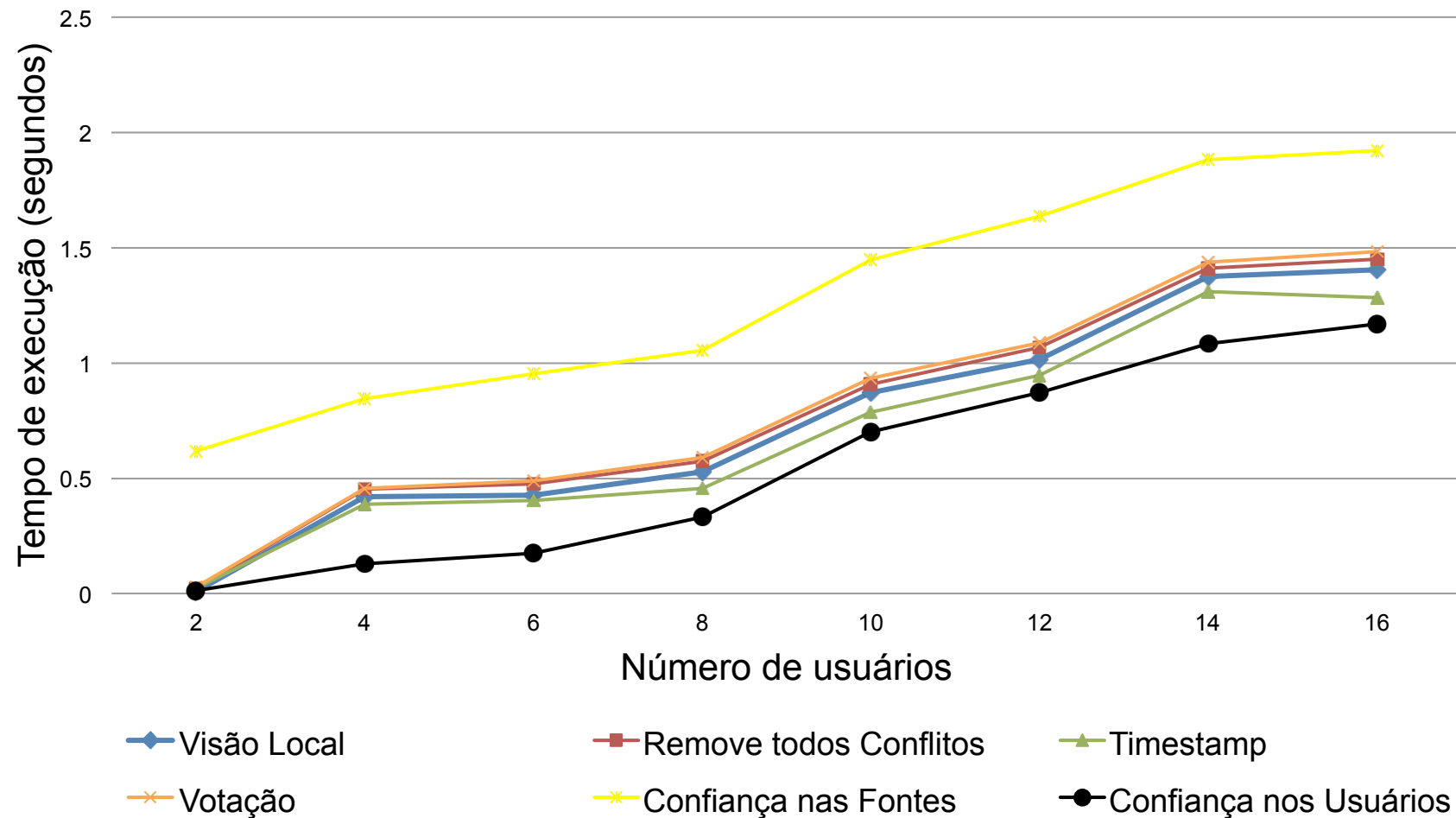
A política baseada em timestamp e a política baseada na confiança nas fontes removeram um número médio de operações do usuário U1

Número de operações removidas do repositório R_1 (4/4)



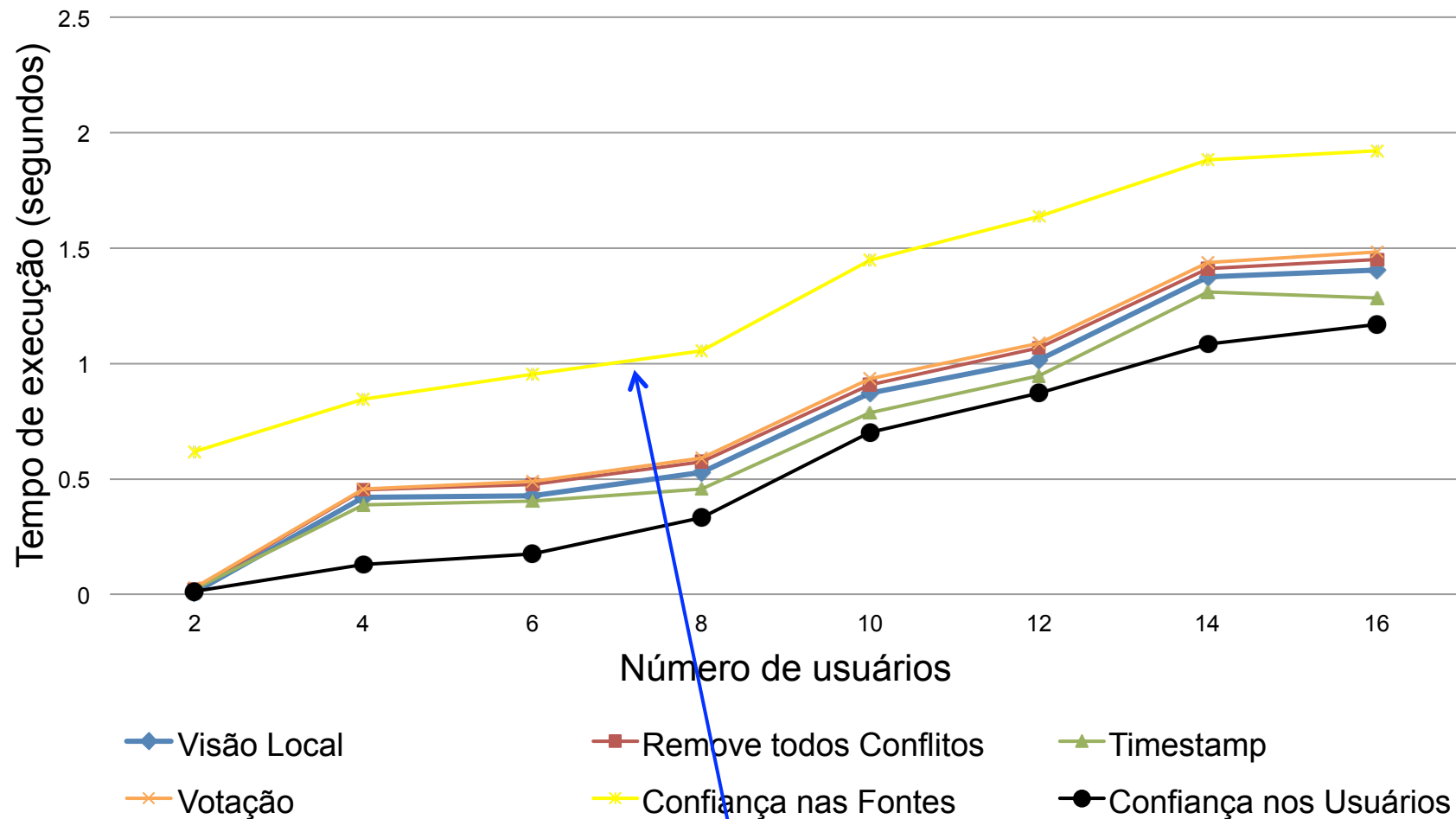
A política que remove todos os conflitos e a política baseada em votação tiveram o mesmo comportamento, removendo um grande número de operações do usuário U_1

Tempos de execução (1/3)



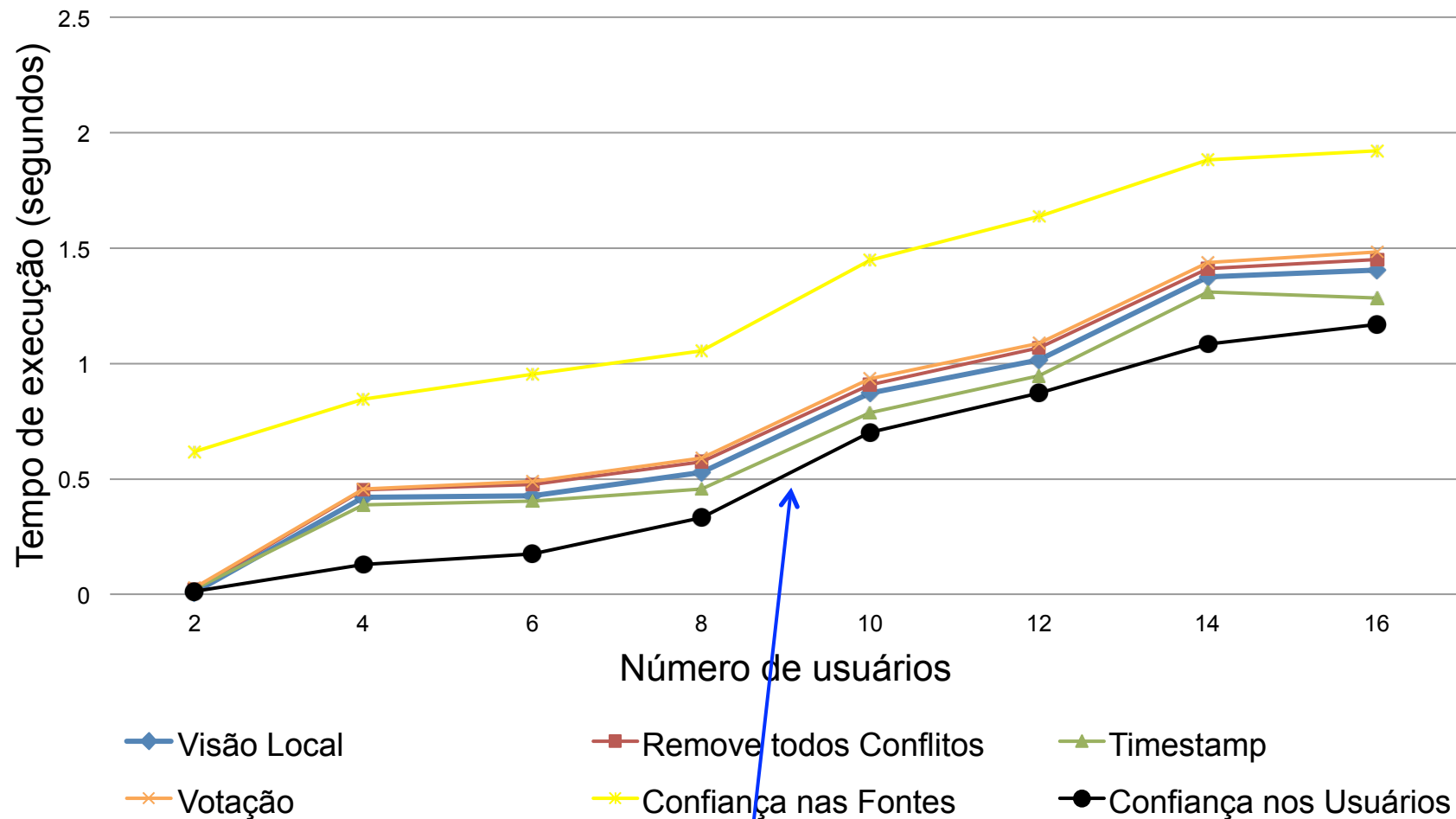
Os tempos de execução aumentaram linearmente com o aumento do número de operações

Tempos de execução (2/3)



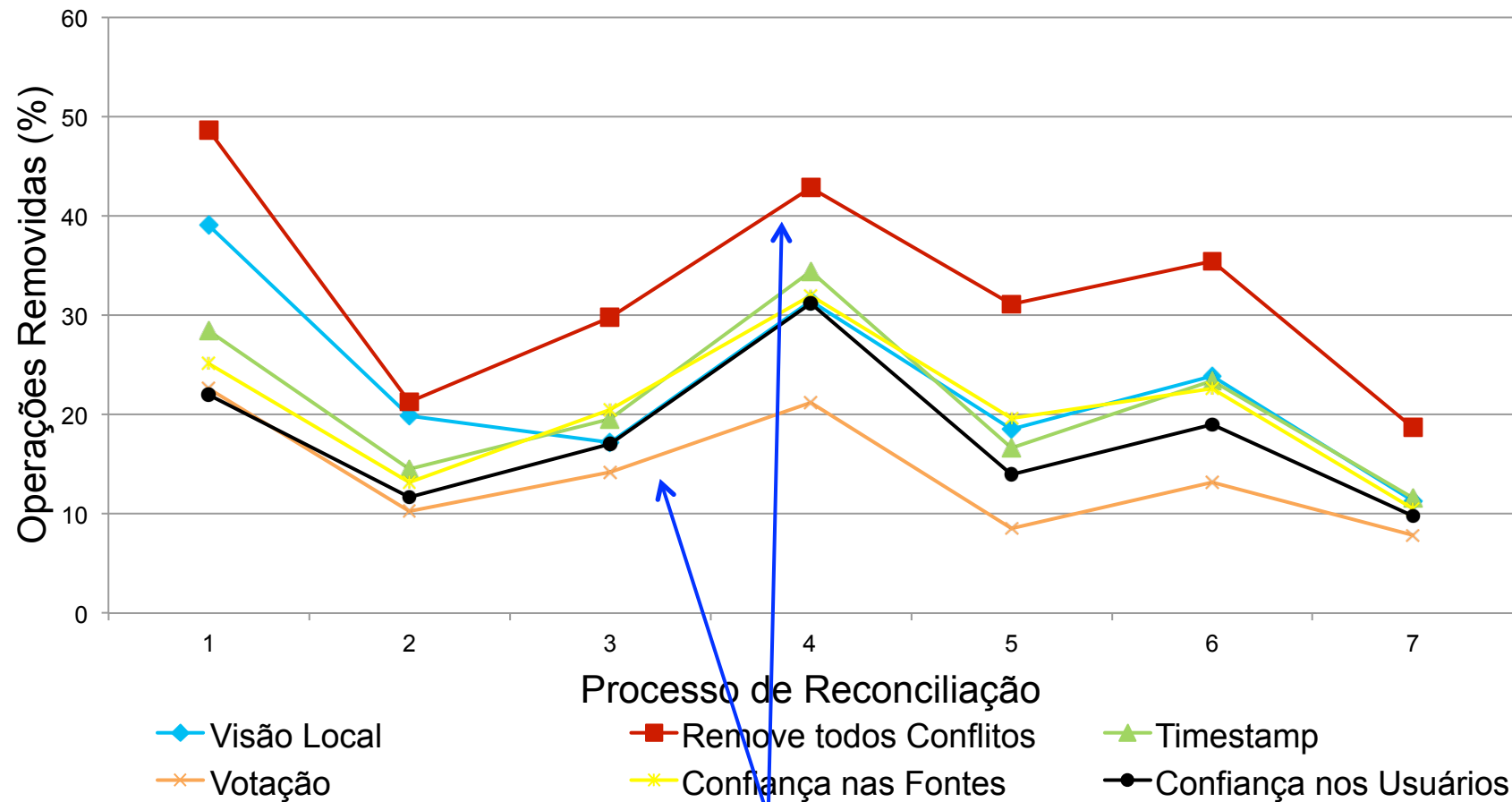
A política baseada na confiança nas fontes teve o tempo aumentado pelo processamento de todas as fontes de dados

Tempos de execução (3/3)



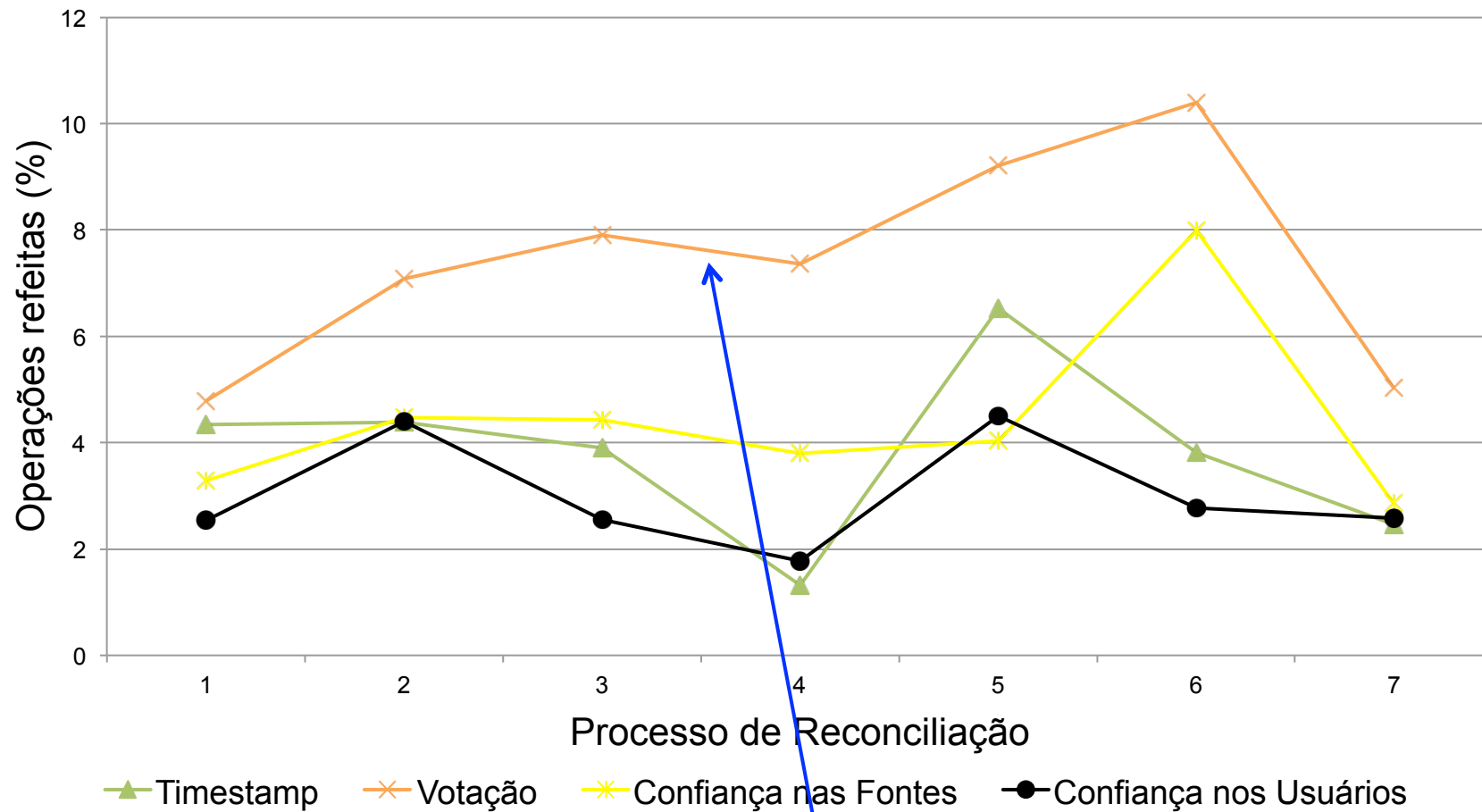
A política baseada na confiança nos usuários obteve os melhores resultados

Porcentagem de operações removidas (1/1)



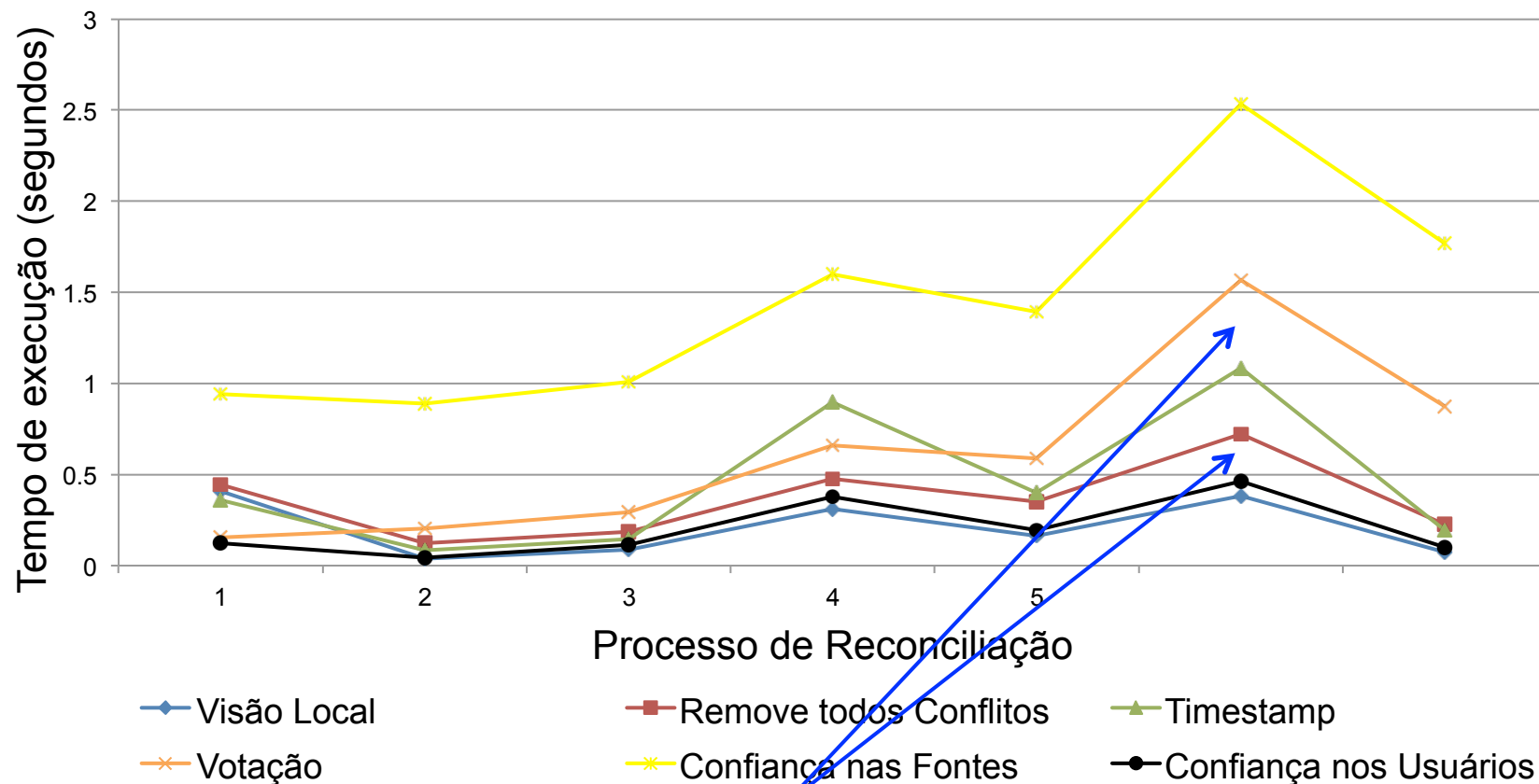
Menor porcentagem de remoções para a *política baseada em votação* e uma maior porcentagem de remoções para a *política que remove todos os conflitos*

Porcentagem de operações refeitas (1/1)



A política baseada em votação obteve os melhores resultados

Tempos de execução



Como nesse teste houve operações vencedoras na votação, a *política baseada em votação* apresentou maiores tempos de execução que a política que remove todos os conflitos

Conclusões

- *Política baseada em votação*

- Menor número de operações removidas
- Maior número de operações refeitas
- Menor tempo de execução

Melhor
classificação

- *Política baseada confiança nos usuários*

- Bons resultados em relação ao número de operações
 - Removidas e refeitas
 - Tempo de execução

Segunda melhor
classificação

Experimento baseado no usuário

Políticas de reconciliação

- 7 usuários da Ciência da Computação
 - Explicação das políticas
 - Exibição dos resultados gerados pelas políticas
 - Aplicação de questionário
 - Aceitabilidade
 - Percepção de consistência
 - Percepção de corretude
 - Percepção de risco
 - Economia de tempo e
 - Satisfação com o resultado
 - Obtenção do nível de concordância
 - Escala de 1 a 7

Experimento baseado no usuário

Políticas de reconciliação

Variável de medição	Política					
	Visão Local	Remove Conflitos	<i>Timestamp</i>	Votação	Conf. Fontes	Conf. Usuários
Aceitabilidade	0%	57,14%	42,86%	100%	71,43%	71,43%
Consistência	14,28%	42,86%	42,86%	100%	100%	85,71%
Corretude	14,29%	33,34%	14,28%	71,43%	100%	85,71%
Risco	28,57%	14,29%	71,43%	42,86%	57,14%	71,43%
Economia de tempo	71,43%	28,57%	57,14%	42,86%	85,71%	85,71%
Satisfação	85,71%	71,43%	28,57%	85,71%	85,71%	85,71%

Políticas mais democráticas

- Consideram operações mais colaborativas
- Garantem a satisfação mesmo quando ele não sabe como resolver um conflito

Experimento baseado no usuário

Políticas de reconciliação

Variável de medição	Política					
	Visão Local	Remove Conflitos	<i>Timestamp</i>	Votação	Conf. Fontes	Conf. Usuários
Aceitabilidade	0%	57,14%	42,86%	100%	71,43%	71,43%
Consistência	14,28%	42,86%	42,86%	100%	100%	85,71%
Corretude	14,29%	33,34%	14,28%	71,43%	100%	85,71%
Risco	28,57%	14,29%	71,43%	42,86%	57,14%	71,43%
Economia de tempo	71,43%	28,57%	57,14%	42,86%	85,71%	85,71%
Satisfação	85,71%	71,43%	28,57%	85,71%	85,71%	85,71%

Políticas com níveis intermediários de percepção

- Geram resultados incorretos ou
- Descartam muitas operações conflitantes

Experimento baseado no usuário

Políticas de reconciliação

Variável de medição	Política					
	Visão Local	Remove Conflitos	<i>Timestamp</i>	Votação	Conf. Fontes	Conf. Usuários
Aceitabilidade	0%	57,14%	42,86%	100%	71,43%	71,43%
Consistência	14,28%	42,86%	42,86%	100%	100%	85,71%
Corretude	14,29%	33,34%	14,28%	71,43%	100%	85,71%
Risco	28,57%	14,29%	71,43%	42,86%	57,14%	71,43%
Economia de tempo	71,43%	28,57%	57,14%	42,86%	85,71%	85,71%
Satisfação	85,71%	71,43%	28,57%	85,71%	85,71%	85,71%

Política que remove todos os conflitos

- Garante satisfação pois se não possível determinar a operação correta, melhor descartar
- Adequada para BDs que não podem conter dados incorretos

Experimento baseado no usuário

Políticas de reconciliação

Variável de medição	Política					
	Visão Local	Remove Conflitos	<i>Timestamp</i>	Votação	Conf. Fontes	Conf. Usuários
Aceitabilidade	0%	57,14%	42,86%	100%	71,43%	71,43%
Consistência	14,28%	42,86%	42,86%	100%	100%	85,71%
Corretude	14,29%	33,34%	14,28%	71,43%	100%	85,71%
Risco	28,57%	14,29%	71,43%	42,86%	57,14%	71,43%
Economia de tempo	71,43%	28,57%	57,14%	42,86%	85,71%	85,71%
Satisfação	85,71%	71,43%	28,57%	85,71%	85,71%	85,71%

Política baseada em timestamp

- Manter as operações mais recentes pode não ser a melhor opção
- Adequada para ambientes que envolvem notícias

Experimento baseado no usuário

Políticas de reconciliação

Variável de medição	Política					
	Visão Local	Remove Conflitos	<i>Timestamp</i>	Votação	Conf. Fontes	Conf. Usuários
Aceitabilidade	0%	57,14%	42,86%	100%	71,43%	71,43%
Consistência	14,28%	42,86%	42,86%	100%	100%	85,71%
Corretude	14,29%	33,34%	14,28%	71,43%	100%	85,71%
Risco	28,57%	14,29%	71,43%	42,86%	57,14%	71,43%
Economia de tempo	71,43%	28,57%	57,14%	42,86%	85,71%	85,71%
Satisfação	85,71%	71,43%	28,57%	85,71%	85,71%	85,71%

Política com baixo nível de percepção

- Não considera as decisões de integração de outros usuários
- Se as operações de U1 são confiáveis: economia de tempo e satisfação

Experimento baseado no usuário

Método de propagação

- 3 usuários realizaram uma operação de cópia, a partir de uma fonte considerada correta para duas outras fontes

Usuário	Tempo
Usuário 1	143s (02m:23s)
Usuário 2	31s (00m:31s)
Usuário 3	83s (01m:23s)
Método de propagação	0,078s

Método de propagação gastou substancialmente menos tempo

Experimento baseado no usuário

Método de propagação

- Usuários sobre 2 objetos presentes em 4 fontes
 - Objeto 1: presente em 3 fontes
 - Objeto 2: presentes em 2 fontes

Usuário	Número de operações	Tempo	Resultado
Usuário 1	28	650s (10m:50s)	Operações não atualizaram todas as fontes
Usuário 2	14	155s (02m:35s)	
Usuário 3	3	140s (02m:20s)	

Experimento baseado no usuário

Método de propagação

- Usuários sobre 2 objetos presentes em 4 fontes
 - Objeto 1: presente em 3 fontes
 - Objeto 2: presentes em 2 fontes

Usuário	Número de operações	Tempo	Resultado
Usuário 1	28	650s (10m:50s)	Faltavam operações
Método de propagação	25	1,451s	Todas as operações

Método de propagação gastou substancialmente menos tempo
Garantiu decisões consistentes para os objetos

Roteiro

- Motivação
- Modelo AcCORD
- Repositório de operações
- Políticas de reconciliação
- Método de propagação de decisões de integração
- Resultados experimentais
- **Conclusões**

Contribuições

- Modelo AcCORD
 - Possibilita o processo de reconciliação multiusuário colaborativo baseado em dados de procedência
- Políticas de reconciliação de dados
 - Garantem flexibilidade permitindo a geração de uma única visão consistente de várias visões distintas
- Método de propagação de decisões de integração
 - Garante um valor consistente para um item de dado em todas as fontes onde ele aparece
 - Diminui o tempo despendido no processo de integração

Análise dos resultados obtidos

- Modelo AcCORD
 - Flexibilidade
 - Eficiência (maior tempo gasto: 14s)
- Políticas de reconciliação
 - Eficácia, mesmo diante de um maior volume de operações
 - Maior aceitação dos usuários para as políticas de reconciliação que permitem maior colaboração
- Método de propagação de decisões de integração
 - Necessidade para evitar inconsistências entre as decisões tomadas
 - Eficiência para diminuir o tempo despendido no processo de integração