

# MultiPrint: um modelo de procedência para integração de dados com suporte para um ambiente multiusuário

---

Disciplina: Procedência de Dados e Data Warehousing

Aluna: Dayse Silveira de Almeida

Professora: Cristina D. A. Ciferri



# Roteiro

- Introdução
- Propostas preliminares para o Modelo MultiPrInt
  - Conceito de transação
  - Métodos para controle de concorrência
    - Repositório centralizado e repositórios distribuídos
  - Momento da integração
    - Cópia locais distribuídas, cópias locais centralizadas e única cópia local centralizada
- Sugestões da banca
- Próximos passos

# Ambiente Multiusuário

- O ambiente multiusuário exige que operações sejam tratadas em termos de transações
- Transações são conjuntos de operações de leitura e escrita que formam uma única unidade lógica de execução, acessando e atualizando diversos itens de dados
- Os principais métodos de controle de concorrência baseados em serialização são
  - Bloqueio, *timestamp*, validação e múltiplas versões

# Ambiente Multiusuário

- Bloqueio – itens de dados acessados de maneira mutuamente exclusiva
  - Ordem de execução de acordo com a solicitação do bloqueio
- *Timestamp* – fixo e exclusivo a cada transação no sistema antes do início da execução
  - Ordem de execução selecionada com antecedência
- Validação – transações executadas em fases: leitura, validação e escrita;
- Múltipla versão – não assegura a serialização abortando ou adiando transações
  - Armazena cópias antigas de cada item de dado
  - Escrita – cria uma nova versão
  - Leitura – seleciona uma versão para ser lida assegurando a serialização

# Roteiro

- Introdução
- Propostas preliminares para o Modelo MultiPrInt
  - Conceito de transação
  - Métodos para controle de concorrência
    - Repositório centralizado e repositórios distribuídos
  - Momento da integração
    - Cópia locais distribuídas, cópias locais centralizadas e única cópia local centralizada
- Sugestões da banca
- Próximos passos

# Propostas Preliminares para o Modelo MultiPrint

- Colaboração nos resultados
- Ambientes multiusuários
  - Definir transação no contexto do modelo
  - Reformular o formato das operações para identificação de transações
  - Definir conflitos entre transações
  - Resolução de conflitos gerados concorrentemente
  - Controle de concorrência sobre o repositório
  - Momento da integração de dados de usuários concorrentes
  - Garantir reaplicação entreprocessos

# Transação

- Ações permitidas ao usuário: inserção, remoção, cópia e edição
- Transação definida em termos de ação do usuário
- Cada ação é englobada em uma única transação
- Definição (transação)
  - Uma transação é uma sequência completa de operações de modificação nas fontes locais, induzidas por uma única ação do usuário, ou seja, uma ação de inserção, remoção, cópia ou edição

# Roteiro

- Introdução
- Propostas preliminares para o Modelo MultiPrInt
  - Conceito de transação
  - Métodos para controle de concorrência
    - Repositório centralizado e repositórios distribuídos
  - Momento da integração
    - Cópia locais distribuídas, cópias locais centralizadas e única cópia local centralizada
- Sugestões da banca
- Próximos passos



# Métodos para Controle de Concorrência

- Quanto ao armazenamento das transações concorrentes no repositório de operações, pretende-se investigar duas abordagens
  1. Repositório centralizado
  2. Repositórios distribuídos

# 1. Repositório Centralizado

1. Bloquear o repositório para escrita
2. Para manter o repositório consistente: *redo*
  - No momento da integração
3. VRT para validação completa e reaplicação

## 2. Repositórios Distribuídos

- Consistência dos repositórios **locais**: *redo*
- Repositórios buscados e validados na origem e destino
  - Individualmente e sequencialmente
- Para transações válidas: busca-se conflitos

# Conflitos Multiusuários

- Definição (operações conflitantes)
  1.  $a = (\text{origem}, \text{destino}, \text{idObjeto}, \text{atributo})$  e  $b = (\text{origem}, \text{destino}, \text{idObjeto}, \text{atributo})$ , ou
  2.  $a = (\text{destino}, \text{idObjeto}, \text{atributo})$  e  $b = (\text{origem}, \text{idObjeto}, \text{atributo})$ , ou
  3.  $a = (\text{origem}, \text{idObjeto}, \text{atributo})$  e  $b = (\text{destino}, \text{idObjeto}, \text{atributo})$

# Conflitos Multiusuários

- Definição (transações conflitantes)
  - Duas transações  $t$  e  $t'$  conflitam se quaisquer de suas operações conflitam

# Conflitos Multiusuários

Usuário	id	op	origem	destino	idObjeto	atributo	valorOrigem	valorDestino	timestamp
X1	1	cp	Bruno	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	15	13:41:47_27/03/2013
X2	1	cp	Bruno	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	15	11:22:33_27/03/2013
	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	11:25:22_27/03/2013
X3	1	cp	Daniel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	15	24	10:45:39_27/03/2013
X4	1	cp	Emanuel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	234	24	08:28:43_27/03/2013
X5	1	ed	null	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	22	15	09:45:57_27/03/2013
	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	22	20	09:50:02_27/03/2013
X6	1	cp	Daniel	Fábio	Artigo [Título=MultiPrInt: Multiuser...]	Página final	15	42	10:13:45_27/03/2013
X7	1	cp	Bruno	Gabriel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	12:24:55_27/03/2013

Figura17: Transações conflitantes em nível multiusuário

## 2. Repositórios Distribuídos

- Transações não conflitantes são reaplicadas
- Para o conjunto de transações conflitantes:
  - Reaplicar a mais recente
  - Detectar transitivas a essa e reaplicar
- Transações conflitantes não escolhidas para reaplicação
  - I. Remoção do repositório
  - II. Tratamento

## 2. Repositórios Distribuídos

- I. Remoção de transações conflitantes
  - Ignorar X ignorar e remover



## 2. Repositórios Distribuídos

### II. Tratamento de transações conflitantes

- Transações conflitantes não escolhidas são consideradas sobrepostas
  - Removidas ou refeitas de acordo com o tipo de sobreposição

## 2. Repositórios Distribuídos

### II. Tratamento de transações conflitantes

- Se a sobreposição detectada for do destino
  - A transação  $t'$  deve ser removida de seu repositório
  - Suas transitivas são refeitas de acordo com o valor de  $t$ , e reaplicadas

## 2. Repositórios Distribuídos

Usuário	id	op	origem	destino	idObjeto	atributo	valorOrigem	valorDestino	timestamp
X1	1	cp	Bruno	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	15	13:41:47_27/03/2013
X2	1	cp	Bruno	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	15	11:22:33_27/03/2013
	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	11:25:22_27/03/2013
X3	1	cp	Daniel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	15	24	10:45:39_27/03/2013
X4	1	cp	Emanuel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	234	24	08:28:43_27/03/2013
X5	1	ed	null	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	22	15	09:45:57_27/03/2013
	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	22	20	09:50:02_27/03/2013
X6	1	cp	Daniel	Fábio	Artigo [Título=MultiPrInt: Multiuser...]	Página final	15	42	10:13:45_27/03/2013
X7	1	cp	Bruno	Gabriel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	12:24:55_27/03/2013

Figura 18: Sobreposição do destino

## 2. Repositórios Distribuídos

Usuário	id	op	origem	destino	idObjeto	atributo	valorOrigem	valorDestino	timestamp
X1	1	cp	Bruno	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	15	13:41:47_27/03/2013
X2	1	cp	Bruno	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	15	11:22:33_27/03/2013
	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	11:25:22_27/03/2013
X3	1	cp	Daniel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	15	24	10:45:39_27/03/2013
X4	1	cp	Emanuel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	234	24	08:28:43_27/03/2013
X5	1	ed	null	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	22	15	09:45:57_27/03/2013
	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	09:50:02_27/03/2013
X6	1	cp	Daniel	Fábio	Artigo [Título=MultiPrInt: Multiuser...]	Página final	15	42	10:13:45_27/03/2013
X7	1	cp	Bruno	Gabriel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	12:24:55_27/03/2013

Figura 18: Sobreposição do destino

## 2. Repositórios Distribuídos

### II. Tratamento de transações conflitantes

- Se a sobreposição detectada for da origem
  - A transação  $t'$  e suas transitivas são refeitas de acordo com o valor de  $t$ , e reaplicadas

## 2. Repositórios Distribuídos

Usuário	id	op	origem	destino	idObjeto	atributo	valorOrigem	valorDestino	timestamp
X1	1	cp	Bruno	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	15	13:41:47_27/03/2013
X2	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	11:25:22_27/03/2013
X3	1	cp	Daniel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	15	24	10:45:39_27/03/2013
X4	1	cp	Emanuel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	234	24	08:28:43_27/03/2013
X5	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	09:50:02_27/03/2013
X6	1	cp	Daniel	Fábio	Artigo [Título=MultiPrInt: Multiuser...]	Página final	15	42	10:13:45_27/03/2013
X7	1	cp	Bruno	Gabriel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	12:24:55_27/03/2013

Figura 18: Sobreposição da origem

## 2. Repositórios Distribuídos

Usuário	id	op	origem	destino	idObjeto	atributo	valorOrigem	valorDestino	timestamp
X1	1	cp	Bruno	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	15	13:41:47_27/03/2013
X2	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	11:25:22_27/03/2013
X3	1	cp	Daniel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	24	10:45:39_27/03/2013
X4	1	cp	Emanuel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	234	24	08:28:43_27/03/2013
X5	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	09:50:02_27/03/2013
X6	1	cp	Daniel	Fábio	Artigo [Título=MultiPrInt: Multiuser...]	Página final	15	42	10:13:45_27/03/2013
X7	1	cp	Bruno	Gabriel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	12:24:55_27/03/2013

Figura 18: Sobreposição da origem

## 2. Repositórios Distribuídos

Usuário	id	op	origem	destino	idObjeto	atributo	valorOrigem	valorDestino	timestamp
X1	1	cp	Bruno	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	15	13:41:47_27/03/2013
X2	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	11:25:22_27/03/2013
X3	1	cp	Daniel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	24	10:45:39_27/03/2013
X4	1	cp	Emanuel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	234	24	08:28:43_27/03/2013
X5	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	09:50:02_27/03/2013
X6	1	cp	Daniel	Fábio	Artigo [Título=MultiPrInt: Multiuser...]	Página final	15	42	10:13:45_27/03/2013
X7	1	cp	Bruno	Gabriel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	12:24:55_27/03/2013

Figura 18: Sobreposição da origem



## 2. Repositórios Distribuídos

Usuário	id	op	origem	destino	idObjeto	atributo	valorOrigem	valorDestino	timestamp
X1	1	cp	Bruno	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	15	13:41:47_27/03/2013
X2	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	11:25:22_27/03/2013
X3	1	cp	Daniel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	24	10:45:39_27/03/2013
X4	1	cp	Emanuel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	234	24	08:28:43_27/03/2013
X5	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	09:50:02_27/03/2013
X6	1	cp	Daniel	Fábio	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	42	10:13:45_27/03/2013
X7	1	cp	Bruno	Gabriel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	12:24:55_27/03/2013

Figura 18: Sobreposição da origem

## 2. Repositórios Distribuídos

### II. Tratamento de transações conflitantes

- Se o destino na transação sobreposta  $t'$  for igual à origem na transação escolhida como correta  $t$ 
  - A transação  $t'$  e suas transitivas devem ser removidas do seu repositório

## 2. Repositórios Distribuídos

Usuário	id	op	origem	destino	idObjeto	atributo	valorOrigem	valorDestino	timestamp
X1	1	cp	Bruno	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	15	13:41:47_27/03/2013
X2	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	11:25:22_27/03/2013
X3	1	cp	Daniel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	24	10:45:39_27/03/2013
X4	1	cp	Emanuel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	234	24	08:28:43_27/03/2013
X5	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	09:50:02_27/03/2013
X6	1	cp	Daniel	Fábio	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	42	10:13:45_27/03/2013
X7	1	cp	Bruno	Gabriel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	12:24:55_27/03/2013

Figura 18: Destino de  $t'$  é igual à origem de  $t$

## 2. Repositórios Distribuídos

### II. Tratamento de transações conflitantes

- Se nenhum dos casos anteriores ocorre
  - A origem na transação  $t'$  é igual à origem na transação escolhida como correta  $t$
  - A transação não é conflitante **com**  $t$  e deve ser reaplicada

## 2. Repositórios Distribuídos

Usuário	id	op	origem	destino	idObjeto	atributo	valorOrigem	valorDestino	timestamp
X1	1	cp	Bruno	Daniel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	15	13:41:47_27/03/2013
X2	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	11:25:22_27/03/2013
X3	1	cp	Daniel	Bruno	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	24	10:45:39_27/03/2013
X5	2	cp	Daniel	Carlos	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	09:50:02_27/03/2013
X6	1	cp	Daniel	Fábio	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	42	10:13:45_27/03/2013
X7	1	cp	Bruno	Gabriel	Artigo [Título=MultiPrInt: Multiuser...]	Página final	24	20	12:24:55_27/03/2013

Figura 18: Origem de t' é igual à origem de t

# Repositório Centralizado e Repositórios Distribuídos

- *Timestamp* não garante a escolha da transação correta
- Transações conflitantes em nível multiusuário
  
- Propostas
  - Remover todas as transações conflitantes
    - Print: garantir a validade das operações e reprodução apenas de operações válidas
  - Recomendação ao usuário no instante da reconciliação
    - Desafio: identificar usuários
    - Para conflitos de um único usuário: política *redo*
    - Para conflitos multiusuários: recomendação
  - Investigação inviável para repositórios distribuídos

# Roteiro

- Propostas preliminares para o Modelo MultiPrInt
  - Conceito de transação
  - Métodos para controle de concorrência
    - Repositório centralizado e repositórios distribuídos
  - Momento da integração
    - Cópia locais distribuídas, cópias locais centralizadas e única cópia local centralizada
- Sugestões da banca
- Próximos passos

# Momento da Integração

- Duas abordagens
  1. Processo definido pelo usuário e cópias locais distribuídas;
    - Uma cópia local para cada usuário
  2. Processo automático e cópia local centralizada
    - Uma cópia local para todos os usuários



# 1. Processo definido pelo usuário, cópias locais distribuídas

- Para repositório distribuído
  - Usuário decide o momento da integração
    - Antes do próximo processo de reconciliação
  - Repositórios de todos os usuários são buscados, integrados e reaplicados

# 1. Processo definido pelo usuário, cópias locais distribuídas

- Para repositório distribuído

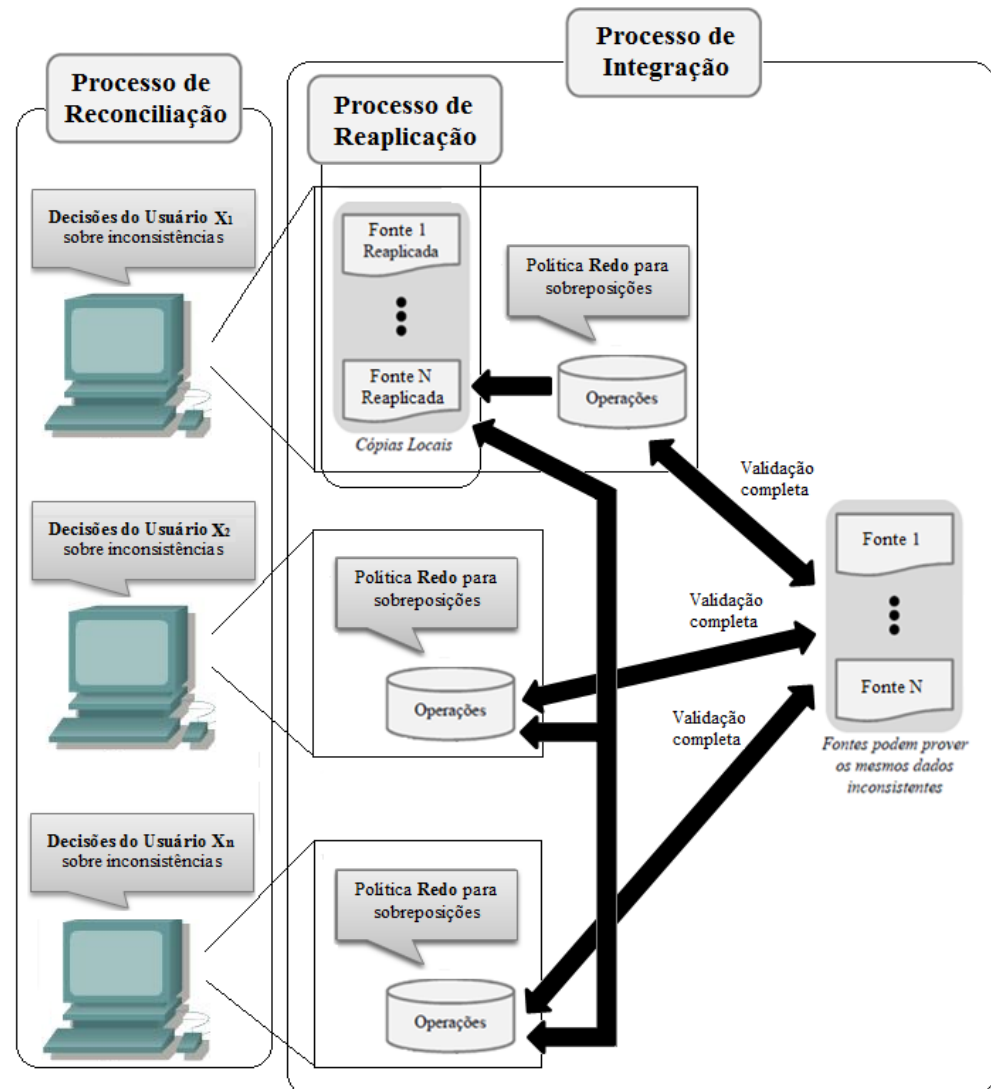


Figura 19: MultiPrint com repositório e cópias locais distribuídas

# 1. Processo definido pelo usuário, cópias locais distribuídas

- Para repositório centralizado
  - Usuário decide o momento da integração
    - Antes do próximo processo de reconciliação
  - Reaplicar até o último processo de reconciliação finalizado

# 1. Processo definido pelo usuário, cópias locais distribuídas

- Para repositório centralizado

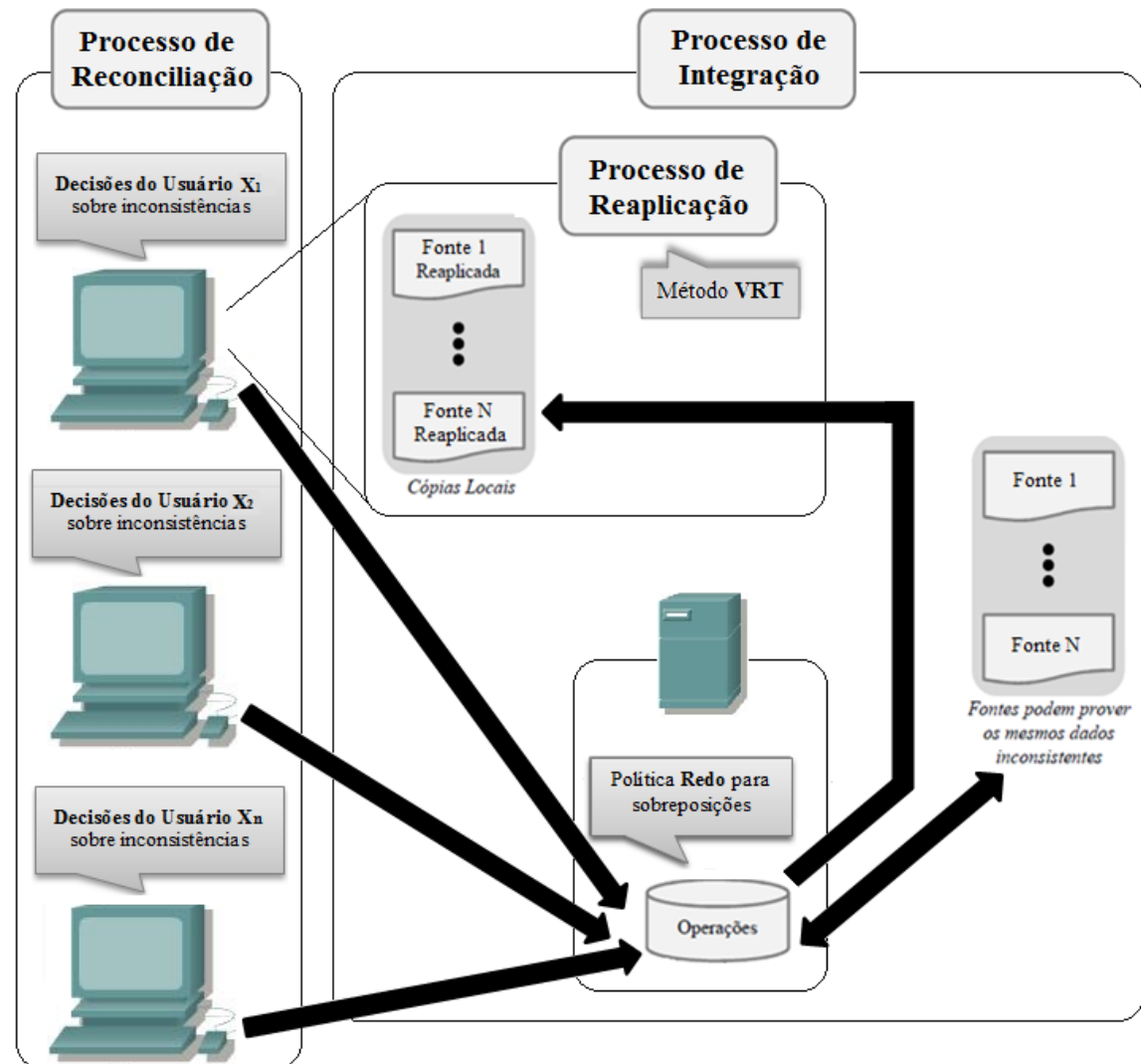


Figura 20: MultiPrint com repositório centralizado e cópias distribuídas

## 1.5. Processo definido pelo usuário, cópias locais centralizadas

- Máquina remota
  - Repositório centralizado
  - Cópias locais centralizadas
    - Uma para cada usuário

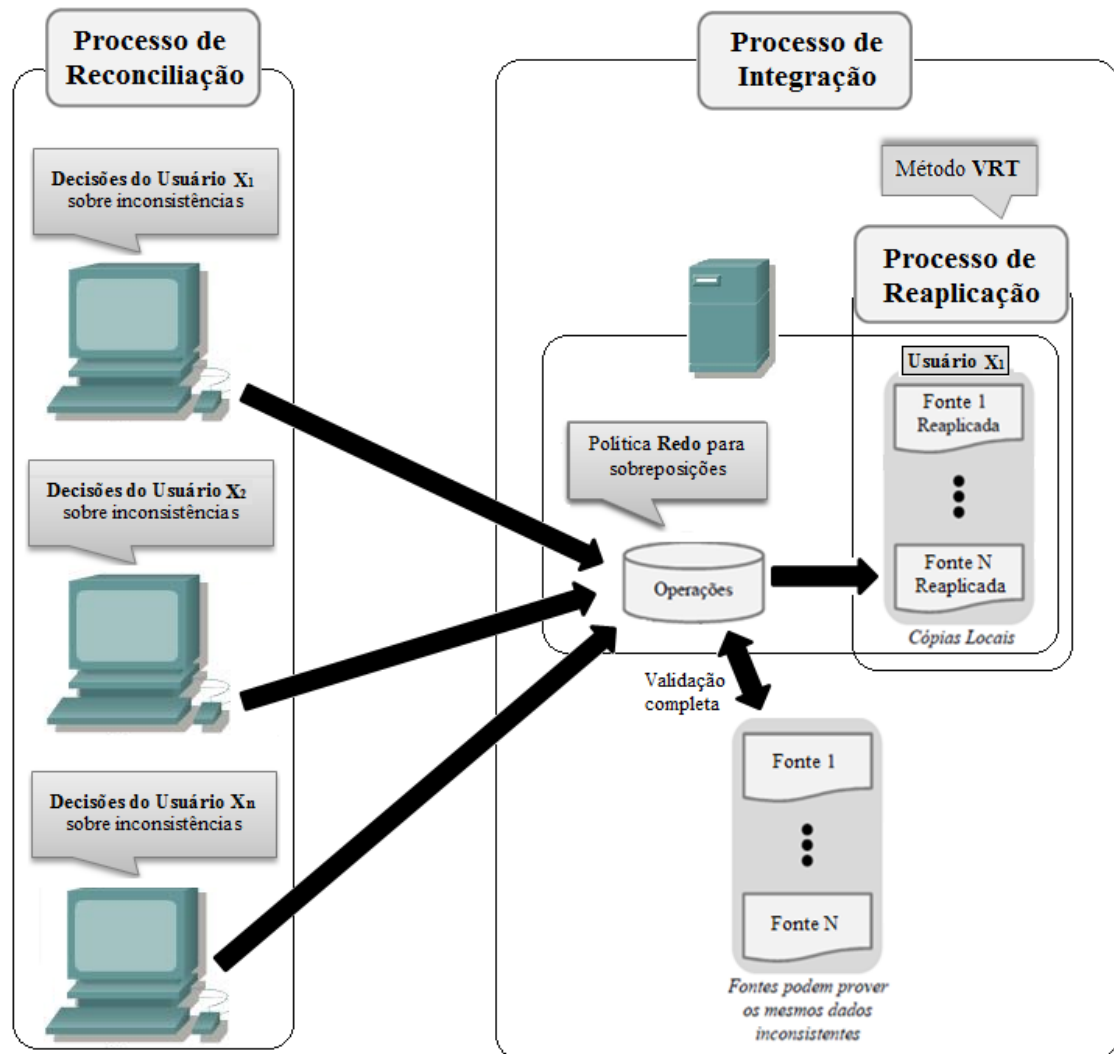


Figura 21: MultiPrint com repositório centralizado e cópias locais centralizadas

## 2. Processo definido automaticamente, cópia local centralizadas

- Momento pré-definido
- Verificação/notificação do estado de todos os processos de reconciliação
  - Antes do início de novos processos de reconciliação de todos os usuários
- Para repositório centralizado
  - Validar e reaplicar usando VRT
- Para repositório distribuído
  - Coletar todos os repositórios, integrar e reaplicar (utilizando os algoritmos mostrados)

# Roteiro

- Propostas preliminares para o Modelo MultiPrInt
  - Conceito de transação
  - Métodos para controle de concorrência
    - Repositório centralizado e repositórios distribuídos
  - Momento da integração
    - Cópia locais distribuídas, cópias locais centralizadas e única cópia local centralizada
- Sugestões da banca
- Próximos passos

# Sugestões da Banca

- Redefinir a transação
  - Transações longas
  - Usuário definir o tamanho da transação – *checkout* do versionamento
- Definição atual de transação interfere na propagação intraprocasso
  - Autor: *Ciferri, C. D. A para Ciferri, Cristina Dutra Aguiar*
  - Propagar para as demais instâncias
  - Uma transação para cada operação de edição
  - Transações diante de conflitos



# Transações para Propagação Intraprocesso

- begin transaction copia2
- **operacao cp** Objeto=Artigo[Título= **MultiPrint: Multiuser...**] Autor[Nome= *Ciferri, C. D. A*] Origem= Ana Destino= Bruno Atributo= nome valorOrigem= **Ciferri, C. D. A** valorDestino= **Ciferri, Cristina Dutra Aguiar**
- operacao **ed** Objeto=Artigo[Título= **X**] Autor[Nome= *Ciferri, C. D. A*] **Origem=null Destino=Bruno** Atributo=nome valorOrigem= *Ciferri, C. D. A* valorDestino= *Ciferri, Cristina Dutra Aguiar*
- operacao **ed** Objeto=Artigo[Título= **Y**] Autor[Nome= *Ciferri, C. D. A*] **Origem=null Destino=Bruno** Atributo=nome valorOrigem= *Ciferri, C. D. A* valorDestino= *Ciferri, Cristina Dutra Aguiar*
- end transaction copia2
  
- begin transaction edicao2
- **operacao ed** Objeto=Artigo[Título= **MultiPrint: Multiuser...**] Autor[Nome= *Ciferri, C. D. A*] **Origem= null Destino= Bruno** Atributo= nome valorOrigem= **Ciferri, C. D. A** valorDestino= **Ciferri, Cristina Dutra Aguiar**
- operacao **ed** Objeto=Artigo[Título= **X**] Autor[Nome= *Ciferri, C. D. A*] **Origem=null Destino=Bruno** Atributo=nome valorOrigem= *Ciferri, C. D. A* valorDestino= *Ciferri, Cristina Dutra Aguiar*
- operacao **ed** Objeto=Artigo[Título= **Y**] Autor[Nome= *Ciferri, C. D. A*] **Origem=null Destino=Bruno** Atributo=nome valorOrigem= *Ciferri, C. D. A* valorDestino= *Ciferri, Cristina Dutra Aguiar*
- end transaction copia2

# Transação

- Definição (transação)
  - Uma transação é uma sequência completa de operações de modificação nas fontes locais, induzidas por uma única ação do usuário, ou seja, uma ação de inserção, remoção, cópia ou edição

# Sugestões da Banca

- Bloqueio do repositório centralizado
  - Criação de um índice
- Resolução de conflitos
  - Consenso quanto à transação escolhida para reaplicação – algoritmos baseados em consenso
  - Reputação
  - Votação – análise dos valores de todas as transações
  - Heurística – suposição de validade

# Próximos Passos

- Repositório centralizado
  - Remover transações conflitantes
  - Usar *timestamp*
    - Selecionar transação como maior *timestamp*, remover as demais transações conflitantes
    - Selecionar transação como maior *timestamp*, tratar as demais transações conflitantes
  - Confiança – resolver conflitos
- Repositórios distribuídos

# Referências

- [Tomazela et al. 2013] Tomazela, B., Hara, C. S., Ciferri, R. R., Ciferri, C. D. A. (2013). Empowering Integration Processes with Data Provenance. *Data & Knowledge Engineering*. Aceito para publicação.

# MultiPrint: um modelo de procedência para integração de dados com suporte para um multiusuário

---

Disciplina: Procedência de Dados e Data Warehousing

Aluna: Dayse Silveira de Almeida

Professora: Cristina D. A. Ciferri

Obrigada.

