



Mark Douglas de Azevedo Jacyntho

Um Modelo de Bloqueio Multigranular para RDF

Tese de Doutorado

Tese apresentada como requisito parcial para obtenção do título de Doutor pelo Programa de Pós-Graduação em Informática da PUC-Rio.

Orientador: Prof. Daniel Schwabe

Rio de Janeiro
Fevereiro de 2012



Mark Douglas de Azevedo Jacyntho

Um Modelo de Bloqueio Multigranular para RDF

Tese apresentada como requisito parcial para obtenção do título de Doutor pelo Programa de Pós-Graduação em Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Daniel Schwabe

Orientador

Depto. de Informática - PUC - Rio

Prof.^a Adriana Pereira de Medeiros

Depto. de Ciência e Tecnologia - UFF

Prof. Arndt von Staa

Depto. de Informática - PUC-Rio

Prof. Marco Antonio Casanova

Depto. de Informática - PUC-Rio

Prof.^a Maria Luiza Machado Campos

Depto. de Ciência da Computação - UFRJ

Prof. José Eugenio Leal

Coordenador Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 02 de Fevereiro de 2012

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Mark Douglas de Azevedo Jacyntho

Graduou-se em Engenharia de Computação, em 1998, pela PUC-Rio, tendo recebido três certificados de excelente desempenho acadêmico. Obteve o título de Mestre em Informática, em 2001, pelo Departamento de Informática da PUC-Rio. Tem experiência profissional em Análise e Design de Sistemas, em especial sob o paradigma de orientação a objetos. É professor da Universidade Cândido Mendes (UCAM-Campos/RJ) e do Instituto Federal Fluminense (IFF) - Campus Campos-Centro.

Ficha Catalográfica

Jacyntho, Mark Douglas de Azevedo

Um modelo de bloqueio multigranular para RDF / Mark Douglas de Azevedo Jacyntho ; orientador: Daniel Schwabe. – 2012.

277 f. : il. (color.) ; 30 cm

Tese (doutorado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2012.

Inclui bibliografia

1. Informática – Teses. 2. Web Semântica. 3. RDF. 4. Transação. 5. Concorrência. 6. Isolamento. 7. Bloqueio. I. Schwabe, Daniel II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Ao meu primeiro e maior mestre, o grande mentor deste trabalho, sem o qual eu nada seria, que onde quer que esteja, tenho plena convicção, foi meu maior orientador. Meu pai, JORGE B. JACYNTHO "in memoriam". E também à minha querida mãe, MARIA IGNEZ, por todo amor e carinho que sempre me deu.

Agradecimentos

A *Deus*, porque sem Ele não seria possível dar nem o primeiro passo.

Ao meu pai, *Jorge Jacyntho* "in memoriam", por tudo que fez por mim em vida e certamente ainda continua fazendo. Eu somente cheguei até aqui graças a Ele.

À minha querida mãe, *Maria Ignez*, por todo amor e carinho que nunca deixou faltar.

À minha noiva, *Lara Romano*, por toda compreensão, paciência, incentivo e carinho que sempre me deu nos momentos mais difíceis.

Ao meu orientador, *Daniel Schwabe*, pelos ensinamentos, atenção, por ter confiado em mim, pela amizade e por todo apoio que me deu.

À minha família, em geral, pelo incentivo e apoio que sempre me deu.

Aos professores do Departamento de Informática da PUC-Rio, pelo conhecimento transmitido.

Aos funcionários do Departamento de Informática, pela atenção e orientações fornecidas.

À minha amiga *Danielle Loyola* pelo incentivo e momentos de descontração que foram verdadeiras fontes de energia para dar seguimento ao trabalho.

Aos meus amigos do Instituto Federal Fluminense, *Fernando, Philippe, Francisco, Rogério e Vinícius* que sempre estiveram prontamente dispostos a ajudar-me, em minhas atribuições, para que tivesse uma dedicação mais continuada ao trabalho.

À coordenadora do curso de Ciência da Computação *Giselle Almeida* e ao diretor da Universidade Cândido Mendes (Campos-RJ) *Dr. Luiz*, por terem compreendido o momento por que passava, permitindo-me concluir esta tese.

Ao CNPq, pelo apoio financeiro fornecido durante o curso.

Resumo

Jacyntho, Mark Douglas de Azevedo; Schwabe, Daniel (Orientador). **Um Modelo de Bloqueio Multigranular para RDF**. Rio de Janeiro, 2012. 277p. Tese de Doutorado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Em aplicações Web, sessões cliente são organizadas em transações, envolvendo requisições que leem e atualizam dados compartilhados. Executando concorrentemente, estas sessões podem invalidar os dados umas das outras. O advento do movimento *Linked Data* vem estimulando sobremaneira a criação de aplicações que empregam o modelo de dados RDF em sua camada de informação. Além de consultas aos dados RDF, também se faz presente o requisito de atualizações *online*, com adequado controle de concorrência. Este trabalho apresenta um modelo de bloqueio, inspirado no protocolo de bloqueio multigranular, voltado para isolamento entre transações que manipulam dados RDF, considerando quatro problemas de concorrência, a saber: *lost updates*, *dirty reads*, *non-repeatable read* e *phantoms reads*. São oferecidos quatro grânulos "hierarquicamente" relacionados, bem como novos tipos de bloqueio de escrita e leitura, especificamente criados para o modelo RDF. Por fim, o desempenho do modelo de bloqueio proposto é avaliado por meio de simulação.

Palavras-chave

Web Semântica; RDF; Transação; Concorrência; Isolamento; Bloqueio.

Abstract

Jacyntho, Mark Douglas de Azevedo; Schwabe, Daniel (Advisor). **A Multigranularity Locking Model for RDF**. Rio de Janeiro, 2012. 277p. DSc Thesis - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Client sessions in Web applications are organized as transactions involving requests that read and write shared data. Executing concurrently, these sessions may invalidate each other's data. The advent of Linked Data is spurring the deployment of applications that use the RDF data model at the information tier. In addition to querying RDF data, there is also the requirement for online updates with suitable concurrency control. This work presents a *locking* model, inspired by the multigranularity *locking* protocol, to address isolation between transactions that manipulate RDF data, considering four concurrency-related issues, namely: *lost updates*, *dirty reads*, *non-repeatable reads* and *phantom reads*. Four "hierarchically" related granules are offered, as well as new read and write *lock* modes, specifically created for the RDF data model. Finally, the performance assessment of the proposed *locking* model is done through simulation.

Keywords

Semantic Web; RDF; Transaction; Concurrency; Isolation; *Lock*; *Locking*.

Sumário

1	Introdução	17
1.1.	Objetivos	18
1.2.	Contribuições	19
1.3.	Trabalhos Relacionados	20
1.3.1.	Transações em RDF	20
1.3.2.	Operações RDF	21
1.3.3.	Transações em Aplicações Web	21
1.4.	Estrutura do Documento	24
2	Fundamentos	25
2.1.	Gerenciamento de Transações	25
2.1.1.	Problemas Comuns de Concorrência	26
2.1.2.	Bloqueio	30
2.2.	Web Semântica	44
2.2.1.	Definição e Requisitos	44
2.2.2.	Modelo RDF	47
2.2.3.	Vocabulários e Ontologias	52
2.2.4.	RDF Schema e OWL	54
2.2.5.	SPARQL	59
2.2.6.	<i>Linked Data</i> e a Web de Dados	65
2.2.7.	Gerenciamento de Transação em <i>Triple Stores</i>	68
3	Modelo de Bloqueio Multigranular para RDF	71
3.1.	Grânulos	72
3.2.	Tipos de Bloqueio	76
3.3.	Protocolo de Bloqueio Multigranular	79
3.4.	Corretude do Protocolo	82
3.5.	Conversão de Bloqueios	93
3.6.	Bloqueio Otimista	99
3.6.1.	Bloqueios de Leitura para Inserção e Remoção	100
3.6.2.	Grânulos	101

4 Implementação	107
4.1. <i>Lock Manager</i>	107
4.1.1. Design	111
4.2. DSL e Vocabulário de Apoio	115
4.2.1. DSL para <i>Workflow</i> de Transações Web	115
4.2.2. Vocabulário para Aquisição de Bloqueios	125
4.3. Avaliação de Desempenho	143
4.4. Diretrizes de Utilização	152
4.4.1. Bloqueio Otimista versus Bloqueio Pessimista	152
4.4.2. Tipos de Bloqueio	152
4.4.3. Grânulos	153
5 Conclusão	155
5.1. Contribuições	155
5.2. Trabalhos Futuros	156
6 Referências	159
Apêndice	163
A.1. Simulação - Estratégia e Resultados	163
A.1.1. Design da Simulação	163
A.1.2. Ambiente da Simulação	170
A.1.3. Resultados da Simulação	171
A.2. Código Fonte do <i>Lock Manager</i>	215
A.3. Vocabulário para Aquisição de Bloqueios	231
A.4. Código Fonte da DSL Interna para <i>Workflow</i> de Transações Web	236
A.5. Código Fonte da Simulação	254

Lista de figuras

Figura 1 - <i>Lost Update</i> : Mr. Smith sem assento [Stroka, 2009]	27
Figura 2 - <i>Dirty Read</i> : Erro no saldo da conta bancária [Stroka, 2009]	28
Figura 3 - <i>Non-repeatable read</i> : Valor total incorreto [Stroka, 2009]	29
Figura 4 - <i>Phantom reads</i> : Mesma consulta, dois resultados [Stroka, 2009]	30
Figura 5 - Exemplo de um <i>Lock Type Graph</i> [Bernstein et al., 1987]	34
Figura 6 - Exemplo de um <i>Lock Instance Graph</i> [Bernstein et al., 1987]	34
Figura 7 - Matriz de compatibilidade do MGL [Bernstein et al., 1987]	35
Figura 8 - Exemplo de múltiplas versões [Bernstein e Newcomer, 2009]	38
Figura 9 - <i>Lock Thrashing</i> [Bernstein e Newcomer, 2009]	39
Figura 10 - Modelo matemático de transações [Bernstein e Newcomer, 2009]	40
Figura 11 - Curva geral de granularidade [Bernstein et al., 1987]	41
Figura 12 - Curva de granularidade - transações longas [Bernstein et al., 1987]	41
Figura 13 - Curva de granularidade - transações curtas [Bernstein et al., 1987]	42
Figura 14 - Um <i>statement</i> RDF	48
Figura 15 - Um <i>statement</i> RDF realista [MANOLA e MILLER, 2004]	50
Figura 16 - <i>Statements</i> sobre um mesmo recurso [MANOLA e MILLER, 2004]	50
Figura 17 - Notação em triplas [MANOLA e MILLER, 2004]	51
Figura 18 - Notação em triplas abreviada [MANOLA e MILLER, 2004]	51
Figura 19 - Uso de "nó em branco" [MANOLA e MILLER, 2004]	51
Figura 20 - Exemplo de serialização em RDF/XML [MANOLA e MILLER, 2004]	52
Figura 21 - Hierarquia de classes RDFS (gráfico) [MANOLA e MILLER, 2004]	56
Figura 22 - Hierarquia de classe RDFS (triplas) [MANOLA e MILLER, 2004]	56
Figura 23 - Vocabulário de veículos em RDFS [MANOLA e MILLER, 2004]	57
Figura 24 - SPARQL SELECT (1) [Prud'hommeaux e Seaborne, 2008]	63
Figura 25 - SPARQL SELECT (2) [Prud'hommeaux e Seaborne, 2008]	64
Figura 26 - LOD cloud diagram em Setembro de 2011 [LOD Project]	68
Figura 27 - <i>Rooted DAG Lock Type Graph</i>	76
Figura 28 - Estruturas de dados do <i>Lock Manager</i>	108
Figura 29 - Modelo de classes do <i>Lock Manager</i>	111
Figura 30 - Exemplo de instanciação do <i>Lock Manager</i>	114
Figura 31 - Metamodelo da DSL para <i>workflow</i> de transações Web	118
Figura 32 - <i>TransactionalActionTriggerDecorator</i>	121
Figura 33 - Transação Web "Simulação Lâmpada" reificada	122

Figura 34 - Statechart da transação Web "Simulação Lâmpada"	123
Figura 35 - <i>OWL Object properties</i> correspondentes aos tipos de bloqueio	126
Figura 36 - Ontologia OWL para o domínio de conferências	127
Figura 37 - Transação de negócio " <i>Document Analysis</i> " adaptado de [Jacyntho e Schwabe, 2010a]	130
Figura 38 - <i>Turnaround</i> (seg) - Monogranular, Tamanho 0.1%, 80% de Escrita	145
Figura 39 - <i>Turnaround</i> (seg) - Monogranular, Tamanho 20%, 80% de Escrita	146
Figura 40 - <i>Turnaround</i> (seg) - Monogranular versus Multigranular, 80% de Escrita	147
Figura 41 - Número de abortos - Monogranular versus Multigranular, 80% de Escrita	148
Figura 42 - Número de abortos - Tipos de Bloqueio, Tamanho 1%, 20% de Escrita	150
Figura 43 - <i>Turnaround</i> (seg) - Tipos de Bloqueio, Tamanho 1%, 20% de Escrita	150
Figura 44 - Número de Abortos - Tipos de Bloqueio, Tamanho 10%, 20% de Escrita	151
Figura 45 - <i>Turnaround</i> (seg) - Tipos de Bloqueio, Tamanho 10%, 20% de Escrita	151
Figura 46 - Modelo de classes da simulação	164
Figura 47 - Cenário 1 - número de bloqueios	172
Figura 48 - Cenário 1 - número de abortos	173
Figura 49 - Cenário 1 - <i>turnaround</i> (seg)	174
Figura 50 - Cenário 2 - número de bloqueios	175
Figura 51 - Cenário 2 - número de abortos	176
Figura 52 - Cenário 2 - <i>turnaround</i> (seg)	177
Figura 53 - Cenário 3 - número de bloqueios	178
Figura 54 - Cenário 3 - número de abortos	179
Figura 55 - Cenário 3 - <i>turnaround</i> (seg)	180
Figura 56 - Cenário 4 - número de bloqueios	181
Figura 57 - Cenário 4 - número de abortos	182
Figura 58 - Cenário 4 - <i>turnaround</i> (seg)	183
Figura 59 - Cenário 5 - número de bloqueios	184
Figura 60 - Cenário 5 - número de abortos	185

Figura 61 - Cenário 5 - <i>turnaround</i> (seg)	186
Figura 62 - Cenário 6 - número de bloqueios	187
Figura 63 - Cenário 6 - número de abortos	188
Figura 64 - Cenário 6 - <i>turnaround</i> (seg)	189
Figura 65 - Cenário 7 - número de bloqueios	191
Figura 66 - Cenário 7 - número de abortos	193
Figura 67 - Cenário 7 - <i>turnaround</i> (seg)	194
Figura 68 - Cenário 8 - número de bloqueios	196
Figura 69 - Cenário 8 - número de abortos	198
Figura 70 - Cenário 8 - <i>turnaround</i> (seg)	199
Figura 71 - Cenário 9 - número de abortos	200
Figura 72 - Cenário 9 - <i>turnaround</i> (seg)	201
Figura 73 - Cenário 10 - número de abortos	202
Figura 74 - Cenário 10 - <i>turnaround</i> (seg)	203
Figura 75 - Cenário 11 - número de abortos	204
Figura 76 - Cenário 11 - <i>turnaround</i> (seg)	205
Figura 77 - Cenário 12 - número de abortos	206
Figura 78 - Cenário 12 - <i>turnaround</i> (seg)	207
Figura 79 - Cenário 13 - número de abortos	208
Figura 80 - Cenário 13 - <i>turnaround</i> (seg)	209
Figura 81 - Cenário 14 - número de abortos	210
Figura 82 - Cenário 14 - <i>turnaround</i> (seg)	211
Figura 83 - Cenário 15 - número de bloqueios	212
Figura 84 - Cenário 15 - número de abortos	213
Figura 85 - Cenário 15 - <i>turnaround</i> (seg)	214
Figura 86 - Pacotes do <i>Lock Manager</i>	215
Figura 87 - Pacotes da DSL interna de <i>workflow</i>	236
Figura 88 - Pacotes da simulação	254

Lista de tabelas

Tabela 1 - Níveis de isolamento e problemas de leitura	37
Tabela 2 - Níveis de isolamento e bloqueios	37
Tabela 3 - Mapeamento dos requisitos da Web semântica em padrões	45
Tabela 4 - Alguns RDF Stores	61
Tabela 5 - Estrutura da consulta SPARQL SELECT	63
Tabela 6 - Vocabulários comuns	67
Tabela 7 - Possíveis configurações "Sujeito-Predicado-Objeto"	73
Tabela 8 - Possíveis configurações "Recurso-Propriedade-Valor"	74
Tabela 9 - Grânulos Candidatos	75
Tabela 10 - Grânulos para transações Web RDF	75
Tabela 11 - Tipos de bloqueio (<i>lock modes</i>) propostos	77
Tabela 12 - Compatibilidade entre os tipos de bloqueio	78
Tabela 13 - Compatibilidade entre os tipos de bloqueio, incluindo bloqueios planejados	80
Tabela 14 - Matriz de conversão entre bloqueios primitivos	94
Tabela 15 - Conversão (downgrade) de bloqueio real em bloqueio planejado	96
Tabela 16 - Matriz de compatibilidade de bloqueios completa	97
Tabela 17 - Matriz de conversão de bloqueios completa	98
Tabela 18 - Bloqueios otimistas de leitura para inserção e remoção	101
Tabela 19 - Exemplo de CBD - Concise Bounded Description [Sticker, 2005]	103
Tabela 20 - Exemplo (código Ruby) de uso do <i>Lock Manager</i>	113
Tabela 21 - Uso da DSL interna para a transação Web "Simulação Lâmpada"	123
Tabela 22 - Cenário 1 - número de bloqueios	172
Tabela 23 - Cenário 1 - número de abortos	173
Tabela 24 - Cenário 1 - <i>turnaround</i> (seg)	174
Tabela 25 - Cenário 2 - número de bloqueios	175
Tabela 26 - Cenário 2 - número de abortos	176
Tabela 27 - Cenário 2 - <i>turnaround</i> (seg)	177
Tabela 28 - Cenário 3 - número de bloqueios	178
Tabela 29 - Cenário 3 - número de abortos	179
Tabela 30 - Cenário 3 - <i>turnaround</i> (seg)	180
Tabela 31 - Cenário 4 - número de bloqueios	181
Tabela 32 - Cenário 4 - número de abortos	182

Tabela 33 - Cenário 4 - <i>turnaround</i> (seg)	183
Tabela 34 - Cenário 5 - número de bloqueios	184
Tabela 35 - Cenário 5 - número de abortos	185
Tabela 36 - Cenário 5 - <i>turnaround</i> (seg)	186
Tabela 37 - Cenário 6 - número de bloqueios	187
Tabela 38 - Cenário 6 - número de abortos	188
Tabela 39 - Cenário 6 - <i>turnaround</i> (seg)	189
Tabela 40 - Cenário 7 - número de bloqueios	191
Tabela 41 - Cenário 7 - número de bloqueios por grânulo	192
Tabela 42 - Cenário 7 - número de abortos	193
Tabela 43 - Cenário 7 - <i>turnaround</i> (seg)	194
Tabela 44 - Cenário 8 - número de bloqueios	196
Tabela 45 - Cenário 8 - número de bloqueios por grânulo	197
Tabela 46 - Cenário 8 - número de abortos	198
Tabela 47 - Cenário 8 - <i>turnaround</i> (seg)	199
Tabela 48 - Cenário 9 - número de abortos	200
Tabela 49 - Cenário 9 - <i>turnaround</i> (seg)	201
Tabela 50 - Cenário 10 - número de abortos	202
Tabela 51 - Cenário 10 - <i>turnaround</i> (seg)	203
Tabela 52 - Cenário 11 - número de abortos	204
Tabela 53 - Cenário 11 - <i>turnaround</i> (seg)	205
Tabela 54 - Cenário 12 - número de abortos	206
Tabela 55 - Cenário 12 - <i>turnaround</i> (seg)	207
Tabela 56 - Cenário 13 - número de abortos	208
Tabela 57 - Cenário 13 - <i>turnaround</i> (seg)	209
Tabela 58 - Cenário 14 - número de abortos	210
Tabela 59 - Cenário 14 - <i>turnaround</i> (seg)	211
Tabela 60 - Cenário 15 - número de bloqueios	212
Tabela 61 - Cenário 15 - número de abortos	213
Tabela 62 - Cenário 15 - <i>turnaround</i> (seg)	214

Demonstre sua capacidade e o seu interesse profissional. Seja persistente e tenha auto-confiança. Somente assim, nenhuma crise econômica e nada que venha ocasionar recessão e desemprego irá lhe derrubar.

Jorge B. Jacyntho.

27 de Outubro de 1998

