

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO



Vinci Pegoretti Amorim

**Uma Arquitetura Flexível para
Replicação de Bases Distribuídas Heterogêneas**

Dissertação de Mestrado

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

Orientador: Marco Antonio Casanova

Rio de Janeiro

Março de 2007



Vinci Pegoretti Amorim

Uma Arquitetura Flexível para Replicação de Bases Distribuídas Heterogêneas

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

Marco Antonio Casanova
Orientador
PUC – Rio

Antonio Luz Furtado
PUC – Rio

Karen K. Breitman
PUC – Rio

José Eugenio Leal
Coordenador Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro
Março de 2007

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

Vinci Pegoretti Amorim

Graduou-se em Ciência da Computação na Universidade Federal de Viçosa - UFV 2004. Hoje é pesquisador no laboratório Tecgraf. Seus interesses incluem banco de dados, engenharia de software, inteligência artificial e sistemas distribuídos.

Ficha Catalográfica

Amorim, Vinci Pegoretti

Uma arquitetura flexível para replicação de bases distribuídas heterogêneas / Vinci Pegoretti Amorim ; orientador: Marco Antonio Casanova. – 2007.

60 f. ; 30 cm

Dissertação (Mestrado em Informática)–Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007.

Inclui bibliografia

1. Informática – Teses. 2. Replicação. 3. Bases heterogêneas. 4. Arquitetura. 5. Framework. 6. Agentes. 7. Arquitetura orientada a serviços. 8. InfoPAE. I. Casanova, Marco Antonio . II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD 004

Este trabalho é dedicado:

Ao meu orientador, por toda tranquilidade e apoio acadêmico.

A meus colegas do Tecgraf, por toda camaradagem e apoio técnico.

À minha família, por toda confiança e apoio motivacional.

Agradecimentos

Primeiramente gostaria de agradecer ao meu orientador, Marco Antônio Casanova, por explicar, facilitar e sugerir em seus devidos momentos.

Um agradecimento para os membros da banca de professores, Antonio Luz Furtado e Karen K. Breitman, que tão prontamente atenderam aos meus pedidos de participarem da avaliação desta dissertação.

Gostaria de agradecer aos colegas do laboratório Tecgraf, em especial ao Aldo Nogueira, Alexandre Gazola, Carlos Coutinho, Demétrius Nunes, Luis Gustavo Ferrão, Marcelo Tílio, Rafael March, Rodnei Couto, Roberto Santos, que estiveram envolvidos com diversos aspectos do desenvolvimento deste trabalho. O apoio de vocês foi vital.

Aos meus pais Rogelio Caetano Amorim e Regina Augusta Pegoretti Amorim e aos meus irmãos Rogelio Pegoretti Caetano Amorim e Darlon Pegoretti Amorim. Mesmo à distância, foram capazes de apoiar, motivar e confiar.

Aos amigos Gleidson, Victor, André, Nara e Anna Carina.

Resumo

Amorim, Vinci; Casanova, Marco Antonio. **Uma Arquitetura Flexível para Replicação de Bases Heterogêneas**. Rio de Janeiro, 2007. 60p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Os diversos serviços de replicação existentes no mercado adquiriram um alto grau de maturidade e desempenho. Entretanto, mostram pouca capacidade de lidar com bases heterogêneas. Este trabalho propõe uma arquitetura focada na replicação de bases heterogêneas. Para conseguir alta escalabilidade e manter a simplicidade, a arquitetura segue uma estrutura multi-agente e adota uma abordagem de projeto orientado a domínio. Em seguida, o trabalho descreve uma implementação de referência e discute decisões técnicas desenvolvidas, enfatizando os problemas de versionamento, verificação de consistência e demais regras de negócio. O trabalho apresenta ainda utilitários que facilitam a configuração e manutenção do sistema de replicação.

Palavras-chave

Replicação; Bancos de Dados; Base Heterogênea; Agentes; Arquitetura Orientada a Serviços; Projeto Orientado a Domínio

Abstract

Amorim, Vinci; Casanova, Marco Antonio. A Flexible Architecture for Replication of Heterogeneous Bases. Rio de Janeiro, 2007. 60p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The replication services available acquire high maturity and performance levels. However, they do not work with heterogeneous data bases. This dissertation first describes a software architecture that focus on how to provide replication services for heterogeneous bases. To obtain high scalability and to maintain simplicity, the architecture follows a multi-agent structure and adopts a domain-driven design approach. Then, the dissertation describes a reference implementation and discusses the technical decisions adopted, focusing on version control problems, consistency verification and specific business-oriented rules. The dissertation also describes utilities that facilitate the configuration and maintenance of the replication system.

Keywords

Replication; Heterogeneous Bases; Agents; Service Oriented Architecture; Domain-Driven Design

Sumário

1	Introdução	12
1.1.	Motivação	12
1.2.	Objetivos	13
1.3.	Organização da Dissertação	13
2	Fundamentos	15
2.1.	Replicação de Dados	15
2.2.	Projeto Orientado a Domínio	16
2.3.	Software Orientado a Agentes	18
3	Arquitetura para Replicação de Bases Heterogêneas	20
3.1.	Base Heterogênea	20
3.2.	Serviço de Replicação de Bases Heterogêneas	21
3.2.1.	Processamento em Ciclos	23
3.2.2.	Versionamento	23
3.2.3.	Controle dos Estados das Bases	24
3.2.4.	Coordenação por Multi-Serviços de Replicação	25
3.2.5.	Processamento Assíncrono	27
3.2.6.	Operações Transacionais	27
3.3.	Componentes da Arquitetura	27
3.3.1.	Observador	29
3.3.2.	Exportador	30
3.3.3.	Importador	31
3.5.1.	Configurador	32
3.5.2.	Copiador	33
3.5.3.	Sincronizador	34
3.5.4.	Analisador	35
3.6.	Agentes, Utilitários e Soluções	36
3.7.	Camadas de um Serviço de Replicação	37

3.7.1. Camada de Fronteira	38
3.7.2. Camada de Negócio	39
3.7.3. Camada de Coordenação	40
4 Um Exemplo de Implementação	41
4.1. InfoPAE	41
4.2. RePAE: Uma Implementação de Referência	43
4.3. Tecnologias Adotadas pelo RePAE	45
4.4. Componentes Físicos do Ambiente	47
4.5. Instalação e Configuração	48
4.6. Testes	51
5 Conclusão	53
5.1. Contribuições	53
5.2. Trabalhos Relacionados	53
5.2.1. SQL Server	54
5.2.2. Daffodil Replicator	54
5.2.3. Delta Replicator	54
5.3. Trabalhos Futuros	55
6 Referências	57
Apêndice A – Vocabulário	60

Lista de figuras

Figura 1. Estratégias de replicação de dados	16
Figura 2. Projeto orientado a domínio	18
Figura 3. Atuação de agentes em um ambiente	19
Figura 4. Base de dados heterogênea	20
Figura 5. Base de dados e serviço de replicação	21
Figura 6. Mapeamento entre os requisitos e as soluções	22
Figura 7. Possíveis trocas de estados de uma base	24
Figura 8. Estratégias de coordenação de fluxo de replicação	25
Figura 9. Possíveis configurações de uma topologia com multi-serviços	26
Figura 10. Nó e topologia	26
Figura 11. Visão geral da arquitetura e utilitários	28
Figura 12. Driver observador	29
Figura 13. Exemplo de registro de alterações	30
Figura 14. Serviços de replicação	32
Figura 15. Utilitário Configurador	33
Figura 16. Utilitário copiador	34
Figura 17. Utilitário sincronizador	34
Figura 18. Utilitário analisador	35
Figura 19. Mapeamento entre as soluções e os agentes e utilitários	37
Figura 20. Camadas	38
Figura 21. Classes do pacote topologia	39
Figura 22. Domínio do InfoPAE	42
Figura 23. Executando uma emergência no InfoPAE	43
Figura 24. Repositórios	44
Figura 25. Base de dados do InfoPAE	44
Figura 26. Tecnologias	46
Figura 27. Exemplo de topologia lógica	47
Figura 28. Exemplo de topologia física	48
Figura 29. Estrutura de arquivos e diretórios	48

Figura 30. Exemplo de configuração de parâmetros

50

Figura 31. Configuração de uma topologia

51