

# Um Mediador para o Processamento de Consultas sobre Bases XML Distribuídas

Autores: Guilherme Figueiredo<sup>1</sup>, Vanessa Braganholo<sup>2</sup>, Marta Mattoso<sup>1</sup>

<sup>1</sup> COPPE/ UFRJ – Programa de Engenharia de Sistemas e Computação

<sup>2</sup> IM/ UFRJ – Departamento de Ciência da Computação



## Objetivos e Justificativa

O objetivo deste trabalho foi o desenvolvimento de uma camada de software para processar, de forma automática e transparente, consultas sobre bases XML distribuídas.

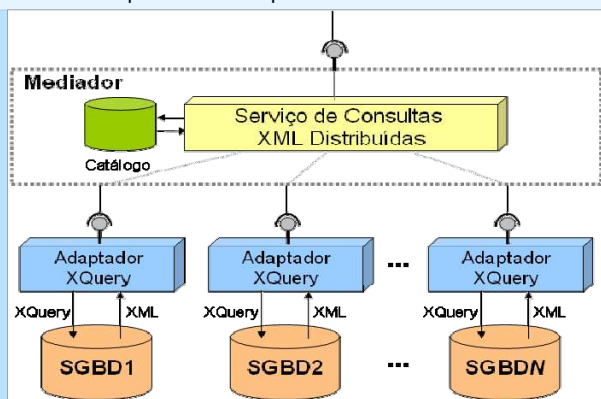
Esta solução pode ser utilizada em pelo menos duas situações distintas: (i) quando uma base centralizada é fragmentada, deve-se ter uma visão centralizada dos fragmentos, para que a fragmentação fique transparente para o usuário; (ii) quando várias bases locais pré-existentes precisam ser acessadas de forma integrada, deve-se poder definir uma visão XML sobre elas, de modo que cada base local possa ser considerado um fragmento desta visão global.

Uma metodologia para processamento de consultas XML distribuídas foi definida em nosso trabalho anterior [1], análoga à metodologia existente para o modelo relacional [2], e implementada na forma de uma arquitetura baseada em um Mediador com Adaptadores acoplados às bases remotas, onde o mediador fornece a visão global dos dados distribuídos.

Através da ferramenta construída, é possível executar consultas sobre uma visão global dos dados distribuídos e também visualizar graficamente todas as etapas de processamento da consulta: decomposição, localização dos dados, e reconstrução do resultado final.

## Arquitetura

A arquitetura implementada é formada por um mediador, que compõe a visão global e serve como ponto único de acesso, conectado a adaptadores acoplados às bases remotas.

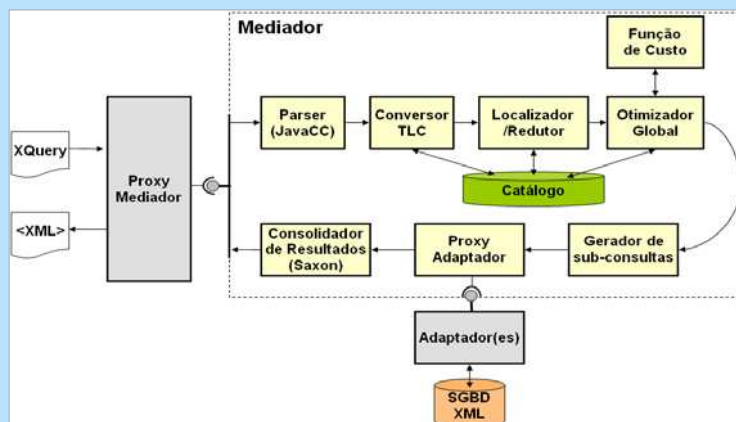


O Mediador é o principal componente da arquitetura, pois ele é responsável pelo processamento da consulta distribuída, realizando as etapas de decomposição, localização e otimização global da consulta, de acordo com a metodologia de processamento de consultas sobre bases XML distribuídas [1].

O mediador possui os componentes básicos para processamento de consulta [3]: um parser; um conversor para álgebra TLC [4]; um componente para localização e redução da consulta; um otimizador global e uma função de custo; um gerador de sub-consultas em XQuery; e um componente para consolidação dos resultados das bases distribuídas no resultado final da consulta.

Um catálogo é utilizado pelo mediador durante o processamento da consulta, onde estão disponíveis informações como os *schemas* das visões locais e da visão global, as definições dos fragmentos, o localização dos adaptadores de cada fragmento e estatísticas dos fragmentos, como número de nodos e tempos de acesso aos seus adaptadores.

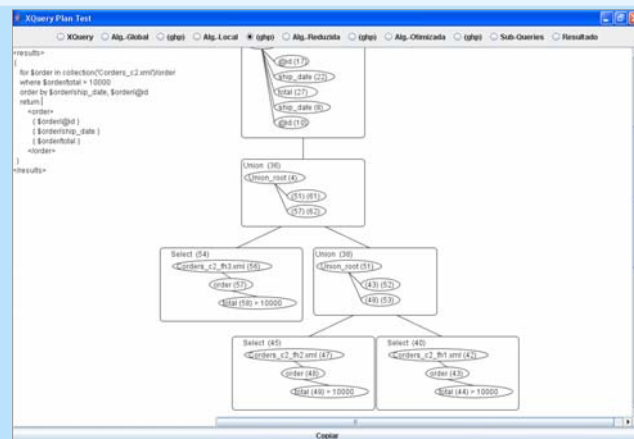
Os adaptadores são componentes bem simples e sua principal responsabilidade é a execução da sub-consulta gerada pelo mediador no SGBD XML.



Para as interfaces entre os adaptadores e o mediador, optamos pelo uso de serviços Web para obter um maior desacoplamento entre os adaptadores e o mediador. O mediador também é um serviço Web disponível aos usuários para execução de consultas XQuery sobre o ambiente distribuído.

## Execução de Consultas

Consultas podem ser executadas através do serviço Web do mediador, ou através de uma interface gráfica desenvolvida para testes do sistema. Esta interface gráfica permite a visualização da representação algébrica da consulta XQuery em TLC durante todas as etapas do processamento. Através da interface, tem-se uma noção muito clara de todas as etapas, o que facilita o entendimento da metodologia.



## Conclusão

Nossa arquitetura torna o processamento de consultas XQuery distribuídas totalmente transparente para o usuário final, além de ser não intrusiva e poder ser utilizada com qualquer SGBD. O fato de utilizarmos adaptadores nas bases locais faz com que seja possível inclusive utilizar bases não XML (relacionais, por exemplo). Basta, para isso, que o adaptador daquela base publique os dados no formato XML esperado por aquele fragmento.

## Principais Referências

- [1] Figueiredo, G. C., Braganholo, V., Mattoso, M. (2007) "A Methodology for Query Processing over Distributed XML Databases", Relatório Técnico ES-710/07 p. 1-24, Rio de Janeiro, Brasil.
- [2] Özsu, M. T., Valduriez, P. (1999) "Principles of Distributed Database Systems". 2ª ed., Prentice Hall.
- [3] Kossman, D. (2000) "The State of the Art in Distributed Query Processing", Em: ACM Computing Surveys, v. 32, p. 422-469.
- [4] Paparizos, S., Wu, Y., Lakshmanan, L. V. S., Jagadish, H. V. (2004) "Tree Logical Classes for Efficient Evaluation of XQuery", Em: SIGMOD, p. 71-82. ACM.