

Consultas sobre visões XML globais de Bases de Dados Distribuídas

Guilherme Coelho de Figueiredo

Orientação: Marta L. Queirós Mattoso e

Orientação: Vanessa de Paula Braganholo

e-mail: {g.coelho, marta, vanessa}@cos.ufrj.br

Nível: Mestrado

Programa de Engenharia de Sistemas e Computação - COPPE/UFRJ

Universidade Federal do Rio de Janeiro

Ingresso: Março / 2004

Previsão de Conclusão: Março /2006

Resumo. Este artigo apresenta uma proposta de uma camada de software para permitir consultas globais sobre bases de dados distribuídas através de visões XML globais dessas bases. Uma consulta XQuery submetida pelo usuário é decomposta em sub-consultas que são enviadas para os *sites* envolvidos do ambiente distribuído. Os documentos XML retornados por cada sub-consulta são consolidados e o resultado é retornado ao usuário na forma de um documento XML.

Palavras-Chave: Bancos de Dados Distribuídos, Visões XML, XQuery.

1 Introdução

Muitas organizações possuem mais de uma base para armazenar dados de diferentes filiais, como por exemplo, uma rede de livrarias. Estas bases de dados, autônomas, mas praticamente homogêneas, podem ser consideradas fragmentos do que seria uma representação global dos dados da organização. No entanto, quando se deseja consultar os dados globais da organização, os usuários têm que submeter consultas a cada site isoladamente, e processar os resultados posteriormente. Para resolver este problema, pode-se criar uma federação de bases de dados distribuídas para a realização de consultas sobre estas bases de forma integrada, facilitando a criação de ferramentas de manipulação dos dados globais da organização.

Os dados na maioria das organizações estão armazenados em bases relacionais devido à maturidade dos recursos de indexação, controle de transação, processamento de consultas e segurança. Entretanto, documentos XML são cada vez mais utilizados como meio de troca de dados entre aplicações, e consultas XQuery estão se popularizando devido a sua capacidade de projetar resultados diretamente em um formato XML definido pelo usuário. Tais recursos facilitam o desenvolvimento de aplicações, principalmente no domínio de comércio eletrônico.

Este trabalho tem como objetivo a criação de uma arquitetura de uma camada de software (Mediador) que permita a execução de consultas XQuery sobre visões XML globais de bases de dados distribuídas. Cada base de dados recebe um adaptador para exportar visões XML locais homogêneas que são utilizadas na composição das visões globais. Consultas XQuery enviadas ao Mediador são decompostas em sub-consultas,

destinadas aos *sites* selecionados pelo plano de execução escolhido. Os resultados destas sub-consultas são re-processados pelo Mediador para obtenção do resultado final.

Este documento foi estruturado da seguinte forma: começamos com uma revisão dos trabalhos relacionados na literatura na seção 2. Na seção 3, fazemos uma apresentação da nossa arquitetura proposta e exemplificamos a sua utilização na seção 4. Na seção 5 apresentamos as considerações finais e as próximas etapas deste trabalho.

2 Trabalhos Relacionados

Diversos trabalhos são encontrados na literatura sobre consultas a bases de dados através de uma arquitetura *mediador – adaptadores* [1], onde grande parte utiliza protocolos ou linguagens proprietárias para a comunicação ou realização das consultas, o que de certa forma limita a utilização da solução ou a restringe a determinados produtos de mercado. Alguns dos trabalhos encontrados apresentam uma abordagem mais semelhante à nossa, através da utilização de protocolos abertos e XML, como o XMF [2], que implementa um mediador baseado em XML e utiliza XPath para a realização de consultas; e o MIX [3] que também utiliza XML, mas a linguagem para consultas é a XMAS. Além de nenhum destes trabalhos suportar XQuery, nenhum trata de consultas com atualização das visões XML globais, ponto este previsto em nossa proposta.

Bases de dados federadas distribuídas [4], e processamento de consultas distribuídas [5] são pontos importantes neste trabalho. As definições de fragmentação de XML consideradas serão as disponibilizadas em [6], onde são definidas fragmentação horizontal, vertical e híbrida de documentos XML, além de sugerir a arquitetura PARTIX para processamento de XQuery distribuída e mostrar resultados experimentais de ganho de desempenho com a fragmentação de documentos XML. Para a construção de visões sobre documentos XML utilizaremos os resultados obtidos por Braganholo em [8].

3 Arquitetura Proposta

Um ambiente de bases de dados distribuídos deve possuir um ponto de acesso capaz de tornar a estrutura distribuída das bases de dados transparente aos usuários e aplicações finais. Acesso totalmente transparente significa que os usuários poderão formular consultas sem qualquer preocupação com a fragmentação, a localização ou a replicação dos dados, e deixar a cargo do sistema a resolução dos mapeamentos necessários [7].

Com isto em mente projetamos a arquitetura exibida na Figura 1, que permite construirmos um ambiente de bancos de dados federados distribuídos, onde as bases de dados, autônomas, são acessadas por um (ou mais) Mediador através de um adaptador XQuery. Os módulos principais da arquitetura são descritos a seguir:

Mediador – ponto de acesso às bases de dados distribuídas onde aplicações e usuários poderão realizar consultas XQuery sobre as visões globais do ambiente. O Mediador contém o Serviço de Consultas XML Distribuídas e um Catálogo com as definições das visões globais e locais de cada *site*.

Serviço de Consultas XML Distribuídas – módulo responsável pelo processamento das consultas XQuery e elaboração do plano de execução global para composição das sub-consultas. O plano de execução global é definido através da minimização de uma função de custo calculada a partir das informações contidas no

Catálogo. As sub-consultas são executadas em paralelo pelos adaptadores XQuery de cada base envolvida e sobre os seus resultados é aplicada uma consulta XQuery derivada da consulta original para a produção do resultado final.

Catálogo – armazena as definições (XQuery) e esquema (XML Schema) das visões XML globais e a relação das bases acopladas ao Mediador. Para cada base de dados são armazenadas as definições das suas visões locais, incluindo o seu predicado de formação (em casos de fragmentação horizontal), custos de acesso, estatísticas e *url* de acesso do seu adaptador XQuery.

Adaptador XQuery – componente responsável pela execução da consulta XQuery sobre as visões XML locais de uma base de dados. O adaptador XQuery deve ser configurado em cada base de dados para que estas forneçam visões XML homogêneas. Implementando uma interface pré-definida para serviço Web, adaptadores XQuery podem ser acessados pelo Mediador através da Internet.

Na implementação deste trabalho estaremos considerando apenas bases de dados locais relacionais (devido a sua maior utilização real) com visões em XML. Para o papel do adaptador XQuery utilizaremos como base o PATAXÓ [8,9], pois possibilita consulta e atualização de visões XML de bases relacionais. No PATAXÓ, as visões XML locais de uma base relacional são definidas pela UXQuery [8], um subconjunto da XQuery. Atualizações na visão XML local são traduzidas para comandos de atualização nas visões relacionais geradas pelo PATAXÓ.

SGBDx – sistemas de banco de dados participante da federação.

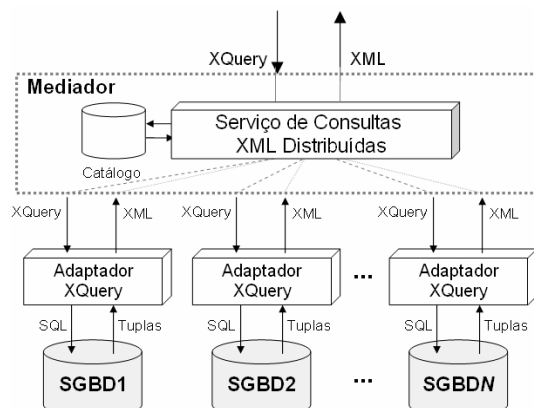


Figura 1: Arquitetura para consultas XML em bases de dados federadas distribuídas.

4 Exemplo de Aplicação

Para exemplificar a aplicação da nossa proposta, consideramos uma rede de livrarias onde cada livraria possui uma base de dados própria com uma réplica do catálogo dos livros e a relação de vendas e estoque disponível na loja, como pode ser visto na Figura 2. Esta relação de vendas pode ser considerada como uma fragmentação horizontal da relação de livros de toda rede de livrarias, já que cada loja contém apenas os dados sobre os seus livros em estoque e suas vendas realizadas.

Todas as lojas da rede de livrarias deverão possuir um adaptador XQuery configurado para produzir visões XML locais de acordo com as especificações no mediador (as visões XML locais devem ser homogêneas). Além disso, as visões XML

locais de um adaptador XQuery devem ser registradas no Catálogo do Mediador, especificando a localização da visão e os predicados de formação no caso de uma fragmentação horizontal (Figura 3).

LOJA 1

LIVRO		
ISBN	NOME	EDITORA
111111	Distributed Databases	XYZ
222222	Banco de Dados	ABC

ESTOQUE_LIVROS				
CD_BARRAS	ISBN	LOJA	DATA_COMPRA	DATA_VENDA
12345678910	111111	1	14/03/05	07/04/05
12345678911	111111	1	14/03/05	
12345678912	222222	1	14/03/05	

LOJA 2

LIVRO		
ISBN	NOME	EDITORA
111111	Distributed Databases	XYZ
222222	Banco de Dados	ABC

ESTOQUE_LIVROS				
CD_BARRAS	ISBN	LOJA	DATA_COMPRA	DATA_VENDA
12345678999	111111	2	18/04/05	20/04/05
12345678998	222222	2	19/04/05	
12345678997	222222	2	19/04/05	02/05/05
12345678996	222222	2	19/04/05	

Figura 2: Bases de dados do exemplo da rede de livrarias.

Visão Global "Catalogo"		Visão Global "RegistroProdutos"	
<IELEMENT catalogo (livros)> <IELEMENT livros (livro*)> <IELEMENT livro (isbn, nome, editora)> <IELEMENT isbn (#PCDATA)> <IELEMENT nome (#PCDATA)> <IELEMENT editora (#PCDATA)>		<IELEMENT RegistroProdutos (produto*)> <IELEMENT produto (cdBarras, isbn, loja, dataCompra, dataVenda)> <IELEMENT cdBarras (#PCDATA)> <IELEMENT isbn (#PCDATA)> <IELEMENT loja (#PCDATA)> <IELEMENT dataCompra (#PCDATA)> <IELEMENT dataVenda (#PCDATA)>	
BASE_ID	URL	BASE_ID	PREDICADO
1	http://www.xyz.com/XRsrv	1	Catalogo
2	http://www.abc.com/XRsrv	1	RegistroProdutos
		2	Catalogo
		2	RegistroProdutos

Figura 3: Catálogo com as definições das visões XML globais e as relações das bases de dados.

* As definições das visões XML estão em DTD, ao invés de XML Schema, devido à limitação de espaço.

O Serviço de Consultas XML Distribuídas irá utilizar as informações do Catálogo para identificar a quais *sites* as sub-consultas deverão ser enviadas. Uma consulta sobre a visão *Catalogo* para buscar a relação de livros vendidos pela rede de livrarias, por exemplo, poderá ser enviada a qualquer um dos *sites* da base distribuída, já que todos fornecem uma réplica completa desta visão. Por outro lado, uma consulta sobre o estoque disponível na loja 1 será enviada apenas à base da respectiva loja, pois esta é a única que possui uma visão *RegistroProdutos* com o predicado que atende a consulta.

Após receber os resultados de todas as sub-consultas, o Mediador aplica uma consulta derivada da consulta original sobre estes resultados para a obtenção do resultado final.

5 Considerações Finais

Este artigo apresentou uma visão geral sobre a nossa proposta para consultas XQuery sobre visões XML globais de bases de dados relacionais distribuídas, através da arquitetura apresentada na Figura 1. Além disso, também pretendemos ser capazes de realizar atualizações nas visões XML globais e, para isso, o adaptador XQuery utilizado

precisa suportar atualizações nas visões locais da base de dados, como é o caso do PATAXO [8].

A principal contribuição deste trabalho é o desenvolvimento de uma abordagem para acesso (consultas e atualizações) a visões XML globais de bases de dados federadas distribuídas. Nossa arquitetura permite certa heterogeneidade entre as bases de dados, desde que as visões XML locais produzidas pelos adaptadores sejam homogêneas.

Os próximos passos deste trabalho incluem o refinamento da especificação de cada módulo da arquitetura proposta, a definição dos algoritmos de decomposição de consultas XQuery, a implementação destes módulos e algoritmos, e a execução de testes funcionais e de carga para validar a generalidade e o desempenho da solução.

Referências

- [1] WIEDERHOLD, Gio, *Mediators in the Architecture of Future Information Systems*, in: Readings in Agents, pp. 185-196. Morgan Kaufmann, San Francisco, CA, USA, 1992.
- [2] LEE, Kangchan; MIN, Jaehong; PARK, Kishik; LEE, Kyuchul: *A Design and Implementation of XML-Based Mediation Framework (XMF) for Integration of Internet Information Resources*. Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02) 7IEEE Computer Society. 2002.
- [3] BARU, Chaitan; GUPTA, Amarnath; LUDAESHER, Bertram; MARCIANO, Richard; PAPAKONSTANTINOU, Yannis; PAVEL, Velikhov; CHU, Vincent: *XML-Based Information Mediation with MIX*. In Demo Session, ACM-SIGMOD'99. 1999.
- [4] SHETH, Amit P. e LARSON, James A.: *Federated database systems for managing distributed, heterogeneous, and autonomous databases*. ACM Computing Surveys (CSUR) 22(3), p.183-236. New York, NY, USA. 1990.
- [5] KOSSMAN, Donald: *The state of the art in distributed query processing*. 32(4), p.422-469. ACM Computing Surveys. 2000.
- [6] ANDRADE, Alexandre; RUBERG, Gabriela; BAIÃO, Fernanda; BRAGANHOLO, Vanessa; MATTOSO, Marta: *PartiX: processing XQuery queries over fragmented XML repositories*. Submetido ao WIDM . 2005.
- [7] ÖZSU, M. T. e VALDURIEZ, Patrick, *Principles of Distributed Database Systems*, Prentice Hall, 1999.
- [8] BRAGANHOLO, Vanessa: *From XML to Relational View Updates: applying old solutions to solve a new problem*. Tese de Doutorado. Instituto de Informática, UFRGS. Porto Alegre, RS, Brasil. 2004.
- [9] BRAGANHOLO, Vanessa; DAVIDSON, Susan; HEUSER, Carlos: *From XML View Updates to Relational View Updates: old solutions to a new problem*. VLDB - International Conference on Very Large Data Bases . Toronto, Canada. 2004.