



## 9º Congresso de Pós-Graduação

# UM PROCESSADOR DE CONSULTAS XQUERY COM TECNOLOGIA RELACIONAL-XML PARA UM SERVIDOR OLAP EM NUVEM

### Autor(es)

---

LUIS GUILHERME SOUTO JARDIM

### Orientador(es)

---

MARINA TERESA PIRES VIEIRA

## 1. Introdução

---

Durante muitos anos podemos verificar o crescimento das técnicas de Análise dos Dados para fins práticos nas empresas. Este movimento, chamado de Business Intelligence, utiliza uma série de tecnologias em paralelo, desenvolvidas com o principal objetivo de simplificar a análise e transportar a responsabilidade de manipulação das informações da área de TI para as mãos dos usuários que necessitam das informações. Podemos confirmar em Kimball (2002), "Temos montanhas de dados nesta empresa, mas não temos acesso aos mesmos, queremos cruzar informações, apenas me mostre o que é importante". Os grandes bancos de dados complexos, originalmente com centenas de tabelas relacionadas, foram reduzidos e simplificados através de diversas técnicas, chegando aos Data Warehouses (DW). Representamos estes DWs através de um modelo multidimensional de Tabelas Fatos, que contem os dados e medidas, e Dimensões, com as variáveis a serem agrupadas e consolidadas. Agora os dados disponíveis estão preparados, limpos, resumidos e organizados de forma sempre a permitir a compreensão e manuseio por pessoas não ligadas à área de TI. A evolução das técnicas de análise levou naturalmente no decorrer dos anos ao surgimento de diversas ferramentas e o processo, como um todo, passou a receber a denominação de OLAP. No entanto a diversificação e o crescimento natural das pesquisas na área acabaram, pouco a pouco, por tornar tais tecnologias novamente complexas e distantes do seu real objetivo inicial. Linguagens de consultas cada vez mais complexas como MDX tentam se impor como padrões, aumentando a complexidade dos sistemas de análise e limitando novamente aos técnicos a sua compreensão. Enquanto isto, o surgimento da Internet forçou o desenvolvimento de novos padrões e uma das tecnologias mais bem sucedidas devido à sua simplicidade e facilidade de leitura por olhos humanos foi a tecnologia XML. Imediatamente a sua utilidade se mostrou de enorme valor e para atingir os Sistemas OLAP foi uma questão de tempo. A linguagem de consulta XQuery, utilizada para se extrair informações de documentos XML, foi rapidamente melhorada com sugestões que visavam a sua utilização em análise de informações. Beyer et al (2005) sugeriram o termo XQuery for Analytics. Hummer et al (2003) propuseram a modelagem multidimensional em XML, o modelo X-Cube, Beyer et al (2005) propuseram extensões à linguagem XQuery, e Boucher et al (2009) propuseram a modelagem dimensional mais avançada em documentos XML hierárquicos. A tecnologia XML, aplicada a sistemas de análise de dados OLAP, representa a esperança na retomada do caminho original de simplificação e usabilidade. Paralelamente, o desenvolvimento de tecnologias ligadas à Internet levou ao movimento Cloud Computing, uma tendência de transporte de Sistemas, antes isolados, para a "nuvem". Tal movimento possui características próprias, como elasticidade de armazenamento e de processamento. Elasticidade de armazenamento é bem compreendida quando nos referimos a Bancos de Dados e o Armazenamento de Dados em Nuvem é um dos temas mais estudados nos dias de hoje. SGBDs Relacionais em nuvem são uma realidade e quando pensamos em utilizar a tecnologia XML, basicamente arquivos de texto, em nuvem, nos deparamos com uma contradição. Documentos XML são excelentes como meio de transporte de informações na Web, representam uma esperança de simplificação em Sistemas OLAP, mas são inadequados como meio de armazenamento elástico em nuvem.

## 2. Objetivos

---

O trabalho aqui proposto se encontra inserido dentro da proposta ampla de projetar e desenvolver os principais componentes de um Servidor OLAP em Nuvem que é tema de dois trabalhos de mestrado. O Servidor OLAP em Nuvem será composto basicamente por dois módulos. O Módulo de Gerenciamento de Consultas em Nuvem (2) e o Módulo XML-DW (1), conforme mostrado na figura 1. O Módulo de Gerenciamento de Consultas em Nuvem (2) é responsável pelo processamento de consultas SQL em paralelo e será tema de outro trabalho de mestrado. O Módulo XML-DW (1), que realizará funções de processador de consultas XQuery, implementará o Webservice responsável por atender as requisições que chegam do usuário, pela Web, e respondê-las através de documentos XML. Este segundo Módulo é o objetivo deste trabalho. FIGURA 1- ESQUEMA DO SERVIDOR OLAP PROPOSTO

### 3. Desenvolvimento

---

Quando se pensa em representar um DW utilizando documentos XML, a primeira constatação é a dificuldade de se estabelecer um padrão. Pouco a pouco um conjunto de pesquisas e novas tecnologias começaram a surgir relacionando OLAP e XML. Álgebras que tentam capturar o processamento XML apareceram como em Hosoya (2003). Também Pedersen et al (2002) descrevem como integrar XML e OLAP. Em conjunto com o crescimento da linguagem de consulta XQuery finalmente amadureceu e cunhou-se o termo XML Data Warehouse para se referir ao conjunto de tecnologias que utilizam documentos XML e linguagem de consulta XQuery para implementar a tecnologia OLAP e representar um Data Warehouse. REPRESENTANDO UM DW EM XML Hummer et al (2003) propuseram o Modelo Xcube. Neste modelo cada tabela fato e cada nível de dimensão são representados por um elemento XML que possui links entre as dimensões e as hierarquias. Estas não são expressas por relacionamentos pai-filho, mas por identificadores. A seguir um exemplo de um DW representado no modelo XCube. ...