



## 6º Congresso de Pós-Graduação

### AUTOMATIZAÇÃO DO PROCESSO DE CRIAÇÃO DE VISÕES PARA MODELAGEM DE DW

#### Autor(es)

---

RICARDO ALEXANDRE NEVES

#### Orientador(es)

---

MARINA TERESA PIRES VIEIRA

#### 1. Introdução

---

Um data warehouse (DW) é um repositório de dados que provê um ambiente integrado para consultas de suporte à decisão e para análises que requerem agregações complexas de grandes quantidades de dados históricos (Hung et al., 2007). Esses dados são integrados a partir de múltiplos bancos de dados operacionais, que podem ser muito grandes, heterogêneos, e distribuídos. A fim de aumentar a eficiência das consultas, uma abordagem comumente usada é armazenar resultados intermediários de agregações resultantes do processamento de consultas no DW. Esses resultados intermediários armazenados no DW são chamados de visões materializadas. Em outras palavras, um DW pode ser visto como um conjunto de visões materializadas relativas aos dados extraídos de bases de dados distribuídas heterogêneas. Uma vez que uma consulta pode produzir muitos resultados intermediários, uma das decisões mais importantes no projeto de Data Warehouse é selecionar as visões mais adequadas.

A materialização de visões é um dos problemas clássicos em um ambiente de Data Warehouse (Valluri et al., 2002). A materialização de visões é a técnica comumente empregada em projetos de DW em busca da minimização dos custos de resposta das consultas. Uma consequência é o aumento dos custos de armazenamento e de manutenção dessas visões, que precisam ser levados em consideração.

Segundo Teodoratos e Simitsis (2006) alguns dos fatores considerados para a materialização de visões são as restrições de espaço e tempo de processamento. E, segundo Gupta (1997), o espaço em disco, o tempo de processamento e custo de manutenção são restrições que justificam a não materializar todas as possibilidades de visões, pois tais restrições devem ser consideradas. Assim, o problema da seleção de um conjunto apropriado de visões para materializar é uma das mais importantes decisões de projeto em um DW.

Um DW pode ser visto como um conjunto de visões materializadas. As visões materializadas podem ser utilizadas de modo parcial ou completo no sentido de avaliar consultas do usuário. (Theodoratos e Simitsis; 2006).

Em Harinarayan et al. (1996), são apresentados conceitos para representação das visões em Data Warehouse, os quais também são necessários para o entendimento deste trabalho:

Notação lattice: Lattice é definido como um conjunto de elementos (consultas ou visões) e sua relação de

dependência. Diagrama lattice: Gráfico em que os elementos do lattice são nós e há uma ligação de a (inferior) para b (superior). Hierarquia: As hierarquias introduzem as dependências de consultas para que se possam determinar quais delas materializar. Em linhas gerais, as dimensões de um cubo de dados consistem em mais de um atributo. Essas dimensões são organizadas como hierarquias desses atributos. Um exemplo simples para melhor entendimento das hierarquias pode ser observado na organização do tempo, levando-se em conta as dimensões: dia, semana, mês e ano.

O presente trabalho está vinculado na área de estudo que compreende a seleção de um conjunto de visões para materializar em um ambiente de Data Warehouse e atua diretamente na fase de projeto do DW, baseado no trabalho de Lima(2006).

## 2. Objetivos

---

O objetivo deste trabalho é automatizar o processo de seleção de visões para materializar durante a concepção do projeto de geração do modelo dimensional para Data Warehouse. Será utilizado como instrumento de validação do projeto a Ferramenta de Lima (2006). Ao trabalhar com tal ferramenta, serão realizados refinamentos no sentido de estabelecer melhorias necessárias com relação às funcionalidades e processos da Ferramenta. Tais refinamentos serão realizados de modo que a ferramenta forneça subsídios necessários para que as visões sejam geradas e selecionadas.

## 3. Desenvolvimento

---

O trabalho encontra-se em andamento e, como proposta, fica estabelecida a criação de um módulo para decidir quais visões serão materializadas, tendo em vista a restrição de espaço, os requisitos de análise do usuário e a frequência das consultas. A ferramenta de Lima(2006) mediante a nova proposta ficará conhecida como DW Designer.

Pretende-se que o projetista de Data Warehouse, quando fizer uso da Ferramenta, após a execução do módulo proposto, tenha condição de projetar o modelo dimensional e cumprir metas de projeto, direcionadas à performance de consultas através do módulo de seleção de visões. Por consequência, a estrutura de metadados do DW ficará organizada de forma que possa facilitar futuramente a manutenção das visões.

O módulo proposto será incorporado à ferramenta DW Designer atuando como complemento às funcionalidades de modelagem dimensional existentes. Esse novo módulo da Ferramenta será chamado de Módulo de Seleção de Visões para Materialização (SVM).

A nova arquitetura da Ferramenta, de acordo com a proposta, é apresentada Figura 1. As metas para concretizar a arquitetura proposta serão divididas em duas partes: a primeira parte compreende o desenvolvimento do módulo de seleção de visões e a segunda trata o refinamento das funcionalidades da Ferramenta para Geração do Modelo Dimensional para DW, que são apresentadas a seguir.

### 3.1 Desenvolvimento do Módulo de Seleção de Visões (SVM)

O módulo de seleção de visões propõe disponibilizar, para o projetista, de maneira interativa, as visões geradas pelo assistente da DW Designer através das etapas de geração do modelo dimensional e que estão armazenadas como metadados do DW. A partir dos resultados apresentados pelo módulo, o projetista terá condição de avaliar e decidir quais visões serão materializadas.

A política adotada para uso no módulo é baseada em restrições tais como: espaço, na frequência de uso das visões e requisitos de usuário.

A solução proposta para o novo módulo consiste em extrair os dados armazenados dos metadados do DW, exibindo-os como visões candidatas para materialização. As visões candidatas são originadas das agregações do modelo dimensional, responsável por atender cada requisito de modelagem do usuário

durante a geração do modelo dimensional, que é executado numa etapa anterior pelo módulo de modelagem. Esse módulo por sua vez, armazenou tais metadados como resultado do modelo do DW, no banco de dados da Ferramenta. Recuperar as visões candidatas, assim como organizá-las por assunto, é uma estratégia que o módulo de seleção irá utilizar para exibir as informações ao projetista para facilitar a análise. As informações disponibilizadas para análise, levando em consideração o custo por espaço e frequência de uso das visões são exibidas em uma primeira versão como mostra o cenário na figura 2. Os parâmetros considerados para este primeiro cenário, como uma primeira proposta a ser avaliada na próxima etapa do trabalho, são:

- Descrição das Visões: Parâmetro que descreve a visão de modo que seja possível identificá-la.
- Tamanho em megabytes (MB) de cada visão: Para que seja possível calcular o tamanho da visão é feita uma seleção nos registros do banco de dados operacional levando em conta os metadados de DW envolvidos.
- Custo por unidade de disco: O custo calculado com ênfase no cálculo do benefício por unidade de disco da abordagem de Gupta (1997).
- Sugestão do algoritmo Greedy: Essa sugestão é o retorno do algoritmo, que seleciona as visões sob o critério apresentado na figura 2.
- Seleção manual: A seleção manual é a possibilidade que o projetista tem de escolher uma visão que o algoritmo não tenha selecionado ou até mesmo retirar uma opção caso o projetista julgue necessário.
- Espaço necessário para materializar todas as visões: Esse parâmetro deve ser calculado automaticamente mediante a lista de visões apresentadas por assunto, de acordo com modelo do DW armazenado na Ferramenta.
- Espaço disponível para projeto: Parâmetro que informa ao algoritmo qual o espaço disponível em disco para materialização, levando em consideração o projeto do DW em questão.

A partir das informações apresentadas pelo cenário da figura 2, a análise para a seleção das visões é feita de acordo com o julgamento do projetista com relação às melhores escolhas para atender a política de espaço e frequência de uso. Daí, então, decidir se concorda com as opções pré-selecionadas pelo algoritmo ou se combina a análise do algoritmo de acordo com as necessidades de projeto, realizando, dessa forma, a seleção das visões para cada assunto.

### 3.2 – Refinamento das Funcionalidades da Ferramenta DW DESIGNER

Em razão da implementação do módulo svm e também para que a Ferramenta possa estar preparada para aceitar módulos futuros foi desenvolvida uma nova versão fazendo uso da tecnologia Java, assim como o refinamento das funcionalidades da Ferramenta proposta por Lima (2006). O refinamento nas funcionalidades não descaracteriza a versão original da Ferramenta, e sim melhora o funcionamento geral da mesma de maneira a reorganizar o conteúdo nas telas, e também proporcionar conexão com novos bancos de dados operacionais tais como: Oracle e Mysql. Uma alteração necessária em um primeiro momento na Ferramenta foi feita nos processos de modelagem, a qual consistiu em acrescentar a geração do modelo parcial (DW'), onde cada modelo parcial representa uma visão no contexto do módulo SVM. Após todas as etapas de modelagem cumpridas, todos os requisitos necessários foram contemplados e então tem-se a geração do Modelo Geral, o DW completo para o contexto esperado.

Após a criação do Modelo Geral, são geradas as agregações e sumarizações para atender aos requisitos informados pelo projetista, em sua interação com a ferramenta DW Designer.

Além de atender aos requisitos, a ferramenta deve fornecer ao projetista a possibilidade de complementar o DW gerado para análises adicionais àquelas definidas pelos usuários, através da análise da parte restante do B.D. operacional que não foi considerada.

## 4. Resultado e Discussão

---

A ferramenta DW Designer encontra-se em estágio de implementação, onde as etapas de conversão da ferramenta de Lima (2006) para a tecnologia Java e o primeiro cenário (Figura 2) estão concluídos. As principais telas podem ser vistas na figura 3. Além da tecnologia Java está sendo utilizado o banco de dados Firebird versão 2.0. Faz-se uso do IDE (ambiente integrado para desenvolvimento de software) Eclipse Europa 3.3.0 para compilar e manipular as tecnologias Java empregadas no projeto da ferramenta. O resultado a ser apresentado pelo Módulo SVM, após o processo de seleção, será no formato da linguagem SQL (Structure Query Language), de modo a facilitar a implementação das visões materializadas. Um exemplo do script SQL a ser produzido como resultado pós seleção das visões é discriminado abaixo:

```
CREATE OR REPLACE MATERIALIZED VIEW VISÃO 1 AS
SELECT
ITENS_VENDIDOS.VALOR_TOTAL,
PRODUTOS.ID_PRODUTO,
PRODUTOS.DESCRICAO,
ESTADO.SIGLA,
ESTADO.NOME
FROM CIDADES,
LOJAS,
PRODUTOS,
ITENS_VENDIDOS,
ESTADO
WHERE CIDADES.ID_CIDADE = LOJAS.ID_CIDADE AND

LOJAS.ID_LOJA = ITENS_VENDIDOS.ID_LOJA AND

PRODUTOS.ID_PRODUTO = ITENS_VENDIDOS.ID_PRODUTO AND

ESTADO.SIGLA = CIDADES.SIGLA;
```

Na versão atual da ferramenta adotou-se o algoritmo Greedy de Gupta (1997) para escolha das visões a materializar, que leva em consideração o custo do espaço utilizado pelas visões. Como continuidade do trabalho estão sendo consideradas outras abordagens, entre elas as de: Gou, Yu e Lu (2006); Kotidis e Roussopoulos (2001) e Yousri, Ahmed e Makky (2005), que contemplam outros parâmetros para a escolha das visões a materializar, além de questões de espaço ocupado pelas visões, tais como tempo de manutenção das visões e custo de processamento das consultas.

## 5. Considerações Finais

---

Como produto final dessa proposta, será fornecido, além do relatório completo com o modelo dimensional gerado a partir dos requisitos informados, o código escrito em SQL para implantação do projeto físico do Data Warehouse. E, como resultado do módulo svm, será gerado um arquivo SQL para que possa ser importado pelo SGBD compatível. O código SQL a ser gerado pela Ferramenta, tanto do DW quanto das visões a materializar, será compatível com o SGBD Oracle.

## Referências Bibliográficas

---

GOU, G., YU, J. and LU, H. **A\* Search: An Efficient and Flexible Approach to Materialized View**

**Selection**, Proceedings of the IEEE Transaction on System, Man, and Cybernetics (Part C), VOL. 36, N. 3, p. 411- 424, 2006.

GUPTA, A., MUMICK I.S. **Maintenance of materialized views: problems, techniques, and applications**, Proceedings of the ACM SIGMOD International Conference on Management of Data, VOL.18, N. 2, p. 3 -18, 1995.

GUPTA, H. **Selection of Views to Materialize in a Data Warehouse**, Proceedings of the 6th International Conference on Database Theory, ICDT '97, p.98-112, 1997.

Harinarayan, V., Rajaraman, A., Ullman, J. D. **Implementing Data cubes Efficiently**. Proceedings of the ACM SIGMOD In Conference Management Of Data, p. 205 -216, 1996.

HUNG, m.,HUANG, M., YANG, D.,HSUEH,N. **Efficient approaches for Materialized views selection in data warehouse**, Information Sciences, Vol.177, p.1333-1348, 2007.

KOTIDIS, Y. and ROUSSOPOULOS, N. **A Case for Dynamic View Management**, Proceedings of the ACM Transactions on Database Systems, LIMA, L. G. **Ferramenta para Geração do Modelo Dimensional para Data Warehouses**, 87p. Dissertação (Mestrado em Ciência da Computação) – Universidade Metodista de Piracicaba, Piracicaba, 2006.

THEODORATOS, D; SIMITSIS, A. **Materialized View Selection for Data Warehouse Design**. Proceedings of the In Encyclopedia of Data Warehousing and Mining, Idea Group, 2005.

VALLURI, S; VADAPALLI, S; KARLAPALEM, K. **View Relevance Driven Materialized Selection in Data Warehousing Environment**. Proceedings of the Australian Database Conference, Vol.5, 2002.

Yousri, N., Ahmed, K. and Makky, N. **Algorithms for Selecting Materialized Views in a Data Warehouse**, Proceedings of the 3rd ACS/IEEE International Conference on Computer Systems and Applications, p.27-34, 2005.

## Anexos

---

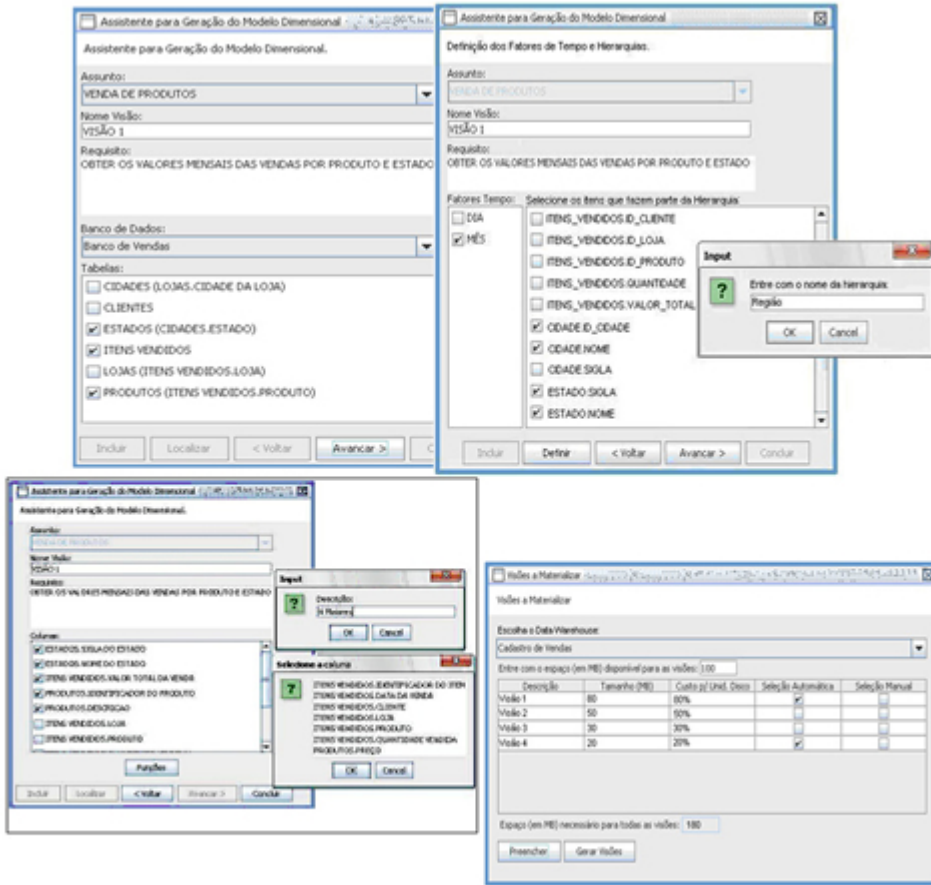


Figura 3 - Telas da DW Designer

Cálculo do benefício
Retorno de M

Descrição das Visões	Tamanho (MB)	Custo por Unidade de disco	Sugestão "Greedy Algoritim"	Seleção Manual	Frequência de uso
Requisito 1	1000		<input type="checkbox"/>	<input checked="" type="checkbox"/>	4
Requisito 2	2000		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	5
Requisito 3	3000		<input type="checkbox"/>	<input checked="" type="checkbox"/>	10

Escolha o assunto do modelo dimensional: ▼

Espaço necessário para todas as visões: 6000 MB

Entre com o espaço disponível para o projeto: 10000 MB

Entre com o espaço temporal para a frequência de uso: diário ▼

Requisitos de usuário

Preencher

Materializar

Figura 2 - Cenário do Módulo SVM

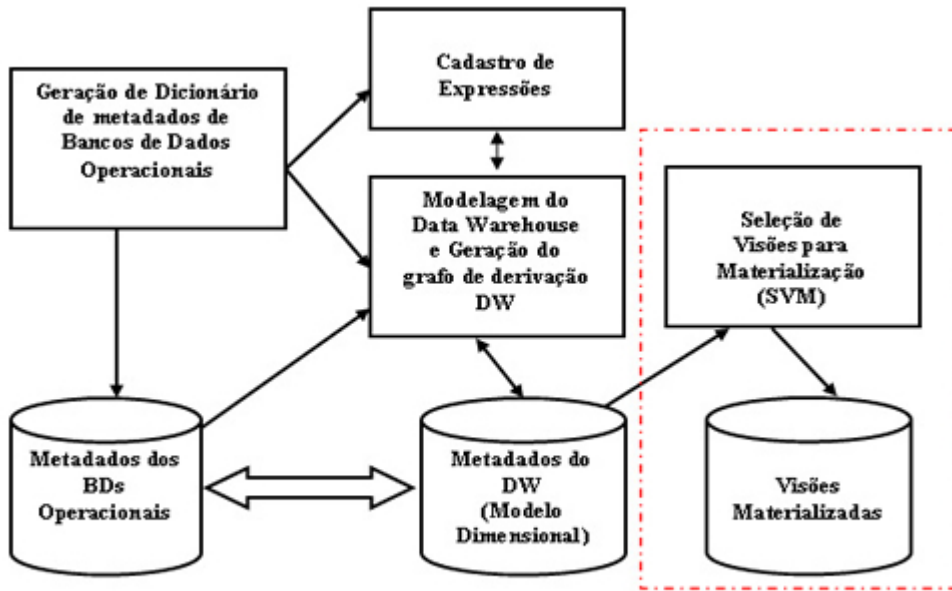


Figura 1 - Arquitetura Proposta para a Ferramenta – DW Designer