

## Uso de interfaces amigáveis para elaboração de consultas a sistemas OLAP

Mariana Duprat, Esdi/Uerj/Andre Monat, Esdi/Uerj

Palavras-Chave: OLAP interface MDX

### Resumo

Esse artigo elabora a proposta de um software que está sendo desenvolvido com o propósito de permitir que pessoas menos experientes em informática consigam acessar bases de dados OLAP. Estes normalmente exigem o uso da linguagem computacional formal MDX. O uso dessa linguagem torna difícil o acesso aos resultados e consultas desejadas por pessoas leigas nessa área. O desenvolvimento de uma interface simples que permita as funções mais comuns fornecidas pelo MDX possibilitaria o acesso desses bancos de dados por outros usuários, usando somente as ferramentas simples que satisfaçam suas necessidades.

### Abstract

This article elaborates on the proposal of a software that is being developed with the purpose of allowing non Information Technology experts to access OLAP data bases, in situations where the usage of a formal computational language, as Microsoft MDX, is normally required. We propose a friendly interface that makes available through graphics components the most common functions provided by MDX. Our intention is to facilitate the access to OLAP databases and therefore to make it available to a broader community of clients.

## 1. INTRODUÇÃO

A análise de dados se insere como um recurso fundamental quando queremos coletar, organizar e comparar informações. Temos, com isso, uma visão mais verdadeira da situação em foco, permitindo que nossas decisões sejam fundamentadas. Como resultado, temos uma noção mais ampla do contexto em que estamos inseridos, o que nos permite prever os resultados de nossas presentes decisões. Resumindo, a análise de dados tem um papel essencial na atual área de gestão.

Para garantir a eficiência, informações estratégicas devem estar disponíveis entre gerentes de altos cargos para que suas decisões sejam consistentes. Esses gerentes não podem se fiar somente em relatórios comuns do setor de Tecnologia da Informação (IT), pois precisam de relatórios para decisões específicas. Normalmente esses demoram a ser construídos e só são entregues quando não são mais necessários. Isso ocorre porque os profissionais dos setores IT são muito ocupados e estão envolvidos em soluções de problemas cotidianos, não podendo parar suas tarefas para fornecer os relatórios necessários para tomada de decisões. Infelizmente, a gestão de empresas é um processo muito competitivo, há um momento certo para cada decisão, e os relatórios devem estar disponíveis quando necessário, senão se tornam irrelevantes.

Para a análise de informações cotidianas, são utilizadas bases de dados denominadas On Line Transactional Process (OLTP). Esses bancos de dados não são eficientes na construção de relatórios estratégicos, pois são desenvolvidos para relatórios que dizem respeito a operações rotineiras. Além disso, qualquer consulta aplicada a esse tipo de sistema deve ser expressa por uma linguagem computacional formal (SQL, por exemplo), fazendo com que pessoas inexperientes em IT não consigam acessar diretamente as informações. Essa é a principal razão para rejeitarmos o sistema OLTP com uma ferramenta estratégica.

Bases de dados OLAP (On Line Analytical Processing) são muito diferentes, são projetadas para que leigos em IT possam construir os relatórios desejados. Essas bases de dados, assim como outros sistemas OLAP, têm capacidade para armazenar uma quantidade gigantesca de informações baseadas em dados históricos. Ao analisar esses dados, os gerentes podem chegar a conclusões sobre tendências, segmentos de mercado, público alvo, entre outras conclusões.

Para possibilitar as análises por usuários menos experientes, bases de dados OLAP, também conhecidas como Data Warehouses, são construídas conforme uma arquitetura bem intuitiva. Essa possui tabelas factuais e dimensionais e é chamada de arquitetura estrela ou esquema estrela. Por exemplo; a figura 1 mostra um esquema estrela para uma rede de supermercados, o exemplo "Food Mart" disponível no sistema OLAP da Microsoft.

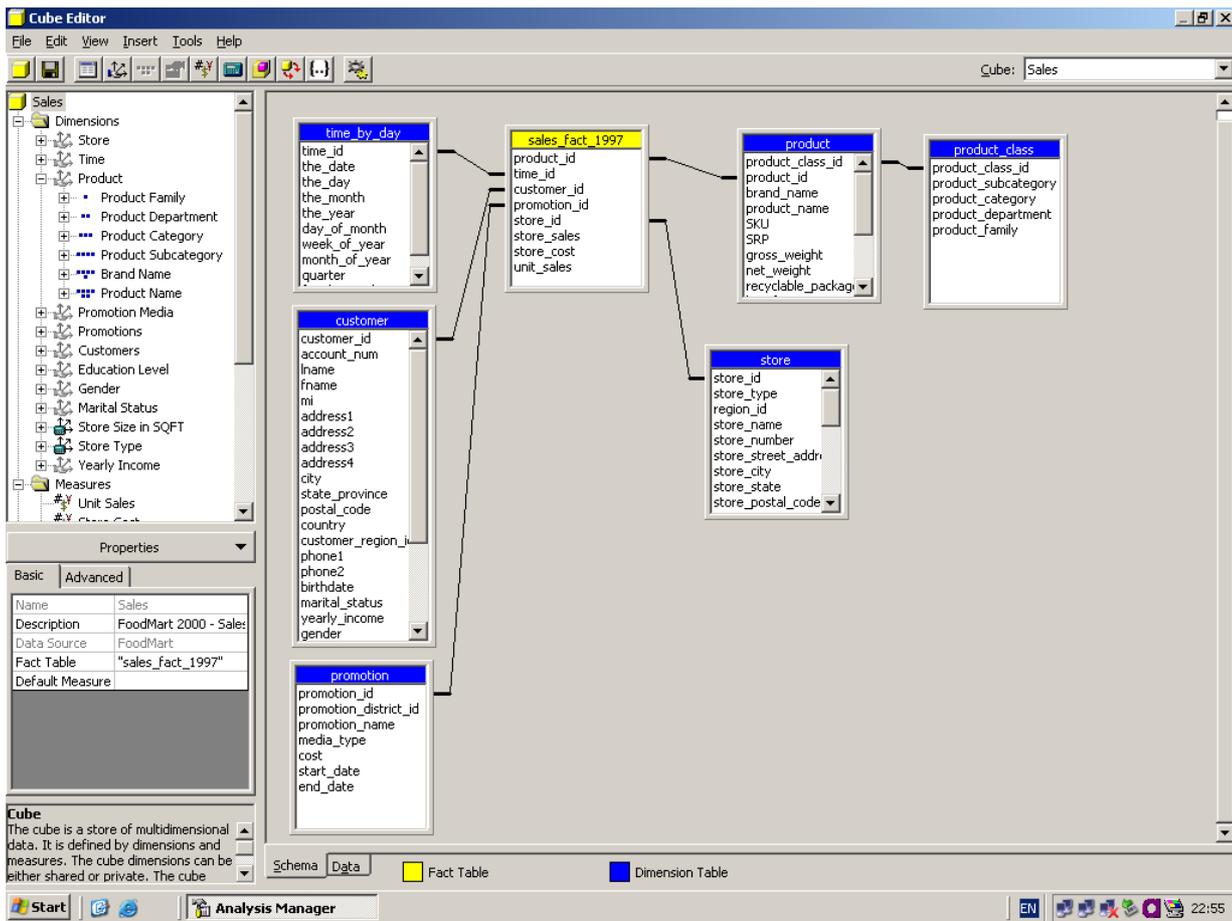


figura 1

As tabelas factuais armazenam valores numéricos denominados medidas. Essas medidas são atributos numéricos de fatos referentes ao negócio em questão. Por exemplo; para um supermercado, a venda de cada produto poderia ser vista como um fato a ser guardado para uma análise posterior, e preço, desconto, peso e número de unidades seriam seus atributos numéricos.

As tabelas dimensionais fornecem a relação entre o fato e suas características. No exemplo da rede de supermercados, poderíamos considerar Consumidor, Loja, Produto, Dado e Tempo como dimensões para o fato que queremos armazenar. Essas dimensões tradicionalmente possuem atributos referentes à venda de um produto. Por exemplo, podemos analisar a venda de produtos comestíveis - atributo da dimensão Produto – de acordo com consumidores casados e solteiros (visto que sexo é um atributo da dimensão Consumidor). Isso nos permite construir relatórios como o demonstrado na figura 2.

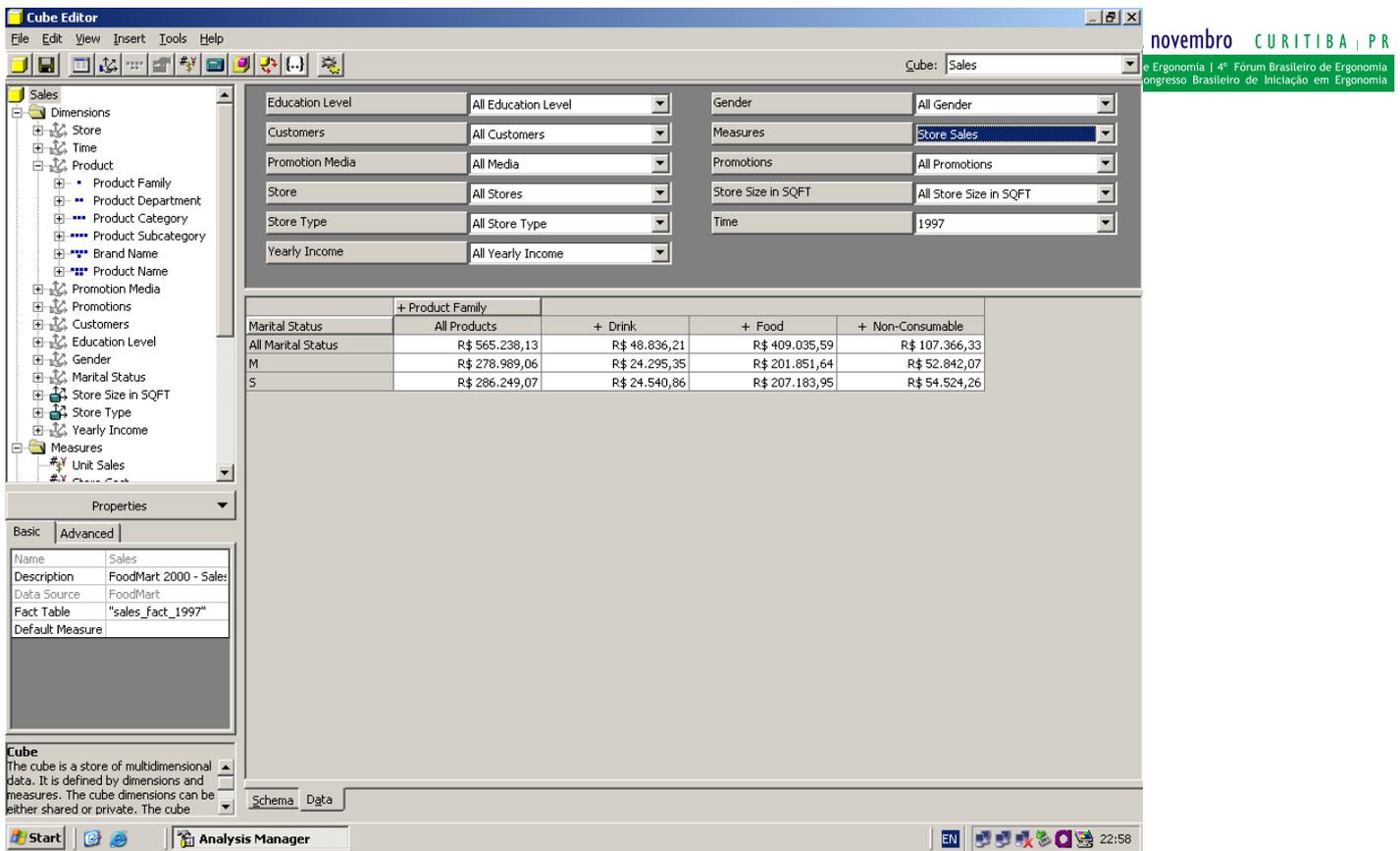


figura 2

## 2 - Analysis Manager

Existem várias ferramentas que criam um sistema OLAP de acordo com um esquema estrela. A Microsoft oferece o programa Analysis Manager como ferramenta para ser usada junto com seu sistema de banco de dados criado pelo Access ou pelo sistema SQL Server. Também é possível trabalhar com dados importados de outros formatos e sistemas. Essa ferramenta, combinada com o Excel ou o FrontPage, se transforma em um sistema front end intuitivo que constrói relatórios a partir dos dados armazenados nas Data Warehouse.

O Analysis Manager usa uma linguagem formal chamada MDX (MultiDimensional eXpression) na construção de relatórios. Na maior parte do tempo, a utilização dessa linguagem é transparente para os usuários, desde que eles permaneçam ignorantes de sua existência. Porém, para utilizarmos todo o potencial do Analysis Manager, o MDX pode se tornar uma ferramenta necessária e muito evidente. O MDX é essencial quando queremos que novas medidas sejam calculadas a partir das medidas já existentes, por exemplo. Assim como outros relatórios mais complexos também podem necessitar dessa linguagem. O MDX tem funções capazes de executar operações muito úteis, como a ordenação dos membros, a criação de novas medidas calculadas baseadas em hierarquias, a exposição dos dados de acordo com uma filtragem por critérios definidos pelo usuário, entre muitas outras.

De qualquer forma, o MDX é uma linguagem computacional formal que requer um técnico de IT, contradizendo nosso desejo primário de “dados e relatórios específicos disponíveis para todos”. O ideal seria concatenar ao máximo as possibilidades oferecidas pelas funções MDX com a facilidade oferecida pelas interfaces desses programas mais conhecidos pelos usuários.

### 2.1 - Como uma pesquisa no Analysis Manager pode ser complexa

Verifiquemos um caso complicado de resolver na interface disponível: Queremos ver somente os dois estados que tiveram as maiores vendas no período entre fevereiro e junho do ano de 1998.

Para começar, o programa não permite que façamos somente a seleção desses meses. Ou selecionamos um só membro de um nível, ou selecionamos o nível como um todo. Para tal, já teríamos que recorrer para uma expressão MDX, como a seguinte:

```
WITH MEMBER [Measures].[Store Sales 1998 Feb-Jun] AS
'SUM([Time].[1998].[Q1].[2]:[Time].[1998].[Q2].[6],[Measures].[Store Sales])'
SELECT
{{Store Sales 1998 Feb-Jun}} ON COLUMNS,
{{Store].[USA].Children} ON ROWS
FROM [Sales]
```

Em outras ferramentas, como o Excel, poderíamos escolher membros específicos em cada nível com bastante facilidade. Então, conseguiríamos chegar até esse ponto de forma simples, mas já teríamos que sair do Analysis Manager. No entanto, para filtrarmos os dois estados com as maiores vendas não teríamos outra opção. O Excel não dispõe de ferramentas para ajudar neste sentido. Logo, teríamos outra expressão com funções ainda mais complexas:

```
{TOPCOUNT ([State]. [USA].CHILDREN, 2,[Store Sales 1998 Feb-Jun])}
```

Com essa expressão (TOPCOUNT) estaríamos filtrando os dois estados com as maiores vendas. Agora, juntando essa expressão com a anterior, configuramos o relatório da forma que desejamos:

```
WITH MEMBER [Measures].[Store Sales 1998 Feb-Jun] AS
'SUM([Time].[1998].[Q1].[2]:[Time].[1998].[Q2].[6],[Measures].[Store Sales])'
SELECT
{{Store Sales 1998 Feb-Jun}} ON COLUMNS,
{TOPCOUNT([Store].[USA].Children,2,[Store Sales 1998 Feb-Jun])} ON ROWS
FROM [Sales]
```

Assim, conseguimos o planejado, no entanto, nosso antigo desejo de “dados e relatórios específicos disponíveis para todos” não seria cumprido.

Já existem alguns programas que servem como interface entre o usuário e um banco de dados complexo como o que estamos analisando, mas normalmente, em um caso como o mostrado anteriormente, a única opção é o uso do MDX.

## 3 – Novo Software

### 3.1 - Parâmetros

Os programas existentes que trabalham com sistemas OLAP com MDX podem parecer complicados para usuários menos experientes, o que torna seus recursos mais limitados. No entanto, a simplificação aqui proposta não entende como necessária a “tradução” de todas as funções já existentes e nem de todas as pesquisas possíveis para suprir as necessidades de tais usuários. Com o intuito de restringir as funções às mais utilizadas, foi feita uma pesquisa a relatórios da Microsoft que indicam quais funções são mais discutidas e empregadas. Além das funções indicadas nessas pesquisas, o senso comum também foi usado para definir algumas das pesquisas mais realizadas.

Uma forma de facilitar essa interação é fazendo uma interface mais amigável para o usuário, sem códigos e com termos mais familiares, fugindo ao máximo da idéia de que estamos lidando com um programador. Da mesma forma, é interessante possibilitar a construção das pesquisas e relatórios através de “Wizards” – ajudantes que guiam o usuário pelo processo. Entretanto, também é imprescindível permitir que essas ações sejam realizadas utilizando-se apenas a interface do programa. Essa também se torna mais familiar por possuir ferramentas com funcionamento semelhante às encontradas nos Wizards.

Com isso em mente, foram desenvolvidas ferramentas simples que satisfazem esses requisitos.

Para facilitar o entendimento das ferramentas, durante seu desenvolvimento as funções MDX foram separadas em duas categorias: visualizações e cálculos de novas medidas.

A categoria de cálculos de novas medidas foi restrita a modelos pré-formulados das medidas mais utilizadas (médias, percentuais, diferença e soma de valores, etc.) Essa medida foi tomada para permitir que o usuário consiga fazer cálculos relevantes sem necessitar de fórmulas. Alguns cálculos mais complexos estarão disponíveis, mas só os necessários para suprir as necessidades básicas.

A categoria visualizações foi posteriormente separada nas seguintes sub-categorias: visualização por dimensões (selecionando as dimensões que entrarão em cada campo da tabela), visualização por membros (selecionando exatamente quais membros e que partes das árvores hierárquicas estarão disponíveis), visualização por critérios (filtrando os dados por restrições como maiores, menores, primeiros, últimos, etc.) e modelos de visualização (sugestões pré-prontas de formas de visualização que o usuário deve tomar como base).

### 3.2 - Interface

A interface construída consiste em uma espécie de “Wizard” que fornece ao usuário uma ferramenta diferente em cada uma de suas abas. A primeira aba consiste em identificar as medidas a serem verificadas - aqui renomeadas de VALORES para facilitar o entendimento. A segunda aba - denominada TABELA - consiste em selecionar as dimensões e cada nível de sumarização aparente nas linhas, colunas e filtros (visualização por dimensões). A terceira aba - denominada MEMBROS - permite ao usuário definir exatamente quais membros de cada nível estarão aparentes (visualização por membros). A quarta aba - denominada RESTRIÇÕES - possui filtros que escondem, destacam ou restringem os membros aparentes através de critérios (visualização por critérios). Haverá ainda uma quinta aba - denominada GERAR VALORES - que permitirá que o usuário crie novas medidas a partir das existentes através de fórmulas prontas e da junção das mesmas (cálculos de novas medidas).

As abas restrições e gerar valores são semelhantes, pois se baseiam na tradução das fórmulas MDX em termos simples. Dessa forma o usuário fornece os dados necessários para suprir as fórmulas ao preencher frases que descrevem suas necessidades.

O Wizard está disponibilizado em uma interface mais ampla. Esta oferece as mesmas ferramentas ao usuário, além de permitir o gerenciamento de consultas e gráficos.

Quando uma tabela ficar pronta, ela é salva como uma consulta dentro do arquivo para permitir posteriores comparações e modificações. Junto com as tabelas serão salvos também vários gráficos referentes às mesmas. As informações de cada consulta podem ser comparadas também por ferramentas de interseção, união ou subtração.

Dessa forma, foi criada uma ferramenta que pode não suprir todas as necessidades de um usuário, mas deve facilitar muito suas consultas.

### 3.3 - Exemplificação da nova interface

Para exemplificarmos o funcionamento dessas novas ferramentas, tomaremos como modelo a pesquisa que fizemos ao demonstrar as dificuldades encontradas no Analysis Manager. Para isto, vamos exemplificar com a consulta em que queremos ver somente os dois estados que tiveram as maiores vendas no período entre fevereiro e junho do ano de 1998.

Primeiramente, selecionamos na primeira aba (VALORES) a medida a ser analisada - vendas das lojas (Store Sales). Ao passarmos para a segunda aba (TABELA), definimos que a dimensão tempo (Time)- com os níveis de

Todo o Tempo até Meses (All Time até Month) - estará visível nas colunas e a dimensão Lojas - com os níveis de Todas as Lojas até Estados (All State até All Store) – estará visível nas linhas. No momento não usaremos os filtros. (fig. 3)

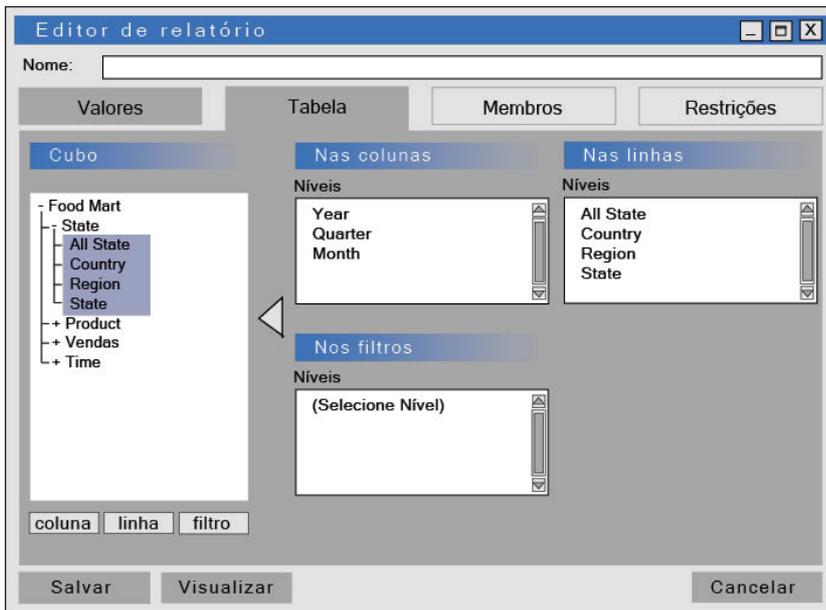


figura 3

Na terceira aba (MEMBROS) selecionamos a dimensão tempo e descemos sua hierarquia até chegar ao nível dos anos. Aqui selecionaremos somente o ano de 1998 e clicaremos na seta para seguir adiante. Passaremos pelo nível dos Trimestres (Quarters) e teremos aparentes somente os meses de 1998, onde selecionaremos os meses de fevereiro a junho (2 até 6), com a tecla “shift” apertada e concluiremos essa etapa. (fig.4)

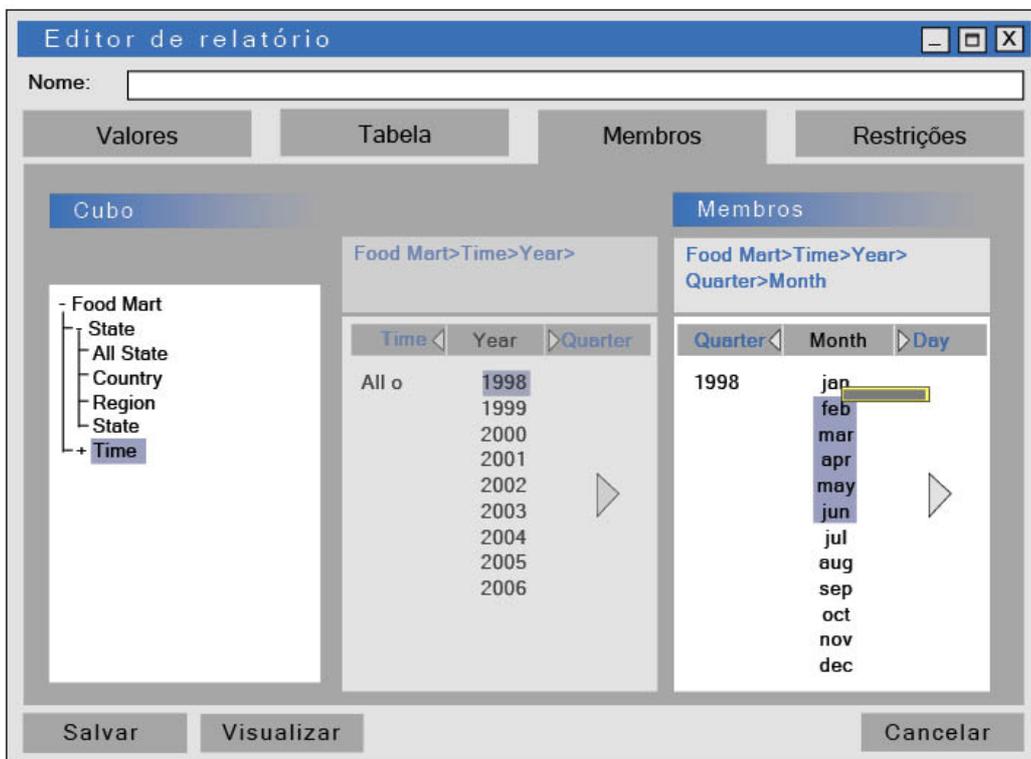


figura 4

Para obtermos a soma das vendas do período selecionado, teremos que gerar uma nova medida calculada. Para isso, selecionamos a quinta aba (GERAR VALORES). Chamaremos esse novo valor de “Vendas das lojas em fev-jun de 1998”, preenchendo o campo “Nome”. Na seção “Analisando...” selecionamos a opção “dados aparentes”, de forma a trabalhar apenas com as informações já selecionadas. Na seção “Calcular...” selecionamos a opção “soma”, o que nos fornece a seguinte frase: “a soma de (valor) dos (membros) da dimensão (selecionar dimensão)”. Na caixa “valor”, selecionamos o valor desejado (“Store Sales”), na caixa “membros”, selecionamos “membros aparentes”, e na caixa “selecionar dimensão”, selecionamos a dimensão Tempo (“Time”). Isso nos dará os valores desejados.

Na quarta aba (RESTRICÇÕES), podemos restringir somente os dois estados com maiores vendas. Para isso, na seção exibir somente, selecionaremos a função n maiores, referente a membros. Feito isso, aparecerá uma frase: membros com “n maiores (valor)”. Na caixa texto, selecionaremos a quantidade de membros que desejamos – no caso “2”; e, na caixa de (valor), selecionaremos o valor desejado – no caso, Vendas das lojas em fev-jun de 1998”. (fig. 5)

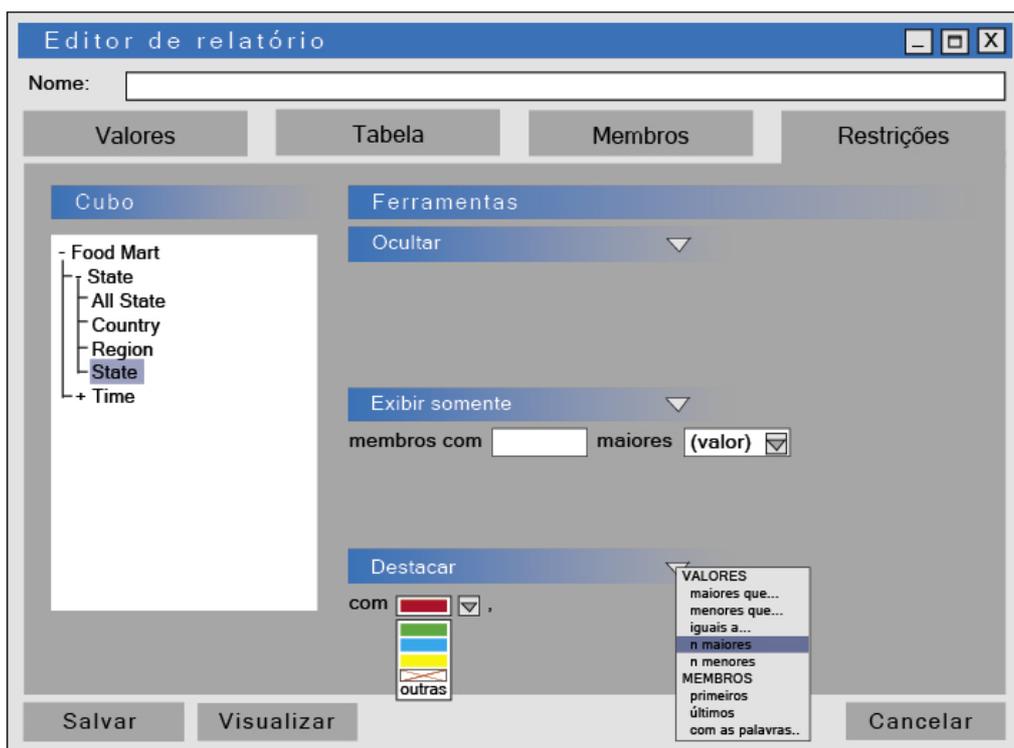


figura 5

Com isso já teremos a tabela da forma que desejamos, sem ter que escrever nenhuma fórmula, chegando à conclusão que os estados Washington (WA) e Califórnia (CA) possuem as maiores vendas do Brasil nesse período.

## CONCLUSÃO

O maior problema encontrado foi o de evitar que o usuário tenha contato com a linguagem MDX, principalmente no uso de funções mais complexas - porém ainda usuais - visto que essas não tinham uma resolução clara na interface comum. Este foi resolvido com a utilização de Wizards que se encarregavam de colocar as expressões de forma transparente para os usuários.

O programa foi separado em componentes conforme suas diferentes ferramentas, permitindo que possam ser utilizadas separadamente, podendo ser posteriormente integrados com outros softwares que permitam extensões (comumente chamadas de add-ins), como o Excel, por exemplo.

A programação desse software foi feita utilizando a linguagem VB.Net. Apesar do programa poder analisar dados recuperados de outras bases de dados OLAP, foi utilizado um sistema OLAP baseado em SQL Server 2005.

Mesmo tendo resolvido muitos problemas nessa interface, ainda há uma grande necessidade de melhorias e avanços com relação à análise de dados. Um novo caminho a ser seguido é o desenvolvimento de interfaces com funções específicas para cada área de gestão. O primeiro campo que está sendo explorado é o de Recursos Humanos (RH), com o desenvolvimento de funções “traduzidas” em interfaces amigáveis que resolvam índices específicos dessa área. Acredita-se que, como cada área tem seus índices e cálculos específicos, a interface deverá ter sempre um núcleo direcionado de componentes para cada uma delas.

## REFERENCIAS

WHITEHORN, Mark.; ZARE, Robert.; PASUMANSKY, Mosha., Fast Track to MDX. EUA: Springer, 2002.

JACOBSON, Reed., Microsoft SQL Server 2005 Analysis Services. EUA: Microsoft Press 2000.

SPOFFORD, George., MDX Solutions. EUA: John Wiley & Sons inc 2001.