

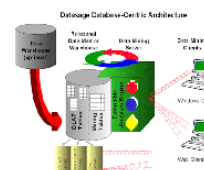


Robert Groth

Relatórios, Query e OLAP



Usama Fayyad et al



Relatórios, Query e OLAP

O propósito do DW, como componente da arquitectura de informação da empresa, é permitir uma visão informacional dos seus dados.

Esta necessidade de visão informacional deriva, por um lado:

- do enorme volume de dados originados na empresa ou provindos do exterior, que impede, na prática, qualquer conclusão e, por outro lado,
- do ambiente de tomada de decisões, em que a informação, é condição fundamental para a obtenção de vantagens competitivas.

Assim, o DW surge, como atrás já foi amplamente descrito, como um repositório de informação com características adequadas, por forma a permitir o suporte à tomada de decisões.





Relatórios, Query e OLAP

A existência do DW, por si só, não é um fim, mas antes um meio; e assim, é fundamental a distribuição da informação armazenada. Esta distribuição pode assumir três formas distintas, dando origem a três tipos de aplicações:

- **relatórios e consultas**
- **produtividade pessoal**
- **planeamento e análise**



Relatórios

Foram, durante muitos anos, o *output* privilegiado dos sistemas operacionais, nos então denominados sistemas de informação para a gestão. Os relatórios assumiam um formato pré - definido, utilizando linguagens de 3GL. Ao ser necessário um novo relatório, o analista de negócio tinha de contactar o pessoal de IS, motivando, em regra, esperas prolongadas, devido ao chamado “*Report Backlog*”. Isto fez com que ele almejasse por acesso directo e autónomo aos dados da empresa. Nesta situação, não poderemos falar propriamente de ferramentas analíticas, dado que as análises permitidas seriam muito incipientes.





Relatórios

Emissão de Relatórios:

- **Ferramentas de emissão de relatórios de produção**
 - permitem gerar relatórios operacionais regulares
 - suportam trabalhos em lotes de grande dimensão
 - linguagens de 3.^a geração e especializadas de 4.^a geração
- **Geradores de relatórios**
 - ferramentas de desktop baratas, concebidas para utilizadores finais
 - permitem ao utilizadores a criação e execução dos relatórios
 - evitam a dependência dos utilizadores do departamento de IS
 - incluem interfaces gráficas e funções de geração de gráficos
 - permitem a extracção de dados de várias fontes, integrando-as num único relatório
 - ex. Crystal Reports, InfoReports, IQ Objects



Aplicações de Query

- **Ferramentas de query**
 - Isolam os utilizadores das complexidades do SQL e estruturas das bases de dados, inserindo uma metacamada entre os utilizadores e a base de dados;
 - A metacamada é o software que proporciona vistas orientadas a assuntos numa base de dados, suportando criação de SQL através da selecção dos diversos objectos; alguns vendedores, denominam esta camada de universo ou catálogo;
 - Muitas destas ferramentas implementam arquitecturas em três níveis, para aumentar a escalaridade;
 - Algumas das ferramentas passaram também a incluir funcionalidades OLAP e de KDD
 - Exemplos: Business Objects, IQ Objects, GQL da Andyne Computing, IBM Decision Server





Aplicação Query / Relatórios: Impromptu

- Solução empresarial para database query e reporting interactivo;
- Escalaridade até 1000 utilizadores;
- Interface gráfica fácil de utilizar que unifica a utilização de aplicações de reporting e query; também apresenta o *look-and-feel* dos produtos *windows* e daí permitir a utilização das competências adquiridas em folhas de cálculo ou processadores de texto;
- Isola os utilizadores da tecnologia de base de dados subjacente, como sintaxe SQL, junções, nomes de campos etc., através do *Catálogo de Informação*;
- Este catálogo é de fácil criação, apresentando a base de dados de uma forma que reflecte a organização do negócio, utilizando a terminologia do negócio;
- Trata-se de uma aplicação de arquitectura orientada a objectos permitindo que alterações a regras de negócio, permissões e actividades de query sejam repercutidas em cascata de forma automática em toda a organização.



Aplicação Query / Relatórios: Impromptu

- Inclui modelos pré-definidos de relatórios independentes da base de dados
- Relatórios de excepção:
 - filtros condicionais - busca valores que estejam fora de intervalos definidos
 - mostra condicional - partes de relatório aparecem, mediante a satisfação de determinada condição
 - Relevos condicional - criar regra para formatação de dados com base nos valores dos dados.
- Através do Request Server, permite que as tarefas sejam executadas no servidor, reduzindo o tráfego na rede, aumentando o desempenho (utilizar as facilidades do servidor e não o PC), poder executar tarefas de noite;
- Suporta uma variedade de bases de dados: Oracle, SQL Server, Sybase, OmniSQL, Informix, DB2, Ingres, Paradox, de entre outras e acessos ODBC.





Aplicações de Produtividade Pessoal

Aplicações de Produtividade Pessoal, por exemplo folhas de cálculo, pacotes estatísticos e ferramentas de gestão gráficas. São úteis para manipular e apresentar dados em PCs individuais. Desenvolvidas para ambientes isolados, estas ferramentas servem aplicações que requerem apenas pequenos volumes de dados. Tomam a forma das chamadas ferramentas *drill* para PC que emergiram no final dos anos 80, como uma alternativa aos EIS baseados em *mainframes*.

Estas ferramentas capitalizaram a adopção explosiva dos PCs e permitiram aos utilizadores a análise dos dados que residiam nos seus sistemas locais. Introduziam um avanço significativo na análise multidimensional, materializada no manejo informacional na forma de um hipercubo. A existência deste modelo multidimensional no PC permitiu uma análise de alto desempenho dos dados sediados localmente.



Aplicações de Produtividade Pessoal

Possuíram e possuem um conjunto de limitações, tais como:

- Procura de recursos técnicos - a complexidade das ferramentas *Pc drill* requer tipicamente a atenção de especialistas IS para a criação e manutenção dos hipercubos, dado que poucos utilizadores terão os conhecimentos técnicos que lhe permitam ser auto-suficientes;
- Relevância dos dados – com a taxa rápida de actualização dos dados nos sistemas operacionais, os dados locais depressa ficam desactualizados, tornando-se irrelevantes ou obsoletos;
- Falta de integração – Estas ferramentas *drill* não são integradas com as ferramentas de consulta e emissão de relatórios, obrigando a migrações sucessivas dos dados, por diversas plataformas e ferramentas. Este processo inviabiliza decisões *on-the-fly*, além dos inevitáveis desperdícios de tempo;
- Origina ilhas de informação – dada a proliferação de extracções estáticas de dados para PC's onde passam a residir, as empresas acumularam muitas ilhas de informação, não havendo partilha.





Aplicações de Produtividade Pessoal

- Mesmo na presença da informação integrada do DW, as análises continuam a revelar-se de âmbito restrito;
- Constituem uma versão actual das velhas aplicações de extracção de dados dos ambientes operacionais;
- Cada nova análise implica, muitas vezes, um novo processo de extracção, solução apenas viável, em situações de análise sectorizada da empresa;
- A sua funcionalidade é reduzida, já que o volume de dados que permite tratar, está sempre limitado à exígua memória do PC e à sua igualmente reduzida capacidade de processamento.



Aplicações OLAP

As Aplicações de Planeamento e Análise vão ao encontro de requisitos essenciais para o negócio como: orçamentação, previsão, rentabilidade de produtos e clientes, análise de vendas, consolidação financeira e análise de linha de produtos produzidos - aplicações que usam dados históricos, projectados e derivados.

“As funções de planeamento e análise obrigam a que a organização olhe não só para o passado, mas, com mais importância, para o desempenho futuro da empresa. A análise dos dados históricos, em conjunto com projecções futuras, é crítica para o sucesso das empresas da actualidade.”.





Aplicações OLAP

- Resultam da combinação de tecnologias multidimensionais e cliente/servidor. Constituem uma chamada segunda e terceira gerações de soluções analíticas.
- A segunda geração é representada por servidores de produtos OLAP:
 - integrados numa arquitectura cliente/servidor
 - os utilizadores, usando os seus computadores, acedem a uma base de dados hospedada num servidor multidimensional
 - permite aos utilizadores tirar benefícios de informação partilhada e da distribuição de recursos proporcionada pela arquitectura cliente/servidor.



Aplicações OLAP

- Os benefícios não ficaram disponíveis para a totalidade dos utilizadores:
 - muitos deles necessitam de efectuar análise do tipo *ad hoc*
 - dispõem de bases de dados relacionais que importa utilizar
- Ideia:
 - adicionar uma camada situada acima da base de dados relacional bidimensional que efectuasse o mapeamento para o formato multidimensional
- A aplicação prática desta ideia faz surgir a terceira geração de ferramentas analíticas.

Vejamos como a resposta dada pela generalidade destas ferramentas é eficaz e como surge o OLAP.





OLAP

- O OLAP (*On-Line Analytical Processing*), assim baptizado por E.F.Codd, não é uma tecnologia isolada.
 - Surge, em simbiose com o DW, a tecnologia relacional e a SQL e como resposta a limitações desta última.
- **Os produtos comerciais RDBMS têm limitações:**
 - funções de suporte à visualização da informação são:
 - quase inteiramente dependentes de produtos de *front-end* para a visualização e consolidação de diferentes tipos de dados da empresa
 - têm limites quanto às possibilidades de
 - agregação, sumarização, consolidação, soma, visualização e análise
 - análise segundo múltiplas dimensões.
- A análise segundo **dimensões múltiplas** é denominada de **Análise de Dados Multidimensional - Multidimensional Data Analysis**, MDA, que constitui uma das características do OLAP.



Soluções OLAP

O OLAP surge primeiro suportado:

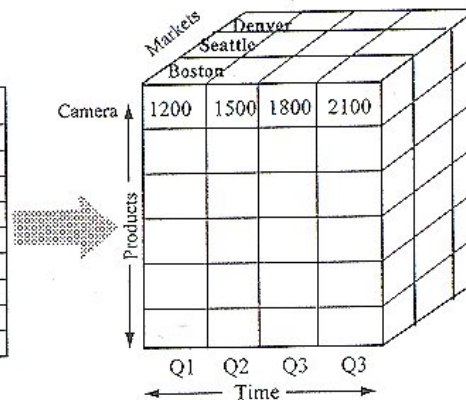
- por bases de dados fisicamente multidimensionais (MOLAP), dada a natureza fisicamente multidimensional das análises necessárias;
- depois, já como *front-end* para DW relacionais (ROLAP), suportado por um motor multidimensional que efectua o mapeamento bi-multidimensional (seguinte)
 - são colocados múltiplos queries SQL, efectuem-se múltiplas junções, agregações e ordenações; o motor multidimensional responde a algumas limitações do SQL quanto a funções matemáticas complexas e séries temporais (ex. média móvel por trimestre).
- actualmente há tendência, em muitas das ferramentas, em promover a fusão das duas tecnologias, reunindo o melhor dos dois mundos, surgindo o que convencionou chamar de HOLAP (Hybrid OLAP)





Mapeamento Bi-Multidimensional

Product	Market	Time	Units
Camera	Boston	Q1	1200
Camera	Boston	Q2	1500
Camera	Boston	Q3	1800
Camera	Boston	Q4	2100
Camera	Seattle	Q1	1000
Camera	Seattle	Q2	1100
.....			
Tuner	Denver	Q1	250
Tuner	Denver	Q2	300



Natureza do Modelo Multidimensional

A questão não é: “Qual o lucro gerado pelo novo produto?”
Mas sim: “Qual o lucro gerado pelo novo produto por mês, na zona norte, por armazém, relativamente à versão anterior do produto?”

E.F.Codd acrescenta: “há tipicamente um número de diferentes dimensões, através das quais um dado conjunto de dados pode ser analisado, de uma forma plural, surgindo uma visão conceptual multidimensional, que é a forma normal, pela qual os analistas do negócio vêem a sua empresa”. Cada uma dessas perspectivas é considerada como que uma dimensão complementar, constituindo a sua análise simultânea, a análise de dados multidimensional.

Cada um destes caminhos de consolidação de dados ou dimensões constitui uma série de níveis de consolidação ou degraus (constituindo hierarquias), que são definidos em termos de parâmetros multinível. Esses parâmetros aplicam-se a valores de qualquer variável, onde cada nível sucessivo representa um nível mais alto de consolidação.





Caminhos de Consolidação e operações

Os caminhos de consolidação irão permitir as operações de *drill-down* e *roll-up*

Drill-Down - traduz o caminhar para um nível de maior detalhe, ou seja, a operação que se refere ao movimento de nível de consolidação mais alto para níveis mais baixos

Roll-Up - consiste na operação inversa, isto é, o movimento de nível de consolidação mais baixo para mais alto.

Drill Anywhere - Operação que incluirá as duas anteriores e ainda a de ***drilling across***, que se traduz na movimentação lateral de um conjunto de dados para outro, ao mesmo nível.



Caminhos de Consolidação e operações

Slice and Dice - Operação que permite visualizar informação consolidada, segundo diversas perspectivas, utilizando o corte e rotação de um pedaço particular de dados agregados (o cubo de dados), segundo uma qualquer dimensão.

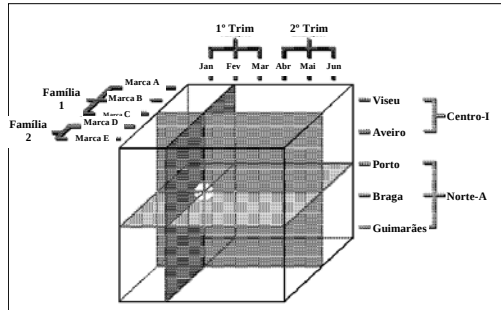
- Por exemplo, um analista de negócio numa companhia de venda de bens de consumo, pode visualizar uma fatia (*slice*) do modelo de dados, que mostra os lucros das vendas para os produtos do tipo bebidas não alcoólicas da cidade de Beja, durante os quatro trimestres de 1997. O analista pode depois focar a sua atenção apenas no produto Coca-Cola, visualizando os lucros respectivos, em todas as cidades da zona Sul, no mesmo período.





Modelo de Dados Multidimensional

- Uma forma de olhar o modelo multidimensional é vê-lo sob a forma de um cubo, multicubo ou hipercubo.



Um cubo não é mais o conceito de uma matriz, expandida a 3 dimensões (ou a n dimensões no caso do hipercubo ou multicubo). Cada elemento de armazenamento do cubo conterá o(s) valor(s) para o(s) facto(s) que se estiver(em) a analisar, relativamente às ocorrências das dimensões que se cruzarem, no elemento de armazenamento em causa.



Características dos Dados Multidimensionais

- O tempo de resposta de um query multidimensional dependerá de quantas células terão de ser adicionadas on the fly.:
 - se n.º de dimensões aumenta, o n.º de células do cubo aumenta exponencialmente
 - mas a maioria das consultas lida com dados agregados:
 - solução: pré-agregar todos os subtotais lógicos, utilizando as hierarquias nas dimensões
 - mas... explosão de tamanho
 - ex. suponhamos 5 dimensões, cada uma com 4 níveis na hierarquia, por cada valor base teríamos de calcular 20 novos valores





Características dos Dados Multidimensionais

Felizmente:

- nem todas as células do cubo de dados estão ocupadas, dado que nem todas as células têm significado através de todas as dimensões.

Contas	Entidades	Períodos	Versões	Tamanho Teórico do Hiper-cubo	Número efectivo de Células Povoadas	Sparcity (Dispersão)
1,000	200	40	4	32,000,000	800,000	98%

- Surge assim a
 - Densidade: percentagem de células do cubo de dados que estão ocupadas, relativamente ao total teórico possível. O seu inverso é a dispersão.
 - Tipicamente, só uma pequena percentagem das células possíveis estarão ocupadas (povoadas), sendo assim de prever mecanismos, na construção física da base de dados, que permitam evitar a inclusão física da representação dessas células.



Execução de queries e em OLAP

- Características mais relevantes para o OLAP:
 - Hierarquias dimensionais
 - Gestão de dados esparsos
 - Pré-agregação
- Permitem o OLAP e são mesmo os seus pontos fundamentais pois que:
 - levam a uma redução do tamanho da base de dados
 - não necessário cálculo de valores, permitindo uma redução de tempo de resposta





Avaliação de Produtos OLAP (1)

E.F.Codd, num artigo de 93, apresenta um conjunto de doze regras para avaliação de um produto OLAP, que apenas se apresenta seguidamente, podendo a descrição aprofundada ser consultada no artigo (a tratar numa aula TP).

Regras para avaliação de um produto OLAP:

1. Visão conceptual multidimensional - Uma ferramenta deve proporcionar aos utilizadores um modelo multidimensional que corresponda aos problemas do negócio, sendo intuitivamente analítico e fácil de utilizar
2. Transparência - A tecnologia OLAP, a base de dados e arquitectura de computação e a heterogeneidade de fontes de dados deve ser transparente para o utilizador, preservando a sua produtividade e utilizando ambientes familiares de front-end (tipo Windows, Excel...)



Avaliação de Produtos OLAP (2)

3. Acessibilidade - A sistema OLAP deve aceder unicamente aos dados requeridos para executar a análise, podendo aceder a todas as fontes heterogéneas de dados.
4. Desempenho constante na execução de relatórios - À medida que o número de dimensões e tamanho da base de dados aumenta, os utilizadores não devem aperceber-se de degradação significativa do desempenho.
5. Arquitectura tipo cliente/servidor - O sistema OLAP deve estar conforme aos princípios arquitecturais tipo cliente / servidor para maximizar o desempenho, flexibilidade, adaptabilidade e interoperacionalidade.
6. Dimensionalidade genérica - Qualquer dimensão de dados deve ser equivalente na sua estrutura e capacidades operacionais.
7. Manuseamento dinâmico de matrizes esparsas - O sistema OLAP deve poder adaptar o seu esquema físico ao modelo analítico específico que optimize o manuseamento de matrizes esparsas para atingir e manter o nível de desempenho requerido.





Avaliação de Produtos OLAP (3)

8. Suporte multiutilizador - O sistema OLAP deve poder suportar um grupo de utilizadores em trabalho concorrente num modelo específico.
9. Operações inter-dimensões não restringidas - O sistema OLAP deve ser capaz de reconhecer as hierarquias dimensionais e efectuar automaticamente os cálculos roll-up associados entre e através das dimensões.
10. Manipulação de dados intuitiva - reorientação de caminhos de consolidação, drill-down e roll-up e outras manipulações devem ser executadas via acções do tipo point-and-click e drag-and-drop sobre células do cubo.
11. Emissão de relatórios flexível - Deve haver habilidade de arranjar linhas, colunas e células numa forma que facilite a análise através de apresentação visual em relatórios analíticos.
12. Número de dimensões e níveis de agregação ilimitados - Um modelo analítico pode ter doze ou mais dimensões, cada uma com hierarquias múltiplas. O sistema OLAP não deve impor quaisquer restrições artificiais no número de dimensões ou níveis de agregação.



Características Complementares

- Ferramentas de gestão de bases de dados - Devem funcionar como ferramenta integrada centralizada
- Habilidade de fazer drill-down até ao detalhe - Isto quer dizer que a ferramenta deve permitir uma transição simples da base de dados multidimensional para o nível de detalhe
- Refrescamento da base de dados incremental - Para que não surjam problemas com o aumento de tamanho da base de dados
- Interface SQL - Um requisito importante para que o sistema OLAP fique perfeitamente integrado no ambiente.





Objectivos Chave das Architecturas Multidimensionais

- Proporcionar um acesso aos dados rápido, de resposta linear e independente do modo como os dados são pedidos, ou seja de onde os dados serão lidos no hipercubo.
- Proporcionar dados calculados, não só sob a forma de agregações, mas todo um conjunto de cálculos, onde qualquer célula do hipercubo possa ser derivada de outras, usando funções standard de negócio ou estatísticas, incluindo lógica condicional.
- Solução:
 - pré-cálculo dos resultados.



Pré-Cálculo dos Resultados

Esta técnica pode ser muito interessante:

- mas não é apropriada, quando o tamanho do modelo completamente calculado é milhares de vezes superior ao volume de dados origem;
- situação que poderá ocorrer, particularmente, quando o número de dimensões é elevado e as hierarquias em cada dimensão são extensas.

O pré-cálculo dos valores agregados:

- não só é consumidor de espaço;
- mas principalmente de tempo.

Para as duas das arquitecturas que descreveremos é sinónimo de desempenho:

- há que efectuar o balanceamento, entre a velocidade de acesso aos dados e a minimização do fenómeno da explosão de dados, através da adição de suporte de cálculos complexos e actualizações em tempo real.





Balanceamento de Pré-Cálculo dos Resultados

Das diversas abordagens, quanto a este balancear e também relativa às tecnologias de armazenamento utilizadas, surgem as três principais arquitecturas OLAP:

- **ROLAP (Relational On-Line Analytical Processing)** tb. denominada de **Multirelacional OLAP**,
- **MOLAP (Multidimensional On-Line Analytical Processing)**, também conhecida por **MD-OLAP** e a recentemente introduzida
- **RAP (Real-Time Analytical Processing)**.

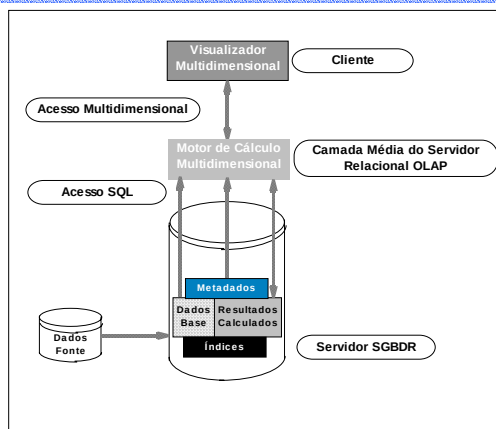


Visão Multidimensional de Dados Relacionais

A visão multidimensional dos dados relacionais conduz à arquitectura representada ao lado.

Máxima adoptada: todos os dados devem ser armazenados em bases de dados relacionais.

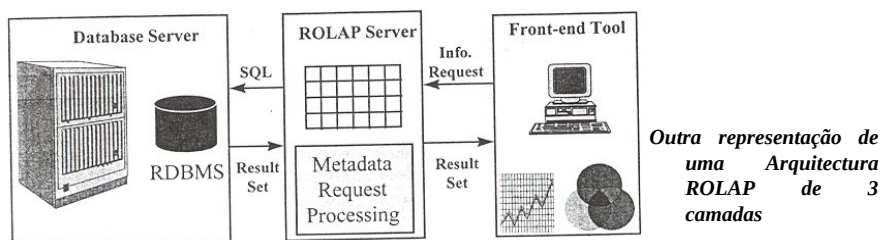
Proporcionam, depois, uma visão multidimensional dos dados armazenados em tabelas relacionais.



Arquitectura simplificada de uma aplicação OLAP relacional de 3 camadas.



Arquitetura ROLAP



Trata-se, normalmente, de uma arquitectura a três níveis: a base de dados relacional, o motor multidimensional e a ferramenta cliente de apresentação e diálogo com o utilizador; os dois últimos podem estar eventualmente fundidos num só, resultando assim numa arquitectura a dois níveis.



Arquitetura ROLAP

Funcionamento deste tipo de arquitectura:

- a ferramenta cliente aceita o pedido do utilizador (normalmente apresentando os dados disponíveis, ferramentas de cálculo e funções, sob a forma de objectos e nomes familiares ao utilizador),
- o pedido é depois convertido nas consultas SQL convenientes, ainda na ferramenta cliente ou já no motor multidimensional,
- o SQL é enviado à base de dados relacional que, a seguir, enviará a resposta ao motor multidimensional,
- depois de manipular adequadamente os dados recebidos, remete-os ao cliente, já sob a forma de dados multidimensionais.





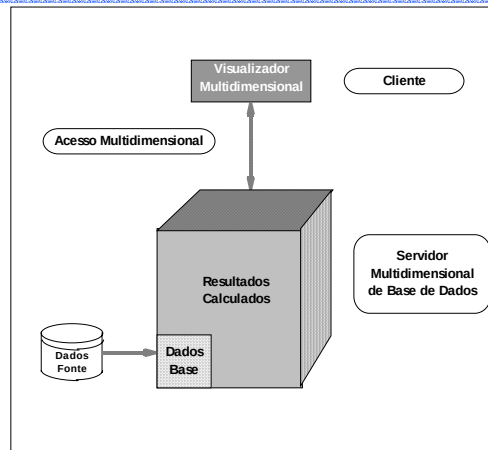
Arquitectura ROLAP: prós e contras

- Virtudes:
 - quase ilimitada escalaridade;
 - utilização do modelo relacional, amplamente conhecido e estabelecido.
- Limitações:
 - “calcanhar de Aquiles” é ser fundamentalmente “só de leitura”;
 - a velocidade:
 - já que o SQL standard, criado com outros propósitos em mente que não OLAP, (mas sim OLTP), não está adaptado às funções analíticas comuns em OLAP;
 - situação está a ser ultrapassada, em vendedores de bases de dados relacionais vocacionadas para DW, que incluem extensões ao SQL standard, além de optimizações no motor para execução das consultas (optimizações de query para esquema em estrela).



Bases de Dados Multidimensionais

- É abandonada a comodidade dum RDBMS e criada uma nova forma de armazenamento, conduzindo à arquitectura representada ao lado.
- Máxima adoptada: “se a análise é feita multidimensionalmente, por que não aproximar a forma de armazenamento ao modo de análise, armazenando então os dados multidimensionalmente?”



Arquitectura simplificada de uma aplicação OLAP Multidimensional



Bases de Dados Multidimensionais

Arquitetura própria, com técnicas, como:

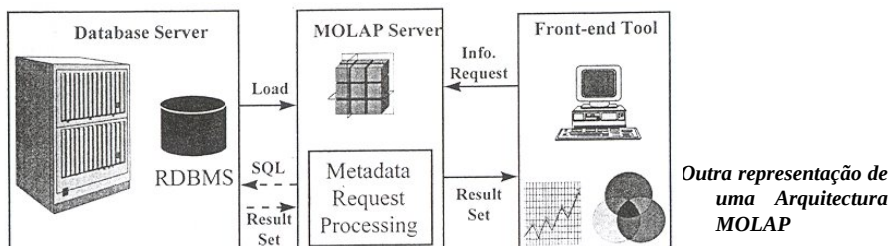
- mapeamento de valores para compressão;
- utilização de índices de apontadores para *arrays* comprimidos de valores;
- algoritmos sofisticados de *caching*.

O armazenamento físico dos dados consiste em objectos tipo *array* comprimidos, normalmente com índices e estruturas de apontadores muito compactas. Cada grupo de objectos do tipo *array* (ou blocos), consistirá em grupos de células que podem ser acedidas individualmente, usando cálculos de *offset* directos, em oposição a qualquer forma de indexação.

Estes blocos são somente armazenados, se contiverem dados (e muitos não os terão, dada a já aludida dispersão dos dados multidimensionais), que, uma vez armazenados, são localizados, usando alguma forma de estrutura de indexação. Não são necessárias chaves para os blocos individuais, dado que o sistema de indexação identifica directamente a localização de cada bloco no disco. Como o índice só necessita de identificar um número relativamente pequeno de blocos e não muitas linhas individuais, as MDDBs utilizam normalmente índices muito pequenos, em média de 2% do tamanho total da base de dados), podendo ficar residente em memória.



Aproximação MOLAP



Estamos em presença duma aproximação, em que é utilizada uma base de dados desenhada especificamente para suportar dados multidimensionais e que pré - calcula todos os valores derivados.

Este facto resulta em tempos de resposta da ordem dos segundos, sendo, no entanto, o menos preparado para resistir ao problema da explosão dos dados.





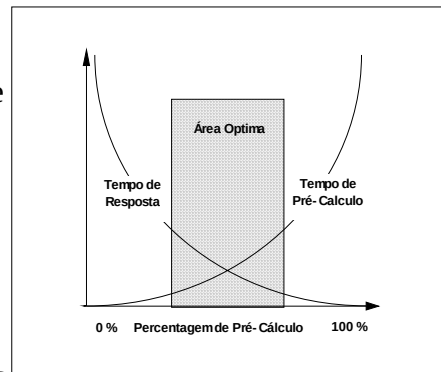
Pré-cálculo de Sumarizações e Tempo de Resposta

Do gráfico ao lado, é mostrado o tempo de resposta e tempo de pré - cálculo, relativamente à percentagem de pré - cálculo:

- é um argumento da tecnologia ROLAP ou RAP.

A favor da abordagem MOLAP:

- teremos o tempo de resposta curto;
- normal completa integração dos diversos componentes que constituem a arquitectura, o que lhe permite uma melhor optimização.



Relação entre o tempo de pré-cálculo de sumarizações e o tempo de resposta.



RAP (Real-Time Analytical Processing)

É uma aproximação que explora o outro extremo da tecnologia de bases de dados multidimensionais:

- **inexistência** de dados **pré - calculados**.

Todos os dados derivados são calculados “a pedido”, evitando-se, assim, a explosão de dados e os tempos prolongados de pré - cálculo.

Reverso da medalha:

- os dados têm de residir em memória, por forma a que os tempos de resposta possam ser rápidos;
- limitação do volume de dados passível de manipulação.



RAP (Real-Time Analytical Processing)

Ajuda para as limitações:

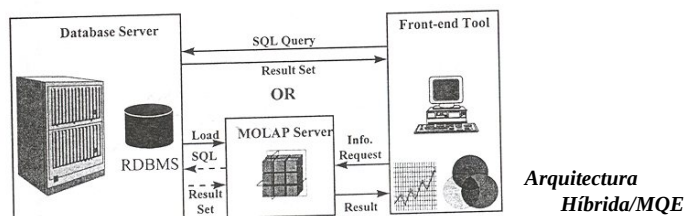
- armazenamento temporário dos valores calculados (enquanto válidos), com vantagens:
 - só as agregações utilizadas são efectivamente realizadas;
 - num ambiente dinâmico, interactivo e com actualizações, como é o caso dum modelo orçamental, os cálculos são sempre os mais actuais.
- armazenar os dados muito eficientemente, geralmente, utilizando 10-15 bytes por número, não armazena identificadores de dimensões;
- permite assim armazenar 4 biliões de células teóricas em 512 Mb, mais do que suficiente (segundo os proponentes da tecnologia) para o tipo de aplicações analíticas, alvo desta aproximação.

APP, MQE ou HOLAP

APP - Aplicações de Produtividade Pessoal

MQE - Managed Query Environment

HOLAP - Hybrid OLAP



Arquitetura
Híbrida/MQE

Alguns produtos de data query e report desenvolveram características para proporcionarem capacidade de análise “datacube” e “slice and dice”. Este cubo de dados pode ser armazenado e mantido localmente para reduzir a sobrecarga traduzida na necessidade de criar a estrutura todas as vezes que uma consulta fosse colocada. Tb. podem trabalhar com servidores OLAP e destes aceder a servidores relacionais.



APP, MQE ou HOLAP

Apelos para esta aproximação:

- Simplicidade de instalação e administração;
- Ideais para proporcionar a utilizadores ocasionais capacidades de análise mais sofisticadas;
- Custos bastante inferiores aos relativos a produtos mais complexos.

Limitações:

- Redundância de dados;
- Sobrecarga de rede;
- Rápida desactualização dos metacubos gerados.

Exemplos:

- PowerPlay da Cognos
- Pablo da Andyne
- Mercury da Business Objects
- CrossTarget da Dimensional Insight's
- Media da Speedware



Apreciação Comparativa entre as Aproximações

	ROLAP	MOLAP	RAP
Versatilidade	Muito boa. Permite balancear tempo de resposta a consultas, face aos requisitos de processamento.	Não versátil. O grau de compilação é elevado, implicando grandes requisitos de pré - processamento.	Muito boa. Não carece de pré - compilação de agregações.
Portabilidade	Muito boa. Pode ligar-se a bases de dados de diversos vendedores e plataformas.	Ainda incipiente. Relativamente difícil de trabalhar com bases de dados relacionais. Normalmente depende inteiramente do motor multidimensional, de que faz parte integrante.	Constitui uma ferramenta dotada de motor multidimensional e interface cliente, pelo que a questão não se coloca.
N.º de Dimensões e Volume de Dados	Grandes. As RDBMS suportam bases de dados até às dezenas de TB. O n.º de dimensões, pode também ser da ordem das várias dezenas, trabalhando com graus de pré - compilação baixos.	Só até algumas dezenas de GB e 5-6 dimensões.	Trabalha em memória. Suporta modelos com poucos dados, podendo no entanto suportar dezenas de dimensões.
Velocidade de Acesso aos Dados	Um pouco inferior a MOLAP, para graus de agregação idênticos.	Trabalha com graus de agregação elevados e com estruturas multidimensionais optimizadas para acesso aos dados, é muito rápida. Tempos de resposta típicos da ordem dos 3-5 segundos.	Muito rápido. Toda a base de dados reside em memória, executando as agregações on-the-fly.
Facilidade de Mudança	Bom, para graus de pré - agregação baixos.	Em regra, obriga a grandes reorganizações, visto que trabalha com graus de pré - agregação elevados.	Muito boa.
Segurança e Controlo de Acesso	Só controla acesso a tabelas e registos. Não reconhece a célula ou dimensão.	Muito boa. Controla o acesso ao nível de célula ou dimensão, já que a MDDB, conhece intrinsecamente esses conceitos.	Semelhante à MOLAP, mas de características mais simplistas, pois que, em regra, não se coloca a questão, numa situação de utilização típica.
Explosão de Dados	Limitada; permite a selecção da zona de funcionamento do modelo, a nível de pré-agregações.	Dramática, para muitas dimensões.	Não há.
Possibilidade de Leitura-Escrita	Normalmente só de leitura.	Leitura / Escrita.	Leitura / Escrita.
Utilização Típica	De utilização genérica. Adaptado idealmente para pesquisas em grandes bases de dados relacionais, com esquemas em estrela. Sectores de telecomunicações, banca, seguros, supermercados, etc.	Mais utilizado a nível de data marts, para consultas de índole pré-definida e com possibilidade de análise de cenários what-if.	Ideal para aplicações que necessitem de entrada de dados frequentes ou análises what-if. Também interessante em ambientes móveis e distribuídos, já que o armazenamento é muito eficiente.





Ferramentas OLAP e a Internet

O casamento entre as tecnologias e conceitos DW e a Internet foi natural e inevitável. Razões:

- A Internet é um recurso livre que proporciona uma conectividade universal dentro e entre empresas;
- A Web facilita tarefas administrativas complexas de gestão de ambientes distribuídos;
- A Web permite às empresas armazenar e gerir os dados e aplicações em servidores que podem ser geridos, mantidos e actualizados centralmente, eliminando problemas relativos a concorrência de software e dados.

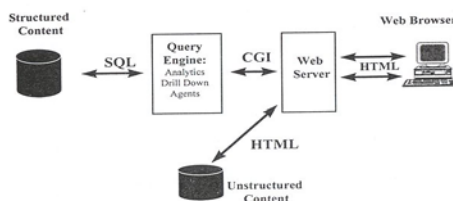
Gerações:

- **Primeira geração: Páginas HTML estáticas**
- **Segunda: Queries interactivos através de gateways**
- **Terceira: Servidores de aplicações Web, utilizando applets Java ou ActiveX**



Gerações de Ferramentas OLAP / Internet

- **1.ª** - Modelo estáticos de distribuição - os clientes acedem a páginas estáticas HTML através de browsers. Acesso a relatórios DSS armazenados como documentos HTML.
 - limitações: inabilidade de proporcionar a clientes Web capacidades analíticas interactivas como drill-down.
- **2.ª** - Suporte de queries interactivos - através de uma arquitectura multinível na qual um cliente Web submete um query na forma de um pedido codificado em HTML a um servidor Web, que o transforma num pedido de dados estruturados (a um gateway p. ex.) que coloca o pedido em SQL à base de dados, recebe os resultados, tradu-los para HTML e envia as páginas para o cliente que fez o pedido.





Gerações de Ferramentas OLAP / Internet

- 3.^a - Os gateways HTML são substituídos por servidores de aplicações baseadas na Web. Applets Java ou ActiveX são transferidas e executadas em clientes ou interactivam com applets correspondentes que correm nos servidores. Permitem-se assim todas as funcionalidades de aplicações OLAP sem obrigar à existência de qualquer software cliente, excepto um browser.



Resumo e Conclusões em OLAP

- **A maioria dos produtos são actualmente compatíveis Web:**
 - **Ferramentas MOLAP** - Essbase da Arbor, Express da Oracle, Gentia da Planning Sciences, Acumate ES da Kenan Technologies, Holos da Holistic Systems e Pilot Server da Pilot Software;
 - **Ferramentas centradas em cliente** - Business Objects, Esperant da Software AG;
 - **Ferramentas centradas em servidor** - Information Advantage e Prodea Beacom da Platinum Technologies;
 - **Ferramentas ROLAP** - DSS Server e DSS Agent com o DSS Web em conjunto com um leque de outras ferramentas de modelação e concepção de EIS.

