

UNIVERSIDADE DO EXTREMO SUL CATARINENSE - UNESC
CURSO DE PÓS-GRADUAÇÃO ESPECIALIZAÇÃO EM
MBA GERENCIAMENTO DE BANCO DE DADOS

RENATO CRAMER

ESTUDO ANALÍTICO DE FERRAMENTAS OPEN SOURCE
PARA AMBIENTES OLAP

CRICIÚMA, AGOSTO DE 2006

RENATO CRAMER

**ESTUDO ANALÍTICO DE FERRAMENTAS OPEN SOURCE
PARA AMBIENTES OLAP**

Monografia apresentada à Diretoria de Pós-Graduação da Universidade do Extremo Sul Catarinense- UNESC, para a obtenção do título de especialista em Gerenciamento de Banco de Dados.

Orientador: Prof. M.Sc. Rafael Gastão Coimbra Ferreira.

CRICIÚMA, AGOSTO DE 2006

Dedico à minha querida Mãe.

RESUMO

O objetivo principal deste trabalho foi apresentar o estado atual das ferramentas open source para ambientes OLAP, oferecendo informações relevantes para os profissionais da área de informática que desejam utilizar estas tecnologias. Foi dedicada uma grande parte do mesmo para apresentação teórica dos conceitos, de OLTP a OLAP. Isso justifica-se pelo fato de que a grande maioria dos profissionais de informática, principalmente os desenvolvedores, trabalham somente com sistemas OLTP, muitas vezes desconhecendo as tecnologias de Data Warehouse e OLAP. Por este motivo também foram descritas as áreas de aplicação e os benefícios do OLAP. Houve uma preocupação em selecionar as ferramentas que pudessem oferecer uma solução OLAP próxima das necessidades dos desenvolvedores, por exemplo, possuindo servidor e cliente OLAP e rodando nos sistemas operacionais mais difundidos, ou seja, Windows e Linux. Por estes motivos esta pesquisa foi basicamente descritiva quanto aos objetivos. Além disso, teve como idéia principal que os dados operacionais dos sistemas OLTP sejam transformados em informações nos sistemas OLAP, como excelentes ferramentas que são, para o apoio à tomada de decisão de gestores e executivos.

Palavras-chave: Business Intelligence; OLAP; Modelagem Multidimensional; Open Source.

LISTA DE ILUSTRAÇÕES

Figura 1 – Arquitetura Funcional de um DW.....	19
Figura 2 – Visão de um Tesseract.....	28
Figura 3 – Cubo tridimensional produtos, meses e vendas.....	29
Figura 4 – Star Schema.....	30
Figura 5 – Entidades do Ambiente DW.....	32
Figura 6 – Snowflake Schema.....	33
Figura 7 – Volume de Vendas por Região.....	35
Figura 8 – Volume de Vendas da Região Sul.....	35
Figura 9 – Slice por Produto.....	36
Figura 10 – Dice por Venda e Tempo.....	37
Figura 11 – Drill Across.....	37
Figura 12 – Pivoting.....	38

LISTA DE TABELAS

Tabela 1 – Comparação entre atividades de processamento de informações operacionais e baseadas em análise, orientadas a decisão.....	15
Tabela 2 – Diferenças entre um Data Mart e um Data Warehouse	22
Tabela 3 – Ferramentas Descartadas	53
Tabela 4 – Características da ferramenta Pentaho – BI	58

LISTA DE ABREVIATURAS E SIGLAS

DBMS – Database Management System ou Sistema de Gerenciamento de Banco de Dados

SGBD – Sistema de Gerenciamento de Banco de Dados

SGBDR – Sistema de Gerenciamento de Banco de Dados Relacional

TI – Tecnologia da Informação

SUMÁRIO

1 INTRODUÇÃO	10
1.1 Tema	10
1.2 Delimitação do Tema	10
1.3 Problema	11
1.4 Objetivos	11
1.4.1 Objetivo Geral	11
1.4.2 Objetivos Específicos	11
1.5 Justificativa	12
2 CONCEITOS E TERMINOLOGIA	13
2.1 Open Source	13
2.2 Flat File	13
2.3 OLTP	14
2.4 Business Intelligence (BI)	15
3 DATA WAREHOUSE	17
3.1 Características	17
3.1.1 Orientado ao Assunto	18
3.1.2 Integrado	18
3.1.3 Não Volátil	18
3.1.4 Variante no Tempo	18
3.2 Granularidade	19
3.3 Arquitetura Funcional	19
3.3.1 Área Interna	20

3.3.2 Área Externa.....	21
3.4 Data Mart	22
4 OLAP.....	23
4.1 O que é OLAP?	23
4.2 O Modelo Multidimensional	24
4.2.1 Fatos	25
4.2.2 Dimensões	25
4.2.3 Membros e Hierarquias	26
4.2.4 Medidas	27
4.2.5 Cubos de Dados	27
4.2.6 Star Schema	30
4.2.7 Surrogate Key	31
4.2.8 Snowflake Schema	33
4.3 Recursos OLAP	34
4.3.1 Drill Down e Roll Up	34
4.3.2 Slice and Dice	36
4.3.3 Drill across	37
4.3.4 Pivoting	38
4.3.5 Ranking	38
4.4 Bancos de Dados Multidimensionais	39
4.5 Servidor e Cliente OLAP	40
4.6 Arquiteturas OLAP	41
4.6.1 ROLAP	42
4.6.2 MOLAP	43
4.6.3 HOLAP	43

4.6.4 DOLAP	44
4.6.5 WOLAP	45
4.7 Benefícios do OLAP	46
4.8 Áreas de Aplicação para OLAP	47
5 CRITÉRIOS PARA AVALIAÇÃO DE FERRAMENTAS OLAP	49
5.1 As 12 Regras de Codd.....	49
5.2 O Teste FASMI	50
5.3 Características Seleccionadas	51
5.3.1 Recursos OLAP	52
5.3.2 Linguagens.....	52
5.3.3 Gerais	52
6 FERRAMENTAS OPEN SOURCE PARA AMBIENTES OLAP	53
6.1 Ferramentas Seleccionadas	54
6.1.1 Pentaho - Business Intelligence.....	54
6.1.2 Mondrian	55
6.1.3 JPivot.....	56
7 METODOLOGIA.....	57
7.1 Tipo de Pesquisa	57
7.1.1 Abordagem.....	57
7.1.2 Objectivos	57
8 APRESENTAÇÃO E ANÁLISE DOS DADOS	58
9 CONCLUSÃO	59
REFERÊNCIAS	61
ANEXOS	64

1 INTRODUÇÃO

O presente trabalho visa apresentar o estado atual das ferramentas open source para ambientes OLAP, analisando as características e funcionalidades específicas de cada ferramenta, com embasamento teórico de conceitos reconhecidos de autoridades no assunto.

Serão analisadas exclusivamente as ferramentas que possuem servidor e cliente OLAP.

Como resultado, esta pesquisa pretende oferecer informações relevantes para os profissionais da área de informática que desejam utilizar a tecnologia OLAP com ferramentas open source.

1.1 Tema

OLAP - Online Analytical Processing (Processamento analítico em tempo real).

1.2 Delimitação do Tema

Ferramentas open source para ambientes OLAP.

1.3 Problema

Não há, de forma explícita e sistematizada, a identificação e caracterização de ferramentas open source para ambientes OLAP.

1.4 Objetivos

A seguir serão apresentados o objetivo geral e os objetivos específicos deste trabalho.

1.4.1 Objetivo Geral

Conhecer as ferramentas Open Source para ambientes OLAP e suas características específicas.

1.4.2 Objetivos Específicos

Possibilitar operações e atividades de análise para diversos tipos de negócios através de ferramentas open source.

Viabilizar o desenvolvimento de sistemas para suporte à tomada de decisão através de custos acessíveis.

1.5 Justificativa

O suporte à tomada de decisão baseado em ferramentas OLAP comerciais possui um custo elevado, que para a maioria das pequenas e médias empresas torna-se proibitivo.

Identificar ferramentas open source para ambientes OLAP capazes de atender os principais requisitos que tais ambientes exigem tornará possível para muitas destas empresas a implementação de sistemas para suporte à tomada de decisão.

Tal possibilidade de implementação será maior para empresas que possuem equipe própria de desenvolvimento de sistemas de informação.

2 CONCEITOS E TERMINOLOGIA

O objetivo deste capítulo é apresentar conceitos relacionados à tecnologia OLAP e esclarecer parte da terminologia utilizada nesse trabalho.

2.1 Open Source

A Open Source Initiative (OSI) é uma corporação sem fins lucrativos dedicada a administrar e promover a definição de software open source. A definição, além da exigência da disponibilidade de acesso ao código fonte dos programas, contém alguns critérios para a distribuição do software, descritos no ANEXO A - The Open Source Definition.

2.2 Flat File

Um Flat File, que pode ser traduzido como arquivo plano, é um tipo de arquivo usado para armazenamento de dados.

Uma definição para Flat File pode ser encontrada em WIKIPEDIA CONTRIBUTORS (2006. Flat file database):

Um flat file é um arquivo contendo registros, geralmente um registro por linha. Os campos podem simplesmente ter uma largura fixa com preenchimento ou serem delimitados por espaço em branco ou tabulações ou vírgulas (CSV) ou outros caracteres. Uma formatação extra pode ser necessária para distinguir um espaço interno de um delimitador. Não existem relacionamentos estruturais. Os dados são “planos” como em uma folha de papel, em contraste com modelos mais complexos tal como um banco de dados relacional.

Uma outra definição para Flat File, que complementa a anterior, é mencionada em Chapple (2006):

Flat files são arquivos de dados que contêm registros sem relacionamentos estruturados. É necessário um conhecimento adicional para interpretar estes arquivos, tal como as propriedades de formato do arquivo. SGBDs modernos usam uma abordagem mais estruturada para administração de arquivos (tal como a definida pela SQL) e portanto têm uma estrutura de armazenamento mais complexa.

2.3 OLTP

OLTP é o acrônimo de “Online Transaction Processing” (Processamento de Transações em Tempo Real). É uma categoria de software utilizado no nível operacional, para armazenar dados detalhados das transações de negócios de uma organização conforme as mesmas ocorrem. Uma aplicação OLTP possui vários usuários concorrentes inserindo, alterando e excluindo dados em tempo real.

As aplicações OLTP atendem a manipulação de dados operacionais, sendo incapazes de efetuarem análises complexas para apoiar o usuário na tomada de decisões estratégicas.

O conceito de OLTP é exemplificado, citando brevemente o contraste com OLAP, em Garcia-Molina (2001, p. 654):

Em geral, as consultas de apoio à decisão usadas em aplicativos de OLAP examinam quantidades muito grandes de dados, ainda que os resultados das consultas sejam pequenos. Em contraste, cada uma das operações comuns de bancos de dados, como depósitos bancários ou reservas de passagens aéreas, toca apenas uma porção minúscula do banco de dados; esse último tipo de operação é chamado frequentemente OLTP (On-Line Transaction Processing).

Erik Thomsen, na tabela a seguir (Tabela 1), resume as diferenças entre as atividades de sistemas operacionais e aquelas baseadas em análise, orientadas

a decisão.

Tabela 1 – Comparação entre atividades de processamento de informações operacionais e baseadas em análise, orientadas a decisão.

ATIVIDADES OPERACIONAIS	ATIVIDADES BASEADAS EM ANÁLISE, ORIENTADAS A DECISÃO
Mais freqüentes	Menos freqüentes
Mais previsíveis	Menos previsíveis
Menores quantidades de dados acessados por consulta	Maiores quantidades de dados acessados por consulta
Consulta principalmente de dados primitivos	Consulta principalmente de dados derivados
Exige principalmente dados atuais	Exige dados passados, presentes e projetados
Pouca ou nenhuma derivação complexa	Muitas derivações complexas

Fonte: Thomsen (2002, p. 13).

Na medida em que uma organização acumula dados operacionais com o passar do tempo, os mesmos podem ser utilizados para suportar tomadas de decisões estratégicas, em nível de gerência. Os dados detalhados das transações de negócios de uma organização devem passar por um processo de transformação para atender este tipo de demanda de informações, resultando em conhecimento estratégico, que auxilia os diversos níveis de uma organização na tomada de decisões.

2.4 Business Intelligence (BI)

O Business Intelligence, que pode ser traduzido como Inteligência de Negócios ou Inteligência Empresarial, é uma classe de tecnologias e ferramentas voltadas para atender o nível gerencial das organizações, transformando dados em

informações significativas. Wayne Eckerson conceitua Business Intelligence, em Eckerson (2002), como: “Os processos, tecnologias e ferramentas necessárias para transformar dados em informação, informação em conhecimento, e conhecimento em planos que conduzem ações de negócio lucrativas”.

Entre estas tecnologias e ferramentas encontram-se Data Warehouses, ferramentas OLAP, e outras, como Decision Support Systems (DSS), Executive Information Systems (EIS), Data Mining e Knowledge Management, por exemplo.

3 DATA WAREHOUSE

O objetivo deste capítulo é fornecer as informações necessárias a respeito da tecnologia de Data Warehouse que serão utilizadas no próximo capítulo, que trata especificamente sobre OLAP.

Os Data Warehouses (DWs) possuem uma arquitetura com o objetivo de oferecer acesso a dados para análises complexas, descoberta de conhecimento e tomada de decisão. No Data Warehouse não existem controles necessários de bancos de dados tradicionais que suportam processamento de transações em tempo real (OLTP), que inclui inserções, atualizações e exclusões, ao mesmo tempo que suportam requisições de consultas. Data Warehouses são projetados para suportar extração, apresentação de dados e processamento eficientes, para finalidades analíticas e de tomada de decisão.

A utilização inicial da expressão "Data Warehouse" é atribuída a William H. Inmon. A expressão é caracterizada em Inmon (2002, p. 31): "Um data warehouse é uma coleção de dados orientados ao assunto, integrados, não-voláteis e variantes no tempo, para fornecer suporte às decisões de gerentes".

3.1 Características

A seguir serão explicadas brevemente as características de um Data Warehouse.

3.1.1 Orientado ao Assunto

Significa que as informações são armazenadas e se relacionam em grupos de assuntos de interesse de uma organização, focando as principais atividades de negócio.

3.1.2 Integrado

Significa que as informações provenientes dos sistemas operacionais de uma organização são consolidadas, de forma que sejam consistentes e tenham um único significado, ou seja, são codificadas de uma única forma.

3.1.3 Não Volátil

Significa que as informações são carregadas em massa no Data Warehouse e não sofrem atualizações como nos sistemas OLTP, sendo disponibilizadas somente para consultas.

3.1.4 Variante no Tempo

Significa que as informações do Data Warehouse sempre estão associadas a um ponto no tempo, compondo séries históricas.

3.2 Granularidade

A granularidade em um Data Warehouse se refere ao nível de detalhamento das informações armazenadas, ou seja, quanto mais detalhadas as informações, menor a granularidade do Data Warehouse.

A granularidade afeta o volume de informações armazenadas e os tipos de consultas que podem ser respondidas aos usuários: quanto menor o nível de granularidade, mais informações serão armazenadas, e maior o detalhamento possível das consultas dos usuários; por outro lado, um alto nível de granularidade possui um volume menor de informações armazenadas e permite maior rapidez nas respostas das consultas, porém limita as mesmas quanto ao seu detalhamento.

3.3 Arquitetura Funcional

A arquitetura funcional de um Data Warehouse divide-se basicamente em duas grandes áreas: interna e externa, conforme ilustrado a seguir:

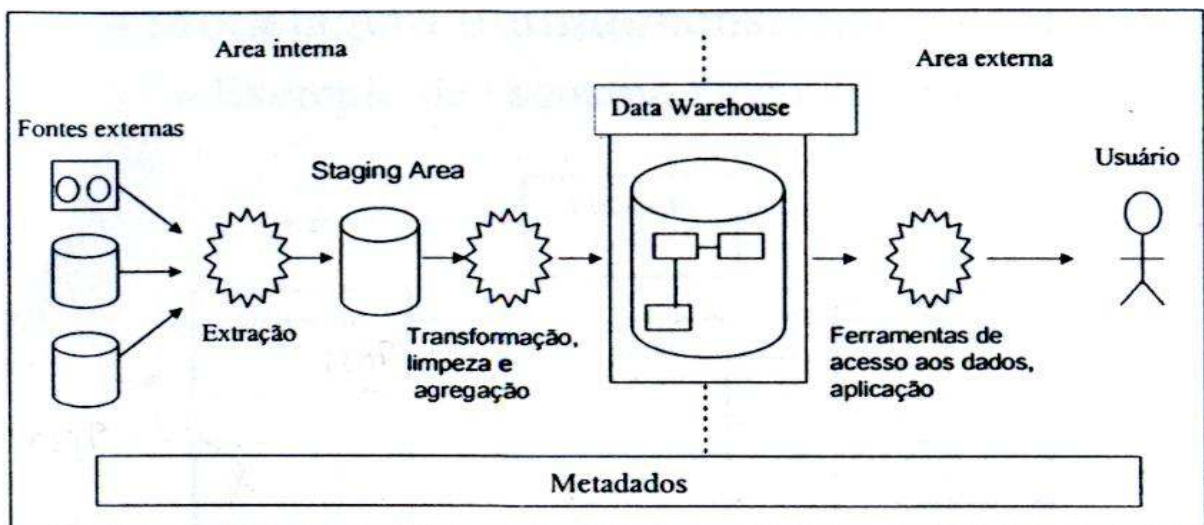


Figura 1 – Arquitetura Funcional de um DW.

Fonte: Ferreira (2004, p. 25).

3.3.1 Área Interna

A área interna é composta pelas fontes externas de dados, staging area, armazenamento de dados do DW e metadados.

As fontes externas armazenam os dados de origem que abastecem o DW, geralmente provenientes de sistemas OLTP legados. Os dados também podem estar armazenados em flat files ou em algum outro meio.

A staging area é uma área intermediária de armazenamento entre as fontes externas e o DW, nela os dados são preparados, limpos e integrados para posterior carga no DW.

O armazenamento de dados, ou repositório do DW contém as informações já consolidadas e disponíveis para serem utilizadas pela área externa, sendo que os dados são carregados em massa.

Os metadados contêm “dados sobre os dados” do DW, ou seja, são informações que descrevem um conjunto de dados. Geralmente são armazenados em um repositório fora do alcance dos usuários. Algumas das informações contidas nos metadados são descritas em Imhoff, Galemme e Geiger (2003, p. 403):

Ele fornece definições para dados, os cálculos usados, informação sobre de onde os dados vieram (quais sistemas fontes), o que foi feito com ele (transformações, rotinas de limpeza, algoritmos de integração, etc.), quem está usando ele, quando o usam, quais as métricas de qualidade existem para várias partes de dados, e assim por diante.

Na área interna ocorre o processo de Extração, Transformação e Carga, denominado ETL (Extraction, Transformation and Load). Este processo é dividido em etapas e é responsável pela integridade, qualidade e consolidação das informações provenientes das fontes externas até o armazenamento do DW. Este processo

ocorre na carga inicial e nas atualizações periódicas dos dados do Data Warehouse, sendo que a periodicidade das atualizações deve levar em consideração o volume de dados e de processamento envolvido.

A ETL é um fator crítico de sucesso na implantação de um Data Warehouse. Deve-se atribuir uma importância fundamental a este processo na fase de projeto de um DW, pois conforme citado em Inmon (2002, p. 295): "O processo de integração e transformação de dados tipicamente consome até 80% dos recursos de desenvolvimento". Esta informação é reiterada por Eckerson e White (2003, p. 4): "De acordo com a maioria dos profissionais, o trabalho de projeto e desenvolvimento com ETL consome de 60% a 80% de um projeto inteiro de BI".

Um processo de ETL mal projetado ou executado pode arruinar completamente a implantação de um DW em uma empresa, tanto pelos recursos e custos necessários, como pela qualidade das informações fornecidas para a área externa, ou seja, para os usuários analistas, tomadores de decisão.

O valor da qualidade de uma informação pode ser medido comparando-se os resultados das ações tomadas a partir de uma informação exata com os resultados obtidos através de uma informação equivocada.

3.3.2 Área Externa

Na área externa encontram-se os metadados e as aplicações e ferramentas de acesso aos dados, possibilitando consultas "ad hoc" e relatórios específicos. Consultas "ad hoc" são consultas circunstanciais, não programadas, pertinentes a um determinado momento de análise. Na área externa localizam-se as ferramentas OLAP além de outras, como por exemplo, Data Mining e Relatórios Analíticos.

3.4 Data Mart

Um Data Mart é uma divisão lógica de um Data Warehouse que atende a uma área específica de negócio de uma organização. Um Data Mart não é um “pequeno Data Warehouse”, mas uma unidade lógica de um DW, podendo ser qualificado como um Data Warehouse departamental. Uma organização pode construir vários Data Marts ao longo do tempo, e vinculá-los a um Data Warehouse da empresa inteira.

Um conceito de Data Mart é citado em Inmon, Tederman e Imhoff (2001, p. 11):

Os data marts são estruturas moldadas pelos dados granulares encontrados no data warehouse corporativo. Os data marts pertencem aos departamentos específicos dentro de uma empresa – geralmente finanças, contabilidade, vendas ou marketing – e são moldados pelos requerimentos dos departamentos. Conseqüentemente, o design de cada data mart é único.

A grande maioria dos DWs atualmente em produção, na realidade, são Data Marts.

Algumas diferenças entre um Data Mart e um Data Warehouse são relacionadas na tabela a seguir, conforme Inmon (1999, p. 4):

Tabela 2 – Diferenças entre um Data Mart e um Data Warehouse.

DATA MART	DATA WAREHOUSE
Departamental	Corporativo
Alto nível de granularidade	Baixo nível de granularidade
Estrutura star-join	Estrutura normalizada
Pequena quantidade de dados históricos	Grande quantidade de dados históricos
Tecnologia otimizada para acesso e análise	Tecnologia otimizada para armazenamento e gerenciamento de grandes volumes de dados
Cada departamento possui uma estrutura diferente	Estrutura adequada ao entendimento corporativo de dados
Altamente indexado	Levemente indexado

Fonte: Inmon (1999, p. 4).

4 OLAP

O objetivo deste capítulo é apresentar a tecnologia OLAP e seus conceitos, a modelagem multidimensional, os recursos e as arquiteturas OLAP, fornecendo o embasamento necessário para o próximo capítulo, que trata dos critérios para avaliação de ferramentas OLAP.

4.1 O que é OLAP?

OLAP é o acrônimo de “Online Analytical Processing” (Processamento analítico em tempo real). Foi criado em 1993 por Edgar F. Codd, que também criou o Modelo Relacional, utilizado por Sistemas Gerenciadores de Banco de Dados Relacionais (SGBDRs).

Basicamente é uma classe de tecnologias projetada para análise de dados multidimensionais, muito utilizada nos sistemas de suporte à decisão, permitindo avaliar séries históricas de dados.

Em Codd, Codd e Saley (1993), quando da criação do termo OLAP, é citada a limitação das ferramentas disponíveis na época e proposta a análise de dados multidimensionais:

As ferramentas de consultas/relatórios e as planilhas eletrônicas têm sido extremamente limitadas nas formas pelas quais os dados (já recuperados do SGBD) podem ser agregados, resumidos, consolidados, somados, visualizados e analisados. A carência mais notada tem sido a capacidade para consolidar, visualizar e analisar dados de acordo com múltiplas dimensões, de maneira que faça sentido para um ou mais analistas específicos em um determinado ponto no tempo. Este requisito é chamado “análise de dados multidimensionais”. Talvez um melhor e mais genérico nome para este tipo de funcionalidade é online analytical processing (OLAP), em que a análise de dados multidimensionais é apenas uma de suas características.

Em janeiro de 1995, THE OLAP COUNCIL (1995), um conselho para padronizar a tecnologia OLAP, publica sua conceituação do termo:

O online analytical processing (OLAP) é uma categoria de tecnologia de software que possibilita que os analistas, gerentes e executivos tenham entendimento sobre os dados de forma rápida, consistente, e com acesso interativo a uma ampla variedade de visões possíveis de informações que foram transformadas a partir de dados brutos para refletir a dimensionalidade real da empresa como entendida pelo usuário.

4.2 O Modelo Multidimensional

A modelagem de dados descreve e representa de uma forma abstrata os dados de uma organização. Data Warehouses e ferramentas OLAP se utilizam do modelo de dados multidimensional, e apresentam informações na forma de cubos de dados (Data Cubes).

Um conceito de modelo multidimensional pode ser encontrado em Imhoff, Galemme e Geiger (2003, p. 400):

Um modelo dimensional é uma forma de modelagem de dados que condiciona os dados de acordo com consultas e processos específicos do negócio. Os objetivos são a compreensibilidade dos usuários de negócio e o desempenho de consultas multidimensionais.

A respeito dos benefícios da modelagem multidimensional, há uma breve descrição em Abelló (2002, p. 11):

Os benefícios da modelagem multidimensional são dois. De um lado, ela torna os esquemas de dados mais compreensíveis para os usuários finais, e por outro lado, ela permite usar armazenamento específico e técnicas de acesso que melhoram o desempenho de queries. A maneira para obter estes benefícios é a simplificação dos esquemas de dados, de forma que eles só contenham as coisas essenciais (i.e. um fato para ser analisado e suas dimensões de análise). Estes esquemas são próximos da concepção de dados dos analistas, e sugerem um tipo específico de queries, de forma que o sistema pode ser personalizado facilmente para resolvê-las com bons tempos de resposta.

O relacionamento entre modelos multidimensionais e OLAP é citado em Kimball e Ross (2002, p. 466):

Os modelos dimensionais são a base de muitos aprimoramentos de desempenho de DBMS, inclusive agregações e métodos de indexação avançados. Também são a base do desenvolvimento incremental e distribuído do data warehouse através do uso de dimensões e fatos em conformidade e a base lógica de todos os sistemas OLAP.

A seguir serão apresentados brevemente os principais elementos do modelo de dados multidimensional.

4.2.1 Fatos

Um fato é um conjunto de itens de dados contextualizados e mensurados em termos numéricos, que representam transações ou eventos de negócio, demonstrando a evolução do mesmo no decorrer do tempo.

No modelo multidimensional um fato é representado por uma tabela fato (Fact Table). Uma tabela fato é constituída de valores quantitativos e numéricos e representa um assunto do DW.

Um conceito de tabela fato é citado em Imhoff, Galemme e Geiger (2003, p. 401): “Uma tabela fato é a tabela dentro de um modelo dimensional que contém as medidas e métricas de interesse”.

4.2.2 Dimensões

Dimensões armazenam as descrições textuais das dimensões do negócio. No modelo multidimensional uma dimensão é representada por uma tabela

dimensão (Dimension Table).

Um conceito de tabela dimensão é citado em Imhoff, Galemme e Geiger (2003, p. 400):

Uma tabela dimensão é um conjunto de tabelas de referência que provêm a base para restringir e agrupar consultas para a informação em uma tabela fato dentro de um modelo dimensional. A chave da tabela dimensão tipicamente é uma parte da chave concatenada da tabela fato, e a tabela dimensão contém informação descritiva e hierárquica.

4.2.3 Membros e Hierarquias

As definições a seguir, de membros e hierarquias, assinalam a possibilidade de haver uma classificação dos dados dentro de uma dimensão.

Um membro de uma dimensão é definido em THE OLAP COUNCIL (1995):

Um membro de uma dimensão é um nome ou identificador discreto usado para identificar uma posição e descrição de um item de dado dentro de uma dimensão. Por exemplo, “janeiro de 1989” ou “primeiro semestre de 1993” são exemplos típicos de membros de uma dimensão Tempo. Atacado, Varejo, etc., são exemplos típicos de membros de uma dimensão Canal de Distribuição.

Uma hierarquia de uma dimensão é caracterizada também em THE OLAP COUNCIL (1995):

Quaisquer membros de uma dimensão podem ser organizados com base em relacionamentos pai-filho, tipicamente onde um membro pai representa a consolidação dos membros que são seus filhos. O resultado é uma hierarquia, e os relacionamentos pai/filho são relacionamentos hierárquicos.

Em uma dimensão geográfica, por exemplo, podem existir hierarquias de Logradouro, Bairro, Cidade, Estado e Região.

4.2.4 Medidas

Medidas são as características específicas e mensuráveis de um fato, relevantes para análises.

Um conceito de medida é citado em Imhoff, Galemme e Geiger (2003, p. 403):

Uma medida é um termo da modelagem dimensional que se refere a valores, geralmente numéricos, que medem algum aspecto do negócio. Medidas residem em tabelas fato. Os termos dimensionais medida e atributo, juntos, são equivalentes ao uso do termo atributo da modelagem relacional.

Fatos, Dimensões e Medidas são inter-relacionados na modelagem multidimensional, como é citado em Singh (2001, p. 146):

Na modelagem, as estruturas de dados são organizadas para descrever “medições” e “dimensões”. Medições fornecem os dados numéricos coletados que ficam armazenados na tabela de fatos central. As dimensões são os parâmetros do negócio que definem cada transação. São armazenadas em tabelas satélites vinculadas à tabela de fatos central. Por exemplo, os dados armazenados em tabelas de fatos incluem dados de vendas, estoque, assinaturas de periódicos, despesas e margem bruta. Tabelas típicas incluem dados de tempo, geografia, conta e produto.

4.2.5 Cubos de Dados

Os cubos de dados são estruturas nas quais os usuários finais fazem análises de dados multidimensionais, ou seja, navegam e exploram os mesmos, extraindo informações e conhecimento a partir dos dados.

Uma definição para cubos de dados é encontrada em Kay (2005):

Um cubo de dados é um tipo de matriz multidimensional que permite que os usuários explorem e analisem uma coleção de dados de muitas perspectivas diferentes, geralmente considerando três fatores (dimensões) de cada vez.

A metáfora visual de um cubo com três dimensões é fácil de se entender, ao passo que com mais dimensões se torna difícil, conforme citado em Thomsen (2002, p. 56): “Tentar usar um cubo como base para visualização de quatro ou mais dimensões pode se tornar rapidamente muito complicado”.

Um cubo com mais de três dimensões é chamado de hipercubo (Hypercube) ou tesseract, um termo matemático para designar um cubo com quatro dimensões. A dificuldade de entendimento de uma metáfora visual com mais de três dimensões pode ser percebida pela visão de um tesseract na figura 2.

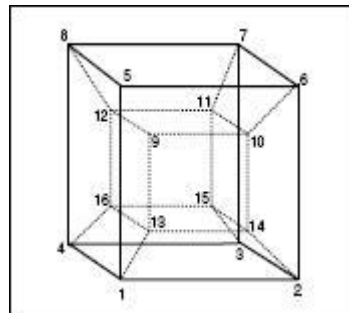


Figura 2 – Visão de um Tesseract

Fonte: Thomsen (2002, p. 57).

Como exemplo, será considerado uma empresa que comercializa determinados produtos e possui revendas em diversas cidades, de diversas regiões do país. A demanda dos clientes é atendida pelas revendas, de acordo com a localização do cliente e a área de cobertura das revendas. Uma representação das vendas desta empresa através de um cubo poderia ser a da figura 3.

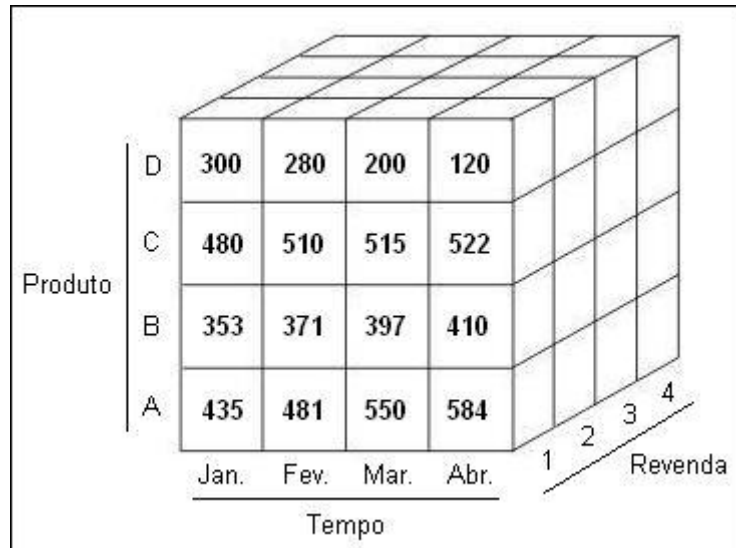


Figura 3 – Cubo tridimensional: produtos, meses e vendas

Fonte: Do autor.

Neste exemplo é apresentado um cubo de três dimensões. Poderiam ser adicionados os clientes - mais uma dimensão - mas se teria um hipercubo de quatro dimensões, não entendido tão facilmente.

Uma unidade de dados é representada em cada célula do cubo, localizada através da interseção das dimensões. Por exemplo, observando o cubo pode-se verificar que o valor das vendas do produto D pela revenda 1 no mês de março é 200.

A relação entre análise multidimensional e estrutura de dados é citada brevemente em Ballard et al. (2006, p. 86):

A análise multidimensional tornou-se uma forma popular de estender as capacidades de consultas e relatórios. Isto é, melhor do que submeter múltiplas queries, os dados são estruturados para possibilitar acesso fácil e rápido para responder as questões que os usuários tipicamente perguntam.

A estruturação dos dados para possibilitar acesso fácil e rápido aos mesmos pode ser percebida no Star Schema, que será visto a seguir.

4.2.6 Star Schema

Conforme citado em Thomsen (2002, p. 638), um Star Schema é:

Um arranjo de tabelas em um banco de dados relacional onde uma tabela de fatos central é conectada a um conjunto de tabelas de dimensão, uma por dimensão. O nome star vem da representação diagramática normal desse esquema, com a tabela de fatos no centro e cada tabela de dimensão mostrada ao seu redor, como as pontas em uma estrela.

Ralph Kimball caracteriza um Star Schema em Kimball (2002, p. 459), como: “Representação genérica de um modelo dimensional em um banco de dados relacional em que uma tabela de fatos com uma chave composta é unida a várias tabelas de dimensão, cada uma com uma chave primária”.

Um Star Schema também é conhecido por Star-Join, Star-Join Schema e Modelo Star Schema.

Para o exemplo, o Star Schema poderia ser o exibido na figura 4.

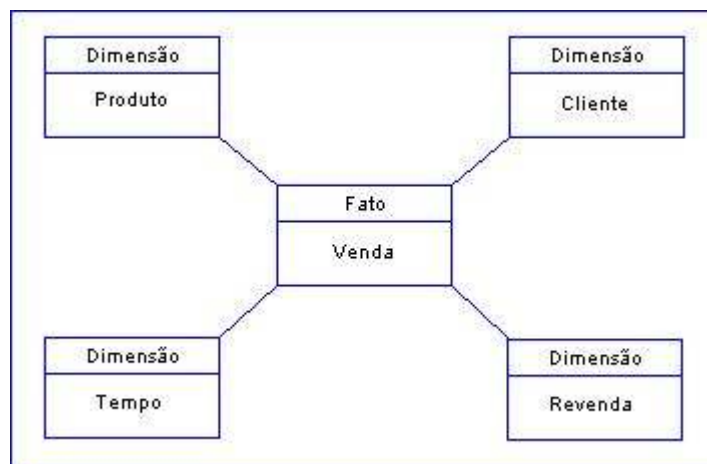


Figura 4 – Star Schema.

Fonte: Do autor.

Neste modelo está representado o armazenamento de dados em tabelas, relacionando o fato da venda com suas dimensões produto, cliente, revenda e tempo.

4.2.7 Surrogate Key

Surrogate keys são componentes das entidades de um DW com o objetivo principal de identificar unicamente os registros de tabelas fato e dimensão.

Por exemplo, com a possibilidade de existirem múltiplas fontes externas de dados, existe também a possibilidade de que uma determinada informação possua identificadores diferentes nas várias fontes externas.

As surrogate keys garantem a unicidade de identificadores por todo o ambiente do DW, e via de regra são inteiros, que ocupam pouco espaço e possibilitam joins mais rápidos do que outros tipos de dados.

A seguir serão apresentadas três definições para surrogate keys.

Uma definição é citada em Imhoff, Galemme e Geiger (2003, p. 406):

Uma surrogate key é uma chave substituta que geralmente é um valor numérico arbitrário atribuído pelo processo de carga ou pelo sistema de banco de dados. A vantagem da surrogate key é que ela pode ser estruturada de forma que seja sempre única por toda a extensão de integração para o data warehouse.

Outra definição é encontrada em DM Review (2006. Glossary: Surrogate Key):

Uma surrogate key é uma parte única, um identificador estabelecido artificialmente para uma entidade. A atribuição de uma surrogate key é um caso especial de dados derivados - aquele onde a chave primária é derivada. Uma forma comum de derivar valores para uma surrogate key é atribuir valores inteiros seqüencialmente.

Denis Howe define surrogate keys, em Howe (2004):

Uma chave primária única gerada pelo SGBDR que não é derivada de quaisquer dados no banco de dados e cujo único significado é funcionar como a chave primária. Uma surrogate key frequentemente é um número seqüencial, mas não é obrigado que seja. Ter a chave independente de todas as outras colunas isola os relacionamentos do banco de dados das alterações nos valores de dados ou do projeto do banco de dados e garante unicidade. Alguns projetistas de banco de dados usam surrogate keys religiosamente apesar da adequabilidade de outras chaves candidatas. De qualquer forma, se uma boa chave já existe, o acréscimo de uma surrogate key somente diminuirá a velocidade de acesso, especialmente se ela for indexada.

Seguindo do Star Schema para um projeto de banco de dados, detalhando as tabelas dimensão e fato, o modelo de entidades do ambiente DW poderia ser ilustrado conforme a figura 5.

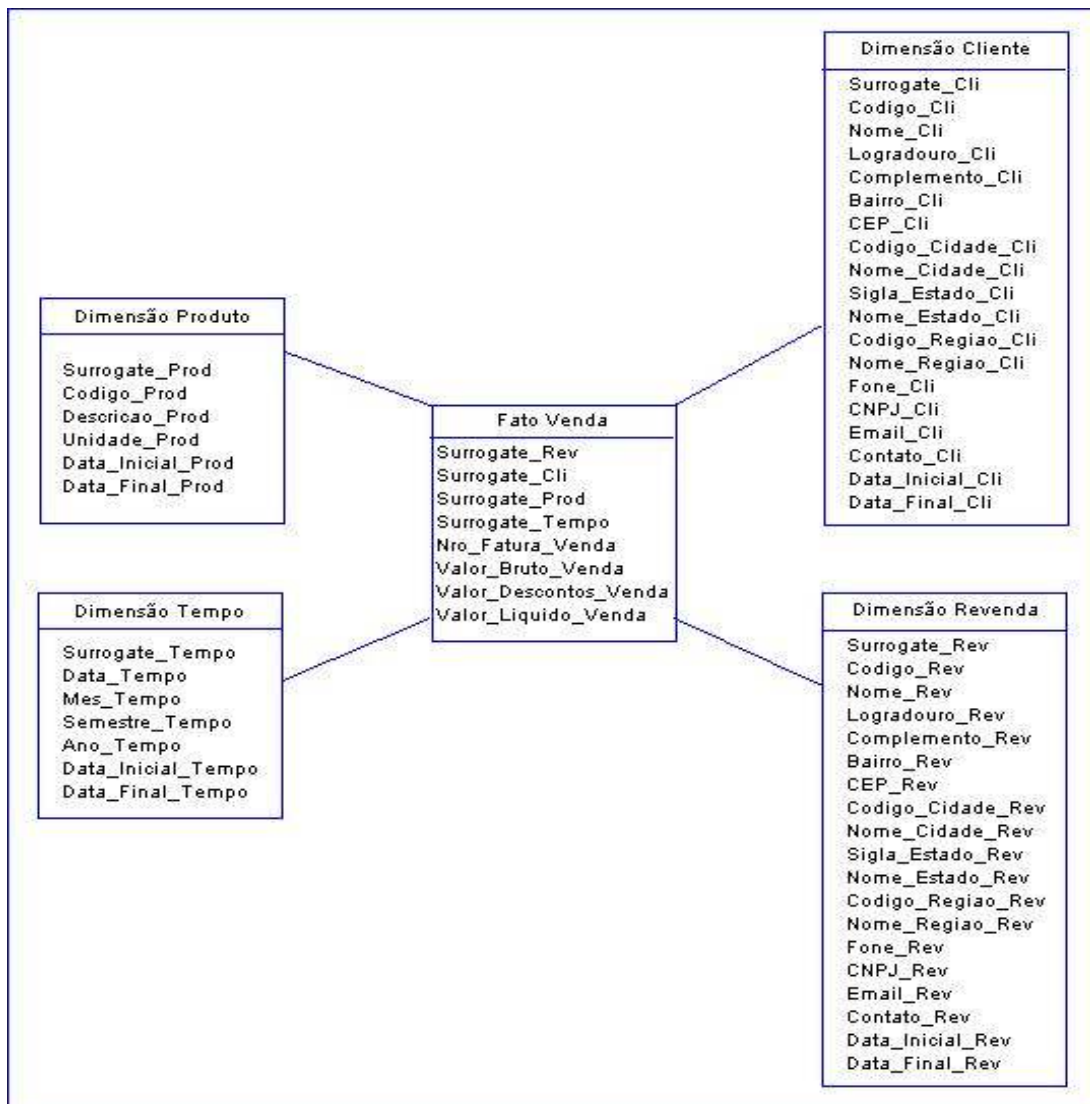


Figura 5 – Entidades do Ambiente DW.

Fonte: Do autor.

As dimensões Revenda e Cliente estão desnormalizadas em relação às informações de Logradouro, Bairro, Cidade, Estado e Região. Esta técnica garante melhor desempenho na recuperação dos dados por evitar joins, operação cara para SGBDs relacionais. Como ponto fraco, comparado ao Snowflake Schema, o volume de informações armazenadas é maior, por haver muita redundância de dados.

4.2.8 Snowflake Schema

O Snowflake Schema é uma extensão do Star Schema, e tem por característica principal a normalização das dimensões. No exemplo, o Snowflake Schema poderia ser o exibido na figura 6.

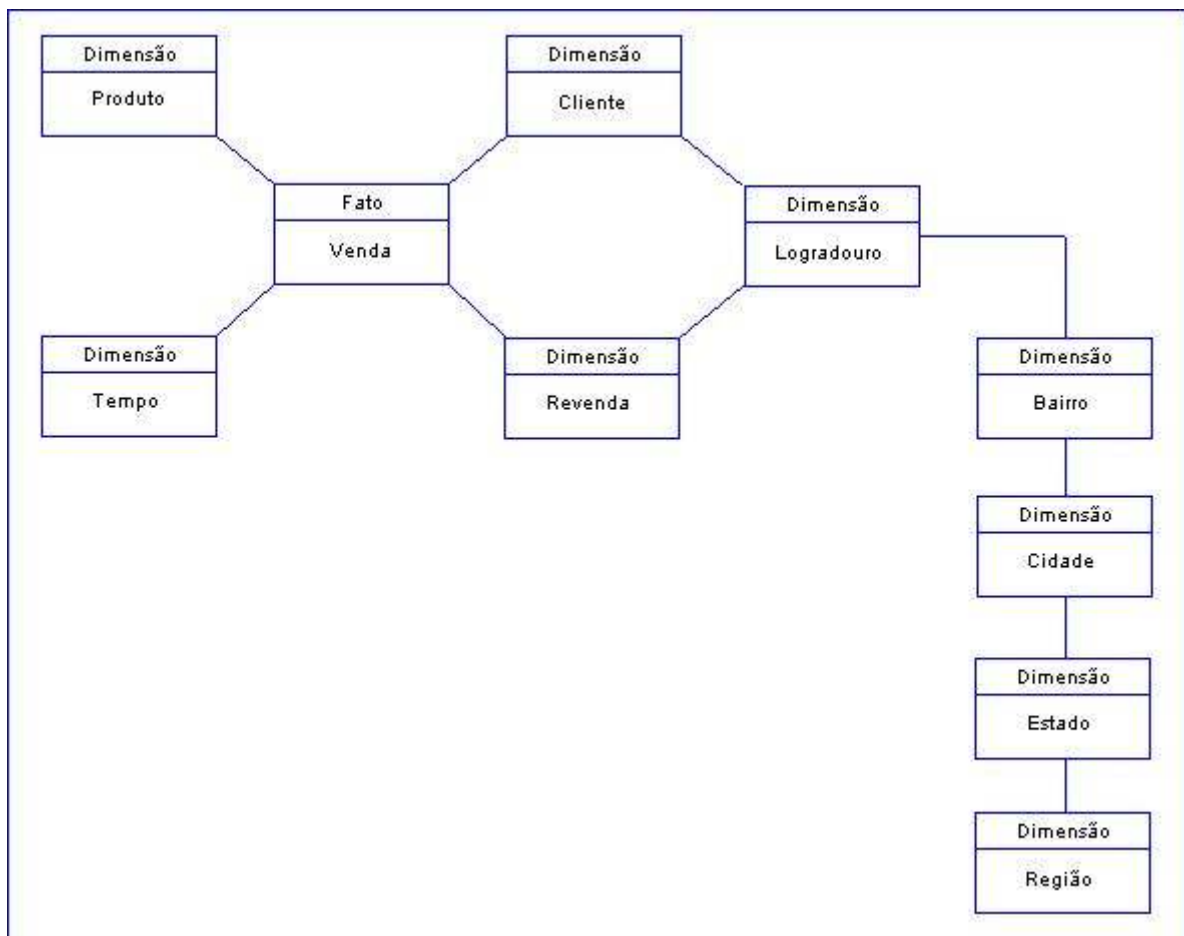


Figura 6 – Snowflake Schema.

Fonte: Do autor.

As dimensões Revenda e Cliente estão normalizadas em relação às informações de Logradouro, Bairro, Cidade, Estado e Região. Esta técnica garante uma otimização do espaço necessário para o armazenamento de dados, porém o desempenho na recuperação dos mesmos é inferior por serem necessários mais joins do que no Star Schema.

4.3 Recursos OLAP

A seguir serão definidos e exemplificados os principais recursos OLAP, ou seja, operações realizadas com cubos de dados que possibilitam visualizações OLAP.

4.3.1 Drill Down e Roll Up

Drill Down e Roll Up são operações onde a visualização dos dados percorre as hierarquias de uma dimensão, ou seja, há uma navegação entre níveis de maior para menor granularidade e vice-versa.

O recurso Drill Down é definido em THE OLAP COUNCIL (1995):

Drill down/up é uma técnica analítica específica através da qual o usuário navega entre níveis de abrangência de dados a partir do mais resumido (up) para o mais detalhado (down). Os caminhos de navegação podem ser definidos pelas hierarquias dentro de dimensões ou outros relacionamentos que podem ser dinâmicos dentro ou entre dimensões. Por exemplo, na visualização de dados de vendas da América do Norte, uma operação drill-down na dimensão Região mostraria Canadá, o leste e o oeste dos Estados Unidos. Um drill-down mais além sobre o Canadá poderia mostrar Toronto, Vancouver, Montreal, etc.

Como exemplo, será considerado o volume de vendas de uma empresa por região nos últimos cinco anos, ilustrado na figura 7.

Volume de Vendas					
Região	2002	2003	2004	2005	2006
Centro-Oeste	23	22	25	24	28
Nordeste	20	28	26	25	25
Norte	15	18	22	19	16
Sudeste	25	22	29	30	35
Sul	30	22	25	35	38

Figura 7 – Volume de Vendas por Região.

Fonte: Do autor.

Uma operação de Drill Down sobre a região sul poderia detalhar as informações de vendas por estados desta região. O resultado poderia ser como o ilustrado na figura 8.

Volume de Vendas					
Estado	2002	2003	2004	2005	2006
Paraná	8	5	6	9	11
Rio Grande do Sul	12	10	11	14	15
Santa Catarina	10	7	8	12	12

Figura 8 – Volume de Vendas da Região Sul.

Fonte: Do autor.

O recurso Drill Down também é chamado de Roll Down por alguns autores.

O recurso Roll Up é o caminho oposto do Drill Down, ou seja, ocorre a partir de uma informação detalhada para uma informação sumarizada. Por exemplo, uma consulta de volume de vendas poderia ser feita por Cidade e logo após ser resumida por Estado e depois por Região.

O recurso Roll Up também é chamado de Drill Up por alguns autores.

4.3.2 Slice and Dice

O termo Slice é conceituado em Ballard et al. (2006, p. 87): “O termo slice na terminologia multidimensional é usado para definir um membro ou um grupo de membros que são separados (de TODAS as outras dimensões) e então avaliado através de todas as dimensões”.

Ou seja, quando se faz um slicing, os membros ou grupo de membros escolhidos de uma dimensão são resumidos através de todas as outras dimensões.

Uma operação de Slicing sobre um grupo de membros da dimensão produto poderia ter o resultado como o ilustrado na figura 9. Neste exemplo, são apresentados dois membros isolados da dimensão produto (Produtos B e D) e os totais de vendas correspondentes, resumidos por todas as revendas, clientes e datas.

Volume de Vendas	
Produto	Total de Vendas
Produto B	2.200
Produto D	812
Total	3.012

Figura 9 – Slice por Produto.

Fonte: Do autor.

O termo Dice é conceituado em Ballard et al. (2006, p. 88):

O conceito de dicing significa que você coloca múltiplos membros de uma dimensão sobre um eixo e então coloca múltiplos membros de uma dimensão diferente sobre um outro eixo. Isto permite que você veja a inter-relação de membros de dimensões diferentes. Dicing é a análise de inter-relações entre dimensões diferentes ou seus membros.

Uma operação de Dicing com os membros da dimensão revenda em um

eixo e os membros da dimensão tempo em outro eixo, poderia ter como resultado o ilustrado na figura 10.

Volume de Vendas			
Revenda	2006		
	Janeiro	Fevereiro	Março
Revenda 1	250	300	280
Revenda 2	280	310	305
Revenda 3	212	220	231
Revenda 4	300	340	330
Total	1.042	1.170	1.146

Figura 10 – Dice por Revenda e Tempo.

Fonte: Do autor.

4.3.3 Drill across

O recurso Drill Across é conceituado em Ballard et al. (2006, p. 92): “Drill-across é um método onde você faz um drill de uma dimensão para outra. Você deve definir o caminho do drill-across. Esta função é freqüentemente usada em ROLAP”.

Uma operação de Drill Across, tomando como base a figura 10 e trocando a dimensão revenda pela dimensão produto com relação à Revenda 1, poderia ter como resultado o ilustrado na figura 11.

Volume de Vendas – Revenda 1			
Revenda	2006		
	Janeiro	Fevereiro	Março
Produto A	75	85	81
Produto B	65	78	72
Produto C	59	73	80
Produto D	51	64	47
Total	250	300	280

Figura 11 – Drill Across

Fonte: Do autor.

Neste exemplo pode-se ver um detalhamento de produtos vendidos pela Revenda 1 no primeiro trimestre de 2006.

4.3.4 Pivoting

O significado do recurso Pivoting (também conhecido como Rotate) é apresentado em Ballard et al. (2006, p. 89): “Pivoting em modelagem multidimensional significa trocar linhas com colunas e vice-versa [...] é simplesmente uma forma rápida de ver os mesmos dados de uma perspectiva diferente”.

Uma operação de Pivoting, tomando como base o Volume de Vendas da Região Sul (ilustrado na figura 8), poderia ter como resultado o ilustrado na figura 12.

Volume de Vendas			
	Estado		
Ano	Paraná	Rio Grande do Sul	Santa Catarina
2002	8	12	10
2003	5	10	7
2004	6	11	8
2005	9	14	12
2006	11	15	12

Figura 12 – Pivoting.

Fonte: Do autor.

4.3.5 Ranking

O recurso Ranking é conceituado em DWBRASIL (2003): “A opção de ranking permite agrupar resultados por ordem de maiores/menores, baseado em objetos numéricos (Medidas). Esta opção impacta somente uma tabela direcionada

(relatório) não afetando a pesquisa (Query)”.

Ou seja, é uma operação simples de classificação ascendente ou descendente das informações de acordo com uma medida escolhida em uma visualização, podendo ser escolhido o número de itens que a classificação terá.

4.4 Bancos de Dados Multidimensionais

Ralph Kimball caracteriza um banco de dados multidimensional em Kimball (2002, p. 450), como: “Banco de dados em que os dados são apresentados em cubos de dados, em oposição a tabelas em uma plataforma de banco de dados relacional”.

W. H. Inmon, destaca a relação entre SGBDs Multidimensionais e Data Warehouses em Inmon (2002, p. 182):

Uma das tecnologias freqüentemente discutidas no contexto do data warehouse é o processamento de SGBDs multidimensionais (algumas vezes chamado de processamento OLAP). SGBDs multidimensionais, ou data marts, fornecem um sistema de informação com a estrutura que permite a uma organização ter acesso muito flexível a dados, para fazer slice and dice de dados de qualquer número de formas, e para explorar dinamicamente o relacionamento entre dados resumidos e detalhados. SGBDs multidimensionais oferecem flexibilidade e controle ao usuário final, e como tal eles ajustam-se bem em um ambiente de sistemas de suporte à decisão.

Erik Thomsen comenta brevemente sobre o uso de bancos de dados relacionais e multidimensionais no OLAP, em Thomsen (2002, p. 7):

[...] os produtos de bancos de dados relacionais são muito mais equipados para lidar com as quantidades imensas de dados, normalmente associadas a iniciativas corporativas de data warehouse. Bancos de dados multidimensionais são muito mais bem equipados para oferecer cálculos rápidos e em estilo dimensional (embora [...] os bancos de dados SQL estejam evoluindo para o suporte mais eficiente aos cálculos em estilo OLAP.

As estruturas físicas de armazenamento de dados de um banco de dados multidimensional, conhecidas por arrays multidimensionais, são diferentes das estruturas utilizadas em bancos de dados relacionais (que dominam grande parte do mercado hoje), justamente para prover as necessidades de multidimensionalidade dos dados.

A linguagem MDX (Multidimensional Expressions), usada em bancos de dados multidimensionais, é a equivalente multidimensional da SQL, usada em bancos de dados relacionais. Sua origem é descrita em WIKIPEDIA CONTRIBUTORS (2006. Multidimensional Expressions):

A MDX foi primeiramente introduzida como parte do OLEDB para a especificação OLAP da Microsoft em 1997. A especificação foi rapidamente seguida pela liberação comercial do Microsoft OLAP Services 7.0 em 1998, e depois pelo Microsoft Analysis Services. Apesar de que não era um padrão aberto, mas particularmente uma especificação de propriedade da Microsoft, ela foi adotada pela ampla extensão de fornecedores OLAP.

4.5 Servidor e Cliente OLAP

Um servidor OLAP é definido em THE OLAP COUNCIL (1995):

Um servidor OLAP é um mecanismo de manipulação de dados multiusuário de alta capacidade projetado para suportar e operar sobre estruturas de dados multidimensionais. Uma estrutura multidimensional é organizada de forma que cada item de dado é localizado e acessado baseado na interseção dos membros de dimensões que o definem. O projeto do servidor e da estrutura de dados são otimizados para rápida recuperação de informações ad-hoc em qualquer orientação, e também para rapidez, cálculos flexíveis e transformação de dados brutos baseados em relacionamentos que consistem de fórmulas. O servidor OLAP pode ou preparar fisicamente as informações multidimensionais processadas para dar rápidos e consistentes tempos de resposta para usuários finais, ou pode povoar suas estruturas de dados em tempo real a partir de bancos de dados relacionais ou outros, ou oferecer uma opção de ambos. Dado o estado atual de tecnologia e o requisito de usuário final para rápidos e consistentes tempos de resposta, preparar os dados multidimensionais no servidor OLAP freqüentemente é o método preferido.

Ou seja, um servidor OLAP processa e atende requisições que envolvem cálculos, consolidação e recuperação de informações multidimensionais. Essas informações podem estar armazenadas em bancos de dados multidimensionais, relacionais ou ambos. No caso de um banco de dados multidimensional, o próprio já incorpora a função de servidor OLAP. Em um banco de dados relacional é necessário um servidor OLAP que acesse as informações do mesmo. A abordagem quanto ao armazenamento OLAP será ilustrada neste trabalho no tópico seguinte sobre arquiteturas OLAP.

Uma característica interessante presente em servidores OLAP é o “Multipass SQL”, que é conceituado em Kimball (1997):

[...] a ferramenta de consulta deve quebrar o relatório em um número de queries simples que são processadas separadamente pelo SGBD. A ferramenta de consulta então automaticamente combina os resultados das queries separadas de uma forma inteligente. Quebrar uma única requisição complexa em várias requisições pequenas é chamado multipass SQL. [...] multipass SQL dá ao navegador agregado uma oportunidade de aumentar a velocidade do relatório, porque cada requisição SQL atômica é simples e facilmente analisada pelo navegador agregado.

Um cliente OLAP é definido em THE OLAP COUNCIL (1995):

Aplicações de usuários finais que podem requerer slices de servidores OLAP e prover exibições bidimensionais ou multidimensionais, modificações de usuários, seleções, classificações, cálculos, etc para propósitos de visualização e navegação. Clientes OLAP podem ser tão simples como um programa de planilha eletrônica recuperando um slice para favorecer o trabalho de um usuário instruído em planilhas eletrônicas ou tão altamente funcional como uma modelagem financeira ou uma aplicação de análise de vendas.

4.6 Arquiteturas OLAP

Erik Thomsen, autor reconhecido em OLAP, destaca a origem do termo e sua utilização pelos vendedores de ferramentas OLAP em Thomsen (2002, p. 7):

Como se o termo de marketing OLAP não fosse suficiente, muitos vendedores e alguns eruditos do setor se sentiram compelidos – especialmente entre 1995 e 1998 – a criar variantes, normalmente na forma de uma única consoante acrescentada na frente do termo OLAP, para distinguir seu tipo de letra de OLAP dos outros.

Thomsen refere-se à proliferação de termos derivados do OLAP para descrever arquiteturas OLAP, como ROLAP, MOLAP, HOLAP, DOLAP e WOLAP. Estes termos derivados serão abordados brevemente a seguir.

4.6.1 ROLAP

ROLAP é o acrônimo de Relational OLAP (OLAP Relacional).

A arquitetura ROLAP consiste de um servidor OLAP que acessa dados armazenados em um banco de dados relacional. Um dos pontos positivos desta arquitetura é o fato de que a tecnologia de bancos de dados relacionais é estabelecida e madura, e seus maiores fornecedores estenderam seus produtos com a finalidade de fornecer um melhor suporte ao OLAP. Isso começou a acontecer por volta de 1994, conforme citado em Powell (2006): “Em 1994, [...] foi também nesta época que os fornecedores de SGBDs reconheceram que suporte a decisão era diferente de OLTP e começaram a implementar reais capacidades OLAP em seus bancos de dados”. O próprio padrão SQL99 ANSI/ISO determina algumas extensões OLAP. Quanto aos pontos negativos, a mesma SQL deixa a desejar tanto no desempenho quanto no conjunto de funções para análises multidimensionais (extensões OLAP).

Uma definição para ROLAP é encontrada em DM REVIEW (2006. Glossary: ROLAP):

Um produto que provê análise multidimensional de dados, agregados e metadados armazenados em um SGBDR. O processamento multidimensional pode ser feito dentro do SGBDR, de um servidor em uma camada intermediária ou no cliente. Um ROLAP 'comercial' é aquele de um fornecedor independente que pode funcionar com qualquer SGBDR padrão.

4.6.2 MOLAP

MOLAP é o acrônimo de Multidimensional OLAP (OLAP Multidimensional).

Uma definição para MOLAP pode ser encontrada em WIKIPEDIA CONTRIBUTORS (2006. MOLAP):

MOLAP é uma alternativa para a tecnologia ROLAP (OLAP Relacional). Enquanto ambas ferramentas analíticas ROLAP e MOLAP são projetadas para permitir análise de dados através do uso de um modelo de dados multidimensional, o MOLAP difere significativamente por requerer a pré-computação e armazenamento de informações no cubo – a operação conhecida como processamento. O MOLAP guarda estes dados em um armazenamento de arrays multidimensionais otimizado, em vez de um banco de dados relacional (i.e. em ROLAP).

Na arquitetura MOLAP os dados são armazenados em um banco de dados multidimensional, que propriamente executa a função de servidor OLAP.

4.6.3 HOLAP

HOLAP é o acrônimo de Hybrid OLAP (OLAP Híbrido).

O conceito de HOLAP é encontrado em DM REVIEW (2006. Glossary: HOLAP): “Um produto que pode prover análise multidimensional simultaneamente de dados armazenados em um banco de dados multidimensional e em um SGBD Relacional. Tornou-se uma arquitetura popular para servidor OLAP.”

Ou seja, HOLAP é uma tecnologia mista de ROLAP e MOLAP que tira

proveito do melhor de cada uma delas: a escalabilidade do ROLAP e o alto desempenho do MOLAP.

4.6.4 DOLAP

DOLAP é o acrônimo de Desktop OLAP (OLAP Desktop).

Uma breve explanação sobre DOLAP pode ser encontrada em Howson (2002):

Ferramentas DOLAP criam dinamicamente um micro cubo ou no PC cliente, ou em uma instalação na internet, ou em um servidor de aplicação em uma camada intermediária. Para criar o micro cubo, os usuários lançam uma query SQL usando uma visão de negócios do data warehouse. A ferramenta DOLAP então envia os resultados de volta para o desktop e os formata em um cubo. O formato de micro cubo permite aos usuários executar certas funções tais como drill-down que historicamente só eram disponíveis com cubos MOLAP. Ele também permite aos usuários lançar instruções SQL complexas, combinando dados de diferentes tabelas fato ou até mesmo planilhas eletrônicas, ainda assim apresentando aos usuários um documento perfeitamente consistente. O principal benefício desta abordagem sobre o MOLAP é a flexibilidade; os usuários não tem que definir suas requisições de informações antes do tempo de uma forma precisa como acontece com o MOLAP. Além disso, um banco de dados relacional pode armazenar mais dados do que cubos MOLAP, dando liberdade aos usuários para analisar grandes conjuntos de dados. O aspecto negativo são os tempos de resposta mais lentos e cálculos multidimensionais limitados.

Ou seja, há uma diminuição da carga de trabalho do SGBD do data warehouse, pois o processamento é feito nas máquinas cliente e não no servidor. Ao servidor compete garantir a integridade dos dados e atualizar os cubos de dados existentes.

Esta arquitetura também é útil para usuários móveis, pois não necessita de uma conexão permanente com o DW.

Um fator a ser considerado é o tamanho do micro cubo, pois se for muito grande as consultas e análises podem tornar-se demoradas ou até mesmo a

máquina cliente pode não suportar a carga de trabalho.

4.6.5 WOLAP

WOLAP é o acrônimo de Web OLAP (OLAP Web).

WOLAP refere-se simplesmente a uma interface Web para acessar a aplicação OLAP, ou seja, os dados OLAP são acessíveis a partir de um Web browser.

Com relação ao que foi exposto anteriormente, a respeito dos mecanismos de armazenamento e acesso a dados, pode-se constatar que a SQL é utilizada nas arquiteturas ROLAP e HOLAP, como também pode-se observar a utilização de bancos de dados multidimensionais nas arquiteturas MOLAP e HOLAP. Além disso, clientes OLAP com processamento multidimensional possivelmente estarão presentes nas arquiteturas ROLAP e WOLAP, e certamente na arquitetura DOLAP.

A respeito de alguns debates na comunidade de Tecnologia da Informação sobre qual arquitetura OLAP seria mais indicada, e sobre a adição de letras ao termo OLAP, há uma menção em Thomsen (2002, p. 7):

Infelizmente, fazer a pergunta “o que é melhor, MOLAP ou ROLAP?” faz tão pouco sentido quanto perguntar “o que é melhor, um carro ou um barco?” Obviamente, isso depende do que você está tentando fazer – atravessar uma cidade ou um lago – e também das suas restrições. A existência do debate ROLAP *versus* MOLAP é baseada na falsa premissa de que a escolha é binária. Na moda, a integração das capacidades multidimensionais e das capacidades relacionais é melhor descrita por um espectro de possibilidades, onde as noções de ROLAP puro e MOLAP puro são limites inatingíveis e teóricos. [...] a maioria das organizações precisa de alguma mistura de capacidades que, se precisasse de uma qualificação por letra, seria HOLAP. Mas, como qualquer conhecimento apropriado de OLAP distinguiria entre a linguagem ou aspectos lógicos do OLAP e sua implementação física, tal conhecimento apropriado do OLAP revela que, fisicamente, ele pode ser de qualquer espécie. Assim, o conceito do H já está incluído nas características físicas do OLAP e nenhuma espécie de letra adicional é necessária.

4.7 Benefícios do OLAP

Um dos maiores benefícios do OLAP é justamente o suporte para tomadas de decisão através do fornecimento de informações de forma rápida e precisa, com alta flexibilidade e desempenho. Com as informações extraídas a partir dos dados corporativos e através de formas simples e intuitivas de navegação, pesquisa e visualização, os executivos são capazes de analisar vários cenários, fazer análises “what if” (o que acontece se), consultas “ad hoc”, previsões e descobrir tendências. Isso pode determinar a competitividade e sucesso de uma empresa.

Na terceira pesquisa chamada “The OLAP Survey”, foram destacados alguns benefícios obtidos com a utilização do OLAP, relatados em Pendse (2006):

- Relatórios mais rápidos ou mais precisos
- Melhores decisões de negócios através de análises mais completas ou oportunas
- Satisfação dos clientes aumentada através da melhoria da qualidade de produtos e/ou serviços
- Economia de outros custos (que não de TI) (e.g. inventário, desperdício, financiamento)
- Rendimentos aumentados através de melhores análises de vendas e marketing
- Economia de pessoal em departamentos comerciais
- Custos externos de TI reduzidos (suporte de hardware, assessoria ou licenciamento de software)
- Economia de pessoal em TI

4.8 Áreas de Aplicação para OLAP

Algumas aplicações para OLAP são descritas em Reinschmidt e Françoise (2000):

Aplicações OLAP abrangem uma variedade de funções organizacionais. Departamentos financeiros usam OLAP para aplicações tais como orçamento, atividades baseadas em custos (destinação de verbas), análise de desempenho financeiro e modelagem financeira. Análise de vendas e previsões são duas das aplicações OLAP encontradas em departamentos de vendas. Entre outras aplicações, departamentos de marketing usam OLAP para análise de pesquisa de mercado, previsão de vendas, análise de promoções, análise de clientes e segmentação mercado/cliente. Aplicações OLAP típicas para manufatura incluem planejamento de produção e análise de defeitos.

Nigel Pendse enumera algumas aplicações para OLAP em PENDSE (2005. OLAP applications):

- Marketing e Análise de vendas
- Análise de Clickstream
- Database marketing
- Orçamento
- Relatórios e consolidação financeiros
- Management reporting
- EIS
- Balanced scorecard
- Análise de Rentabilidade
- Análise de Qualidade

Ralph Kimball ilustra a criação de modelos multidimensionais com estudos de caso e aborda algumas áreas de aplicação para OLAP em Kimball e Ross (2002):

- Vendas a varejo
- Estoque
- Compras
- Gerenciamento de Pedidos
- Gerenciamento de relacionamento com o cliente (CRM)
- Contabilidade
- Gerenciamento de recursos humanos
- Serviços financeiros
- Telecomunicações e utilitários
- Transporte
- Educação
- Assistência médica
- Comércio eletrônico
- Seguro

5 CRITÉRIOS PARA AVALIAÇÃO DE FERRAMENTAS OLAP

O objetivo deste capítulo é apresentar os critérios conhecidos para avaliação de ferramentas OLAP, finalizando com a seleção das características para a análise das ferramentas open source, motivo deste trabalho.

5.1 As 12 Regras de Codd

Edgar F. Codd foi quem definiu os primeiros critérios, suas famosas “12 Regras para avaliar produtos OLAP” em Codd, Codd e Salley (1993):

1. Visão Conceitual Multidimensional
2. Transparência
3. Acessibilidade
4. Desempenho Coerente de Relatório
5. Arquitetura Cliente-Servidor
6. Dimensionalidade Genérica
7. Tratamento Dinâmico de Matriz Esparsa
8. Suporte para Multiusuário
9. Operações Irrestritas de Dimensão Cruzada
10. Manipulação de Dados Intuitiva
11. Relatórios Flexíveis
12. Dimensões e Níveis de Agregação Ilimitados

Quando da publicação de suas 12 regras para OLAP, Codd era um pesquisador contratado por uma empresa que possuía um produto OLAP e notou-se uma parcialidade das mesmas em relação ao produto. Por esse motivo, suas 12 regras (às quais posteriormente foram adicionadas mais 6) foram muito criticadas. Houveram alegações de que eram específicas demais e também que não podiam ser atendidas por sistemas que poderiam ser considerados OLAP.

Este assunto é abordado em WIKIPEDIA CONTRIBUTORS (2006. OLAP):

O primeiro produto que executava queries OLAP foi o *Express* da empresa IRI, que foi liberado em 1970 (e adquirido pela Oracle em 1995). Porém, o termo não apareceu até 1993 quando foi cunhado por Ted Codd, que havia sido descrito como “o pai do banco de dados relacional”. Mas o artigo de Codd foi financiado pela antiga Arbor Software (agora Hyperion Solutions), como uma espécie de golpe de marketing: a companhia havia liberado seu próprio produto OLAP – *Essbase* – um ano antes. Como resultado, as “doze regras de processamento analítico online” de Codd eram explícitas em sua referência ao *Essbase*. Houve alguma controvérsia como consequência disso, e quando a *Computerworld* descobriu que Codd era pago pela Arbor, ela retratou o artigo.

5.2 O Teste FASMI

O “teste FASMI” de Nigel Pendse é uma alternativa às 12 regras de Codd, conforme relatado em Pendse (2005. What is OLAP?):

Nossa definição é projetada para ser curta e fácil de lembrar – 12 regras ou 18 características são demais para a maioria das pessoas levar em suas cabeças; nós estamos satisfeitos em ser capazes de resumir a definição OLAP em simplesmente cinco palavras chave: Fast Analysis of Shared Multidimensional Information – ou FASMI para abreviar.

Dessa forma, a sigla FASMI apresenta cinco características para OLAP, descritas brevemente a seguir e relatadas em PENDSE (2005. What is OLAP?):

Fast – significa que o sistema deve apresentar respostas aos usuários

dentro de aproximadamente cinco segundos, com as análises mais simples não demorando mais de um segundo e pouquíssimas análises demorando mais de vinte segundos.

Analysis – significa que o sistema pode suportar qualquer lógica de negócio e análise estatística que é relevante para a aplicação e para o usuário e que possam ser feitas com suficiente facilidade pelo usuário final.

Shared – significa que o sistema implementa todas as condições de segurança para a confidencialidade, e se múltiplos acessos de escrita não necessários, existe um bloqueio de atualizações concorrentes em um nível apropriado.

Multidimensional – é o requisito chave. O sistema deve prover uma visão conceitual multidimensional dos dados, incluindo total suporte a hierarquias e múltiplas hierarquias, certamente a maneira mais lógica para análise de negócios e organizações.

Information - São todos os dados e informações derivadas necessárias, sempre que ela seja relevante para a aplicação. A medição da capacidade de vários produtos é em termos de quantos dados eles poder tratar, e não de quantos gigabytes eles podem armazenar.

5.3 Características Seleccionadas

A seguir serão descritas as características que serão verificadas nas ferramentas open source para ambientes OLAP, divididas em três categorias.

5.3.1 Recursos OLAP

Será verificada a existência ou não dos seguintes recursos OLAP:

Drill Down e Roll UP;

Slice and Dice;

Drill Across;

Pivoting;

Ranking.

5.3.2 Linguagens

Serão verificadas as seguintes características relativas à linguagens:

Linguagem MDX – Verificação se a ferramenta implementa ou não a MDX;

Consultas “ad hoc” – Verificação da possibilidade de execução ou não;

Multipass SQL – Verificação de existência ou não (aplicável para ROLAP e HOLAP);

Suporte a Java – Verificação da existência ou não;

Suporte a JDBC – Verificação da existência ou não.

5.3.3 Gerais

Serão verificadas as seguintes características gerais:

Suporte a Hierarquias – Verificação da existência ou não;

Acesso Multiusuário – Verificação se permite acesso multiusuário ou não;

Arquitetura OLAP – Verificação da arquitetura OLAP das ferramentas.

6 FERRAMENTAS OPEN SOURCE PARA AMBIENTES OLAP

Conforme relatado na introdução deste trabalho, as ferramentas analisadas devem obrigatoriamente possuir servidor e cliente OLAP.

As ferramentas foram selecionadas principalmente através do SourceForge.net, o maior repositório de projetos de software open source da internet, e também através de pesquisas na própria internet. A data de fechamento desta seleção foi 18/07/2006. O SourceForge.net foi consultado através da URL http://sourceforge.net/search/?type_of_search=soft&words=olap.

A amostragem foi feita considerando-se o percentual de atividade da ferramenta no repositório, com um limite mínimo desejado não inferior a 90%.

As ferramentas devem rodar pelo menos nos sistemas operacionais Windows e Linux. Este aspecto é importante devido à portabilidade das aplicações que possam ser desenvolvidas com estas ferramentas.

Considerando estes requisitos, foram descartadas deste trabalho algumas ferramentas que não atendem aos mesmos. Estas ferramentas e o motivo principal do descarte estão relacionados na tabela a seguir (Tabela 3).

Tabela 3 – Ferramentas Descartadas

FERRAMENTA / SITE DO PROJETO	MOTIVO DO DESCARTE
BEE http://sourceforge.net/projects/bee	Não roda em Windows
BIRT Project - Business Intelligence and Reporting Tools http://www.eclipse.org/birt/phoenix/gratis-OLAP (gOLAP)	Ferramenta voltada para geração de relatórios e não para OLAP
http://sourceforge.net/projects/golap	Atividade muito baixa no SourceForge.net: 12,49%
Hydracube http://sourceforge.net/projects/hydracube	Não possui cliente OLAP

JasperReports - Java Reporting http://sourceforge.net/projects/jasperreports	Não possui servidor OLAP
JMagallanes Olap and Report http://sourceforge.net/projects/jmagallanes	Não possui servidor OLAP
OpenI: Web-based BI Application for OLAP http://sourceforge.net/projects/openi	Não possui servidor OLAP
OpenOLAP for MySQL http://sourceforge.net/projects/r-kajiyama	Não possui cliente OLAP
PALO - Open Source MOLAP http://www.opensourceolap.org/	Não possui cliente OLAP
pocOLAP - the little OLAP project http://sourceforge.net/projects/pocolap	Não possui servidor OLAP
Rubik http://sourceforge.net/projects/rubik	Não possui servidor OLAP
XML-OLAP http://sourceforge.net/projects/xml-olap	Não possui desenvolvimento ativo desde 02/06/2005
Fonte: Do autor.	

6.1 Ferramentas Seleccionadas

Somente uma ferramenta foi seleccionada: Pentaho - Business Intelligence. A ferramenta Pentaho possui um módulo chamado Pentaho Analysis que incorpora o servidor OLAP Mondrian e o cliente OLAP JPivot. Por este motivo, serão analisadas como uma única ferramenta.

A seguir serão apresentadas algumas características relevantes desta ferramenta e seus dois componentes principais para OLAP, coletadas no repositório SourceForge.net.

6.1.1 Pentaho - Business Intelligence

- Descrição do projeto: Uma completa plataforma de BI que inclui relatórios, análises (OLAP), dashboards, data mining e ETL. Use-a como uma suíte completa ou como componentes individuais que são acessíveis via web services. Classificada como número 1 em BI open source.

- Atividade no repositório: 99,95%;
- Status de desenvolvimento: 4 - Beta, 5 - Produção/Estável;
- Sistema Operacional: Independente de SO (Escrita em uma linguagem interpretada);
 - Linguagem de Programação: Java, JavaScript, JSP, XSL (XSLT/XPath/XSL-FO);
 - Interface de Usuário: Java SWT, Eclipse, Web-based;
 - Licença: Mozilla Public License 1.1 (MPL 1.1);
 - Categoria: Enterprise, Data Warehousing, ERP, OLAP;
 - Site do Projeto: <http://sourceforge.net/projects/pentaho>.

6.1.2 Mondrian

• Descrição do projeto: Mondrian é um banco de dados OLAP escrito em Java. Ele implementa a linguagem MDX, a XML for Analysis e as especificações JOLAP. Ele lê a partir de SQL e outras fontes de dados, e agrega dados em uma memória cache.

- Atividade no repositório: 99,87%;
- Status de desenvolvimento: 5 - Produção/Estável;
- Sistema Operacional: (Nenhum listado);
- Linguagem de Programação: Java;
- Interface de Usuário: (Nenhuma listada);
- Licença: Common Public License;
- Categoria: Database Engines/Servers, Enterprise, OLAP;
- Site do Projeto: <http://sourceforge.net/projects/mondrian>.

6.1.3 JPivot

- Descrição do projeto: Um cliente OLAP baseado em JSP. JPivot é uma biblioteca customizada JSP que desenha uma tabela e gráfico OLAP. Os usuários podem executar típicas navegações OLAP como drill down e slice and dice. Ele usa Mondrian e XMLA como seus engines OLAP;

- Atividade no repositório: 99,03%;

- Status de desenvolvimento: 5 - Produção/Estável;

- Sistema Operacional: Independente de SO (Escrita em uma linguagem interpretada);

- Linguagem de Programação: Java;

- Interface de Usuário: Web-based;

- Licença: Common Public License;

- Categoria: Front-Ends, Dynamic Content, Financial, Software Development;

- Site do Projeto: <http://sourceforge.net/projects/jpivot>.

7 METODOLOGIA

Este capítulo descreve a metodologia utilizada neste trabalho de pesquisa.

7.1 Tipo de Pesquisa

Os itens seguintes descrevem o tipo desta pesquisa, quanto à abordagem e aos objetivos.

7.1.1 Abordagem

A pesquisa possui uma abordagem qualitativa considerando-se que a mesma propõe-se a identificar e caracterizar seus objetos de pesquisa. Também é uma pesquisa aplicada visto que seus objetos serão operacionalizados na prática.

7.1.2 Objetivos

Quanto a seus objetivos, de acordo com Thomsen (2002, p. 543) no capítulo “Orientações multidimensionais”, esta pesquisa é descritiva:

A finalidade destas orientações é descrever as principais áreas da funcionalidade multidimensional em termos de categorias de especialidade que qualquer ferramenta precisa oferecer e em termos das características que as ferramentas podem ou não oferecer.

8 APRESENTAÇÃO E ANÁLISE DOS DADOS

As características analisadas da única ferramenta open source que possui servidor e cliente OLAP, que possui uma atividade no SourceForge.net não inferior a 90% e que roda nos sistemas operacionais Windows e Linux são exibidas na tabela a seguir (Tabela 4).

Tabela 4 – Características da ferramenta Pentaho – BI.

CARACTERÍSTICA	RESULTADO
Recursos OLAP	
Drill Down e Roll UP	Sim
Slice and Dice	Sim
Drill Across	Não
Pivoting	Sim
Ranking	Sim
Linguagens	
Linguagem MDX	Sim
Consultas “ad hoc”	Sim (MDX Query Editor)
Multipass SQL	Sim
Suporte a Java	Sim
Suporte a JDBC	Sim
Gerais	
Suporte a Hierarquias	Sim (Agregação)
Acesso Multiusuário	Sim
Arquitetura OLAP	ROLAP

Fonte: Do autor.

9 CONCLUSÃO

O panorama atual de ferramentas OLAP proprietárias mostra fornecedores e produtos estabelecidos, maduros e consistentes. Sendo que os principais fornecedores de SGBDs proprietários oferecem ferramentas OLAP e estenderam seus produtos para um melhor suporte ao OLAP, tudo isso já a alguns anos.

Tais características ainda não estão presentes nas ferramentas OLAP open source. Como exemplo pode-se citar a ferramenta Pentaho BI que não suporta o recurso OLAP Drill-across, que é uma operação OLAP básica. No site da ferramenta observa-se que este recurso está a mais de um ano para ser implementado. Quanto à extensão dos SGBDs open source para um melhor suporte ao OLAP a cena se repete. Existem alguns movimentos isolados iniciais, como o caso do SGBD Bizgres, que é uma extensão do PostgreSQL para um melhor suporte a Data Warehouse.

Observa-se uma expectativa da comunidade de TI sobre ferramentas open source não só para OLAP, mas para Business Intelligence de uma maneira geral, e também que está havendo um movimento dos desenvolvedores open source nesse sentido; mas também observa-se que há um grande caminho a percorrer até chegar ao estágio atual das ferramentas proprietárias.

Através deste estudo também é possível concluir que várias ferramentas não oferecem uma solução completa (ou algo próximo disso) para OLAP, mas somente alguns componentes, como servidor ou cliente OLAP. A própria ferramenta Pentaho BI incorporou dois projetos open source independentes – Mondrian e JPivot

– para montar sua solução OLAP, chamada de Pentaho Analysis.

No repositório SourceForge.net também encontram-se ferramentas OLAP com baixíssimo percentual de atividade, ou então com pouquíssimos desenvolvedores trabalhando em alguns projetos.

Por tudo que foi visto, pode-se afirmar que as ferramentas open source para ambientes OLAP ainda estão em um estágio de desenvolvimento um tanto quanto inconsistente, e que os profissionais de TI que desejam usufruir de seus benefícios ainda terão um tempo de espera pela frente até que as mesmas estejam em um patamar mais sólido.

REFERÊNCIAS

ABELLÓ, Alberto. **YAM²: A Multidimensional Conceptual Model**. 2002. 179 f. Tese (Programa de Doctorat de Software) – Universitat Politècnica de Catalunya, Barcelona.

BALLARD, Chuck et al. **Dimensional Modeling: In a Business Intelligence Environment**. International Business Machines Corporation (IBM Corp.), 2006. Disponível em: <<http://www.redbooks.ibm.com/redbooks/pdfs/sg247138.pdf>>. Acesso em: 06 jun. 2006.

CHAPPLE, Mike. **Databases Glossary: "Flat File"**. About, Inc., 2006. Disponível em: <<http://databases.about.com/cs/administration/g/flatfile.html>>. Acesso em: 02 maio 2006.

CODD, E. F.; CODD, S. B.; SALLEY, C. T. **Providing OLAP to User-Analysts: An IT Mandate**. E. F. Codd & Associates, 1993. Disponível em: <http://dev.hyperion.com/resource_library/white_papers/providing_olap_to_user_analysts.pdf>. Acesso em: 10 out. 2005.

DM REVIEW. **Glossary: HOLAP**. DM Review and SourceMedia, Inc., 2006. Disponível em: <<http://www.dmreview.com/resources/glossary.cfm?keywordId=H>>. Acesso em 18 jul. 2006.

DM REVIEW. **Glossary: ROLAP**. DM Review and SourceMedia, Inc., 2006. Disponível em: <<http://www.dmreview.com/resources/glossary.cfm?keywordId=R>>. Acesso em 18 jul. 2006.

DM REVIEW. **Glossary: Surrogate Key**. DM Review and SourceMedia, Inc., 2006. Disponível em: <<http://www.dmreview.com/resources/glossary.cfm?keywordId=S>>. Acesso em 18 jul. 2006.

DWBRASIL. **OLAP**. 2003. Disponível em: <<http://www.dwbrasil.com.br/html/olap.html>>. Acesso em 11 jul. 2006.

ECKERSON, Wayne. **A New Old Market**. The Data Warehouse Institute, 2002. Disponível em: <<http://www.tdwi.org/research/display.aspx?id=6475>>. Acesso em 05 abr. 2006.

ECKERSON, Wayne; WHITE, Colin. **Evaluating ETL and Data Integration Platforms**. The Data Warehouse Institute, 2003. Disponível em: <http://download.101com.com/tdwi/research_report/2003ETLReport.pdf>. Acesso em 06 abr. 2006.

FERREIRA, Rafael. G. C. **Data Warehouse e Sistema OLAP**. Porto Alegre: [s/e], 2004. (mimeo)

GARCIA-MOLINA, Hector; ULLMAN, J. D.; WIDOM, Jennifer. **Implementação de Sistemas de Bancos de Dados**. Tradução de Vandenberg D. de Souza. Rio de Janeiro: Campus, 2001. 685 p.

HOWE, Denis. **Surrogate Key Definition**. Learnthat.com, 2004. Disponível em: <<http://www.learnthat.com/define/view.asp?id=7032>>. Acesso em 04 maio 2006.

HOWSON, Cindi. **MOLAP and DOLAP: Apples and Oranges**. The Data Warehouse Institute, 2002. Disponível em: <http://www.askcindi.com/TDWI_DOLAP_MOLAP07_02.htm>. Acesso em 12 jul. 2006.

IMHOFF, Claudia; GALEMMO, Nicholas; GEIGER, J. G. **Mastering Data Warehouse Design: Relational and Dimensional Techniques**. Indianapolis: Wiley Publishing, Inc., 2003. 438 p.

INMON, W. H. **Building the Data Warehouse**. Third Edition. New York: John Wiley & Sons, Inc., 2002. 412 p.

_____. **Data Mart Does Not Equal Data Warehouse**. DM Review, 1999. Disponível em: <http://dmreview.com/article_sub.cfm?articleId=1675>. Acesso em 05 abr. 2006.

INMON, W. H.; TEDERMAN, R. H.; IMHOFF, Claudia. **Data Warehouse: Como transformar informações em oportunidades de negócios**. Tradução de Melissa Kassner. São Paulo: Berkeley Brasil, 2001. 266 p.

KAY, Russell. **Data Cubes**. Computerworld Inc., 2005. Disponível em: <<http://www.computerworld.com/databasetopics/data/story/0,10801,91640,00.html>>. Acesso em 25 abr. 2006.

KIMBALL, Ralph. **Features for Query Tools**. Miller Freeman, Inc., 1997. Disponível em: <<http://www.dbmsmag.com/9702d05.html>>. Acesso em 21 jul. 2006.

KIMBALL, Ralph; ROSS, Margy. **The Data Warehouse Toolkit: Guia completo para modelagem dimensional**. Tradução de Ana Beatriz Tavares e Daniela Lacerda. 2. ed. Rio de Janeiro: Campus, 2002. 494 p.

OPEN SOURCE INITIATIVE. **The Open Source Definition**. Version 1.9. 2006. Disponível em: <<http://www.opensource.org/docs/definition.php>>. Acesso em 27 fev. 2006.

PENDSE, Nigel. **OLAP applications**. Londres: The OLAP Report, 2005. Disponível em: <<http://www.olapreport.com/Applications.htm>>. Acesso em 13 jul. 2006.

_____. **What is OLAP?** Londres: The OLAP Report, 2005. Disponível em: <<http://www.olapreport.com/fasmi.htm>>. Acesso em 10 out. 2005.

_____. **Drilling Into OLAP Benefits.** DM Review and SourceMedia, Inc., 2006. Disponível em: <http://www.dmreview.com/article_sub.cfm?articleId=8174>. Acesso em 13 jul. 2006.

POWELL, Ron. **A 10 Year Journey.** DM Review and SourceMedia, Inc., 2006. Disponível em: <http://www.dmreview.com/article_sub.cfm?articleId=3003>. Acesso em 18 jul. 2006.

REINSCHMIDT, Joerg; FRANCOISE, Allison. **Business Intelligence Certification Guide.** IBM Corp., 2000. Disponível em: <<http://www.redbooks.ibm.com/redbooks/pdfs/sg245747.pdf>>. Acesso em 14 jul. 2006.

SINGH, H. S. **Data Warehouse: Conceitos, Tecnologias, Implementação e Gerenciamento.** Tradução de Monica Rosemberg. São Paulo: Makron Books, 2001. 382 p.

THE OLAP COUNCIL. **OLAP and OLAP Server Definitions.** 1995. Disponível em: <<http://dssresources.com/glossary/olaptrms.html>>. Acesso em 17 out. 2005.

THOMSEN, Erik. **OLAP: Construindo sistemas de informações multidimensionais.** Tradução de Daniel Vieira. 2. ed. Rio de Janeiro: Campus, 2002. 708 p.

WIKIPEDIA CONTRIBUTORS. **Flat file database.** Wikipedia, The Free Encyclopedia, 2006. Disponível em: <http://en.wikipedia.org/wiki/Flat_file_database>. Acesso em: 18 jul. 2006.

WIKIPEDIA CONTRIBUTORS. **MOLAP.** Wikipedia, The Free Encyclopedia, 2006. Disponível em: <<http://en.wikipedia.org/wiki/MOLAP>>. Acesso em 18 jul. 2006.

WIKIPEDIA CONTRIBUTORS. **Multidimensional Expressions.** Wikipedia, The Free Encyclopedia, 2006. Disponível em: <http://en.wikipedia.org/wiki/Multidimensional_Expressions>. Acesso em: 18 jul. 2006.

WIKIPEDIA CONTRIBUTORS. **OLAP.** Wikipedia, The Free Encyclopedia, 2006. Disponível em: <<http://en.wikipedia.org/wiki/Olap>>. Acesso em: 18 jul. 2006.

ANEXOS

ANEXO A - The Open Source Definition

A definição do termo open source é encontrada em OPEN SOURCE INITIATIVE (2006):

1. Redistribuição Livre

A licença não deve restringir qualquer parte de vender ou dar o software como um componente de uma distribuição de software agregado contendo programas de várias fontes diferentes. A licença não deve exigir direitos autorais ou outro pagamento para tal venda.

2. Código Fonte

O programa deve incluir o código fonte e permitir a distribuição do mesmo, também na forma compilada. Onde alguma forma de um produto não é distribuída com o código fonte, devem existir, bem publicados, os meios de obtenção do código fonte, que não sejam mais do que uma reprodução razoável de melhor custo, por download pela internet sem pagamento algum. O código fonte deve ser a forma preferida pela qual um programador modificará o programa. Código fonte deliberadamente confuso não é permitido. Formas intermediárias tais como a saída de um pré-processador ou tradutor não são permitidas.

3. Trabalhos Derivados

A licença deve permitir modificações e trabalhos derivados, e deve permitir que os mesmos sejam distribuídos sob os mesmos termos da licença do software original.

4. Integridade do Código Fonte do Autor

A licença pode restringir o código fonte de ser distribuído na forma modificada somente se a licença permitir a distribuição de “arquivos de correções”

com o código fonte para o propósito de modificar o programa em tempo de construção. A licença deve permitir explicitamente a distribuição de software construído a partir de código fonte modificado. A licença pode exigir trabalhos derivados para produzir um nome ou número de versão diferente do software original.

5. Nenhuma Discriminação Contra Pessoas ou Grupos

A licença não deve discriminar contra qualquer pessoa ou grupo de pessoas.

6. Nenhuma Discriminação Contra Área de Aplicação

A licença não deve restringir qualquer um de fazer uso do programa em uma área específica de aplicação. Por exemplo, ela não pode restringir o programa de ser usado em um negócio ou de ser usado para pesquisa genética.

7. Distribuição de Licença

Os direitos anexados ao programa devem ser aplicados a todos para quem o programa é redistribuído sem a necessidade de execução de uma licença adicional para aquelas partes.

8. A Licença Não Deve Ser Específica para um Produto

Os direitos anexados ao programa não devem depender de partes do programa de uma distribuição de software particular. Se o programa é extraído desta distribuição e usado ou distribuído dentro dos termos da licença do programa, todas as partes para quem o programa é redistribuído devem ter os mesmos direitos como aqueles que são concedidos em conjunção com a distribuição de software original.

9. A Licença Não Deve Restringir Outro Software

A Licença não deve estabelecer restrições sobre outro software que é distribuído junto com o software licenciado. Por exemplo, a licença não deve insistir

que todos os outros programas distribuídos na mesma mídia devem ser software open source.

10. A Licença Deve Ser Tecnicamente Neutra

Nenhum fornecimento da licença pode ser estabelecido sobre qualquer tecnologia individual ou estilo de interface.