



Fundação Educacional do Município de Assis
Instituto Municipal de Ensino Superior de Assis
Campus "José Santilli Sobrinho"

FRANCIELE DE OLIVEIRA DIAS

**FERRAMENTAS OLAP: UM ESTUDO BASEADO NO SISTEMA
PENTAHO**

ASSIS
2014

FRANCIELE DE OLIVEIRA DIAS

**FERRAMENTAS OLAP: UM ESTUDO BASEADO NO SISTEMA
PENTAHO**

Trabalho de conclusão de curso apresentado ao curso de Análise e Desenvolvimento de Sistemas do Instituto Municipal de Ensino Superior de Assis - IMESA, como requisito parcial à obtenção, do Certificado de Conclusão.

Orientador: Prof. Dr. Osmar Aparecido Machado

Área de Concentração: Sistema de Informação

ASSIS
2014

FICHA CATALOGRÁFICA

DIAS, Franciele de Oliveira.

FERRAMENTAS OLAP: Um estudo baseado no Sistema Pentaho/Franciele de Oliveira Dias. Fundação Educacional do Município de Assis, 2014.

Orientador: Prof. Dr. Osmar Aparecido Machado

Trabalho de Conclusão de Curso

Instituto Municipal de Educação Superior de Assis – IMESA

1. Ferramentas Olap 2. Sistema Pentaho

CDD: 001.61

Biblioteca da FEMA

FERRAMENTAS OLAP: UM ESTUDO DE BASEADO NO SISTEMA PENTAHO

FRANCIELE DE OLIVEIRA DIAS

Trabalho de Conclusão de Curso apresentado ao Instituto Municipal de Ensino Superior de Assis, como requisito do Curso Superior de Análise e Desenvolvimento de Sistemas, analisando pela seguinte comissão examinadora:

Orientador: Prof. Dr. Osmar Aparecido Machado

Analisador: Domingos de Carvalho Villela Junior

ASSIS

2014

DEDICATÓRIA

Dedico este trabalho primeiramente a Deus que me iluminou todos esses anos e não deixou que eu desistisse do meu sonho, aos meus pais que me apoiaram nesta jornada. Ao meu namorado também, pelo seu carinho, compreensão e companheirismo.

AGRADECIMENTOS

Em primeiro lugar, agradeço á Deus, que me deu forças e clareou o meu caminho, ajudando-me superar as dificuldades e os obstáculos, mas no final, fui presenteado por este momento. Aos meus pais Valdeci e Luís que me incentivaram durante toda a caminhada, não deixando que eu desistisse me incentivando e mostrando a segurança de uma verdadeira família.

Ofereço também ao meu namorado Alan que durante a minha vida sempre esteve junto comigo me ajudando e apoiando ele foi meu alicerce essencial para este projeto.

Quero agradecer também aos professores que passaram pela minha vida, desde o ensinamento das primeiras letras até os algoritmos mais complexos. Neste momento gostaria de enfatizar meu agradecimento ao professor Osmar Aparecido Machado, que acreditou no meu potencial e, com dedicação e empenho ajudou-me a realizar este trabalho além de ser a figura fundamental nesta orientação, pela sua paciência e por nunca negar seu conhecimento.

A todas as pessoas que passaram pela minha vida e me acrescentaram algum conhecimento, ficam aqui meu sentimento e minha gratidão.

“Habilidade é o que você é capaz de fazer.

Motivação determina o que você faz.

Atitude determina a qualidade do que você faz.”

(Lou Holtz)

RESUMO

Embora o conceito de BI (*Business Intelligence*) esteja sendo utilizado a décadas pelas empresas, muitas ainda utilizam ferramentas como planilhas eletrônicas para a análise dos seus dados, cujas limitações impossibilitam o cruzamento de muitas informações, desconsiderando dados importantes para a tomada de decisão. O objetivo deste trabalho é apresentar um estudo sobre as ferramentas OLAP, baseado no aplicativo *Pentaho*, oferecendo informações relevantes para os profissionais da área de informática que desejam utilizar estas tecnologias, buscando também atender as necessidades das empresas no sentido de gerenciamento de informações, assim com o fornecimento de mecanismos moderno e dinâmico para a busca e manipulação de informações. Isso se justifica pelo fato de as empresas buscarem constantemente novas ferramentas para melhorias de seus processos, para assim, sobreviverem no mercado altamente competitivo e globalizado. Neste projeto a abordagem será qualitativa, onde se busca fundamentar os argumentos quanto ao nível de usabilidade ou complexidade operacional do sistema *Pentaho*. Como resultado, esta pesquisa pode abrir uma nova oportunidade de negócio para os profissionais de informática, além dos conhecimentos acadêmicos adquiridos ao decorrer do projeto, isso sem dúvida acrescenta a necessidade de acompanhar os avanços desta tecnologia que está em crescente evolução.

Palavras-chave: Business Intelligence; OLAP; Pentaho; Tomada de decisão.

ABSTRACT

Although the concept of Business Intelligence (BI) is being used for decades by companies, many still use tools like spreadsheets for analysis of their data , which cannot cross a lot of information , disregarding important for decision-making data. The objective of this paper is to present a study of the OLAP tools , based on the Pentaho application , providing relevant information for professionals in the computer area who wish to use these technologies , also seeking to meet the needs of companies towards information management , so with providing modern and dynamic mechanisms for searching and manipulating information . This is justified by the fact that businesses have the need to be always improving their tools used by them, so as to survive in the highly competitive, global marketplace. This project will be qualitative approach, where elements aimed to support an argument about the level of usability and operational complexity of the Pentaho system. As a result, this study may open a new business opportunity for IT professionals beyond the academic knowledge acquired during the project, it said without any doubt the need to monitor the progress of this technology is in increasing trend.

Keywords: Business Intelligence; OLAP; Pentaho, Making decision.

LISTA DE ILUSTRAÇÕES

Figura 1 - Ferramentas de um ambiente Business Intelligence.....	17
Figura 2 - Exemplo de unificação de informações de vendas via propriedade orientada ao assunto.....	31
Figura 3 - Exemplo de tratamento de dados).....	31
Figura 4 - Formas de Atualização.....	32
Figura 5 - Modelo Multidimensional (Star Schema).....	34
<i>Figura 6 - Exemplo genérico do modelo Estrela (Star Schema).....</i>	<i>34</i>
<i>Figura 7 - Modelo Multidimensional Snow Flake.....</i>	<i>35</i>
Figura 8 - Conceito de Data Mart.....	38
<i>Figura 9 - Arquitetura do Pentaho.....</i>	<i>43</i>
<i>Figura 10 - Imagem Pentaho User Console (PUC).....</i>	<i>45</i>
<i>Figura 11 - Imagem Pentaho Metadata Editor (PME).....</i>	<i>46</i>
<i>Figura 12 - Imagem Pentaho Report Designer (PRD).....</i>	<i>47</i>
<i>Figura 13 - Imagem Pentaho Schema Workbench (PSW).....</i>	<i>48</i>
Figura 14 - Mensagem aviso verificador de versão.....	50
Figura 15 - Tela do login do Pentaho.....	51

SUMÁRIO

1- INTRODUÇÃO	14
1.1 OBJETIVOS	14
1.2 JUSTIFICATIVAS / MOTIVAÇÕES	14
1.3 ESTRUTURA DO TRABALHO	15
2. FUNDAMENTAÇÃO TEÓRICA	16
2.1 BUSINESS INTELLIGENCE	16
2.2 SISTEMAS DE INFORMAÇÃO ASSOCIADOS A BI	17
2.2.1 Sistema de gestão empresarial	18
2.2.2 Sistema Transacional ou <i>On-line Transactional Processing</i> (OLTP).....	19
2.2.3 Sistema de Apoio à decisão (SAD) ou <i>On-line Analytical Processing</i> (OLAP)	20
2.2.4 Diferenças entre Sistemas OLTP e OLAP.....	20
3. FERRAMENTAS OLAP	22
3.1.2 Estrutura Multidimensional.....	24
3.2.1 Arquiteturas OLAP.....	27
4.1 Orientado pelo assunto	30
4.2 Integração.....	31
4.3 Não Volátil	32
4.4 Variável no tempo:.....	32
4.5 MODELAGEM DE DADOS NO DATA WAREHOUSE	33
4.5.1 Modelo Estrela (<i>Star Schema</i>)	33
4.5.2 Modelo Floco de Neve (<i>Snow Flake</i>).....	35

4.6 Arquitetura do Data Warehouse	36
4.6.1 Granularidade.....	36
4.6.2 Data Mining	37
4.6.3 Data Mart	37
5 - FERRAMENTAS ETL	39
5.1-Etapas do processo ETL.....	39
5.2 Ferramentas KETTLE (PENTAHO).....	40
6- PLATAFORMA PENTAHO.....	42
6.1 Módulos do Pentaho.....	43
6.2.2 Ferramentas do Pentaho.....	44
6.2.1 Pentaho Administration Console (PAC):	44
6.2.2 Pentaho User Console (PUC).....	44
6.2.3 Pentaho Metada Editor (PME).....	45
6.2.4 Pentaho Report Design (PRD)	46
6.2.5 Pentaho Schema Workbench (PSW).....	47
7. CONCLUSÃO	52
REFERENCIAS BIBLIOGRÁFICAS	53

LISTA DE ABREVIATURAS E SIGLAS

BD	Banco de Dados
BI	<i>Business Intelligence</i>
DM	<i>Data Mart</i>
DOLAP	<i>Desktop On line Analytical Processing</i>
DW	<i>Data WareHouse</i>
ELP	<i>Extract Transform Loader</i>
ER	Entidade Relacionamento
ERP	<i>Enterprise Resources Planning</i>
HOLAP	<i>Hybrid On line Analytical Processing</i>
MOLAP	<i>Multidimensional On line Analytical Processing</i>
OLAP	<i>On line Analytical Processing</i>
OLTP	<i>On line Transaction Processing</i>
ROLAP	<i>Relational On line Analytical Processing</i>
SAD	Sistema de Apoio à decisão
SGBD	Sistema Gerenciador de Banco de Dados
WOLAP	<i>Web On line Analytical Processing</i>
SGBD	Sistema Gerenciador de Banco de Dados

1- INTRODUÇÃO

O processo decisório nas empresas tem sido objeto de investigação de diversos estudos teóricos e executivos. Sua relevância tem despertado o interesse de profissionais de diversas áreas envolvidas em processos de tomada de decisão. Isto porque no atual contexto econômico, político, social, de globalização e intensa concorrência, o desenvolvimento tecnológico tem exigido cada vez mais desses profissionais decisões acertadas, em espaço de tempo cada vez mais reduzido.

As mudanças estão cada vez mais rápidas influenciadas pela evolução tecnológica e da complexidade das operações, as empresas são obrigadas a aperfeiçoarem seus processos de tomada de decisão, tornando-se mais rápidas, mais eficientes e mais seguras.

1.1 OBJETIVOS

Neste contexto, o presente trabalho tem como objetivo explorar os conceitos em torno do ambiente dos sistemas OLAP e utilizar o sistema PENTAHO a fim de identificar sua relevância para os processos de tomada de decisão em organizações de pequeno porte.

1.2 JUSTIFICATIVAS / MOTIVAÇÕES

Com o mercado cada vez mais competitivo as empresas se veem na necessidade de tomar decisões de forma e segura. Nesse sentido, elas buscam por ferramentas que auxiliam na sua gestão e especialmente, que auxiliem nos processos de tomada de decisão (Machado, 2000).

Este trabalho busca trazer uma contribuição teórica sobre ferramentas de apoio e suporte à decisão, com destaque para os sistemas de apoio a decisão. Diversas empresas já adotaram tais ferramentas e, embasados no seus resultados, pode se afirmar que essas ferramentas quando bem

implementadas podem resultar em benefícios significativos para as organizações.

Assim, a necessidade de implementação de ferramentas com esta orientação, são uma tendência especialmente em organizações de grande porte.

1.3 ESTRUTURA DO TRABALHO

A organização deste trabalho está dividida em sete capítulos descritos a seguir:

O primeiro capítulo faz uma breve introdução sobre o objetivo, a metodologia e motivação do conteúdo abordado neste trabalho.

O segundo capítulo apresenta a fundamentação teórica com os principais conceitos e tecnologias que estão direta ou indiretamente ligados ao processo de BI, também serão apresentados conceitos sobre o SAD ou OLAP que serão ferramentas que auxiliam nas tomadas de decisões.

No terceiro capítulo serão abordados os conceitos da ferramenta OLAP, visando sua implementação, sua estrutura e suas dimensões. Além disso, trata da viabilidade de uso para suporte em processos de tomadas de decisão.

No quarto capítulo são descritos os conceitos básicos, modelagens, a arquitetura, elementos, estruturas e fundamentos sobre os *data warehouse*.

No quinto capítulo, são definidos os conceitos, as características e funções das ferramentas ETL, sua abrangência e etapas, assim como uma breve referencia sobre as ferramentas *CloverETL Talend e Kettle/PENTAHO*.

Finalmente, o capítulo sete apresenta os resultados e as considerações finais deste trabalho, destacando os benefícios da abordagem escolhida, as contribuições à conclusão dos resultados obtidos e sugestões para trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

Sistemas de informação são de grande relevância para as organizações, haja vista que uma empresa, na atualidade, não pode operar sem um sistema, seja para cumprir suas rotinas, seja para dar apoio à tomada de decisão.

Nesse sentido, este capítulo trata dos principais conceitos sobre o *business intelligence* – BI e, da mesma forma, faz uma descrição das principais tecnologias envolvidas nos processos de desenvolvimento e utilização de sistemas BI.

2.1 BUSINESS INTELLIGENCE

Business Intelligence (BI) é um termo denominado “guarda-chuva”, pois inclui arquiteturas, ferramentas, banco de dados, aplicações e metodologias. Os principais objetivos do BI são permitir o acesso interativo aos dados (às vezes em tempo real), proporcionar a manipulação desses dados e fornecer aos gerentes e analistas de negócios a capacidade de realizar uma análise adequada das informações a fim de prover suporte para os processos de tomada de decisão.

A idéia do Business Intelligence engloba todas as pesquisas feitas em uma empresa sobre seu desempenho passado que ajudam a moldar o seu desempenho no futuro. Business Intelligence por sua definição é a reunião de todos os dados estatísticos necessários para executar com êxito qualquer negócio, e tomar decisões adequadas com base nesses dados. Estes ambientes tem possibilitado às organizações, compreender e explorar seus dados para dar suporte à tomada de decisões. O tempo do ciclo dos negócios, na atualidade, está extremamente apertado. Por isso, identificar ferramentas que possibilitem melhorias nos processos de decisão, tornando-os mais rápidos, é mais que uma obrigação, é uma necessidade competitiva para a organização.

O ambiente de *Business Intelligence* e suas estruturas pode ser sintetizado na figura 1.

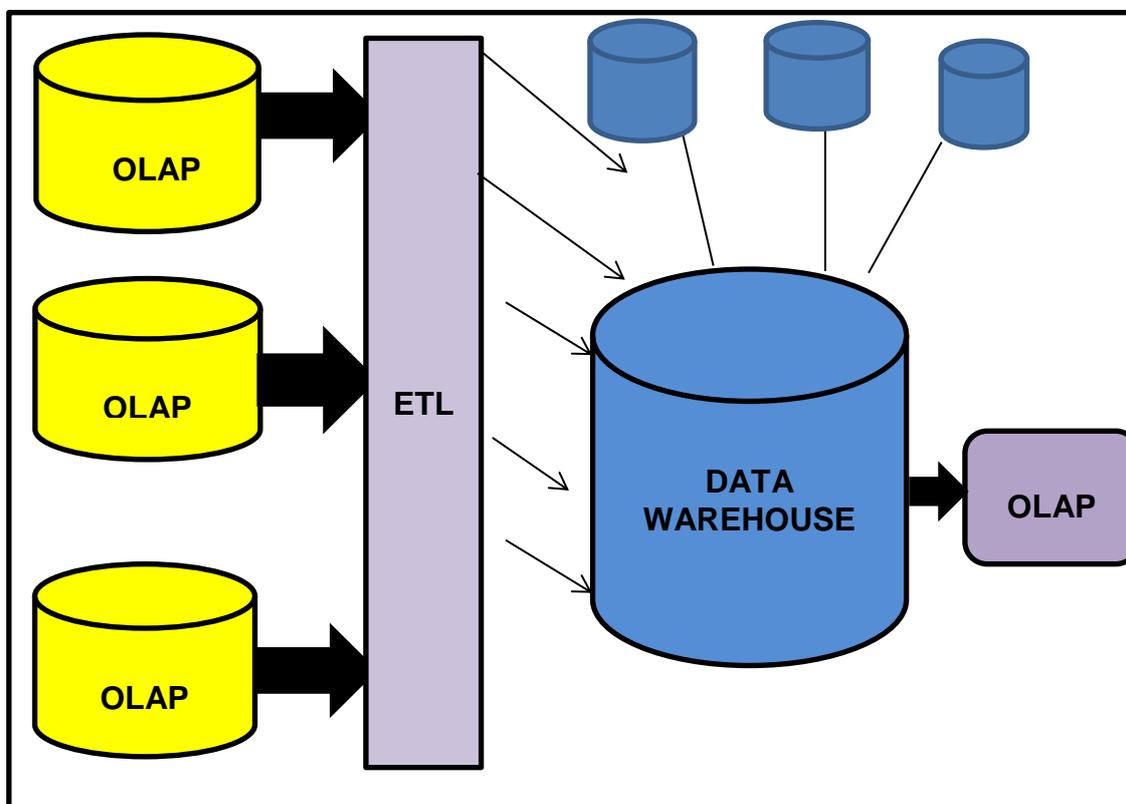


Figura 1 - Ferramentas de um ambiente Business Intelligence, Fonte adaptada do autor segundo (BARBIERI, 2001).

2.2 SISTEMAS DE INFORMAÇÃO ASSOCIADOS A BI

Quando se fala em *Business Intelligence* (BI), intrinsecamente são relacionados alguns tipos de sistemas de informação, como sistemas de gestão empresarial ou ERP (*Enterprise Resource Planning*), sistemas transacionais ou OLTP (*On-line Transactional Processing*) e sistemas de apoio á decisão (SAD) ou OLAP (*On-line Analytical Processing*).

Desta forma, para o pleno entendimento do que é *Business Intelligence* é importante descrever o que são e quais são as funcionalidades de cada um destes tipos de sistemas.

2.2.1 Sistema de gestão empresarial

Desde o início dos anos 2000, em função do bug do milênio, os ERPs tornaram-se uma tendência. São sistemas de gestão integrados, que proporcionam a integração dos setores de uma organização, entretanto, ainda hoje, muitas empresas não os utilizam.

Uma definição bem aceita de ERP, é a que abrange o termo de uma forma didática o conceitua como um sistema de informação adquirido na forma de pacotes comerciais de *softwares*, que permitem a integração entre dados dos sistemas de informação transacionais e dos processos de negócios de uma organização. Ele é definido também uma arquitetura de *software* que facilita o fluxo de informações entre todas as atividades da empresa como fabricação, logística, finanças, e recursos humanos. É um sistema amplo de soluções e informações. Ele possui uma base de dados única, operando em uma plataforma comum que interage com um conjunto integrado de aplicações, consolidando todas as operações do negocio em um ambiente computacional.

Uma das vantagens de um sistema ERP é a garantia de que haverá apenas entrada de informações, numa única vez. Por exemplo, um representante de vendas grava um pedido de compra no sistema ERP da empresa, quando a fabrica começa a processar a ordem, o faturamento e a expedição podem checar o status da ordem de produção e estimar a data de embarque, o estoque pode checar se a ordem pode ser suprida pelo saldo e podem então notificar a produção com uma ordem que apenas complemente a quantidade de itens requisitados. Uma vez expedida, a informação vai direto a relatório de vendas para gerenciamento superior. (MACHADO, 2000)

O ERP emprega a tecnologia cliente/servidor. Isto significa que o usuário do sistema (cliente) roda uma aplicação (rotina de um módulo do sistema) que acessa as informações de um sistema de gerenciamento de uma base de dados única (servidor). Tem algumas empresas que optam por desenvolver uma solução ERP específica para o seu negócio, outras preferem sistemas de ERP prontos no mercado, onde a empresa escolhe que módulos deseja comprar. Para isso ela tem que adaptar o seu negócio ao ERP escolhido e, fazendo isso, ela acaba por padronizar o seu processo operacional, assim facilitando o seu sistema e padronizando ao que será utilizado.

A adoção de um sistema ERPs certamente garante maior agilidade na realização das transações organizacionais. Quanto à forma de adoção, seja por meio de desenvolvimento próprio, seja pela aquisição de pacotes, em cada situação existem vantagens e desvantagens, que devem ser analisadas e consideradas no processo de implantação de um sistema desta natureza.

2.2.2 Sistema Transacional ou *On-line Transactional Processing* (OLTP)

O conceito denominado OLTP (*On-line Transaction Processing*) refere-se aos sistemas de banco de dados que coletam e armazenam os dados de forma on-line. Isto é, tratam da captura dos dados das transações de negócios de um Sistema e os armazenam no banco de dados.

Por não salvar históricos dos dados, isso não o qualifica como uma base de dados ideal para ajudar nos processos de tomada de decisões. Do modo que, a base de dados necessita de um backup regularmente, pois caso o banco de dados seja perdido os dados não podem ser reaproveitados.

2.2.3 Sistema de Apoio à decisão (SAD) ou *On-line Analytical Processing* (OLAP)

OLAP é um *software* cuja tecnologia de construção permite aos analistas de negócios analisarem e visualizar dados corporativos de uma forma rápida, consistente e principalmente interativa. A funcionalidade do OLAP é inicialmente caracterizada pela análise dinâmica e multidimensional dos dados consolidados de uma organização permitindo que as atividades do usuário final.

As ferramentas OLAP são geralmente desenvolvidas para trabalhar com banco de dados desnormalizados ou seja, desnormalizar é um processo consciente com um objetivo claro. Um banco de dados bem normalizado costuma ser mais rápido que um banco de dados não normalizado e estas ferramentas são capazes de navegar pelos dados de um *Data Warehouse*, possuindo uma estrutura adequada tanto para a realização de pesquisas como para a apresentação de informações.

2.2.4 Diferenças entre Sistemas OLTP e OLAP

Segundo (HAN et al., 2011) as principais diferenças entre os sistemas transacionais e os sistemas de suporte a decisão, são:

- Um sistema OLTP é orientado ao cliente, ao processo diário de trabalho de uma organização, enquanto que um sistema OLAP é orientado ao mercado, a concorrência, a obtenção de diferencial competitivo.
- O nível operacional e o nível administrativo utilizam OLTP (vendas, compra, RH, chão de fábrica, etc.) para operações que ocorrem no dia-dia da empresa.
- O nível de conhecimento e o nível estratégico da empresa (os administradores, por exemplo) utilizam o OLAP para as tomadas de decisões e assim traçar um planejamento estratégico.

- Os dados manuseados em sistemas OLTP são atuais, representam, por exemplo, o endereço atual de certo cliente. O sistema OLAP armazena uma grande quantidade de dados históricos, o que significa contem todos os endereços de determinado cliente associado ao período em que o cliente residiu em cada local. Além disso, o sistema OLAP possibilita o fornecimento de dados abstraído o nível de detalhe de um sistema OLTP.
- A modelagem do banco de dados de um sistema OLTP é relacional, utilizando o modelo entidade-relacionamento (ER). O sistema OLAP utiliza a modelagem multidimensional, seguindo o modelo estrela (*star*) ou floco de neve (*snowflake*).
- Operações de consulta, alteração, inclusão e exclusão de registros são realizadas a todo o momento em um sistema OLTP. Já em um ambiente OLAP, o usuário final, apenas realiza consultas, embora muitas vezes, essas consultas são bem complexas.

Os dois fornecem uma base sólida para análise e apresentam tanto vantagens quanto desvantagens. Para se escolher entre os dois métodos devem-se levar em consideração os requisitos e a abrangência do aplicativo a ser desenvolvido

Juntamente ao OLAP, o OLTP é uma solução eficaz e inteligente que proporciona condições favoráveis ao gerenciamento empresarial, tornando pesquisas muito mais fáceis, ágeis e seguras de se fazer, permitindo que haja uma redução considerável de tempo na hora de se fazer consultas aos bancos de dados.

3. FERRAMENTAS OLAP

O OLAP é uma ferramenta de *Business Intelligence* utilizada para apoiar as empresas em análises de informações, visando obter novos conhecimentos que são empregados na tomada de decisão. O termo OLAP possui vários significados, pois sua tecnologia está presente em varias camadas como: armazenamento, acesso, compiladores, linguagens e conceitos.

O conceito de OLAP inclui a ideia de dimensão com hierarquia e referencias cruzadas, assim o seu banco de dados é desnormalizado, fugindo das regras convencionais de normalização, que ocasionam valores redundantes.

O termo OLAP refere-se a um conjunto de ferramentas voltadas para acesso e análise de dados, com objetivo final de transformar dados em informações capazes de dar suporte às decisões gerenciais de forma amigável e flexível ao usuário em tempo hábil. OLAP trouxe uma grande capacidade de efetuar cálculos complexos como previsões, percentuais de crescimento e médias diversa considerando-se a variável tempo.

Para leitão (2000, p.9),

O termo OLAP – *On-line Analytical Processing* – refere-se a um conjunto de tecnologias voltadas para acesso e análise *ad-hoc* de dados. Sendo assim, o objetivo final de uma ferramenta OLAP é transformar dados em informações capazes de dar suporte a decisões gerenciais de forma amigável e flexível ao usuário e em tempo hábil.

As ferramentas OLAP proporcionam condições de análise de dados on-line necessárias para responder as possíveis perguntas dos analistas, gerentes e executivos (MACHADO, 2004). Elas são aplicações que os usuários finais têm

acesso para extraírem os dados de suas bases e construir relatórios capazes de responder as suas questões gerenciais.

Dentre as características que devem estar presentes nas ferramentas OLAP, estão listadas algumas operações, descritas pela DWBRASIL OLAP (2005):

- *Drill Across*: é quando o usuário pula um nível intermediário dentro de uma mesma dimensão. Exemplo, a dimensão tempo é composta por ano, semestre, trimestre, mês e dia. O *Drill Across* é executada quando o usuário passa de ano direto para trimestre ou mês;
- *Drill Down*: ocorre quando o usuário aumenta o nível de detalhe da informação, diminuindo a granularidade (A granularidade determina quais os tipos de consultas podem ser feitas no Data Warehouse. Ela influencia diretamente na velocidade do acesso às informações e no volume de dados armazenados);
- *Drill Up*: é o contrario do *Drill Down*, ocorre quando o usuário aumenta a granularidade, diminuindo o nível de detalhamento da informação;
- *Drill Throught*: ocorre quando o usuário passa de uma informação contida em uma dimensão para outra. Exemplo: Inicia na dimensão do tempo e no próximo passo analisa a informação por região;
- *Dlice and Dice*: é uma das principais características de uma ferramenta OLAP. Como a ferramenta OLAP recupera o microcubo (No OLAP, as informações armazenadas em cubos multidimensionais, que gravam valores quantitativos e medidas, permitindo visualização através de diversos ângulos. Essas medidas são organizadas em categorias descritivas, chamadas de dimensões e formam, assim, a estrutura do cubo), surgiu a necessidade de cria um modulo, que se convencionou de *Slice and Dice*, ficando responsável por trabalhar esta informação, ela serve para modificar a posição da informação, trocando linhas por colunas de uma maneira que possa facilitar a compreensão dos usuários.

Nas seções a seguir, serão descritas as principais características da estrutura do ambiente OLAP, de como os dados são armazenados e acessados pelos usuários.

3.1.2 Estrutura Multidimensional

O modelo multidimensional é uma técnica de modelagem conceitual de negócios, que facilita a investigação, o resumo e a organização de dados para a análise de negócios, ele permite visualizar dados abstratos de forma simples e relacionar informações de diferentes setores da empresa de forma muito eficaz.

Ele relaciona tabelas de fatos com tabelas de dimensões em um banco de dados do tipo OLAP, possibilitando inúmeras combinações. Este tipo de modelagem possibilita um desempenho melhor nas consultas, possuindo regras diferentes das relacionais em sua modelagem.

A modelagem multidimensional visa somente consultas analíticas.

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 sistemas pode ser personalizado facilmente para resolvê-las com bons tempos de resposta. (ABELLÓ, p.11).

O autor apresenta os dois benefícios da modelagem multidimensional, a qual é fundamental para analisar de forma simples os dados necessários.

3.1.2.1 Fatos

Um fato é um conjunto de itens de dados contextualizados e mensurados em termos numéricos, A tabela central é chamada fato e as outras tabelas são chamadas de dimensão, que representam transações ou eventos de negocio, 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, além disso, “uma tabela fato é a tabela dentro de um modelo dimensional que contem as medidas e métricas de interesse” (IMHOFF; GALEMMO; GEIGER, 2003, P.401).

3.1.2.2 Dimensões

Uma tabela dimensão é definida como:

Uma tabela dimensão é um conjunto de tabelas de referencia 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. (IMHOFF; GALEMMO, GEIGER, 2003, P.400)

Dimensões armazenam as descrições textuais das dimensões do negocio. No modelo multidimensional uma dimensão é representada por uma tabela dimensão (Dimension Table).

3.1.2.3 Membros e Hierarquias

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

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. (THE OLAP COUNCIL 1995)

O autor esclarece a importância de estabelecer membros dentro de uma dimensão, levando em consideração a praticidade que pode ser oferecida pela declaração deste identificador.

3.1.2.4 Medidas

São medidas as características específicas e mensuráveis de um fato, relevantes para análise.

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. (IMHOF; GALEMMO; GEIGER, 2003, p.403).

Objetivo apresentar os dados altamente redundantes para se obter um melhor desempenho. Estas tabelas representam transações, acontecimentos, ou outras atividades, que são utilizadas para medir o funcionamento e os resultados dos processos de negócio.

3.1.2.5 Cubos de Dados

KAY(2005) define o conceito de cubos de dados como :

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. (KAY, 2005).

Os cubos de dados são estruturas nas quais os usuários finais fazem análise de dados multidimensionais, ou seja, navegam e exploram os mesmos,

extraindo informações e conhecimento a partir dos dados. Os cubos armazenam dados tornando-os mais fáceis de analisar.

3.1.2.6 Star Schema

O conceito de Esquema Estrela (*em inglês: Star Schema*), propõem uma visão para modelagem de base de dados para sistemas de apoio a decisão, sua característica principal é a presença dos dados altamente redundantes, melhorando o desempenho. O esquema em estrela é uma metodologia de modelagem de dados utilizada do desenho de um *Data Warehouse*.

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. (THOMSEN, 2002, p.638)

Portanto é bom utilizar um modelo estrela, pois fornece um acesso, mas rápido aos dados e mais fácil de navegar, criando tabelas auxiliares para dimensões.

3.2.1 Arquiteturas OLAP

Apesar de obedecer a uma estrutura cliente/servidor multiusuário, as ferramentas OLAP podem ser implementadas de diversas formas, classificadas em cinco tipos a seguir (INMON, 1995).

3.2.1.1 Rolap (Relational On line Processing)

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 banco 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 banco de dados.

3.2.1.2 Molap (Multidimensional On line Analytical Processing)

No MOLAP dos dados ficam armazenados em um banco de dados multidimensional, onde o servidor MOLAP atua e o usuário trabalha, monta e manipula os dados diferentes no servidor (CARVALHO, 2004). Os dados de um banco multidimensional são armazenados em um espaço menor que o utilizado para armazenar os mesmos dados em um banco de dados relacional.

3.2.1.3 Holap (Hybrid On line Analytical Processing)

O HOLAP, ou processamento híbrida, está se tornando a mais popular para os produtos atuais, porque consegue combinar a capacidade e a estabilidade das ferramentas ROLAP com o desempenho superior dos bancos de dados multidimensionais (THOMSEN, 2002).

3.2.1.4 DOLAP (Desktop On line Analytical Processing)

A DOLAP é a arquitetura *desktop* do OLAP, ou seja, é uma ferramenta para usuários que possuam uma cópia da base dimensional ou de um subconjunto dela ou ainda que queiram acessar um repositório de dados central localmente. O usuário ao acessar este repositório, dispara uma instrução SQL e acessa os cubos já existentes no banco de dados multidimensional residente no servidor OLAP e obtém de volta um para ser analisado em sua estação de trabalho. (THOMSEN, 2002)

3.2.1.5 WOLAP (Web On line Analytical Processing)

O WOLAP é a utilização de uma ferramenta OLAP a partir de um *browser*. Esta arquitetura tem duas tecnologias que estão em constante evolução, a primeira é a *Web* e a segunda são as ferramentas OLAP. A diferença desta ferramenta para as outras é que ela utiliza a *Web*, facilitando assim, a

distribuição da ferramenta, o acesso remoto dos dados a serem analisados. (INMON, 1995).

Portanto, a ferramenta OLAP traz ganhos significativos de produtividade, devido ao auxílio na tomada de decisões, disponibilidade de informações estratégicas. Tudo isto de forma instantânea e consistente, com respostas rápidas às consultas e perguntas de gerentes e analistas.

Esta ferramenta traz benefícios é que a torna tão poderosa num ambiente empresarial. A organização dos dados sob uma estrutura de cubos permite a análise multidimensional das informações e fornece a informação precisa trazendo grandes vantagens ao empreendimento.

4- DATA WAREHOUSE

Os extensos volumes de dados gerados pelas organizações representam na maioria das vezes, problemas de acessos e de redundâncias, dentre outros. Nesse sentido, algumas técnicas e conceitos foram elaborados com objetivos de tentar facilitar a extração de informações nas grandes bases de dados existentes nas organizações. Dentre as técnicas e conceitos mais conhecidos estão os *data warehouses*, que têm por objetivos disponibilizar os dados em uma modelagem de fácil entendimento para os usuários (GONÇALVES, 2003, p 3).

Segundo Corey (2001, p 9), um Data Warehouse é,

Um Banco de Dados reunido a partir de muitos sistemas destinados a suportar a produção de relatórios gerenciais e a tomada de decisão. De uma forma geral, quando as pessoas falam a respeito de armazenamento de dados, na verdade elas estão se referindo aos armazenamentos de dados e às ferramentas que, juntos, formam os modernos ambientes de produção de relatórios.

Ele é considerado um grande *container* de dados divididos por assunto, detalhamento, não voláteis e históricos.

“*Data Warehouse* é uma coleção de dados orientada por assuntos, integrada, variante no tempo e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão”.
(INMON, 1997)

Apresentamos a seguir as principais características da tecnologia DW que são: orientados pelo assunto, integração, não volátil, variável do tempo.

4.1 Orientado pelo assunto

Um *Data Warehouse* armazena as informações agrupando- as por assunto de interesse da empresa, ao contrario de um sistema operacional onde as informações estão organizadas de acordo com as aplicações. Esses assuntos são os processos principais de uma empresa (MACHADO, 2004, P. 28-29).

A figura 2 ilustra o procedimento da propriedade orientada ao assunto

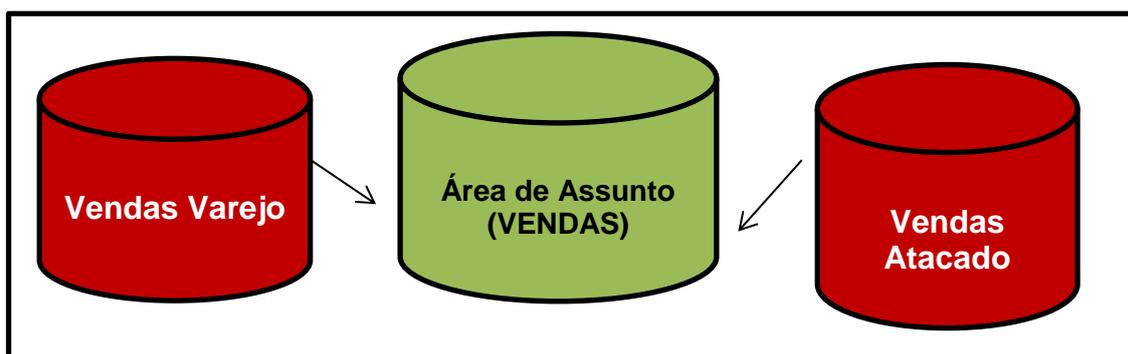


Figura 2 - Exemplo de unificação de informações de vendas via propriedade orientada ao assunto, retirada de COREY; ABBEY; ABRAMSON; TAUB, 2001, p.11.

4.2 Integração

É uma das características mais importantes de um *Data Warehouse* e se refere à questão da unificação de tipos de dados, valores de atributos e padrões de informações descritivas, que podem estar diferentes nos diversos sistemas que compõem uma empresa (MACHADO, 2004, P. 31). A figura 3 abaixo ilustra a integração.

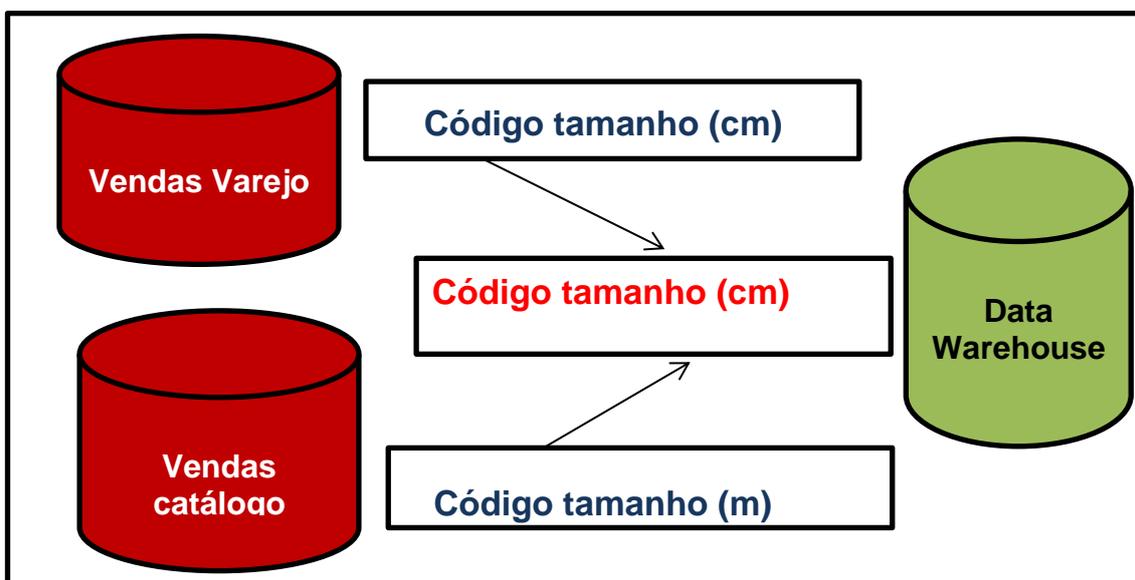


Figura 3 - Exemplo de tratamento de dados, inspirado em GONÇALVES (2003, p.14).

4.3 Não Volátil O fato de um *Data Warehouse* ser “não volátil” significa que os dados uma vez carregados nunca mais são alterados pelos usuários, apenas consultados, a não ser em novas cargas de dados. Na verdade, podem-se realizar somente inclusões (cargas) e consultas em um *Data Warehouse*, a opção de modificação (alteração) esta descartada.

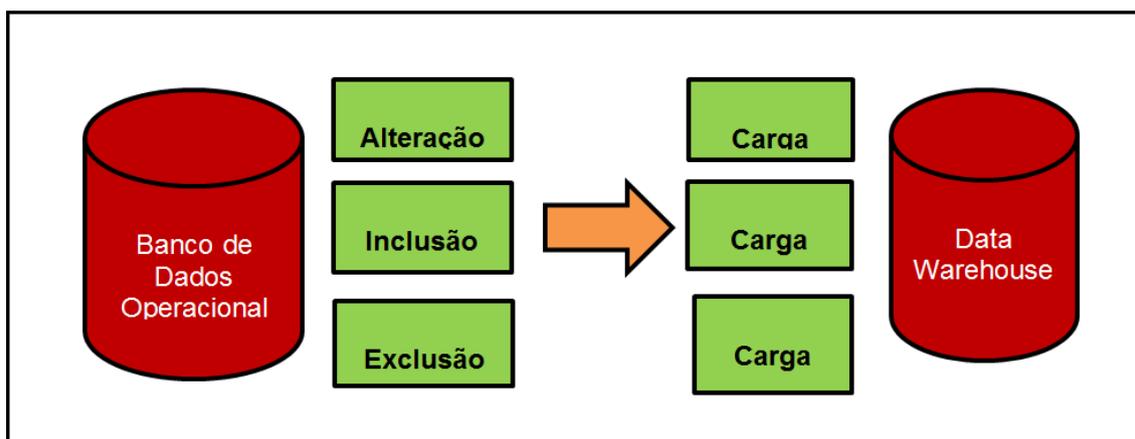


Figura 4 - Formas de Atualização, inspirada em GONÇALVES (2003, p.15)

4.4 Variável no tempo:

Conforme, INMON (1997 *apud* GONÇALVES, 2003, p.15) um ambiente de *Data Warehouse* é variável com o tempo devido ao fato de que os dados contidos no mesmo referem-se a algum momento específico, o que significa dizer que o dado quando carregado recebe, na sua chave, uma unidade de tempo e nunca mais é atualizado. Isso permite que os analistas de negócios façam análise de tendências.

Um exemplo bem simples sobre o conceito de variável no tempo, pegue uma foto sua quando era recém nascido, depois, pegue outra quando você tinha cinco anos, e compare. Com certeza, muitas modificações ocorreram durante o tempo, porém as fotos retratam exatamente a sua situação naquele exato momento no tempo, e isso acontece no ambiente de *Data Warehouse*.

4.5 MODELAGEM DE DADOS NO DATA WAREHOUSE

Segundo MACHADO (2004), para criar o modelo de dados para um *Data Warehouse* não basta copiar o modelo de dados transacional para um banco de dados separado. Isso não funcionara, principalmente, por que:

- As consultas se tornarão muito complexas aos usuários (infinidade de detalhes e particularidades da área operacional que não interessam na tomada de decisão).
- Como os bancos de dados são definidos respeitando os conceitos de normalização, para realização de consultas será necessário acessar uma grande quantidade de tabelas e conseqüentemente, haverá uma grande quantidade de junções (*joins*).

Na modelagem de um DW se abre mão dos conceitos de normalização e se utiliza a modelagem dimensional ou multidimensional. O modelo multidimensional é composto por três elementos (MACHADO, 2004 p. 79):

- **Fatos:** representam um evento de negocio da empresa.
- **Dimensões:** representam elementos que compõem um fato.
- **Medidas:** são os atributos numéricos de um fato.

Existem alguns modelos utilizados geralmente nos processos de modelagem de um DW. As próximas seções detalham estes modelos.

4.5.1 Modelo Estrela (*Star Schema*)

O modelo Estrela é assim denominado por apresentar a tabela de fatos no centro do esquema e as tabelas de dimensões nas extremidades (GONÇALVES, 2003, p.112). As figuras 5 e 6 ilustra genericamente o modelo estrela.

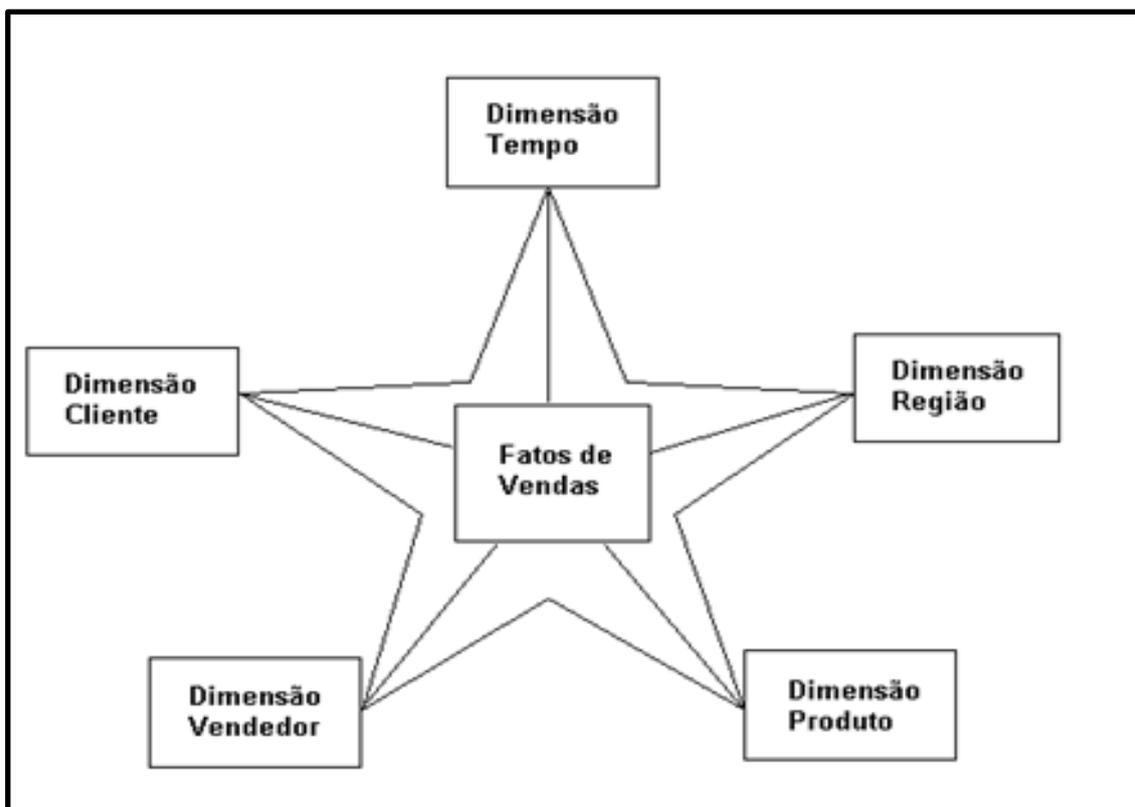


Figura 5 - Modelo Multidimensional (Star Schema), MACHADO (2008, p.93)

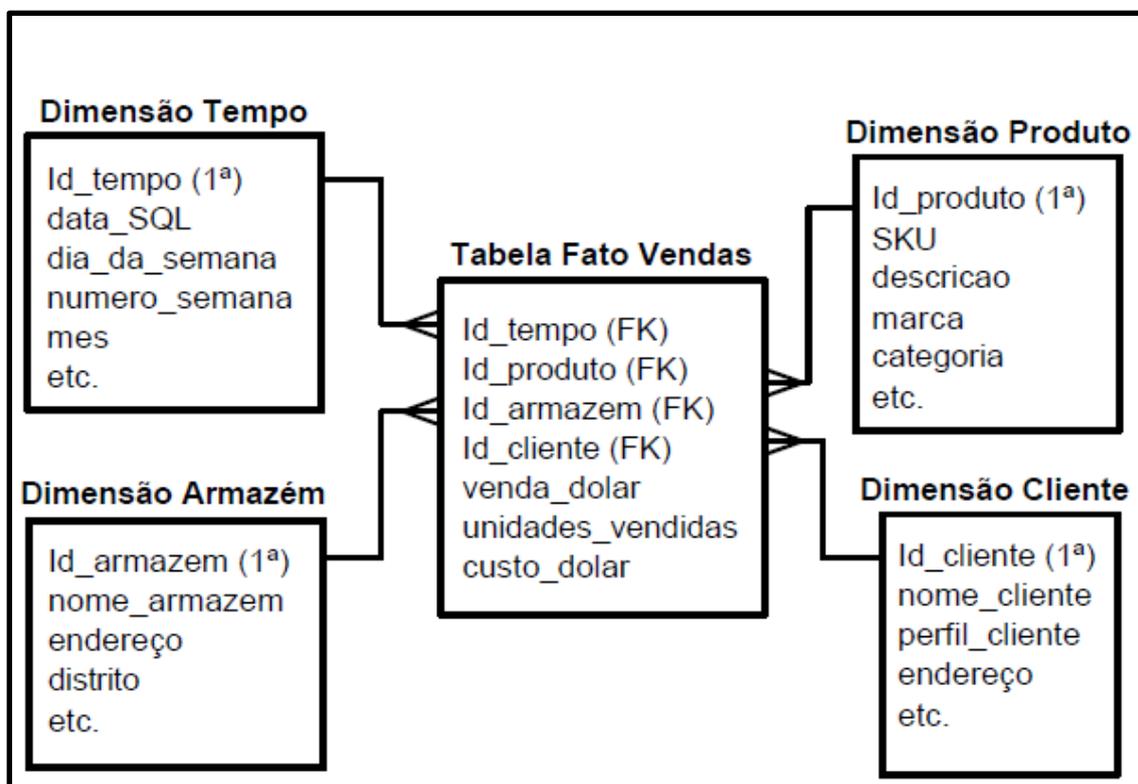


Figura 6 - Exemplo genérico do modelo Estrela (Star Schema)

4.5.2 Modelo Floco de Neve (*Snow Flake*)

O modelo Floco de Neve (*Snow Flake*) é uma extensão do modelo estrela (*Star Schema*) em que cada uma das pontas da estrela (dimensões) passa a ser o centro das outras estrelas (GONÇALVES, 2003, p.113).

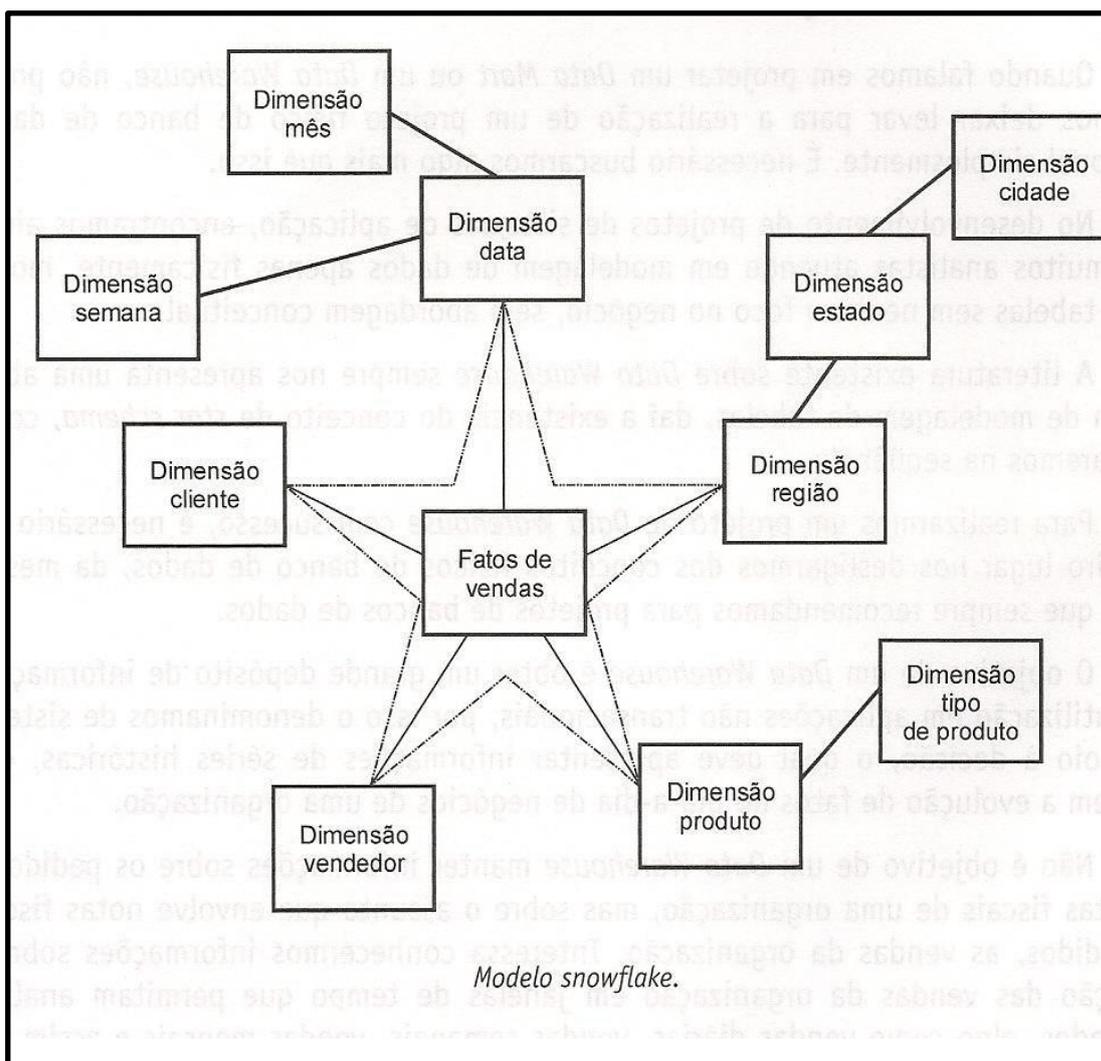


Figura 7 - Modelo Multidimensional Snow Flake, MACHADO (2008, p. 95).

4.6 Arquitetura do Data Warehouse

Segundo Machado (2008), a arquitetura define o modelo lógico da instalação, independente da sua estrutura física do Data Warehouse. A escolha da arquitetura passa por uma decisão gerencial que está relacionada a fatores relativos à infraestrutura disponível pelo formato da instalação, se central ou distribuída em instalações remotas ou locais. Para Willian H. Inmon (1997) Data Warehouse é uma coleção de dados orientados por assunto, integrado, variável com o tempo e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisões.

Podemos dizer também dizer, em outras palavras, que um Data Warehouse é um banco de dados contendo dados extraídos do ambiente de produção da empresa, que foram selecionados e depurados, tendo sido otimizados para processamentos de consultas e não para processamento de transações. Em geral, Data Warehouse necessita da consolidação de outros recursos de dados, além dos armazenados em bancos de dados relacionais, incluindo informações contidas em planilhas eletrônicas, documentos de texto, etc.-.

4.6.1 Granularidade

A granularidade pode ser considerada o nível de detalhe ou de resumo de dados que pode ser encontrada no ambiente de *Data Warehouse*. Ela é um componente interessante, mas ao mesmo tempo pode afetar os volumes de dados e conseqüentemente os tipos de consultas que poderão, ou não, serem atendidas, dependendo do nível de granularidade se encontra. Um exemplo, quanto menor o nível de granularidade maior será o nível de detalhamento (detalhes), podendo assim atender a qualquer consulta, mas necessitando de mais espaço para armazenar os dados por completo, por outro lado, quanto maior o nível de granularidade menor será o nível do detalhamento (menos detalhes) podendo prejudicar alguns tipos de consultas.

A escolha dos níveis de granularidade apropriados para o ambiente projetado é vital para o sucesso. O método normal para a escolha dos níveis de granularidade consiste em usar o bom senso, criar uma pequena parte do warehouse e deixar o usuário acessar os dados. A seguir, ouvir atentamente o usuário, colher o feedback que ele proporcionar, e ajustar, adequadamente, os níveis de granularidade. (INMON, 1997, p.156).

A mais importante questão que o desenvolvedor deve ter em um projeto DW diz respeito à definição da granularidade do *Data Warehouse*. Quando a granularidade é propriamente estabelecida, os demais aspectos do projeto e implementação fluem tranquilamente.

4.6.2 Data Mining

De acordo com BARBIERI (2001) o conceito de *Data Mining* está relacionado a busca de correlações escondidas em altos volumes de dados, que nem sempre evidentes no tratamento cotidiano de sistemas de informações.

Para BRAIN (2010, p.291):

O *software* de *data mining* analisa grandes conjuntos de dados históricos das empresas que foram preparados para análise em *data warehouses*. O *data mining* procura descobrir padrões, tendências e correlações ocultas nos dados, o que pode propiciar uma vantagem competitiva estratégica a uma empresa.

PRIMAK (2008) explica que o *Data Mining* é uma técnica que combina estatística convencional com inteligência artificial que possibilita a transformação de dados em informação de forma automatizada num pequeno espaço de tempo.

4.6.3 Data Mart

Um *Data Mart* é uma divisão lógica de um *Data Warehouse*, que atende uma área específica da organização. Os *Data Mart's* pertencem às áreas específicas da empresa contabilidade, finanças, vendas e etc. E são modelados de acordo com as especificações de cada departamento, possuindo assim, *design* único (INMON; TEDERMAN; IMHOFF, 2001).

O objetivo do *Data Mart* é aumentar a velocidade na consulta de informações, pois possui um escopo menor de informações e, como é segmentado determinada áreas, possibilita a análise multidimensional.

Segundo Barbieri (2001) o termo *Data Mart* (Mercado de Dados) significa depósito de dados que atende a certas áreas específicas da empresa voltadas para o processo decisório gerencial. *Data Mart* é um subconjunto de um *Data Warehouse* e é utilizado para um determinado assunto como por exemplo ao setor de Vendas, Estoque e Financeiro entre outros. A figura abaixo demonstra os vários *Data Marts* sendo ligados em uma única base de dados. O *Data Warehouse*.

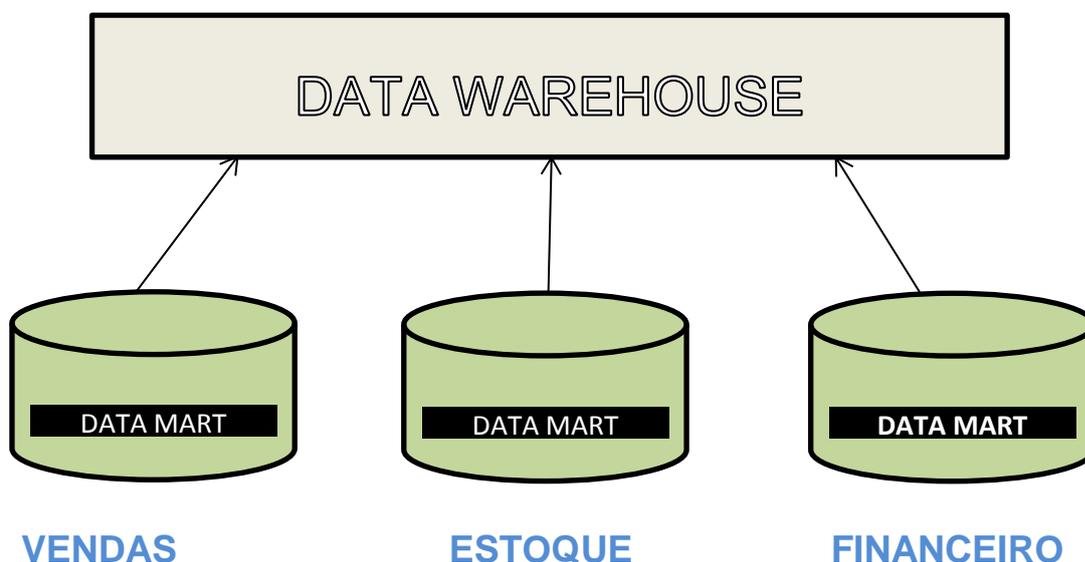


Figura 8 - Conceito de Data Mart, fonte Barbieri, 2001, adaptada pelo autor.

Portanto conclui-se que estas novas tecnologias como o Data Warehouse, permitem aos administradores descobrirem novas formas de diferenciar suas empresa numa economia globalizada, deixando-os mais seguros para definirem as metas e adotarem diferentes estratégias em sua organização, conseguindo assim visualizarem, antes de seus concorrentes, novos mercados e oportunidades, atuando de maneiras diferentes conforme o perfil de seus consumidores.

5 - FERRAMENTAS ETL

O processo de ETL (*Extract, Transform and Load*) destina-se á extração, transformação e carga de dados de uma ou mais bases de dados de origem para uma ou mais bases de destino (*Data Warehouse*). ETL é o processo mais demorado e critico na construção de um DW. A maior parte do esforço exigido no desenvolvimento de um DW é consumida neste momento e não é incomum, que oitenta por cento de todo os esforço seja empregado no processo de ETL (INMON, 1997).

Esta ferramentas devem fazer parte de um projeto de DW, assim como um gerenciamento central dará apoio nas tarefas de movimentação de dados e, preparando uma boa arquitetura do ETL, vai possibilitar uma melhor implementação do projeto.

Estas etapas de levantamento das necessidades de informações, requisitos do projeto e definição da modelagem dos dados a serem apresentados no *Data Warehouse*, o passo seguinte é identificar a origem dos dados, local onde são processados nos sistemas transacionais da organização.

5.1-Etapas do processo ETL

O processo de ETL se resume basicamente em três principais grandes partes, que são:

- **Extração:** onde serão definidas as origens de dados e realizada a extração deles. Estas origens tanto podem vir de bases de dados distintas como possuir formatos diferentes desde planilhas Excel á arquivos de textos.
- **Limpeza:** nesta fase é realizada a limpeza das informações, onde são retiradas inconsistências que possam existir.

- **Transformação:** neste processo é realizada a transformação de dados. Caso os dados sejam de diferentes sistemas, em alguns casos faz-se necessário padronizar os mesmos. Um exemplo que pode ser citado é quando numa base de dados o tipo do sexo do cliente está como “H” para masculino e em outras bases foi utilizado o “M” para masculino.

5.2 Ferramentas KETTLE (PENTAHO)

De acordo com o Pentaho (2011), o *Pentaho Data Integration* – PDI, também conhecido como o *Kettle*, é a ferramenta de ETL da Pentaho. Esse produto foi desenvolvido em Java e é totalmente baseado em metadados (camada física, camada de negócios e camada de visualização). No *Kettle*, o conceito de processo ETL clássico (extração, transformação e carregamento) foi modificado, porque é composto por quatro elementos ETTL, que significam:

- Extração de dados das bases de dados de origem;
- Transporte de dados;
- Transformação de dados;
- Carregando (*Loading*) dados em um *Data Warehouse*.

Assim, *Kettle* é um conjunto de ferramentas e aplicativos que permitem manipulações de dados através de múltiplas fontes. Abaixo são descritos os principais componentes do *Pentaho Data Integration* – PDI (PENTAHO, 2011):

- *Spoon* – Ferramenta gráfica utilizada para modelar o fluxo de dados de um processo de transformações ETTL. Desempenha as funções de fluxo de dados em uma variedade de diferentes fontes de dados e destinos, as transformações projetadas em *Spoon* podem ser executadas com *Pan* e *Kitchen*.
- *Pan* – é uma aplicação dedicada para executar transformações de dados projetados em *Spoon*.

- *Chef*- uma ferramenta para automatizar tarefas de atualização do banco de dados de uma forma complexa.
- *Kitchen*- é um aplicativo que ajuda a executar as tarefas em modo de lote, geralmente usando uma programação que torna fácil para iniciar e controlar o processamento ETL.
- *Carte*- um servidor web que permite o monitoramento remoto dos processos ETL através de um navegador web.

Embora as ferramentas de ETL sejam frequentemente usadas em ambientes de *Data Warehouse*, o Kettle também pode ser aplicado para outros fins:

- Migrar dados entre aplicações ou bases de dados;
- Exportar ou importar dados entre banco de dados e arquivos simples;
- Limpar e analisar dados;
- Integrar aplicações.

Segundo a Pentaho (2011), o PDI/Kettle é fácil de usar e todo processo é criado com uma ferramenta gráfica onde você especifica o que fazer sem escrever código para indicar como fazê-lo; por isso pode-se dizer que o PDI/Kettle é orientado por metadados.

O aperfeiçoamento do conhecimento é o um fator primordial para a utilização da ferramenta, tendo em vista tratar-se de uma ferramenta complexa, a fim de evitar possíveis erros de utilização indevida.

O capítulo a seguir tem como objetivo dissertar sobre a ferramenta Pentaho, apresentando suas características, relevância e estrutura.

6- PLATAFORMA PENTAHO

Em função da escassez de materiais sobre o assunto, o estudo pautou-se nas guias e manuais eletrônicos disponibilizados juntamente como o próprio Pentaho.

A plataforma Pentaho, é composta pelas seguintes ferramentas PAC, PUC, PME, PRD, PSW. A plataforma baseia-se na construção de soluções para BI e trata-se de uma suíte de soluções de Inteligência de Negócios, que disponibiliza benefícios como: relatórios, tabelas dashboarding, processamento OLAP (BOUMAN; DONGEN, 2009), dentre outros.

Desenvolvido utilizando a tecnologia Java o software oferece uma arquitetura voltada para WEB e executa nos principais servidores como: Tomcat e Jboss O sistema Pentaho dispõe de recursos de apoio ao usuário, como fóruns oficiais, para que sejam sanadas as dúvidas dos usuários e profissionais, cursos para desenvolvimento de BI utilizando a ferramenta, artigos acadêmicos, além de sua facilidade de uso, possuindo um ambiente intuitivo (FERREIRA, 2008).). Possui uma estrutura integrada para as soluções BI, trata-se de um sistema unificado proporcionando redução na sobrecarga de desenvolvimento de soluções para inteligência empresarial.

A solução Pentaho define-se a si mesma como uma plataforma de BI orientada para a solução e centrada em processos. Ou seja, não só apresenta os resultados de uma forma única e dando uma visão geral do estado da empresa, como implementa os próprios processos (workflow) para a resolução de problemas. A figura apresenta a arquitetura do Pentaho composta pelos módulos de apresentação, de infraestrutura, de integração de dados e origem dos dados.

A figura a seguir apresenta a arquitetura do Pentaho BI Suite.

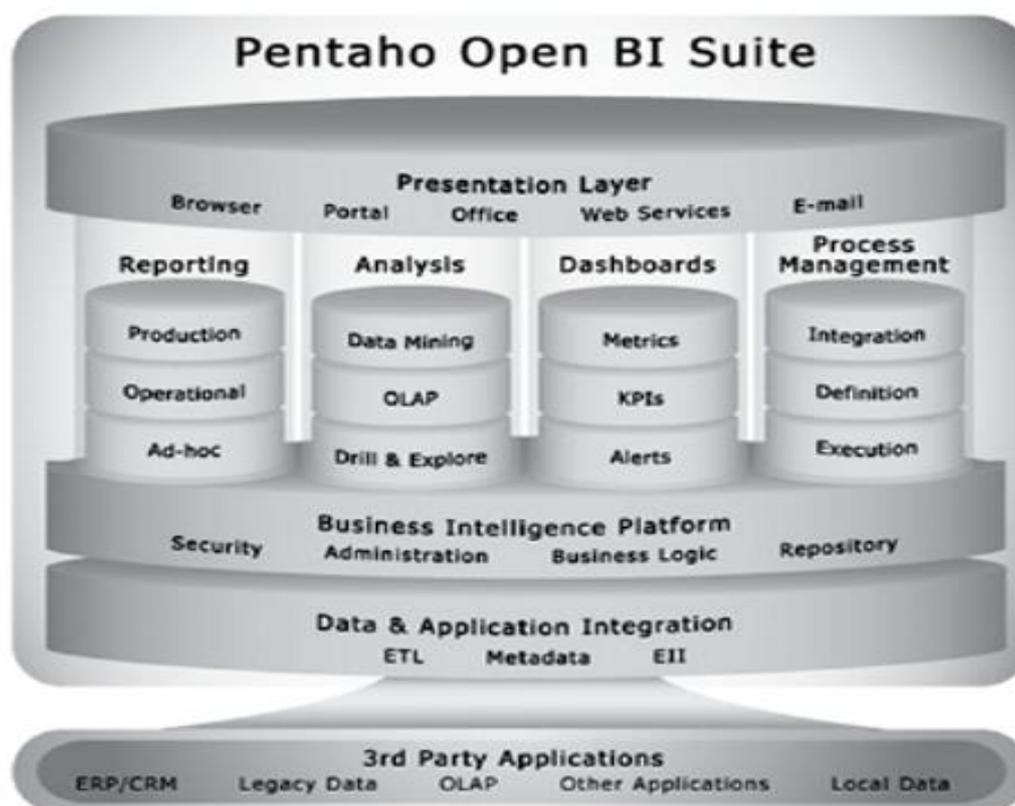


Figura 2: os componentes da Suite

Figura 9 - Arquitetura do Pentaho, fonte: BOUMAN; DOUGEN (2009)

6.1 Módulos do Pentaho

- **Pentaho Data Integration - PDI (Kettle):** utiliza as técnicas de ETL, para obtenção dos dados. Ela é capaz de se conectar com vários banco de dados como *Oracle*, *PostgreSQL*, *SQLServer*, *MySql*, entre outros. Responsável em popular o DW, migração da base de dados e integração entre as aplicações;
- **Pentaho Analysis View (Mondrian):** conhecida como *Mondrian*, permite operações *slice-and-dice*, *drill-down*, entre outras, além da exploração gráfica de cubos OLAP.
- **Pentaho Reporting (Reporting):** esta ferramenta é capaz de gerar relatórios de alta qualidade com grande simplicidade;

- **Pentaho Dashboards:** permite a criação de painéis de controle, sendo possível reunir numa mesma tela os principais indicadores de um departamento ou de toda a empresa;
- **Weka (Data Mining):** permite a criação de testes de hipóteses na área de dados, utilizando a mineração de dados.

6.2.2 Ferramentas do Pentaho

A plataforma Pentaho é constituída por uma série de ferramentas que viabilizam o desenvolvimento, a implementação e uso da ferramenta pelas organizações. São as seguintes ferramentas PAC, PUC, PME, PRD, PSW, descritas nas seções seguintes.

6.2.1 Pentaho Administration Console (PAC):

A PAC foi utilizada para cadastrar a conexão. A figura 10 apresenta a tela do PAC na funcionalidade de cadastro de conexões, onde há duas cadastradas: a base exemplo do Pentaho, cadastrada sob o nome SampleData, e a base DW, criada neste trabalho como DWPentaho. O PAC é onde foi feita a configuração com o login e senha do usuário, e as conexões com o banco de dados que utilizaremos para a aplicação.

6.2.2 Pentaho User Console (PUC)

Interface para o usuário final, é a partir do PUC que os usuários terão acesso às análises e relatórios criados, poderão ainda – de acordo com as permissões definidas – criar relatórios Ad-Hoc utilizando o WAQR, criar novas visões analíticas utilizando o JPivot, executar relatórios criados previamente com o PRD (ou outro gerador de relatórios suportado), visualizar dashboards, agendar a execução de relatórios e compartilhar qualquer um desses artefatos com outros usuários.

O PUC é onde vai gerar o relatório do cubo de dados, esta ferramenta é para a interface do usuário, dentro dela contem o relatório que sera gerado através da

ferramenta Pentaho Schema Workbench, onde implementaremos e criaremos o cubo de dados.

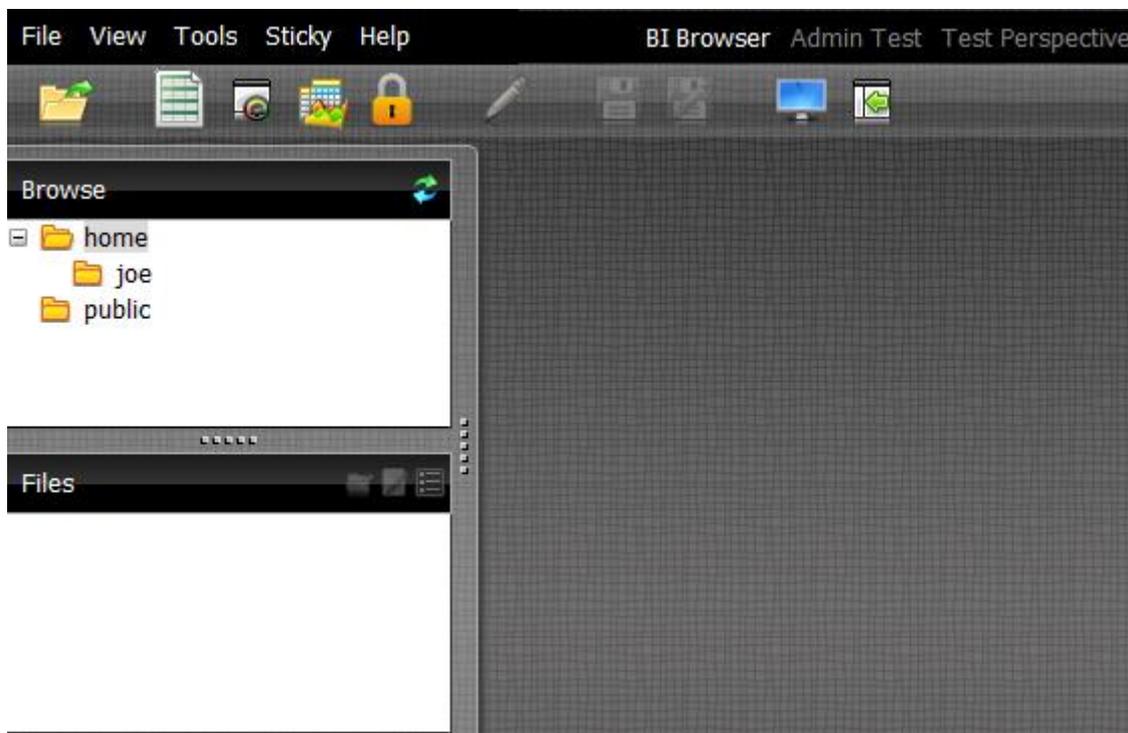


Figura 10 - Imagem Pentaho User Console (PUC), fonte autoria própria.

6.2.3 Pentaho Metada Editor (PME)

Através da PUC (ferramenta do Pentaho para usuários finais), possam gerar seus próprios relatórios, é necessário criar metadados através do PME.

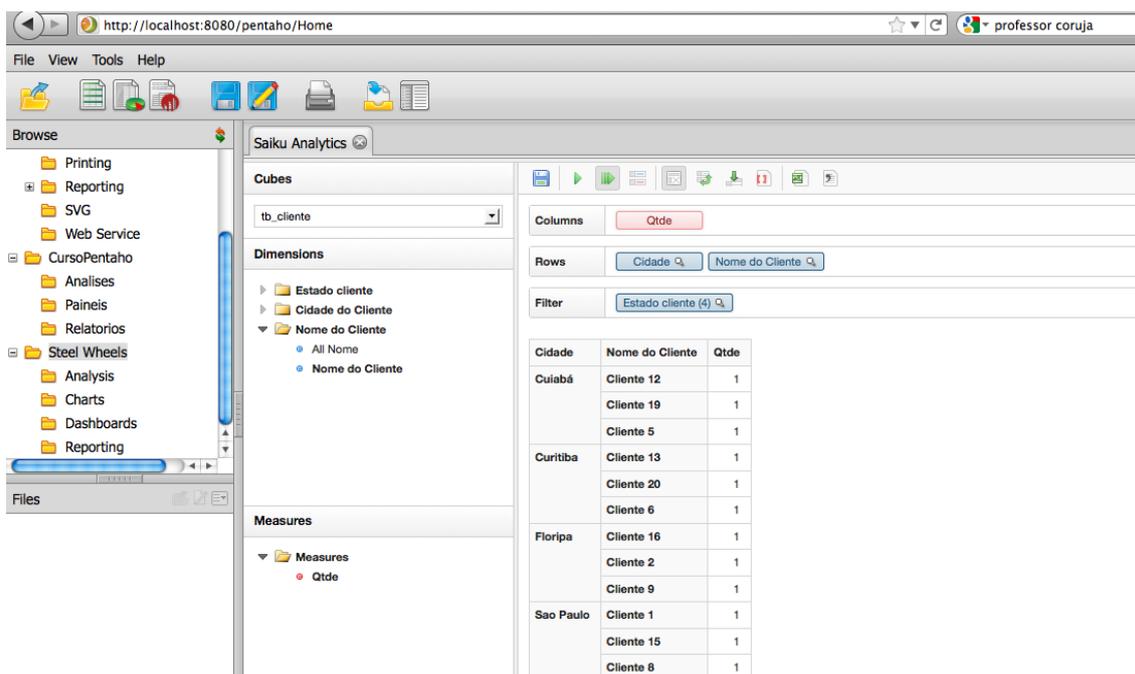


Figura 11 - Imagem Pentaho Metadata Editor (PME), fonte blog Prof° Coruja.

6.2.4 Pentaho Report Design (PRD)

Com essa ferramenta são desenvolvidos relatórios que ficam acessíveis aos usuários na PUC. Os relatórios são construídos com base em consultas SQL, por isso podem expressar um nível maior de complexidade do que os relatórios diretamente na PUC.

Essa ferramenta apresenta um facilitador (*wizard*), onde no primeiro momento escolhe-se o visual do relatório (*template*), depois se seleciona a base de dados e se consulta SQL, depois são escolhidas as informações agrupadas no relatório, sobre as quais serão apresentados os subtotais e total, e as informações detalhe. Os tipos de dados, formatação e alinhamento das informações do relatório são definidos na etapa seguinte, por fim o esboço do relatório é apresentado. Esta ferramenta, não sera feita a implementação dentro do trabalho, só uma breve introdução.



Figura 12 - Imagem Pentaho Report Designer (PRD), fonte autoria própria.

6.2.5 Pentaho Schema Workbench (PSW)

Esta ferramenta possibilita a criação de cubos OLAP. Para o desenvolvimento da solução tratada neste trabalho, foram criados três *schemas*, cada um contendo um cubo. Cada cubo representado uma tabela fato: Compras, Vendas e Orçamento. Cada um desses *schemas* resulta em um arquivo XML contendo as definições do cubo, relacionamentos e medidas. É com esta ferramenta que será feita a implementação dentro do sistema Pentaho, nela conectaremos ao banco e começamos a criar os cubos ligando com o banco de dados SQL, que utilizaremos para este trabalho.

A imagem da ferramenta com a estrutura dos cubos criados é ilustrada na figura abaixo.

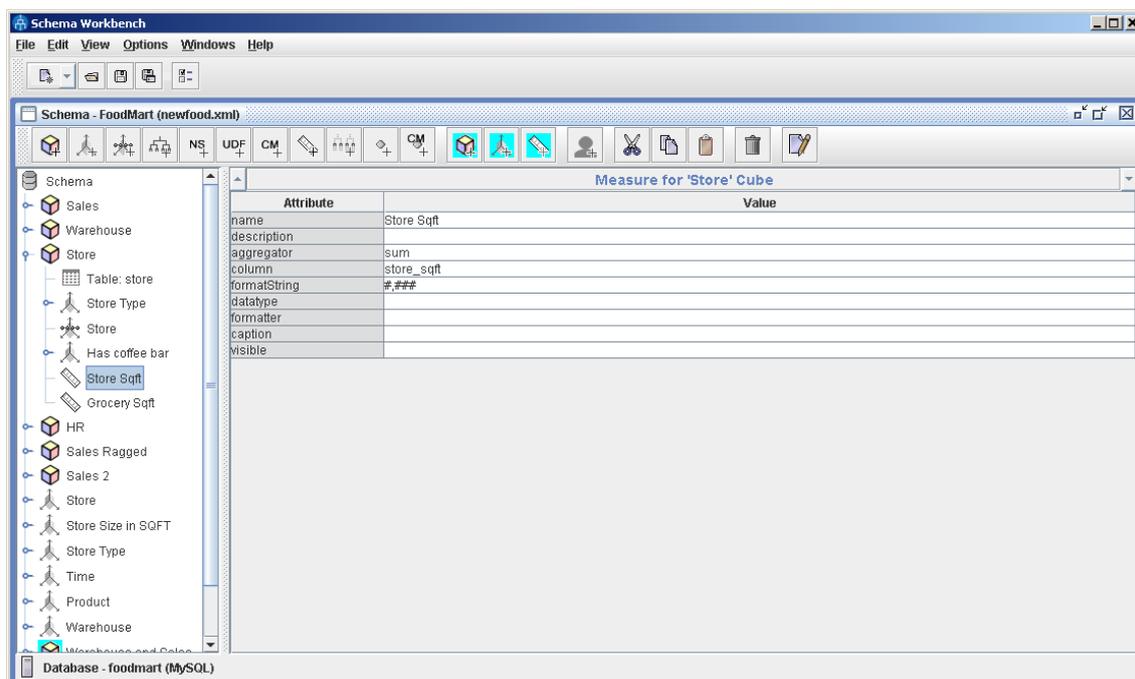


Figura 13 - Imagem Pentaho Schema Workbench (PSW), fonte autoria própria.

6.3 Instalação e Configurações

O processo de instalação do Pentaho será abordado neste capítulo com detalhes importantes para colocar no ambiente de produção ou mesmo para seus testes iniciais. A instalação da plataforma BI Pentaho é bem amigável e intuitiva para o usuário.

Configurações do Java

Faça o Download de

<http://www.oracle.com/technetwork/java/javase/downloads/index.html> do JDK e siga os procedimentos do instalador.

Configurando as variáveis de ambiente Java.

O Java utiliza das seguintes variáveis de ambiente para execução do Pentaho:

JAVA_HOME

JRE_HOME

PATH

CLASSPATH

- Entre no painel de controle do computador.
- Depois de um duplo clique em sistema você e encontrara nos menus a aba de configurações avançadas selecione;
- Posteriormente clique em “variáveis de ambiente” e que lhe permitira configurara as varieaveis a seguir;
- Clique em novo em variáveis de usuários.
- Nome da variável: JAVA_HOME
Valor da variável: C:\Program Files\Java\jdk1.6.0_20 (esse local é o local onde esta instalado o seu Java)
- A variável PATH deve ser complementada conforme abaixo, mantenha o que está configurado na mesma.
Nome da variável: PATH
Valor da varievel: XXXXXXXX;C:\Programas\Java\jdk1.6.0_22\bin (\bin é o diretório de binários do java);
- A variável CLASSPATH já vem correta após a instalação do JDK.

Instalando o Pentaho BI Server

- 1- Acesso site para download
<http://sourceforge.net/projects/pentaho/files/>
- 2- Baixe o arquivo do Pentaho BI Server (biserver-ce-3.7.0 –stable.zip)
- 3- Descompacte o arquivo.
- 4- Serão descompactados dois diretórios (admistration-console biserver-ce)
- 5- Criar um diretório para o Pentaho no seu HD.

- 6- Mova as pastas descompactadas(administration-console biserver-ce) para este diretório.
- 7- Iniciando o Pentaho BI Server, rode o script start-pentaho.bat
- 8- Na primeira vez é iniciada uma mensagem de aviso sobre o verificador de versão é emitido conforme a imagem abaixo, de um ok para continuar.

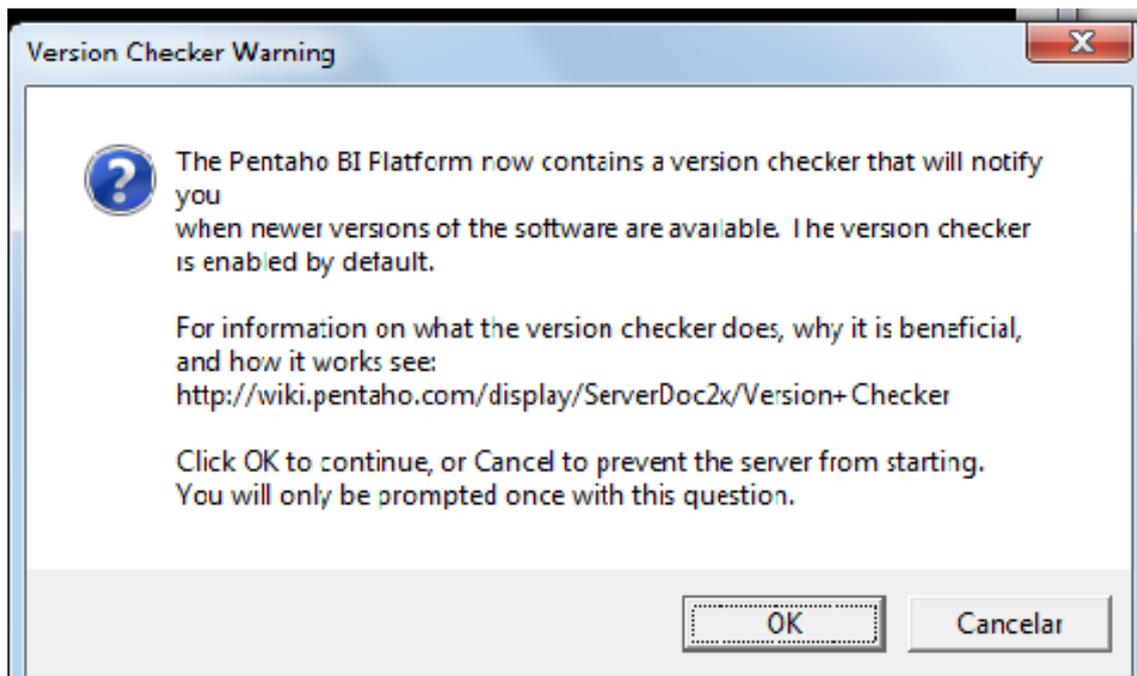


Figura 14 - Figura 14 - Mensagem aviso verificador de versão, fonte autoria própria.

Para testar acesso <http://localhost:8080/pentaho>

Inicie o Pentaho Administrator console acesse <http://localhost:8099>

O usuário padrão é **admin** e a senha é **password**, conforme a figura abaixo.

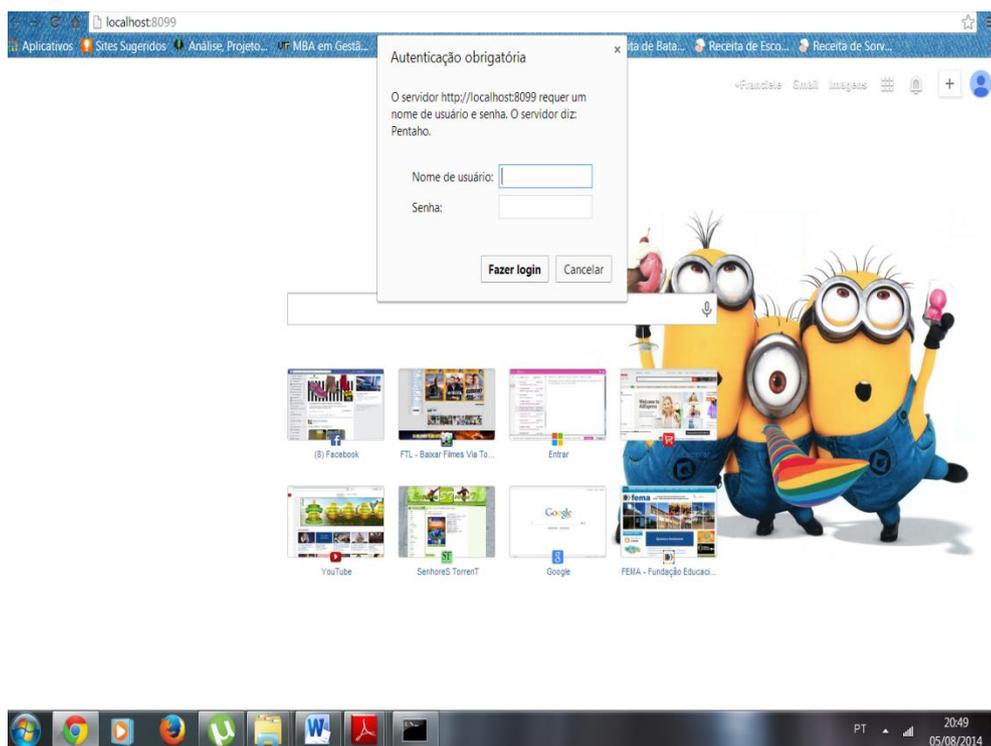


Figura 15 - Tela do login do Pentaho., fonte autoria própria.

Tereso (2011) relata em uma pesquisa que empresas de pequeno e médio porte estão buscando inovações para auxiliar gestores na tomada de melhores decisões para se destacarem no mercado, com isso empresas optam por ferramentas livres e dentre as ferramentas mais escolhidas no mercado destaca-se o Pentaho. E visando todas essas informações concluímos que esta ferramenta dispõe recursos acessíveis para todos os usuários terem acesso a essa ferramenta desde de sua instalação ate sua implementação, pois este aplicativo tem a disponibilidade de diversas ferramentas que permite a empresa e ao usuário decidir qual aplicativo irá utilizar.

7. CONCLUSÃO

O intuito deste estudo foi apresentar recursos, para os sistemas de apoio a decisão, e as tecnologias de *Business Intelligence*, que possibilitam a transformação das informações em armas potentes a serem usadas pelas organizações.

Existem no mercado inúmeras ferramentas de BI, mas o que se tem encontrado são ferramentas que necessitam inúmeras instalações e configurações de ambiente para que estejam aptas para serem usadas, o que acaba desestimulando a continuidade do uso pelas empresas do processo de BI. Isto com certeza é um motivo para que muitas empresas não utilizem os recursos disponibilizados pelos ambientes de BI. Apesar do Pentaho ser uma ferramenta gratuita é necessário para sua instalação um profissional que tenha conhecimentos avançados, pois sua instalação é complexa.

Na criação deste estudo houve um erro na visualização do cubo na ferramenta Pentaho e isto gerou um desgaste, pois foram realizadas inúmeras procuras para tentar solucionar o problema sem solução. Outras dificuldades encontradas durante o desenvolvimento do estudo foi na conexão e nos procedimentos com o banco de dados. Porém as maiores dificuldades foram com os processos de instalação, configuração e utilização das ferramentas do Pentaho.

A falta de experiência com os sistemas OLAP, somada a gama de opções que essas ferramentas apresentam aos desenvolvedores e usuários, acabaram por dificultar na aprendizagem intuitiva dessas ferramentas. É, contudo, importante destacar facilidade de visualização de indicadores que são muito importantes nos processos decisórios e, incentivando, desta forma, a utilização de uma ferramenta de BI na empresa.

Entretanto, o estudo não esgota o assunto, mas deixa para estudos futuros um caminho aberto, uma proposta lançada para novas implementações e análises da ferramenta Pentaho em relação a sua implementação e uso tanto em meio acadêmico como organizacional.

REFERENCIAS BIBLIOGRÁFICAS

BLOG PROFESSOR CORUJA. Disponível em:
<<http://blog.professorcoruja.com>>. Acesso em: 17/05/2014.

CARVALHO, B.F. **Arquiteturas de Ferramentas OLAP**. SQL Magazine, Rio de Janeiro, ano 1, ed. 9, p.12-16, 2004.

CAVALCANTI, M. C.; OLIVEIRA, P. C. S.; MONTEIRO, R. R.; SOARES, V. J. **Análise Comparativa de Ferramentas OLAP**. UFRJ, 1998. DW Brasil. OLAP. 2007. Disponível em <<http://www.dwbrasil.com.br/html/olap.html>> Acesso em: 06/03/2014.

CHOO, Chun Wei. **A Organização do Conhecimento**. São Paulo: Editora SENAC São Paulo, 2003. SAP- Disponível em:
<http://www.sap.com/brazil/sme/index.epx>. Acesso em: 12/03/2014.

DW Brasil **OLAP** Disponível em: <<http://www.dwbrasil.com.br/html/olap.html>>
Acesso em: 19/05/2014.

FERREIRA, Manuela. Manual de Utilização Plataforma Pentaho Business Intelligence.2008.

GUIDE. Guide to Data Mining. **Introdução ao Data Mining**. Disponível em:<<http://www.data-mining-guide.net/>>. Acesso em: 04/05/2014.

INMON, W.H., **What Is A Data Warehouse?** 1995. Disponível em http://www.cait.wustl.edu/papers/prism/vol1_no1/. Acesso 25/04/2014

IMHOFF, Claudia; GALEMMO, Nicholas; GEIGER, J. G. **Mastering Data Warehouse Design: Relational and Dimensional Techniques**. Indianapolis: Wiley & Sons, Inc., 2002. 412p.

KIMBALL, Ralph. **Data warehouse toolkit**. Tradução Mônica Rosemberg. São Paulo: Makron Books, 1998. 388p.

MACHADO, F.N.R. **Projeto de Data Warehouse: Uma Visão Multidimensional**, São Paulo: Érica, 2000.

MACHADO, Felipe Nery Rodrigues. **Tecnologia e projeto de Data Warehouse**. São Paulo: Editora Érica, 2004.

MACHADO, Felipe Nery Rodrigues. **Tecnologia e Projeto de Data Warehouse: Uma visão multidimensional**. Tatuapé: Érica, 2006.

Editora Érica, 2004. MONDRIAN. **Mondrian OLAP Server**. Disponível em: <<http://mondrian.pentaho.org>>. Acesso em: 21/05/2014

ORACLE. Oracle **Data Mining**. Disponível em:<<http://www.oracle.com/technology/products/bi/odm/index.html>>. Acesso em: 04/05/2014

PENTAHO BI SUITE – Enterprise Edition Intro. Disponível em <<http://www.youtube.com/watch?v=Vy2coaxJfCo>>. Acesso em: 24/05/2014.

PENTAHO COMMUNITY. Disponível em <<http://community.pentaho.com>> . Acesso em :17/05/2014.

PENTAHO. Disponível em <<http://www.pentaho.com>>. Acesso em: 24/05/2014.

PENTAHO **História**. Disponível em <http://www.multilingualarchive.com/ma/enwiki/pt/Pentaho>. Acesso em: 24/05/2014

PENTAHO. **Open Source BI**. Disponível em <<http://www.pentaho.com/about/>>. Acesso em: 21/05/2014.

SISNEMA. **A Tecnologia do OLAP**. 2007. Disponível em <<<http://sisnema.com.br/Materias/idmat002228.htm> > Acesso em: 11/03/2014.

SISNEMA Entendendo o OLAP Disponível em:
<<http://sisnema.com.br/Materias/idmat014802.htm>> Acesso em: 08/05/2014

SISNEMA Tecnologia OLAP: Aprenda mais sobre essa importante ferramenta Disponível em: <<http://sisnema.com.br/Materias/idmat014812.htm>> Acesso em: 08/05/2014.

TERESO, M. Open source business intelligence tools for SMEs. Iberian Conference on, Junho 2011.

THOMSEN, E. **Construindo Sistemas de Informações Multidimensionais.** 2ª ed. São Paulo: Campus, 2002.