

**UNIVERSIDADE DO ESTADO DE MINAS GERAIS
FUNDAÇÃO DE ENSINO SUPERIOR DE PASSOS
FACULDADE DE INFORMÁTICA DE PASSOS**

PAULO HENRIQUE PASSOS MOREIRA

**CRIAÇÃO DE RELATÓRIOS MULTIDIMENSIONAIS UTILIZANDO A
FERRAMENTA PENTAHO COMMUNITY EDITION**

PASSOS-MG

2012

PAULO HENRIQUE PASSOS MOREIRA

**CRIAÇÃO DE RELATÓRIOS MULTIDIMENSIONAIS UTILIZANDO A
FERRAMENTA PENTAH COMMUNITY EDITION**

Monografia apresentada à Faculdade de Informática de Passos –
Universidade do Estado de Minas Gerais – Campus de Passos,
como parte das exigências para obtenção do título de Bacharel em
Sistemas de Informação.

Orientador: Prof. Espec. Anderson Ferreira Souza

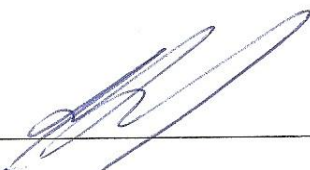
PASSOS-MG

2012

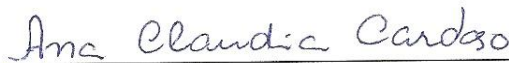
PAULO HENRIQUE PASSOS MOREIRA

**CRIAÇÃO DE RELATÓRIOS MULTIDIMENSIONAIS UTILIZANDO A
FERRAMENTA PENTAHO COMMUNITY EDITION**

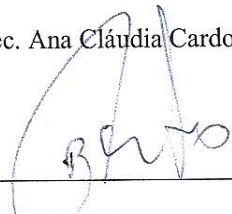
Monografia apresentada à Faculdade de Informática de Passos – Universidade do Estado de Minas Gerais – Campus de Passos, como parte das exigências para obtenção do título de Bacharel em Sistemas de Informação.



Prof. Orientador: Espec. Anderson Ferreira Souza



Prof.^a Espec. Ana Cláudia Cardoso



Prof. Espec. Eduardo Linares Porto

Passos, 6 de Junho de 2012.

Catálogo na fonte Biblioteca Central da FESP

Moreira, Paulo Henrique Passos

M838c Criação de relatórios multidimensionais utilizando a ferramenta Pentaho Community Edition [manuscrito] / Paulo Henrique Passos Moreira. - 2012.

61 f. : il., graf., tab.

Orientador : Anderson Ferreira Souza.

Monografia (graduação) – Faculdade de Informática. Fundação de Ensino Superior de Passos, Universidade do Estado de Minas Gerais, para obtenção do título de Bacharel em Sistemas de Informação.

Bibliografia : f. 57-61.

1. Tomada de decisão. 2. Sistema de apoio a decisão. 3. Business Intelligence. 4. Pentaho Community Edition. I. Souza, Anderson Ferreira. II. Faculdade de Informática. Fundação de Ensino Superior de Passos. Universidade do Estado de Minas Gerais. III. Título.

CDU- 005.53

DEDICATÓRIA

Dedico primeiramente a Deus por tudo que me proporciona na vida, por me dar forças para enfrentar os obstáculos em meu caminho. Aos meus pais, os quais amo muito, pelo exemplo de vida e família, por me acolherem em todos os momentos e por me darem condições de viver. Aos meus irmãos, por tudo que me ajudaram até hoje. À minha filha Bianca Caroline, pela alegria e diversão. À minha namorada Talita, pelo carinho, compreensão e companheirismo.

AGRADECIMENTOS

Agradeço ao Prof. Anderson Ferreira de Souza, orientador e amigo, pela dedicação e pela ajuda durante esse ano de trabalho. Um muito obrigado aos meus colaboradores Prof. Eduardo Linares Porto e Prof.^a Ana Cláudia Cardoso pelo apoio prestado. Ao Msc. Prof. Gualberto Rabay, pelo aprendizado e dedicação durante esses anos em que fui seu aluno e ao Msc. Prof. Kleber Resende por seu profissionalismo. Agradeço a minha amiga Veilda, que me fez acreditar que esse projeto seria possível. Agradeço também aos meus amigos Elton Silva e Hétory Ferreira pela dedicação no apoio prestado, pelas dicas, ajuda no material e nas discussões. Ao Especialista em *Pentaho*, Sr. Caio Moreno de Souza (Professor Coruja). Obrigado também aos membros das comunidades “Business Intelligence – BI” e “BI Brasil” da rede social LinkedIn®, e aos integrantes dos grupos “*Pentaho Brasil*” e “*PentahoBr*” do Google Grupos e Grupos Yahoo, respectivamente. Um agradecimento a todos os professores em geral da Faculdade de Informática de Passos (FIP). Agradeço à minha namorada Talita, pelo carinho e pelas palavras de conforto. Aos meus pais, irmãos e minha filha por serem meus amigos, companheiros e acima de tudo, minha família. Vocês são tudo na minha vida! E finalmente, agradeço a todos que me ajudaram diretamente ou indiretamente para o desenvolvimento deste projeto. Muito obrigado.

EPÍGRAFE

“A informação só tem valor quando gera conhecimento e o conhecimento não aplicado é tão inútil como a desinformação.”

Paulo Rubini

RESUMO

O processo de tomada de decisão é extremamente importante para qualquer empresa, independente do seu segmento de mercado ou de seu tamanho, pois a todo o momento elas se deparam com situações em que os envolvidos no negócio (analistas, gestores, empresários, equipe, etc.) devem aplicar a solução mais correta para algum problema específico. A inexistência de métodos que auxiliem a tomada de decisão pode causar muitos transtornos a pequeno, médio ou até longo prazo, como a escolha de uma alternativa errada, a interpretação ou elaboração incorreta de um relatório de vendas, consulta de uma informação sem crédito, entre outros. Com o propósito de amenizar tais problemas o projeto apresenta o conceito de Sistemas de Apoio à Decisão (SAD), cujo objetivo é aprimorar os processos de tomada de decisão mas foca principalmente na apresentação das metodologias de Business Intelligence (BI). Será utilizada a Suíte *open source* (software livre) mais utilizada no mundo todo chamada *Pentaho Community Edition* (Versão Comunidade) que possui diversas ferramentas, onde quando agregadas tem o objetivo de transformar dados em conhecimento. A integração desses fatores irá determinar o sucesso do projeto.

Palavras-chave: Tomada de decisão. Sistemas de Apoio à Decisão. Business Intelligence. Pentaho Community Edition.

ABSTRACT

The process of decision making is extremely important for any company, regardless of their market segment or its size, because at every moment they are faced with situations where those involved in business (analysts, entrepreneurs, managers, staff, etc.) should apply the most suitable solution for any specific problem. The lack of methods that assist decision-making can cause much disruption to small, medium or even long term, as the wrong choice of an alternative, incorrect interpretation or preparation of a sales report, query information without a credit and other. In order to minimize such problems, the project introduces the concept of Decision Support Systems (DSS), whose objective is to improve the processes of decision making but focuses mainly on the presentation of the methodologies of Business Intelligence (BI). Suite will be used most widely used *open source* worldwide called *Pentaho Community Edition* that has many tools, which when aggregated aims to transform data into knowledge. The integration of these factors will determine the success of the project.

Keywords: Decision-making. Decision Support Systems. Business Intelligence. Pentaho Community Edition.

LISTA DE SIGLAS

BD – *Banco de Dados*

BI – *Business Intelligence* (Inteligência de negócios)

CPF – *Cadastro de Pessoas Físicas*

CRM – *Customer Relationship Management* (Gestão de Relacionamento com o cliente)

DW – *Data Warehouse* (Armazém de dados)

DM – *Data Mart*

ERP – *Enterprise Resource Planning* (Sistema de Gestão Empresarial)

ETL – *Extract, Transformation and Load* (Extração, Transformação e Carga)

FK – *Foreign Key* (Chave Estrangeira)

MDX – *Multidimensional Expressions* (Expressões Multidimensionais)

OLAP – *On Line Analytical Processing* (Análise de Processamento On Line)

OLTP – *On Line Transaction Processing* (Processamento de transações On Line)

PDI – *Pentaho Data Integration*

PK – *Primary Key* (Chave Primária)

PSW – *Pentaho Schema Workbench*

ROLAP – *Relational On-Line Analytical Processing* (Análise de Processamento On Line)

SAD – *Sistemas de Apoio a Decisão*

SGBD – *Sistema Gerenciador de Banco de Dados*

SI – *Sistemas de Informação*

SIE – *Sistemas de Informação Executivos*

SIG – *Sistema de Informação Gerencial*

SQL – *Structured Query Language* (Linguagem de Consulta Estruturada)

TI – *Tecnologia da Informação*

XML - *Extensible Markup Language* (Linguagem de Marcação Extensível)

LISTA DE FIGURAS

Figura 1: Arquitetura simples de um SGBD.....	19
Figura 2: Teia causada pela falta de integração.	20
Figura 3: Dado X Informação X Conhecimento X Decisão.	22
Figura 4: Cubo multidimensional.	31
Figura 5: Modelo Star Schema.	32
Figura 6: Operações Drill-down e Roll-up	34
Figura 7: Processo de ETL.....	36
Figura 8: Passos para implantação do ETL.	36
Figura 9: Modelo do Data Mart	47
Figura 10: Modelagem do cubo multidimensional no Pentaho Schema Workbench	49
Figura 11: Transformação para Fato Vendas.....	50
Figura 12: Transformações para Dimensões Cliente e Produto.....	50
Figura 13: Tela inicial Mondrian - Cliente JPivot - Cubo OLAP	51
Figura 14: Botões de navegação JPivot	51
Figura 15: Drill-down na dimensão período.....	52
Figura 16: Gráfico de vendas durante o ano de 2011	52
Figura 17: Gráfico comparativo dos 5 clientes que mais compram com os outros	53
Figura 18: Exploração do cubo OLAP	53
Figura 19: Gráfico dos 5 produtos que mais arrecadaram em 2011	54
Figura 20: Variação da venda de leite em 2011	54
Figura 21: Comparação das vendas entre Janeiro de 2011 e 2012	55

LISTA DE TABELAS

Tabela 1: Diferenças entre Banco de Dados Operacionais e Data Warehouse	26
Tabela 2: Características das ferramentas OLAP.....	29
Tabela 3: Características que diferenciam os sistemas OLAP e OLTP	30
Tabela 4: Dimensão Produto	47
Tabela 5: Dimensão Cliente	48
Tabela 6: Dimensão Período	48
Tabela 7: Fato Venda.....	48

SUMÁRIO

1 INTRODUÇÃO.....	14
1.1 Tema e Problema.....	14
1.2 Objetivo.....	15
1.3 Justificativa	16
1.4 Organização do Projeto	16
2 REVISÃO BIBLIOGRÁFICA	18
2.1 A evolução dos Sistemas de Informação	18
2.1.1 Sistemas de Informação Gerencial	20
2.1.2 Sistemas de Informação Executivos	21
2.1.3 Sistemas de Apoio a Decisão	21
2.2 O ciclo da tomada de decisão	22
2.3 Business Intelligence	23
2.3.1 Os benefícios do Business Intelligence.....	24
2.4 Data Warehouse.....	25
2.4.1 Definições de Data Warehouse	26
2.4.2 Data Mart	27
2.5 OLAP	28
2.5.1 OLAP versus OLTP	29
2.5.2 Cubo Multidimensional	30
2.5.3 Modelo Star Schema ou Estrela	32
2.5.4 ROLAP (Relational On Line Analytical Processing)	33
2.5.5 Ferramentas de OLAP.....	33
2.6 ETL	35
2.6.1 Fases de um processo ETL	35
3 METODOLOGIA	38
3.1 Etapas realizadas	38
3.1.1 Pesquisa bibliográfica	38
3.1.2 Elaboração do Modelo Entidade Relacionamento (MER) do Data Mart.....	39
3.1.3 Criação do Data Mart e extração dos dados gerenciais	39
3.1.4 Desenvolvimento do cubo multidimensional	40
3.1.5 Desenvolvimento e apresentação dos relatórios multidimensionais	40
3.2 Ferramentas utilizadas	41
3.2.1 Mysql.....	41
3.2.2 A Pentaho	41

3.2.3 Pentaho BI Server	42
3.2.4 Mondrian.....	43
3.2.5 Kettle	43
3.2.6 Pentaho Schema Workbench (PSW)	44
3.2.7 JPivot	44
3.2.8 SQL Power Architect.....	44
4 DESENVOLVIMENTO DO PROJETO.....	46
4.1 Modelagem do Data Mart Multidimensional	46
4.2 Dicionário de dados.....	47
4.3 Modelagem do cubo multidimensional	49
4.4 Transformações ETL com Kettle	50
4.5 Visualização do cubo no JPivot	51
5 CONCLUSÃO.....	55
REFERÊNCIAS BIBLIOGRÁFICAS.....	57

1 INTRODUÇÃO

Os negócios atualmente estão se tornando muito mais complexos e têm sofrido várias mudanças ao longo do tempo. As organizações sejam elas pequenas, médias ou grandes, forçam seus líderes e suas respectivas equipes, a buscarem alternativas para apoiar as decisões com mais rapidez, agilidade e eficácia. Isso quer dizer que tomar decisões exige quantidades consideráveis de dados, além de informação e conhecimento. O processamento desse grande volume de informações deverá ocorrer de forma rápida, em tempo real e na maioria das vezes, com apoio computacional. No projeto será usado o Sistema de Gerenciamento de Banco de Dados (SGBD) MySQL onde estão os dados gerenciais e a Plataforma *Pentaho Community Edition* em ambiente Windows 7, para desenvolvimento do projeto e apresentação de relatórios multidimensionais. O estudo foi realizado em uma distribuidora de produtos alimentícios de Passos-MG, mas pode servir de apoio para outras empresas interessadas.

1.1 Tema e Problema

A pedido do gerente, a identificação da empresa e também seus dados privados, ficarão no anonimato. Assim, durante o projeto será denominada **Empresa Teste**. Atualmente, a referida empresa possui um sistema gerencial desenvolvido em base de dados relacionais, mas tem dificuldades em obter e visualizar de forma inteligente as informações necessárias sobre as suas vendas, pois os relatórios são limitados, fazendo

que seu gerente não consiga uma melhor interpretação de seus dados gerenciais e consequentemente a tomada de decisão.

Por esses motivos, foram desenvolvidos relatórios multidimensionais, utilizando a suíte *Pentaho Community Edition*, com o intuito de auxiliar o gerente a visualizar melhor os seus resultados para poder, assim, tomar melhores decisões.

1.2 Objetivos

Criar relatórios multidimensionais para auxiliar o gerente de uma distribuidora de produtos alimentícios a tomar decisões mais objetivas através das metodologias de Business Intelligence (BI).

Estudar as ferramentas da suíte BI *Pentaho Community Edition* (Versão Comunidade) que será utilizada para o desenvolvimento do projeto. As ferramentas estudadas serão: *Kettle*, *Mondrian*, *JPivot* e *BI Server*, além do SGBD MySQL e o SQL Power Architect.

O projeto também apresentará os conceitos e definições de BI, seus principais componentes, ferramentas e metodologias de coleta de informações, podendo servir de apoio para outras empresas interessadas.

1.3 Justificativa

A tomada de decisão através de um sistema de BI, ainda não é uma prática conhecida pelos empresários e até mesmo pelos profissionais da área de tecnologia. Através da utilização de sistemas BI é possível evitar problemas como: falta de informação sobre os seus clientes, fornecedores e até mesmo seus concorrentes, a não compreensão de seu próprio ambiente interno e organizacional, no que diz respeito à análise de dados, etc. (LIMA¹, LIMA², 2011). Com a implantação de um sistema BI, estes problemas citados poderão ser amenizados, pois esse sistema irá auxiliar a tomada de decisão.

1.4 Organização do Projeto

O capítulo um contém a introdução do projeto. O capítulo dois contém a revisão bibliográfica, onde está sendo apresentada a evolução dos Sistemas de Informação, o ciclo da tomada de decisão, os conceitos e benefícios do Business Intelligence, entre outros. No capítulo três contém a metodologia que foi utilizada, as etapas, ferramentas e técnicas utilizadas no desenvolvimento do projeto.

O capítulo quatro contém a apresentação do projeto: o modelo do Data Mart, o dicionário de dados do Data Mart e os relatórios gerados nesse estudo, ou seja, os resultados obtidos durante o desenvolvimento.

Por fim, o capítulo cinco contém a conclusão do projeto. E logo em seguida as referências bibliográficas utilizadas durante o estudo e desenvolvimento do projeto.

2 REVISÃO BIBLIOGRÁFICA

Para o desenvolvimento do projeto foram necessárias fundamentações teóricas, como as definições e características de Data Warehouse e Data Mart, consultas OLAP (*On Line Analytical Process* – Processo de Análise On Line), a técnica de extração e transformação de dados ETL (*Extract, Transformation and Load* – Extrair, Transformar e Carregar) e a plataforma *Pentaho Community Edition* e suas ferramentas, com o intuito de adquirir maior conhecimento teórico e técnico (PRIMAK, 2008). A seguir, serão apresentadas as principais fundamentações teóricas necessárias para o seu desenvolvimento.

2.1 A evolução dos Sistemas de Informação

Nos últimos anos, a concorrência entre as empresas, fez com que elas se adaptassem às mudanças para se manterem no mundo dos negócios. Para tanto, era primordial que toda a estrutura “mercadológica” (JUNIOR, 2004, p. 2), no que diz respeito a buscar novos clientes, inovar na qualidade, ampliar o ramo de negócios, fosse reformulado, forçando a cúpula administradora a analisar o seu cenário atual para tomar as decisões corretas (JUNIOR, 2004).

Na década de 60, a estrutura de um SI (Sistemas de Informação) e de recursos de hardware eram muito limitados. Diante deste cenário a visão da empresa tornava-se singular no qual não trazia ganhos na área de decisão e dos negócios (PRIMAK, 2008).

No início da década de 70, surgiram o armazenamento em disco, o SGBD e o OLTP (*On Line Transaction Process* – Processo de Transação On Line). Através do armazenamento em disco, os dados eram acessados diretamente em um só local, fazendo com que o tempo de processamento fosse mais rápido, enquanto que os sistemas OLTP vieram para processar dados operacionais ou transacionais garantindo maior segurança, agilidade e eficiência. (JUNIOR, 2004)

Ainda no pensamento de Júnior (2004, p. 3), os SGBD tinham como principais objetivos “potencializar o gerenciamento dos dados, eliminando redundâncias de informação. (Figura 1). Mas mesmo com os SGBD, as empresas continuam criando sistemas isolados em termos de compartilhamento de dados em comum”. (Figura 2)

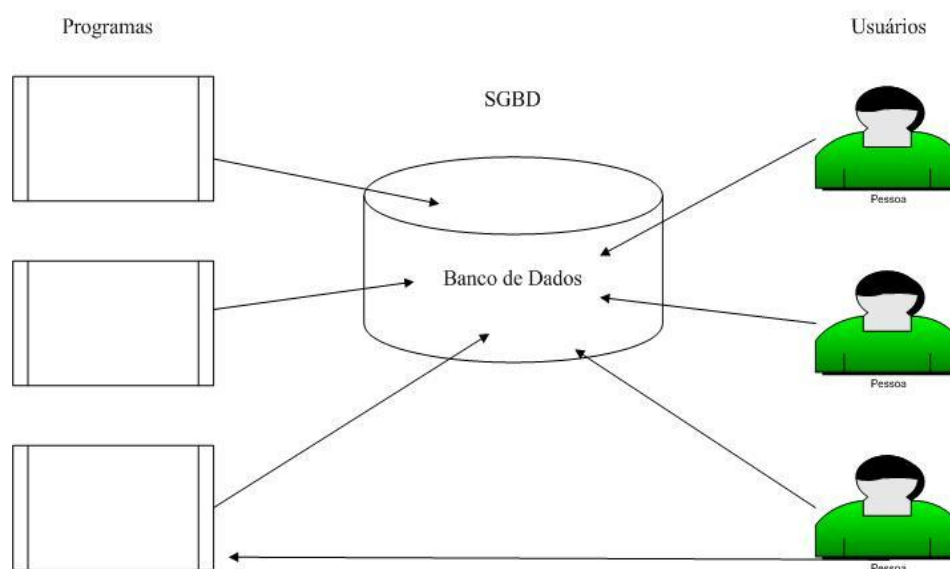


Figura 1: Arquitetura simples de um SGBD (Fonte: JUNIOR, 2004, p.4).

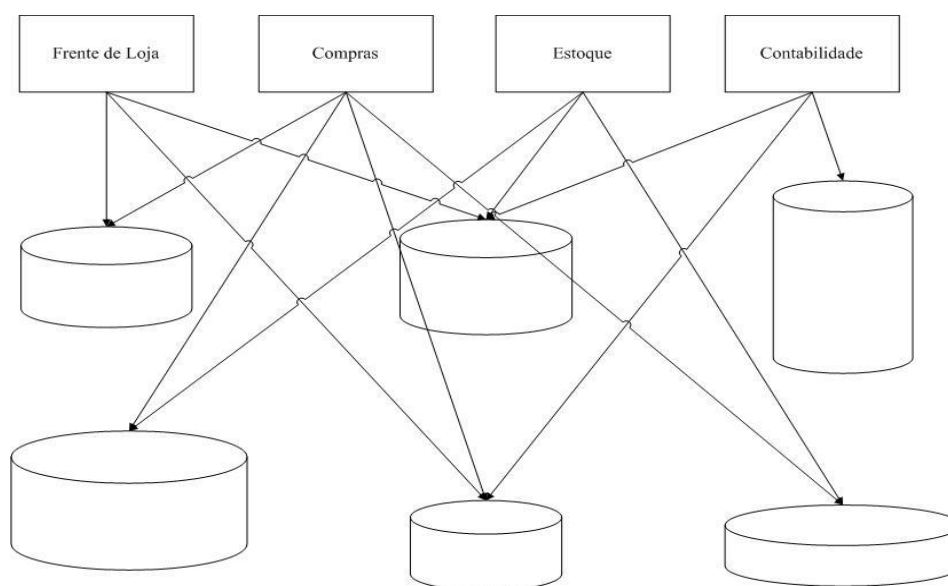


Figura 2: Teia causada pela falta de integração (Fonte: JUNIOR, 2004, p.5).

2.1.1 Sistemas de Informação Gerencial

O Sistema de Informação Gerencial (SIG) oferece funções de planejamento, controle e organização de uma empresa, as quais são de extrema importância para a gerência da empresa. “O SIG é representado pelo conjunto de subsistemas, visualizados de forma integrada e capaz de gerar informações necessárias ao processo decisório” (MAGALHAES, GEDRES 2008).

Um SIG verdadeiro deverá ter a capacidade de fornecer informações para os planejamentos operacional, tático e estratégico (JUNIOR, 2004).

2.1.2 Sistemas de Informação Executivos

Os Sistemas de Informação Executivos (SIE) são desenvolvidos para atender a necessidade dos executivos de gerenciarem o processo de gestão. Um SIE deverá ter uma interface bastante simples de ser utilizada, pois seus usuários são pessoas que não tem muito tempo disponível para analisar as informações (PRIMAK, 2008).

Segundo Pozzebon e Freitas (1996, p.29, *apud* BEUREN; MARTINS, 2011, p. 10), o SIE “é uma solução em termos de informática que disponibiliza informações corporativas e estratégicas para os decisores de uma organização, de forma a aperfeiçoar sua habilidade para tomar decisões de negócios importantes”.

2.1.3 Sistemas de Apoio a Decisão

Sistemas de Apoio à Decisão (SAD) passou a ser a ferramenta ideal para auxílio da tomada de decisão. SAD são sistemas complexos que permitem total acesso à base de dados corporativos, modelagem de problemas, simulações, e que possuem uma interface amigável. Um SAD deve ser desenvolvido com ativa participação do usuário, além de permitir mudanças com rapidez e sem transtornos. Sendo assim, após um curto prazo (semanas), o sistema pode ser modificado de acordo com as necessidades reais (PRIMAK, 2008).

Um SAD apoia os processos de tomada de decisão nos níveis tático, estratégico e operacional. Entretanto, somente essas características não são suficientes, pois qualquer sistema de informação (SI) pode ser gerencial, mas nem por isso todo SI será um SAD (JUNIOR, 2004)

Segundo Júnior (2004), um SAD é formado por Banco de Dados (BD), Sistema Gerenciador de Banco de Dados (SGBD) e Business Intelligence (BI).

2.2 O ciclo da tomada de decisão

“Dado” é qualquer observação bruta que, sozinha, não traz nenhum significado, ou sua interpretação pode passar um duplo sentido (OLIVEIRA et al, 2009). “Informação” é o que podemos fazer com o dado, ou seja, como podemos aplicar. O resultado que se obtêm da interpretação de uma informação dá-se o nome de “conhecimento” (PRIMAK, 2008). E por fim, a “decisão” que consiste na ação através dos dados, informações e conhecimentos que será executada.

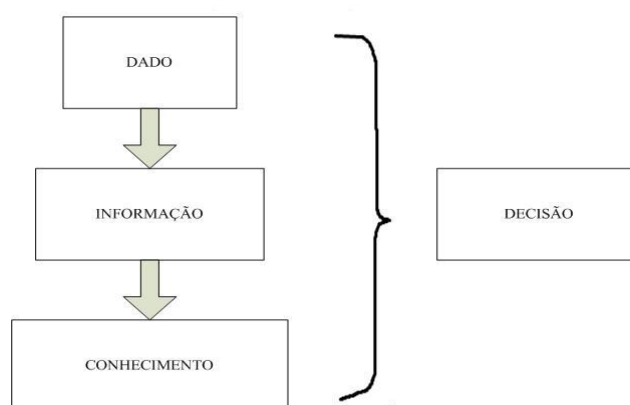


Figura 3: Dado X Informação X Conhecimento X Decisão
(Fonte: PRIMAK, 2008, p.19).

2.3 Business Intelligence

BI pode ser definido como Inteligência de Negócios ou Inteligência Empresarial. É um método que tem como ponto chave auxiliar organizações a tomarem decisões mais inteligentes, explorando informações, analisando-as e desenvolvendo percepções e entendimentos a seu respeito, através de dados coletados de Sistemas Transacionais (SIG, ERP, CRM, etc.). Esse método visa a interpretar dados armazenados em imensos Bancos de Dados, gerando assim informações de qualidade. De acordo com Primak:

[...] O termo Business Intelligence, surgiu na década de 80 no Gartner Group e faz referência ao processo inteligente de coleta, organização, análise, compartilhamento e monitoração de dados contidos em Data Warehouses/ Data Marts, gerando informações para o suporte e a tomada de decisões no ambiente de negócios (PRIMAK, 2008, p.5).

BI é desenvolvido sobre estruturas de Banco de dados e permite que os envolvidos no negócio tenham acesso às informações sem depender da intervenção direta da área de Tecnologia da Informação (TI), concentrando os esforços na construção e aperfeiçoamento de um ambiente fácil para seus usuários.

Turban et al (2009, p. 27), dizem que: “Business Intelligence ‘é um termo guarda-chuva’, que inclui arquiteturas, ferramentas, bancos de dados, aplicações e metodologias”.

Ainda na visão destes autores, “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 a análise adequada” (TURBAN et al, 2009, p. 32).

Hoje em dia é possível implantar o BI em qualquer empresa, independente de seu tamanho, para que elas possam ser mais organizadas, diminuir os custos, prever acontecimentos, entre outras coisas (PRIMAK, 2008).

“BI baseia-se na transformação de dados em informações, depois em decisões e finalmente em ações” (TURBAN et al, 2009, p. 27).

BI é composto pelos seguintes componentes: DW (*Data Warehouse*), ETL (*Extract, Transformation and Load* – Extrair, Transformar e Carregar), OLAP (*On-Line Analytical Process* – Processo de Análise On-Line) e *Data Mining* (Mineração de dados).

2.3.1 Os benefícios do Business Intelligence

A implantação de BI pode trazer alguns dos seguintes benefícios (ECKERSON, 2003, *apud* TURBAN et al, 2009):

- Resultados mais rápidos;
- Melhores estratégias e planos;
- Baixo custo;
- Melhores decisões táticas.

“É preciso o alinhamento entre o objetivo do projeto com os interesses e as estratégias da empresa, levando-se em conta a importância da lei ‘Custo X Benefício’ (PRIMAK, 2008, p. 12).

2.4 Data Warehouse

O conceito Data Warehouse (DW) existe desde o final da década de 80. Para compreender o que é uma DW, é necessário fazer uma comparação com a definição de Banco de dados:

[...] Um Banco de dados é uma coleção de dados operacionais armazenados e utilizados pelo sistema de aplicações de uma empresa, chamados de operacionais ou primitivos. Pode-se dizer que um DW é, na verdade, uma coleção de dados derivados dos dados operacionais para Sistemas de Apoio à Decisão. Esses dados são referidos como dados ‘gerenciais’, ‘informacionais’ ou ‘analíticos’ (PRIMAK, 2008, p. 37).

Os responsáveis pela tomada de decisão necessitam de informações confiáveis sobre operações atuais, tendências e mudanças. É possível verificar que um DW é uma ferramenta que tem um papel fundamental na tomada de decisão, pois todos os dados, sejam eles de muitos anos ou recentes, serão analisados e interpretados desse repositório de dados. Concluindo, um DW é supostamente um local onde os dados são armazenados de tal forma que qualquer pessoa ligada ao processo tenha total acesso aos dados.

2.4.1 Definições de Data Warehouse

Um DW consiste em um repositório de dados, desenvolvido em uma ampla estrutura que disponibiliza dados atuais e históricos, com o objetivo de auxiliar a o processo analítico na tomada de decisão (TURBAN et al, 2009).

Na tabela a seguir, estão relacionadas algumas diferenças, e os dados que são manipulados por um Banco de Dados Operacional e um DW.

Tabela 1: Diferenças entre Banco de Dados Operacionais e DW

Características	Bancos de dados operacionais	Data Warehouse
Objetivo	Operações diárias do negócio	Analisar o negócio
Uso	Operacional	Informativo
Tipo de processamento	OLTP	OLAP
Unidade de trabalho	Inclusão, alteração, exclusão	Carga e consulta
Números de usuários	Milhares	Centenas
Tipo de usuário	Operadores	Comunidade gerencial
Interação do usuário	Somente pré-definida	Pré-definida e ad-hoc
Volume	Megabytes - gigabytes	Gigabytes - terabytes
Histórico	60 a 90 dias	5 a 10 anos
Granularidade	Detalhados	Detalhados e resumidos
Estrutura	Estática	Variável
Manutenção desejada	Mínima	Constante
Acesso a registros	Dezenas	Milhares
Atualização	Contínua	Periódica
Número de índices	Poucos/simples	Muitos/complexos

Fonte: PRIMAK, 2008, p. 38-39.

Além de possuir estrutura relacional ou multidimensional, um DW possui algumas outras características:

- **Orientado por assunto:** Os dados mais importantes são armazenados e organizados por assunto, de acordo com as regras de negócio (MACHADO, 2000);
- **Integrado:** O DW integrado consiste em agrupar dados de bases de dados heterogêneas em um formato único na saída (TURBAN et al, 2009);
- **Variação no tempo:** As informações fazem parte de uma determinada época: dia, semana, mês, ano, etc. (JUNIOR, 2004);
- **Não volátil:** Os dados, uma vez inseridos, não podem ser alterados somente consultados (TURBAN et al, 2009);

2.4.2 Data Mart

Um DW possui informações de toda empresa, um Data Mart (DM) é reduzido e é concentrado em uma área ou departamento específico, sendo assim, um subconjunto de um DW. Um DM pode ser dependente ou independente. Um DM dependente é um subconjunto criado a partir de um DW, o que é uma grande vantagem, pois suporta um único modelo de dados em toda a organização. Um DM independente é um DW pequeno, projetado diretamente para um departamento ou para uma unidade estratégica de

negócios. (PRIMAK, 2008). O referido autor diz ainda que, “enquanto um DM trata de um problema departamental ou local, um DW envolve o esforço de toda companhia para que o suporte às decisões atue em todos os níveis da organização” (PRIMAK, 2008, p.48).

Muitas empresas aderem ao DM, não somente por ser um DW menor, mas principalmente pelo fato de que um DW requer bastante tempo para implantação e adaptação, investimentos financeiros e muito esforço gerencial (PRIMAK, 2008).

É importante frisar que *metadados* são todos os dados que estão armazenados no DW e que fazem parte de seu ambiente (desenvolvimento e utilização).

2.5 OLAP

Processamento analítico on-line, ou OLAP, é um conjunto de ferramentas de BI, usadas em sistemas on-line, para apoiar empresas na análise de dados na tomada de decisão através de várias perspectivas, cujo resultado final é a geração de consultas, relatórios, gráficos ad-hoc e sua execução, apresentações visuais e consultas estatísticas, utilizados na maioria das vezes por usuários finais.

Turban et al diz que “os produtos OLAP oferecem recursos de modelagem, análise e visualização de grandes conjuntos de dados, ou para sistemas de gerenciamento de Banco de dados, ou para sistemas baseados em DW” (TURBAN et al, 2009, p. 109).

Tabela 2: Características das ferramentas OLAP

Características das ferramentas OLAP
Visão conceitual multidimensional
Transparência
Acessibilidade
Informações de desempenho consistente
Arquitetura cliente/servidor
Dimensionalidade genérica
Manipulação de matrizes dinamicamente
Suporte a multiusuários
Operações ilimitadas em referências cruzadas
Manipulação dos dados intuitivamente
Consultas flexíveis
Níveis de dimensões e agregações ilimitadas

Fonte: PRIMAK, 2008, p. 57

2.5.1 OLAP versus OLTP

Por muito tempo, o foco dos analistas foi criar sistemas que pudessem gerar processamento de transações corporativas, em que deveriam ser tolerantes a falhas e fornecer respostas rapidamente, em curto espaço de tempo. Anzanello (2002), diz que o avanço tecnológico possibilitou o armazenamento de grande volume de dados em meio magnético, gerando a necessidade de recuperá-los e analisá-los. As aplicações OLTP (*On Line Transaction Processing* - Processamento de Transações On Line) atendem apenas à manipulação de dados operacionais, o que não era suficiente para apoiar o usuário na tomada de decisões estratégicas.

Turban et al (2009, p.109) afirma que “OLTP ofereceu uma solução realmente eficaz, que gira em torno de atividades repetitivas e de rotina, usando um ambiente de Banco de dados relacional, distribuído em grandes quantidades e manipulações simples”.

Tabela 3: - Características que diferenciam os sistemas OLAP e OLTP

Características	OLTP	OLAP
Operação típica	Atualização	Análise
Telas	Imutável	Definida pelo usuário
Nível de dados	Atomizado	Altamente sumarizado
Idades dos dados	Presente	Histórico, anual e projetado
Recuperação	Poucos registros	Muitos registros
Orientação	Registro	Arrays
Modelagem	Processo	Assunto

Fonte: ANZANELLO, [s.d.], p. 3.

2.5.2 Cubo Multidimensional

A característica principal dos sistemas baseados em OLAP é mostrar os dados de uma empresa de forma multidimensional, os quais são apresentados por um cubo, onde cada lado ou dimensão representa as áreas mais importantes como, produto, cliente, funcionário, etc. (JUNIOR, 2004).

As consultas baseadas na utilização do cubo multidimensional são rápidas, pois na maioria das vezes os dados já estão calculados, dando um salto muito grande na obtenção de resultados (SILVA, 2012).

Dimensão é a unidade de análise que relaciona e agrupa os dados do negócio, ou seja, são basicamente os cabeçalhos compostos de colunas e linhas (SILVA, 2012). Como o exemplo da Figura 4: os *produtos* movimentados (o quê aconteceu), o *tempo* (quando aconteceu) e a *região* (onde aconteceu).

Métricas são os dados quantitativos ou valores que serão calculados ou medidos. As métricas são armazenadas na tabela fato (SOUZA, 2010).

Fato é a tabela onde são armazenadas as chaves que relacionam as dimensões e métricas (SOUZA, 2010).

Hierarquia são as possibilidades de detalhamento de uma análise dentro de uma dimensão, por exemplo, detalhar a data em: mês, semana, bimestre, ano, etc. (PEREIRA, 2010?).

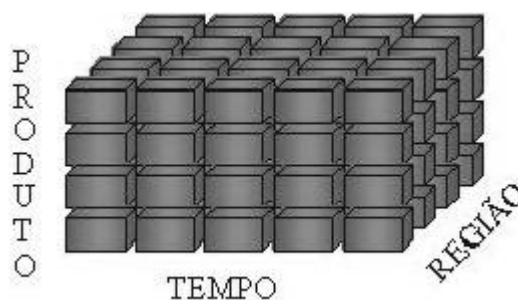


Figura 4: Cubo multidimensional (Fonte: RUBINI, 1999 *apud* DALFOVO et al, [s.d.], p. 3).

2.5.3 Modelo Star Schema ou Estrela

Dentro da modelagem multidimensional existem 2 tipos de modelos: *Star Schema* (modelo Estrela) e *Snowflake* (Modelo Flocos de Neve). O modelo adotado no Data Mart do projeto foi o modelo Star Schema.

Esse modelo tem esse nome, pois o seu formato faz lembrar uma estrela, e tecnicamente falando, a tabela fato se conecta com as dimensões por múltiplos relacionamentos, e cada dimensão se conecta por apenas um relacionamento na tabela fato. A figura 5 a seguir nos mostra o modelo Star Schema (SOUZA, 2010).

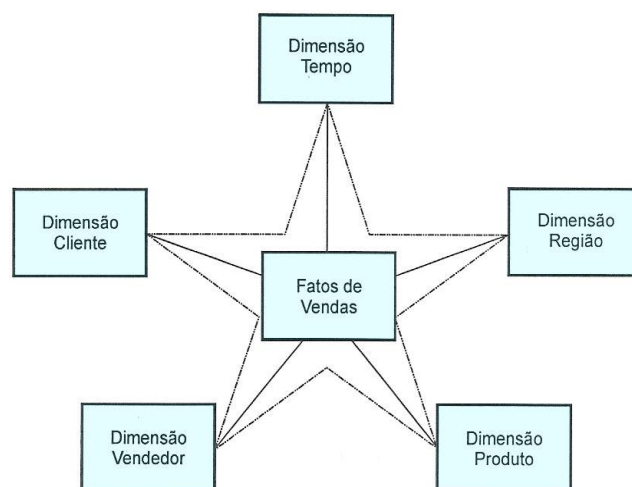


Figura 5: Modelo Star Schema. Fonte: MACHADO *apud* BORGES

(2010, p. 26).

2.5.4 ROLAP (Relational On Line Analytical Processing)

ROLAP ou OLAP Relacional ocorre quando o OLAP é criado sobre um BD relacional, extraindo dados de bancos de dados relacionais. (TURBAN et al, 2009). Nesse tipo de OLAP, o pedido é enviado e processado no servidor de Banco de dados relacional, onde o cubo permanece armazenado.

2.5.5 Ferramentas de OLAP

Nos dias atuais, existem muitas ferramentas de OLAP e isso faz com que mudanças aconteçam em uma velocidade considerável. E essas ferramentas, em sua maioria, possuem dois componentes importantes: ferramenta do administrador e ferramenta do usuário final, sendo que a parte administradora cria e gera os cubos, enquanto que o usuário acessa e extrai informações para tomada de decisão (ANZANELLO, 2003).

Primak (2008) descreve alguns termos mais utilizados na arquitetura OLAP:

- Consultas ad-hoc: o próprio usuário gera consultas de acordo com suas necessidades de cruzar as informações que o levem a encontrar aquilo que procura;

- Slice-and-dice/Pivot table: usado para mudar de lugar uma informação, trocando linhas por colunas, de forma que os seus usuários entendam o que está proposto, e também girar o cubo sempre que precisar;
- Drill-Down/Roll-Up: arquitetura usada para explorar informações em diversos níveis, de forma detalhada, como por exemplo: anual, semestral, trimestral, mensal, diário, entre outros.

Para melhor entendimento, Costa afirma que a operação drill-down detalha uma ‘fatia’ do cubo de dados decompondo e formando um novo cubo, passando a ter o detalhamento em um maior nível. Já a operação roll-up faz o inverso da operação drill-down gerando um cubo mais abrangente do que o anterior.

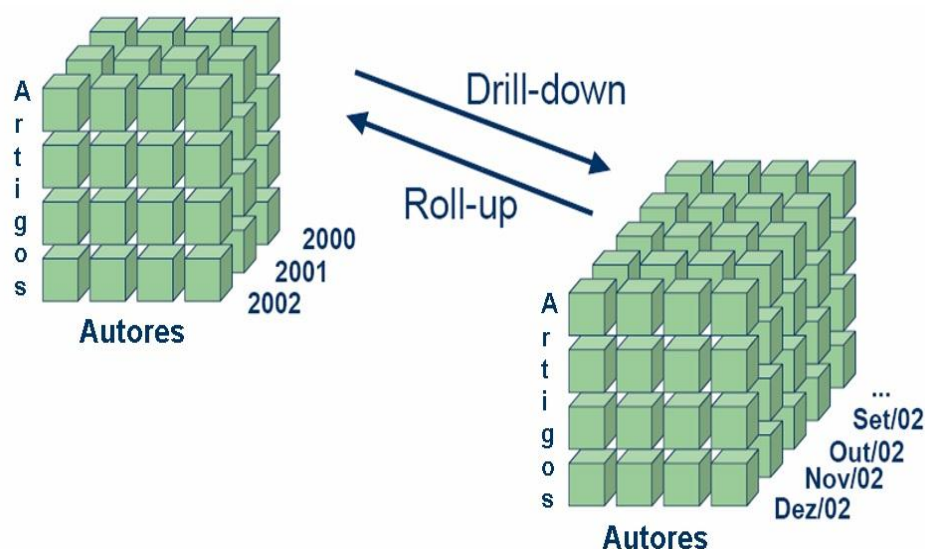


Figura 6: Operações Drill-down e Roll-up (Fonte: COSTA, s.d., p.5).

2.6 ETL

O processo de ETL consiste na leitura dos dados, transformação e carga dos dados de um banco de dados origem, para outros bancos de dados de destino (DW, Excel ou outro), sendo que, o processo de transformação não é obrigatório. No desenvolvimento de um DW, o processo de ETL é o mais crítico e demorado do projeto, pois além de coletar dados, ele faz a transformação e a limpeza (ABREU, 2008).

Os dados podem ser originários de vários sistemas e, por isso, uma mesma informação poderá ter diferentes representações.

2.6.1 Fases de um processo ETL

Extração ou *Extract*: Nessa primeira fase, é feita a seleção da base de dados e enviados ao DW, mas somente serão enviados os dados que estiverem dentro do padrão de DW.

Transformação ou *Transformation*: Nessa fase os dados selecionados na extração, passam por um processo de padronização e também limpeza, corrigindo inconsistências existentes para padronizar um dado que tem o mesmo significado, só que representado de forma diferente (COSTA, s.d.).

Carga ou Load: O objetivo dessa fase é simplesmente carregar os dados para o DW ou DM, mantendo a integridade dos dados para serem analisados posteriormente. Conforme as necessidades da empresa, o processo de ETL pode crescer gradualmente.

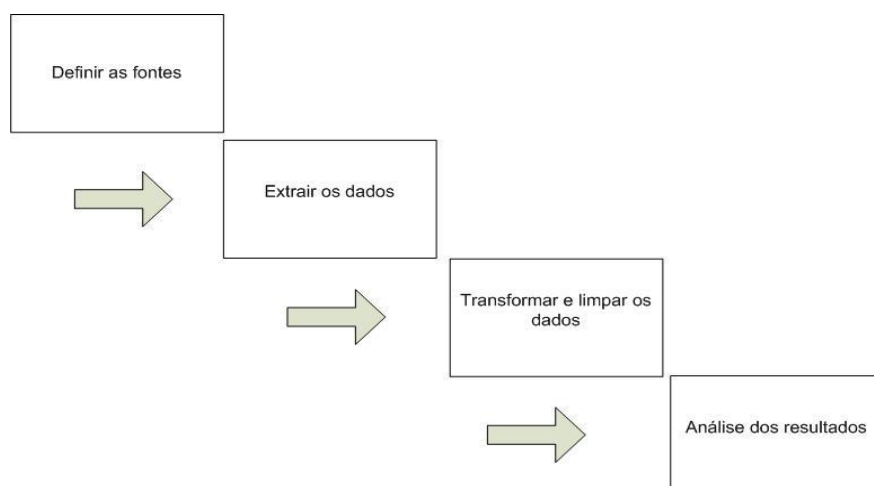


Figura 7: Processo de ETL (Fonte: PRIMAK, 2008, p.64).

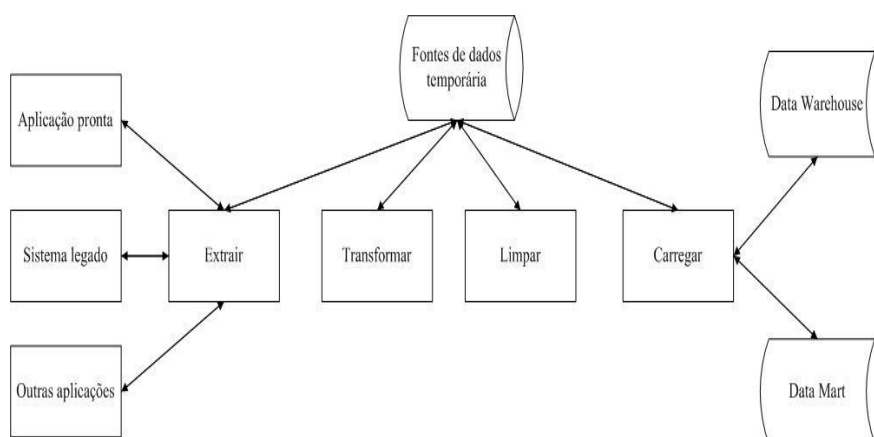


Figura 8: Passos para implantação do ETL (Fonte: TURBAN et al, 2009, p. 72).

Para facilitar a carga, todos os dados migrados para um DW são gravados em tabelas temporárias. Um DW possui várias regras de negócio que definem a forma com que os dados serão utilizados: regras de consumo, padronização dos atributos codificados e regras de cálculo. Se os dados possuírem algum problema na qualidade, eles precisarão ser corrigidos antes de serem carregados no DW (TURBAN et al, 2009).

Brown (2004, *apud* TURBAN et al, 2009) nos mostra os critérios mais importantes para escolher qual ferramenta de ETL deve-se usar:

- Ler e gravar dados em vários tipos de bases de dados;
- Capturar e entregar automaticamente os metadados;
- Possuir interface amigável, tanto para o usuário quanto para o programador.

3 METODOLOGIA

3.1 Etapas realizadas

Durante o desenvolvimento do projeto, várias etapas foram realizadas. A seguir essas etapas serão apresentadas:

3.1.1 Pesquisa bibliográfica

Durante o desenvolvimento do projeto, foram necessárias várias fundamentações teóricas para melhor entendimento, tanto técnico quanto teórico sobre os assuntos tratados no percurso do estudo. Foi necessário um estudo aprofundado sobre diversas tecnologias utilizadas no desenvolvimento do projeto, como estudo das principais funcionalidades da suíte BI *Pentaho Community Edition* e SGBD MySQL.

3.1.2 Elaboração do Modelo Entidade Relacionamento (MER) do Data Mart

Para uma melhor visão do Data Mart desenvolvido, foi criado um MER (Modelo Entidade Relacionamento) do mesmo através da ferramenta SQL Power Architect. Escolheu-se essa ferramenta, pois foi possível criar de uma forma totalmente gráfica as tabelas dimensões e a fato, e em seguida executar via comando SQL a criação da estrutura diretamente no MySQL (UFBA, 2010?).

3.1.3 Criação do Data Mart e extração dos dados gerenciais

O banco de dados operacional da empresa e o Data Mart onde os metadados da **Empresa Teste** estão armazenados, foram desenvolvidos em MySQL. Foi escolhido o MySQL por possuir licença livre para estudo e também pela facilidade de se conectar com os componentes da suíte do *Pentaho*.

Utilizando o conceito de ETL, através do Kettle, foi possível extrair os dados gerenciais da empresa, fazer a transformação e finalmente populá-los nas tabelas dimensões e fato do Data Mart.

As dimensões: *DimCliente* e *DimProduto*, simplesmente receberam os metadados de descrições e algumas características dos itens como: nome, endereço, bairro e cidade, tanto ao clientes, e nome e grupo quanto aos produtos. Para carregar os metadados na dimensão *DimPeriodo* (referente a hierarquia de tempo: ano, mês e dia) e a tabela

FatoVenda, foi necessário um estudo mais complexo de como armazenar os metadados na dimensão tempo e nas métricas da tabela *FatoVenda* (PEIXOTO, 2011).

3.1.4 Desenvolvimento do cubo multidimensional

Após a extração dos dados para o Data Mart, foi necessário criar o *cubo multidimensional*. Nessa etapa as relações entre as dimensões, fatos, metadados, hierarquias e métricas foram elaboradas. Após essa etapa concretizada foi possível publicar o cubo no *Pentaho BI Server* (UFBA, 2010?).

3.1.5 Desenvolvimento e apresentação dos relatórios multidimensionais

Após a publicação do cubo no *Pentaho BI Server*, os relatórios foram desenvolvidos através do *JPivot*, elemento utilizado pelo servidor para geração de relatórios multidimensionais, tornando possível a sua navegação e exploração.

3.2 Ferramentas utilizadas

A seguir serão apresentadas as ferramentas e metodologias que foram utilizadas no desenvolvimento do projeto. Estão divididas em sistema de gerenciamento de banco de dados e suíte de aplicações BI e seus componentes.

3.2.1 Mysql

Mysql é um sistema de gerenciamento de banco de dados relacionais que utiliza a linguagem SQL (*Structure Query Language*) e possui licença *open source* (software livre) e é um dos bancos de dados mais poderosos e utilizados por milhares de profissionais ao redor do mundo. MySQL AB Limited Company é a empresa que desenvolve e disponibiliza os arquivos de instalação para download (NEVES, RUAS, 2005).

3.2.2 A Pentaho

A Pentaho Corporation é uma empresa formada por antigos integrantes de outras organizações de inteligência de negócios como Cognos, Oracle, SAS, entre outros, que nasceu em 2004 em Orlando, Flórida, Estados Unidos, mas que hoje já se encontra

espalhada por todo planeta. A própria Pentaho Corporation desenvolve suas ferramentas de BI (SOUZA, 2010).

A suíte de ferramentas da Pentaho possui 2 versões: *Enterprise Edition* e *Community Edition*. A versão *Enterprise Edition* é usado por empresas mais robustas, onde as mesmas recebem suporte através do pagamento de uma parcela anual. A versão *Community Edition* possui uma licença *open source* (software livre), ou seja, é possível fazer o download das ferramentas e utilizá-las normalmente, porém não possui suporte da Pentaho. Atualmente existem vários grupos de discussões na internet, onde profissionais e aprendizes debatem questões importantes sobre a plataforma.

Os principais projetos que integram a suíte da *Pentaho* são: Mondrian: o motor OLAP, Kettle: também conhecido como PDI (*Pentaho Data Integration* – Integração de Dados da *Pentaho*) que executa as cargas e transformações com ETL, parte mais importante do projeto, Weka: que é responsável pela parte de Mineração de Dados, JPivot: componente de interface gráfica para visualização de cubos multidimensionais e a ferramenta de relatórios: JFreeReporting (SOUZA, 2010).

3.2.3 Pentaho BI Server

Pentaho BI Server é o servidor principal da suíte, responsável pela apresentação dos resultados obtidos durante todo o desenvolvimento e também análises desses resultados. Para acessá-lo, foi necessário a instalação do pacote de desenvolvimento Java (JDK), onde as instancias das variáveis do TomCat são inicializadas. O servidor roda

através de qualquer navegador de internet, como Internet Explorer, Mozilla Firefox, etc., através do endereço: localhost:8080 (PENTAHO,2012).

3.2.4 Mondrian

Mondrian é conhecido como motor ou servidor OLAP, desenvolvido em Java e incorporado ao Pentaho BI Server. O Mondrian executa consultas na linguagem MDX e converte para SQL, fazendo leitura das bases de dados relacionais e gerando dados multidimensionais (PENTAHO, 2012).

3.2.5 Kettle

Conhecido também como PDI (Pentaho Data Integration), Kettle é o componente gráfico da suíte *Pentaho* que utiliza as técnicas de ETL, compatível com vários SGBD do mercado (MySQL, SQL Server, Oracle, etc.). O *Kettle* executa leituras de banco de dados relacionais e carrega os dados para o DW ou DM e podendo também exportar e importar para arquivos Excel, CSV e banco de dados ODBC (baseado em Windows). Como dito anteriormente, o trabalho é feito totalmente de forma gráfica (PENTAHO, 2012).

“O Kettle pode integrar dados entre empresas e sistemas, substituindo a criação de camadas de programas para integração por operações visuais” (BORGES, p. 33, 2010).

3.2.6 Pentaho Schema Workbench (PSW)

O PSW é usado para criar os cubos multidimensionais e gerá-los em formato XML, utilizando a linguagem MDX (*Multidimensional Expressions* - Expressões Multidimensionais). Através dessa ferramenta é possível criar as dimensões, hierarquias e métricas, e navegar no cubo de forma visual (BORGES, 2010).

3.2.7 JPivot

A ferramenta utilizada para estabelecer a comunicação entre o usuário e o servidor OLAP Mondrian, é o visualizador OLAP JPivot. Essa ferramenta permite ao usuário apresentar as informações em diversos tipos de gráficos, podendo também gerá-los em Excel (SOUZA, 2010).

“É através dele que são feitas as consultas no cubo e também onde o cubo é visualizado” (BORGES, p. 31,2010).

3.2.8 SQL Power Architect

SQL Power-Architect é uma ferramenta *open source* (software livre) utilizada para construir o modelo multidimensional e gerar o script SQL para ser executado no banco de dados criando as tabelas, campos e chaves correspondentes. Através dela

também foi elaborado o MER (Modelo Entidade Relacionamento) do DM (UFBA, 2010?).

4 DESENVOLVIMENTO DO PROJETO

A **Empresa Teste** possui atualmente um sistema gerencial que faz o controle de toda a parte de compras e vendas dos produtos. Para melhor entendimento dos negócios, foi necessário conhecer algumas das regras de negócios do ambiente empresarial e também entender o ciclo de vendas da empresa.

Durante a pesquisa, foi observada a necessidade do gerente em obter relatórios mais detalhados e com maior expansão, onde as informações sejam apresentadas de forma simples de ser visualizada, onde ele possa tomar decisões mais eficientes.

Após analisar e discutir a situação atual foi decidido que a aplicação das metodologias BI usando a ferramenta *Pentaho* tornará a tomada de decisão mais inteligente.

4.1 Modelagem do Data Mart Multidimensional

A imagem a seguir mostra o Modelo Entidade Relacionamento do Data Mart onde os metadados de decisão foram carregados, criado no SQL Power Architect:

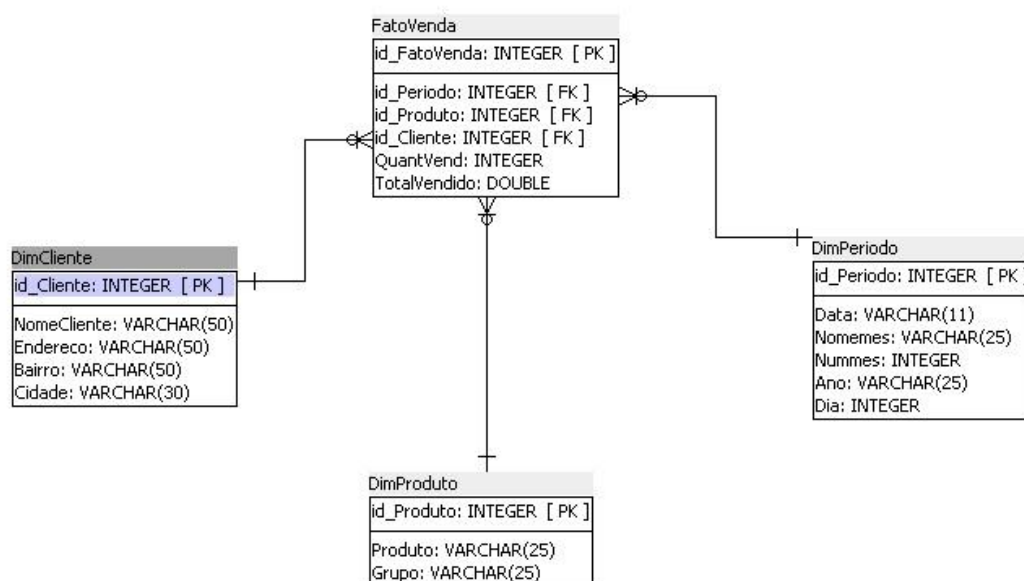


Figura 9 – Modelo do Data Mart (Fonte: Próprio Autor).

4.2 Dicionário de dados

A seguir estão descritas a fato, dimensões, campos e métricas do Data Mart:

Nome da Tabela: DimProduto			
Definição: Tabela dimensão que armazena os dados dos produtos.			
Nome do Atributo	Significado	Tipo de dado	Descrição
id_Produto	Código do produto	Integer	Chave primária (PK) da tabela e auto-incremento e chave externa que referencia-se com a tabela FatoVenda
Produto	Descrição do Produto	Varchar(100)	Armazena o nome do produto.
Grupo	Descrição do Grupo	Varchar(25)	Armazena o nome do grupo do produto.

Tabela 4 – Dimensão Produto (Fonte: Próprio Autor).

Nome da Tabela: DimCliente			
Definição: Tabela dimensão que armazena os dados dos clientes.			
Nome do Atributo	Significado	Tipo de dado	Descrição
id_Cliente	Código do Produto	Integer	Chave primária (PK) da tabela e auto-incremento e chave externa que referencia-se com a tabela FatoVenda
NomeCliente	Nome do Cliente	Varchar(50)	Armazena o nome do cliente.
Endereco	Descrição do Grupo	Varchar(50)	Armazena o endereço do cliente.
Bairro	Descrição do Bairro	Varchar(50)	Armazena o bairro do cliente.
Cidade	Descrição do Cidade	Varchar(30)	Armazena a cidade do cliente.

Tabela 5 – Dimensão Cliente (Fonte: Próprio Autor).

Nome da Tabela: DimPeriodo			
Definição: Tabela dimensão que armazena os dados dos clientes.			
Nome do Atributo	Significado	Tipo de dado	Descrição
id_Periodo	Código do Período	Integer	Chave primária (PK) da tabela e auto-incremento que referencia-se com a tabela FatoVenda
Data	Data	Varchar(11)	Armazena a data.
Ano	Ano	Varchar(4)	Armazena o ano.
Nomes	Nome do Mês	Varchar(25)	Armazena o nome do mês.
Nummes	Número do Mês	Integer	Armazena o número do mês.
Dia	Dia	Integer	Armazena o dia.

Tabela 6 – Dimensão Período (Fonte: Próprio Autor).

Nome da Tabela: FatoVenda			
Definição: Tabela dimensão que armazena as métricas dos cálculos.			
Nome do Atributo	Significado	Tipo de dado	Descrição
id_FatoVenda	Código da FatoVenda	Integer	Chave primária (PK) da tabela e auto-incremento.
id_Produto	Código do Produto	Integer	Chave estrangeira (FK) que referencia-se com a dimensão DimProduto.
id_Cliente	Código do Cliente	Integer	Chave estrangeira (FK) que referencia-se com a dimensão DimCliente.
id_Periodo	Código do Período	Integer	Chave estrangeira (FK) que referencia-se com a dimensão DimPeriodo.
QuantVend	Métrica Quantidade Vendida	Integer	Calcula a quantidade vendida.
TotalVend	Métrica Total Vendido	Double	Calcula o total R\$ vendido.

Tabela 7 – Fato Venda (Fonte: Próprio Autor).

4.3 Modelagem do cubo multidimensional

A figura 10 está apresentando a modelagem do cubo multidimensional do projeto, desenvolvido através do Pentaho Schema Workbench.

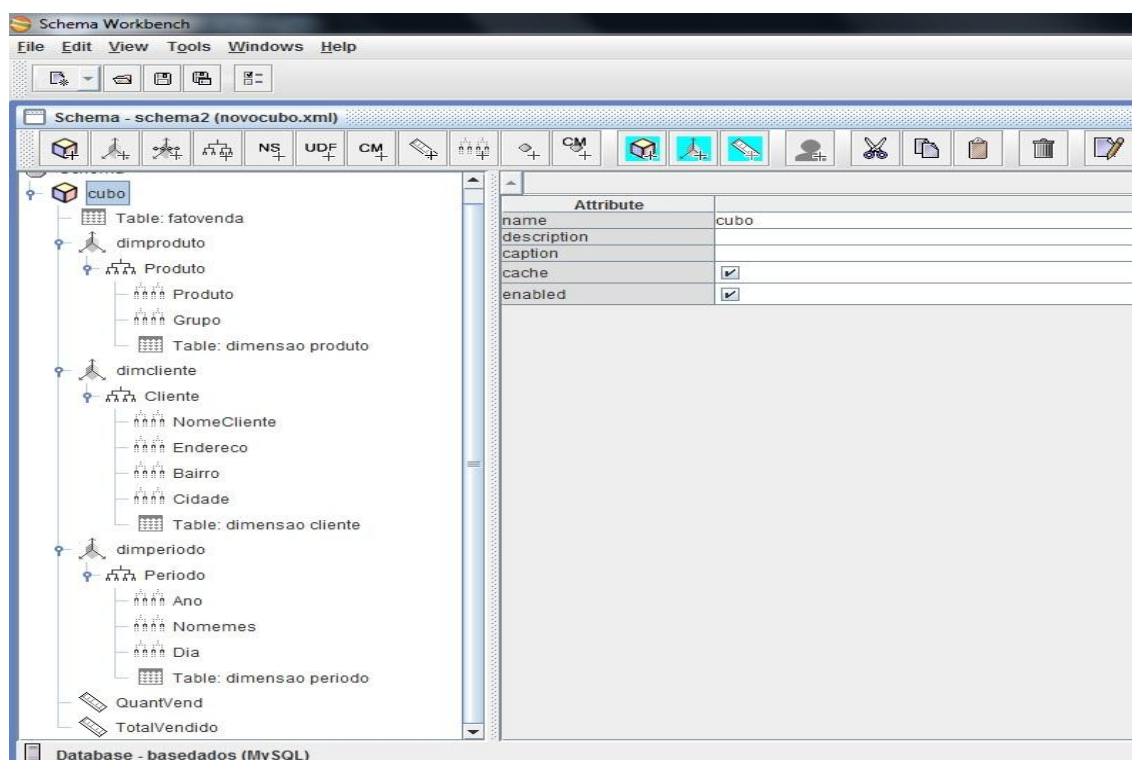


Figura 10 – Modelagem do cubo multidimensional no Pentaho Schema Workbench (Fonte: Próprio Autor).

4.4 Transformações ETL com Kettle

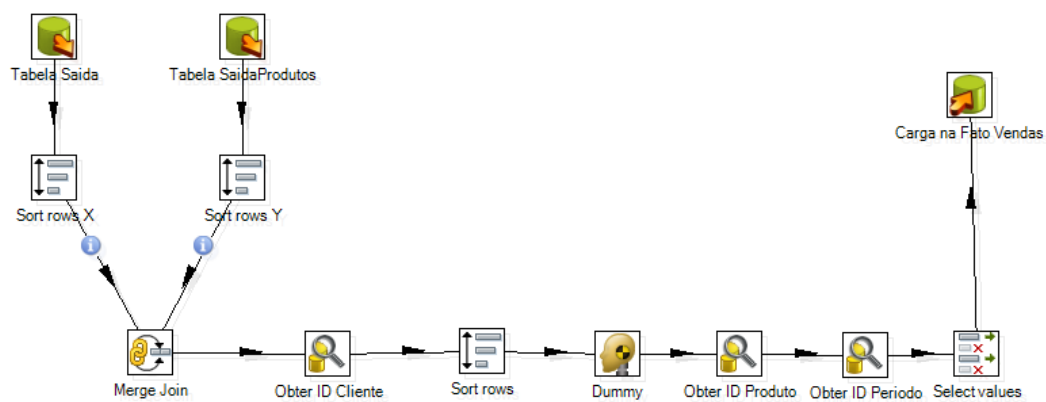


Figura 11 – Transformação para Fato Vendas (Fonte: Próprio Autor).

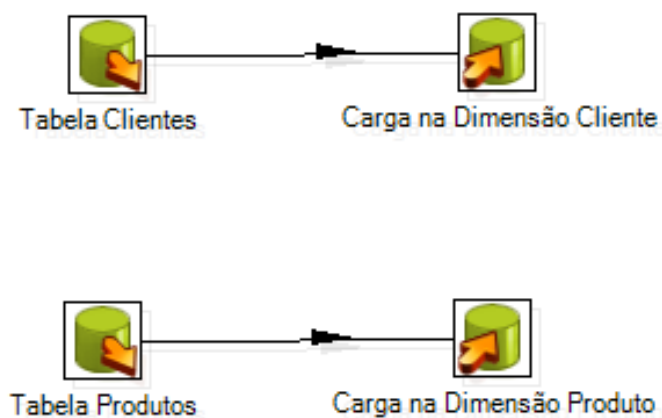


Figura 12 – Transformações para Dimensões Cliente e Produto (Fonte: Próprio Autor).

4.5 Visualização do cubo no JPivot

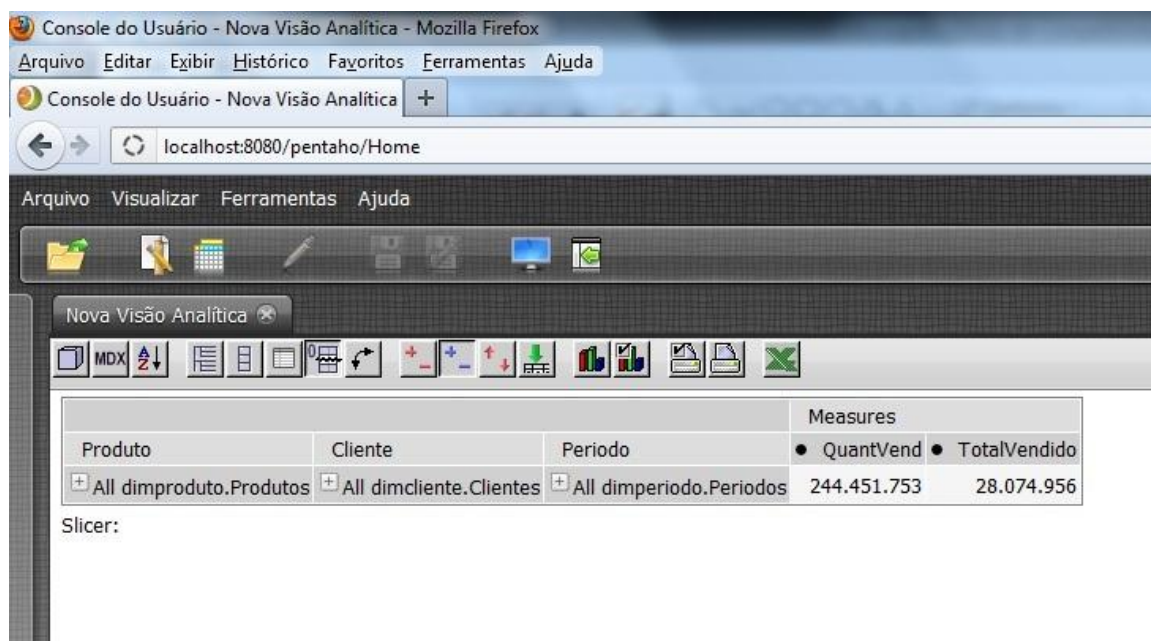


Figura 13 – Tela inicial Mondrian - Cliente JPivot - Cubo OLAP (Fonte: Próprio Autor).

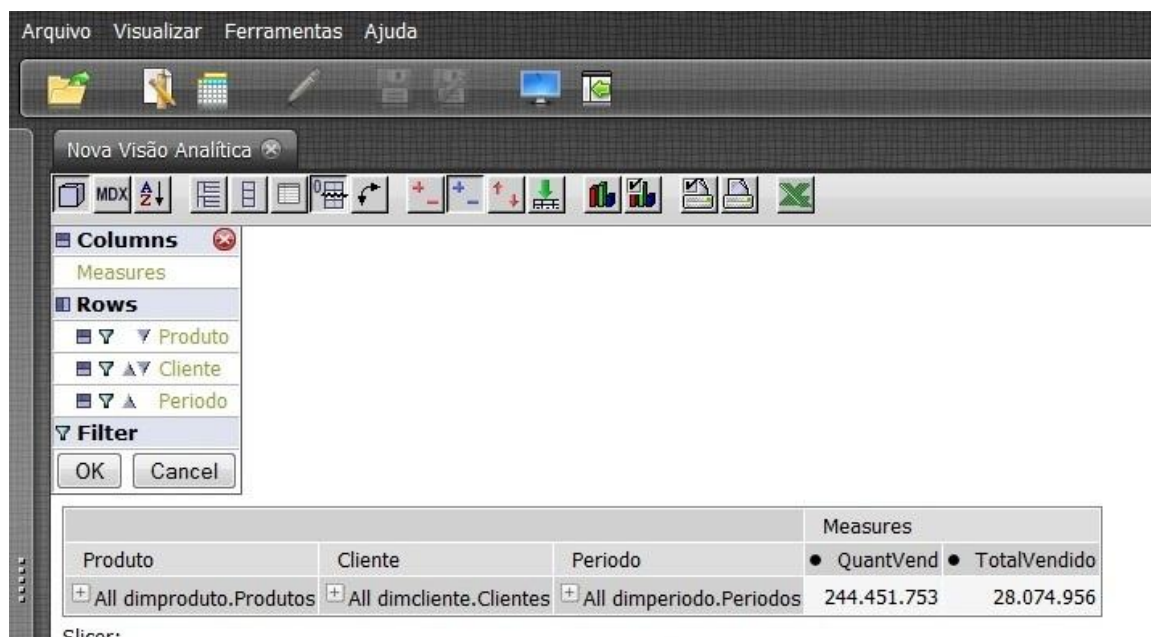


Figura 14 – Botões de navegação JPivot (Fonte: Próprio Autor).

Arquivo Visualizar Ferramentas Ajuda

Nova Visão Analítica

Produto	Cliente	Período	Measures	
			QuantVend	TotalVendido
All dimproduto.Produtos	All dimcliente.Cientes	All dimperíodo.Períodos	244.451.753	28.074.956
		2011	223.391.730	25.379.008
		Janeiro de 2011	16.450.672	1.668.989
		Fevereiro de 2011	17.795.411	1.730.256
		Março de 2011	18.990.951	2.053.823
		Abril de 2011	17.425.369	1.883.866
		Maio de 2011	17.359.203	1.813.568
		Junho de 2011	17.422.203	1.780.777
		Julho de 2011	18.879.242	2.242.846
		Agosto de 2011	20.386.920	2.371.520
		Setembro de 2011	20.058.341	2.227.877
		Outubro de 2011	20.521.308	2.446.457
		Novembro de 2011	19.629.726	2.474.366
		Dezembro de 2011	18.472.384	2.684.663
		2012	21.060.023	2.695.948

Slicer:

Figura 15 – Drill-down na dimensão período (Fonte: Próprio Autor).

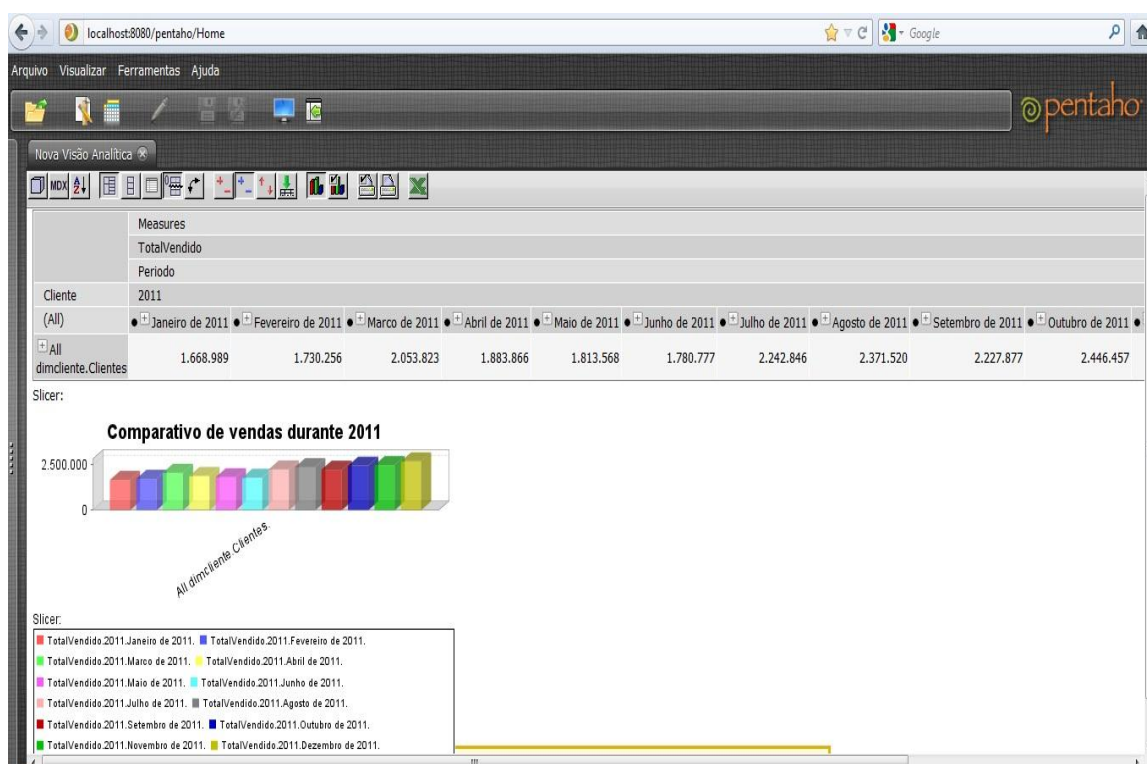


Figura 16 – Gráfico de vendas durante o ano de 2011 (Fonte: Próprio Autor).

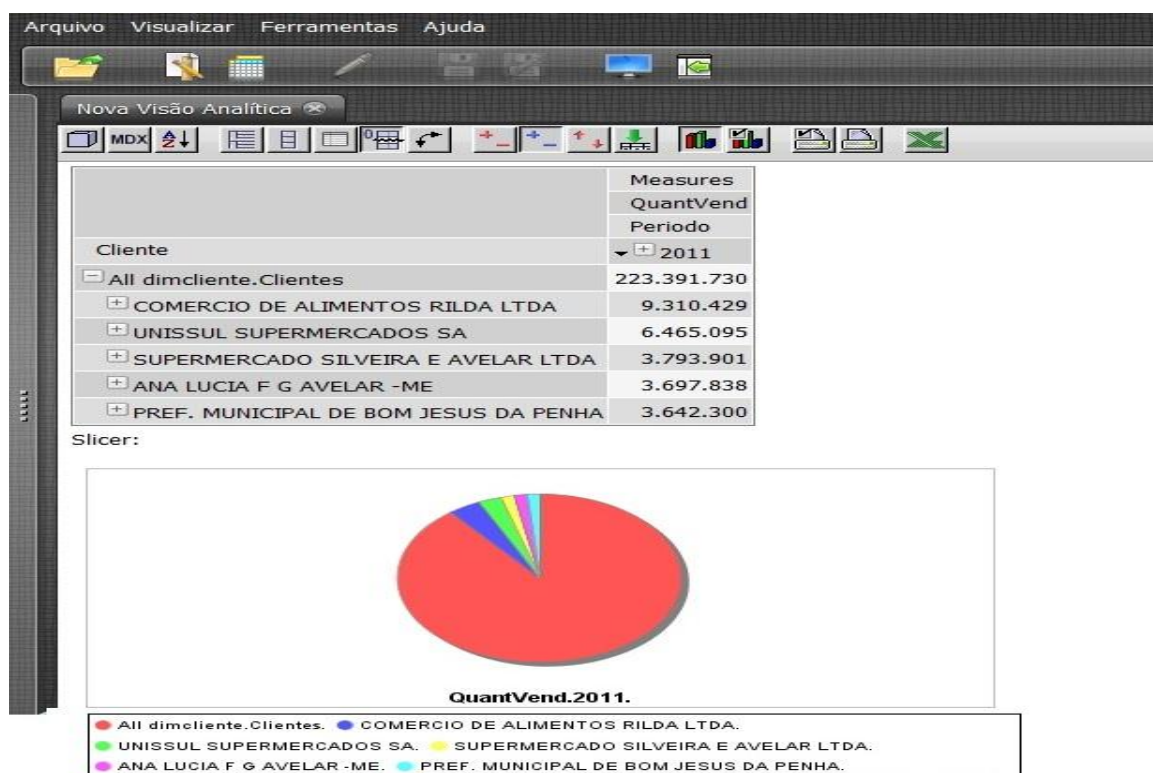


Figura 17 – Gráfico comparativo dos 5 clientes que mais compram com os outros (Fonte: Próprio Autor).

The screenshot shows an OLAP tool interface with a menu bar (Arquivo, Visualizar, Ferramentas, Ajuda) and a toolbar. The main window displays a detailed table of product sales by client and period.

Produto	Cliente	Periodo	QuantVend	TotalVendido
All dimproduto.Produtos	All dimcliente.Cientes	All dimperiodo.Periodos	244.451.753	28.074.956
BALDE PALITO LARA	All dimcliente.Cientes	All dimperiodo.Periodos	6.300	75.600
		2011	5.900	69.300
		Janeiro de 2011	1.900	6.300
		Fevereiro de 2011	1.200	8.400
		Marco de 2011	400	8.400
		Abril de 2011	700	14.700
		Mai de 2011	500	10.500
		Junho de 2011	100	2.100
		Julho de 2011	600	8.400
		Agosto de 2011	300	6.300
		Novembro de 2011	200	4.200
		2012	400	6.300
	ERNANDO FUNC	All dimperiodo.Periodos	100	2.100
	IZALTINO JOSE LTDA	All dimperiodo.Periodos	5.800	65.100
	SUPERMERCADO FONSECA LTDA	All dimperiodo.Periodos	100	2.100
	SUPERMERCADO TONIN	All dimperiodo.Periodos	300	6.300
BEBIDA LACTEA COCO 1 LITRO	All dimcliente.Cientes	All dimperiodo.Periodos	4.351.900	900.075
BEBIDA LACTEA COCO 150 ML	All dimcliente.Cientes	All dimperiodo.Periodos	3.327.400	43.364

Figura 18 – Exploração do cubo OLAP (Fonte: Próprio Autor).

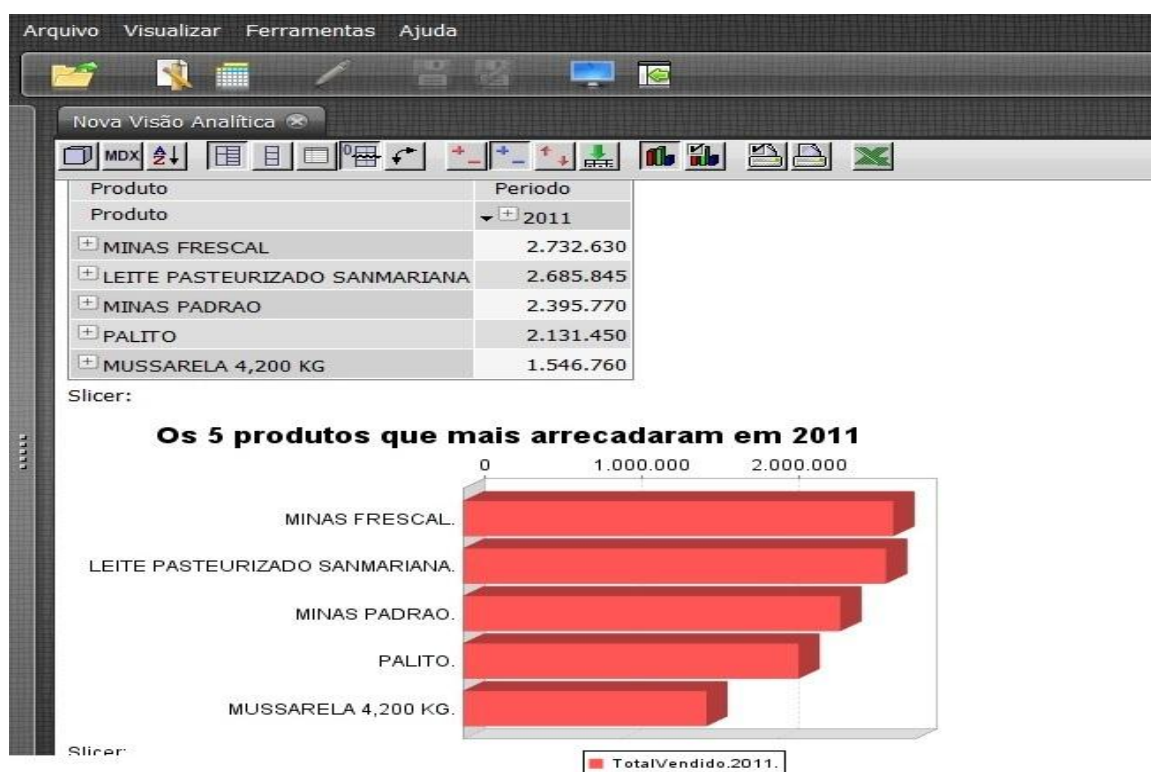


Figura 19 – Gráfico dos 5 produtos que mais arrecadaram em 2011 (Fonte: Próprio Autor).



Figura 20 – Variação da venda de leite em 2011 (Fonte: Próprio Autor).

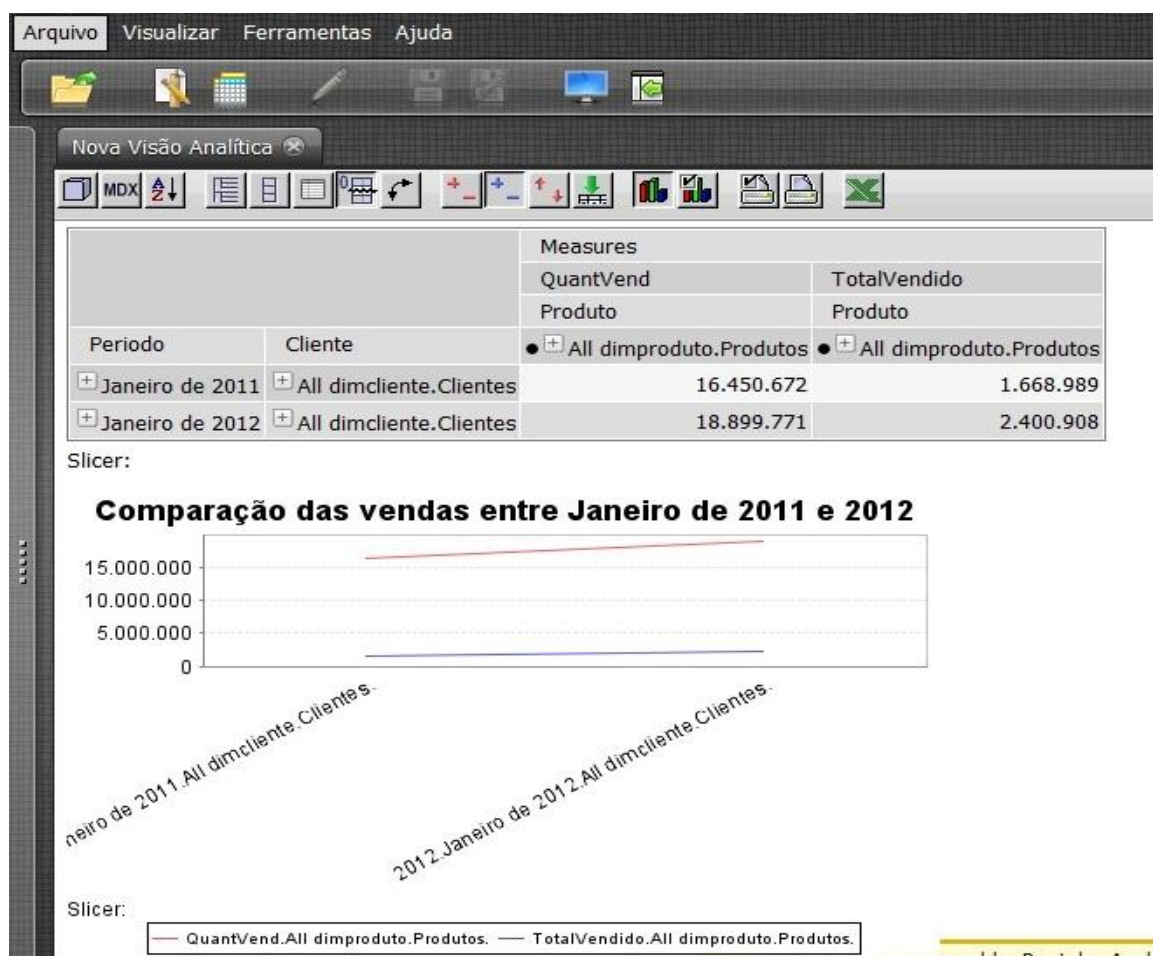


Figura 21 – Comparação das vendas entre Janeiro de 2011 e 2012 (Fonte: Próprio Autor).

5 CONCLUSÃO

Neste projeto realizou-se uma pesquisa sobre as metodologias, técnicas e ferramentas para a criação de um Data Mart. Através da utilização da suíte *Pentaho Community Edition* foi possível criar o Data Mart com o intuito de auxiliar a cúpula administrativa a tomar decisões mais inteligentes.

Foi realizado um estudo das necessidades de uma distribuidora de produtos alimentícios, denominada no projeto como **Empresa Teste**, com o objetivo de obter informações relevantes para o suporte à decisão. Devido a falta de recursos financeiros, utilizou-se de uma ferramenta *open source* (software livre) durante o processo de criação do cubo de dados, ETL e visualização do cubo de dados carregados no Data Mart. Estas ferramentas possuem um nível de integração entre elas satisfatório, facilitando seu uso.

A fonte de dados disponibilizada para o estudo é referente às vendas da empresa, pois de acordo com o seu gerente, era a que mais necessitava de informações de natureza decisória. As consultas no cubo multidimensional realizadas no servidor OLAP foram satisfatórias, pois atendem as necessidades da empresa. Com todas essas informações obtidas, foi possível tomar decisões levando em conta os fatos ocorridos.

O desenvolvimento e aplicação do projeto foi realizado no período de 6 meses, sendo que as maiores dificuldades foram o curto prazo para desenvolvimento, implantação e análise dos resultados do projeto e pesquisas para conhecimento para manipular as ferramentas do *Pentaho*.

Futuramente, o projeto poderá ser expandido, onde novas dimensões podem ser inseridas, fazendo com que a capacidade analítica e o poder de tomar decisões sejam ainda mais inteligentes.

REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, Fábio S. G. G. e Abreu, *Desmitificando o conceito de ETL*, 2008. Disponível em: <http://www.fsma.edu.br/si/Artigos/V2_Artigo1.pdf>. Acesso em: 01 de Maio de 2011.

ANZANELLO, Cynthia Aurora. *OLAP Conceitos e utilização*, [s.d.]. Disponível em: <<http://www.fag.edu.br/professores/limanzke/Administra%E7%E3o%20de%20Sistemas%20de%20Informa%E7%E3o/OLAP.pdf>>. Acesso em: 29 de Maio de 2011.

BEUREN, Ilse Maria; MARTINS, Luciano Waltrick. *Sistemas de Informações Executivas: Suas características e reflexões sobre sua aplicação no processo de gestão*, 2001. Disponível em: <<http://www.scielo.br/pdf/rcf/v12n26/v12n26a01.pdf>>. Acesso em: 20 de Maio de 2011.

BORGES, Cairo da Silva. *Criação de um ambiente de exploração OLAP para analisar dados das vendas do grupo de postos de combustíveis Pioneiro*. Trabalho de conclusão de curso de Sistemas de Informação, Faculdade de Balsas. Balsas, 2010.

BROWN, 2004. In: TURBAN, Efraim; SHARDA, Ramesh; ARONSON, Jay; KING, David. ***Business Intelligence – Um enfoque gerencial para a inteligência do negócio***. Editora Bookman, 2009.

COSTA, Clayton M. ***Uma generalização do processo ETL em Sistemas Data Warehouse***, [s.d.]. Disponível em: <<http://www.lia.ufc.br/~claytonmaciel/dw.pdf>>. Acesso em: 22 de Maio de 2011.

DALFOVO, Oscar; PERFEITO, Juarez; AZAMBUJA, Ricardo Alencar de; DIAS, Paulo Roberto. ***A ferramenta OLAP para auxílio à tomada de decisão em um Data Warehouse***, [s.d.]. Disponível em: <http://www.professorgersonborges.com.br/site/pdf/apostila/BI/ferramentas_olap1.pdf>. Acesso em: 15 de Abril de 2012.

ECKERSON, W. ***Performance Dashboards***. Editora: Hoboken, NJ: Wiley, 2003. In: TURBAN, Efraim; SHARDA, Ramesh; ARONSON, Jay; KING, David. ***Business Intelligence – Um enfoque gerencial para a inteligência do negócio***. Editora Bookman, 2009.

JUNIOR, Methanias Colaço – ***Projetando sistemas de apoio à decisão baseados em Data Warehouse***. Editora: Axcel Books do Brasil Editora, 2004.

LIMA¹, Álvaro Vieira; LIMA², Davi Marques. *Business Intelligence como ferramenta gerencial no suporte ao processo de business performance management*, 2011. Disponível em: <publicacoes.uniceub.br/index.php/gti/article/download/1201/1071.pdf>. Acesso em: 27 de Abril de 2012.

MACHADO, Felipe Nery Rodrigues – *Tecnologia e projeto de Data Warehouse*. Editora: Érica, 2000.

MAGALHÃES, Keli Daiana Martins; GEDRES, Carlos Alberto Pouey. *Tópicos emergentes do Sistema de Informação Gerencial*, 2008. Disponível em: <<http://www.administradores.com.br/informe-se/artigos/sistema-de-informacao-gerencial/23741/>>. Acesso em: 20 de Fevereiro de 2012.

NEVES, Pedro M. C.; RUAS, Rui P. F. *O guia prático do Mysql*, 2005. Disponível em: <<http://www.centroatl.pt/titulos/tecnologias/imagens/excerto-e-book-ca-oguiapraticodomysql.pdf>>. Acesso em 16 de Março de 2012.

OLIVEIRA, Bruno Gama de; CAVAZZINI, Etti; GONÇALVES, José Carlos, 2009. *Comparativo entre Dado x Informação x Conhecimento*. Disponível em: <<http://gestoresdofuturo2009.blogspot.com/2009/09/comparativo-entre-dado-x-informacao-x.html>>. Acesso em 24 de Maio de 2011.

PEIXOTO, Luciano. *Populando Fatos e dimensões no PDI (Kettle)*, 2011. Disponível em: < <http://kettlebeginners.blogspot.com.br/2011/09/populando-fatos-e-dimensoes-no-pdi.html>>. Acesso em: 10 de Abril de 2012.

PENTAH0, 2012. Disponível em: <<http://www.pentaho.org>>. Acesso em: 29 de Março de 2012.

PEREIRA, Otacílio José. *Tópicos Avançados – Business Intelligence*, [s.d.]. Disponível em: <<http://www.ceunes.ufes.br/downloads/2/otaciliopereira-EC.BD%20-%20Unidade%2011%20-%20TopAv%20-%20Business%20Intelligence%20v3.pdf>>. Acesso em: 29 de Março de 2012.

PRIMAK, Fábio Vinícius. *Decisões com BI*. Rio de Janeiro: Ed. Ciência Moderna, 2008.

SILVA, Elton A. *Indicadores de Qualidade para o Gerenciamento de Ambiente de Data Warehouse*. Iniciação Científica – USP. São Paulo, 2012.

SOUZA, Caio Moreno. *Integração de Ferramentas de Código Aberto (Java, Pentaho e Android) e mapas, aplicada a projetos de Inteligência de Negócios*. Trabalho de conclusão de curso de Especialização em Tecnologia Java, Universidade Tecnológica Federal do Paraná. Curitiba, 2010.

TURBAN, Efraim; SHARDA, Ramesh; ARONSON, Jay; KING, David. ***Business Intelligence – Um enfoque gerencial para a inteligência do negócio***. Editora Bookman, 2009.

UFBA, Universidade Federal da Bahia. ***Plataforma Pentaho de Business Intelligence – Manual de Utilização***, [2010?]. Disponível em: <
http://wiki.softwarelivre.org/pub/PentahoBrasil/Documentos/Pentaho_3_5.pdf>. Acesso em: 01 de Março de 2012.