



**CURSO DE SISTEMAS DE INFORMAÇÃO**

**RONALDO ARAÚJO DA SILVA**

**UM ESTUDO DE CASO ENVOLVENDO BUSINESS INTELLIGENCE  
COMO INSTRUMENTO DE APOIO À DECISÃO**

Salvador  
2009

**RONALDO ARAÚJO DA SILVA**

**UM ESTUDO DE CASO ENVOLVENDO BUSINESS INTELLIGENCE  
COMO INSTRUMENTO DE APOIO À DECISÃO**

Trabalho de conclusão de curso apresentado à disciplina Projeto de Diplomação II do 8º semestre do curso de Sistemas de Informação como requisito final para a obtenção do grau em Bacharel em Sistemas de Informação pelo Centro Universitário da Bahia Estácio/FIB.

Orientador: Antônio José dos Santos Santana

Salvador  
2009

**RONALDO ARAÚJO DA SILVA**

**UM ESTUDO DE CASO ENVOLVENDO BUSINESS INTELLIGENCE  
COMO INSTRUMENTO DE APOIO À DECISÃO**

Este Trabalho de Conclusão de Curso foi julgado adequado e aprovado em sua forma final pelos seguintes professores, no dia 10 de dezembro de 2009.

**BANCA AVALIADORA**

---

Profº. Antônio José dos Santos Santana Bacharel em Administração com Habilitação em Análise de Sistema e Especialista em Banco de dados e Informação Estratégica – Faculdade Ruy Barbosa. Professor do Centro Universitário da Bahia - FIB

---

Profº. Antônio Cláudio L. Araújo Bacharel em Sistemas de Informações e Especialista em Sistemas Distribuídos – Universidade Federal da Bahia. Professor do Centro Universitário da Bahia - FIB

Aos meus pais e irmão, obrigado pela paciência e pelas palavras de apoio e confiança, fazendo-me sempre acreditar que este momento seria possível.

A minha namorada, pela atenção e carinho no meu dia-a-dia.

## **AGRADECIMENTOS**

A DEUS, espírito de amor e força. Obrigado a toda minha família em especial meus pais por me darem esta oportunidade e por me proporcionar este grande momento na minha vida e ao professor Antônio Santana.

## Resumo

O BI, também conhecido por Inteligência de Negócios ou Inteligência Empresarial, é o conjunto de tecnologias orientadas a disponibilizar informação e conhecimento. Incluem-se ferramentas como *Customer Relationship Management* (CRM), *Data Mining* (Mineração de dados), *Data Warehouse* (Armazém de dados), entre outras. O BI tinha aplicação limitada em empresas de médio e pequeno porte, principalmente devido ao alto custo das aplicações comerciais. Contudo, com o surgimento de ferramentas que podem ser obtidas gratuitamente (ou a custos reduzidos), as médias e pequenas empresas começam a se beneficiar da utilização dessa tecnologia. Neste trabalho, além de alencarmos os conceitos sobre BI, seus componentes e recursos, abordaremos também sua aplicabilidade em uma empresa de médio porte a “FTC EAD”. Para tanto, será utilizado o Pentaho (ferramenta baseada em licença de software livre) e o SQL Server, Sistema Gerenciador e Banco de Dados da Microsoft.

**PALAVRAS-CHAVES:** BI. *Data Warehouse*. Informação. Conhecimento. Inteligência Empresarial.

## **ABSTRACT**

BI, also known as Business Intelligence and Business Intelligence, is the set of technologies aimed to provide information and knowledge. Include tools such as Customer Relationship Management (CRM), Data Mining (Data Mining), Data Warehouse (Data Warehouse), among others. BI had limited application in medium and small, mainly due to the high cost of commercial applications. However, with the emergence of tools that can be obtained for free (or reduced), the medium and small companies are beginning to benefit from using this technology. In this work, and alencar on BI concepts, their components and resources, we will also its applicability in a midsize company to "FTC EAD. Therefore, it will use the Pentaho (tool based on free software license) and SQL Server, System Manager Database and Microsoft.

**KEY WORDS:** BI. Data Warehouse. Information. Knowledge. Business Intelligence.

## LISTA DE ILUSTRAÇÕES

<i>Figura1 – Evolução do Dado</i>	18
<i>Figura2 - Pirâmide dos Sistemas de Informação</i>	21
<i>Figura3 – Sistema de informação transacional</i>	25
<i>Figura4 – Representação do modelo dimensional</i>	42
<i>Figura5 – Representação do modelo estrela</i>	43
<i>Figura6 – Representação do modelo floco de neves</i>	43
<i>Figura7 – Esquema ETL</i>	47
<i>Figura8 – Diagrama de uma solução de BI</i>	59
<i>Figura9 – Arquitetura do Pentaho BI</i>	61
<i>Figura10 – Diagrama físico do banco de dados do estudo de caso.</i>	63
<i>Figura11 - Arquitetura sistema transacional</i>	64
<i>Figura12 – Diagrama físico do banco de dados para o Data Warehouse</i>	66
<i>Figura13 - Arquitetura do processo de ETL</i>	67
<i>Figura14 - DTS Designer</i>	67
<i>Figura15 – Novo Pacote</i>	69
<i>Figura16 – Propriedades de Conexão</i>	69
<i>Figura17 – Executar Propriedades da Tarefa</i>	70
<i>Figura18 - Esquema de ETL</i>	71
<i>Figura19 - Esquema de ETL completo</i>	71
<i>Figura20 - Conexão com o cubo de dados no Schema Workbench</i>	72
<i>Figura21 – Novo Cubo</i>	73
<i>Figura22 - Nova Medida</i>	74
<i>Figura23 - Nova Dimensão</i>	74

<i>Figura24 - Nova Hierarquia</i>	75
<i>Figura25 – Novo Nível</i>	75
<i>Figura26 - Visualização do cubo criado</i>	76
<i>Figura27 – Navegador OLAP</i>	77
<i>Figura28 – Editor de consulta MDX</i>	77
<i>Figura29 – Configuração da tabela OLAP</i>	78
<i>Figura30 – Mostra membros com parentescos</i>	78
<i>Figura31 – Ocultar informações repetidas</i>	79
<i>Figura32 - Mostrar Propriedades</i>	79
<i>Figura33 – Reprimir linhas e colunas vazias</i>	80
<i>Figura34 – Inversão de eixos</i>	80
<i>Figura35 – Expandir membros</i>	81
<i>Figura36 – Expandir posição</i>	81
<i>Figura37 – Expandir substituindo</i>	82
<i>Figura38 – Expandir com detalhe</i>	82
<i>Figura39 - Visualizar Gráfico</i>	83
<i>Figura40 – Configuração do gráfico</i>	83
<i>Figura41 - Configurar Definições de impressão</i>	84
<i>Figura42 - Imprimir esta página através de PDF</i>	84
<i>Figura43 - Iniciar o Excel</i>	85

## LISTA DE TABELAS

<i>Tabela 1- Dado, informação e conhecimento</i>	<u>15</u>
<i>Tabela 2- Tabela comparativa entre data mart e data warehouse</i>	<u>46</u>

## LISTA DE ABREVIATURAS E SIGLAS

BI	-	Business Intelligence
CRM	-	Customer Relationship Management
DM	-	Data Mart
DOLAP	-	Desktop On Line Analytical Processing
DSS	-	Decision Support Systems
DW	-	Data Warehouse
EDP	-	Electronic Data Processing
EIS	-	Execute Information System
ERP	-	Enterprise Resource Planning
ETC	-	Extração, Transformação e Carga
ETL	-	Extraction Transformation Loading
HOLAP	-	Hybrid On Line Analytical Processing
MDS	-	Management Decision System
MDX	-	Multi-Dimensional Expressions
MOLAP	-	Multidimensional On Line Analytical Processing
OLAP	-	On-line Analytical Processing
OLTP	-	On-line Transaction Processing
PPL	-	Pentaho Public License
ROLAP	-	Relational On Line Analytical Processing
SAD	-	Sistemas de Apoio à Decisão
SGBD	-	Sistema Gerenciador de Banco de Dados
TI	-	Tecnologia da Informação

## Sumário

<b>1 INTRODUÇÃO</b>	<b>13</b>
<b>2 DADOS, INFORMAÇÃO E CONHECIMENTO</b>	<b>15</b>
<b>3 SISTEMAS DE INFORMAÇÃO</b>	<b>20</b>
3.1 SISTEMAS DE PROCESSAMENTO DE TRANSAÇÕES (SPT)	21
3.2 SISTEMAS DE INFORMAÇÕES GERENCIAIS (SIG)	22
3.3 SISTEMAS DE APOIO À DECISÃO (SAD)	23
3.4 SISTEMAS DE INFORMAÇÕES EXECUTIVAS (SIE)	24
3.5 SISTEMAS TRANSACIONAIS (OLTP)	25
<b>4 SISTEMAS DE APOIO DA DECISÃO</b>	<b>27</b>
4.1 A DECISÃO NA ANTIGÜIDADE	27
4.2 A EVOLUÇÃO DA DECISÃO NESTE SÉCULO	28
4.3 CONCEITOS BÁSICOS SOBRE DECISÃO	29
4.4 ETAPAS DO PROCESSO DECISÓRIO	30
4.5 OS FATORES DE INFLUÊNCIA DO PROCESSO DECISÓRIO	31
4.6 A DECISÃO E OS SISTEMAS	33
4.7 A EVOLUÇÃO DOS SISTEMAS DE APOIO À DECISÃO	34
<b>5 DATA WAREHOUSE</b>	<b>36</b>
5.1 ARQUITETURA DE UM DATA WAREHOUSE	37
5.2 DADOS DO DATA WAREHOUSE	38
5.3 MODELAGEM DIMENSIONAL	41
5.4 DATA MART	44
5.5 ETL (ETL – Extraction, Transformation, Loading)	46
5.5.1 EXTRAÇÃO	47
5.5.2 TRANSFORMAÇÃO	50
5.5.3 CARGA	51
5.6 OLAP	52
5.7 LINHAS DO OLAP	53
<b>6 BUSINESS INTELLIGENCE</b>	<b>55</b>
6.1 VISUALIZAÇÃO DE DADOS	58
<b>7 ESTUDO DE CASO</b>	<b>59</b>
7.1 PLATAFORMA PENTAHO DE BUSINESS INTELLIGENCE	59
7.1.2 ARQUITETURA	61

<b>7.2 AMBIENTE OPERACIONAL DA EMPRESA</b>	<b>62</b>
<b>7.3 CRIAÇÃO DO DW (DATA WAREHOUSE)</b>	<b>64</b>
<b>7.4 PROCESSO ETL</b>	<b>67</b>
<b>7.5 CRIAÇÃO DO CUBO</b>	<b>72</b>
<b>7.6 VISUALIZAÇÃO</b>	<b>76</b>
<b>8 CONCLUSÃO</b>	<b>86</b>
<b>REFERÊNCIAS</b>	<b>87</b>
<b>APÊNDICE A</b>	<b>93</b>
<b>APÊNDICE B</b>	<b>96</b>

## 1 INTRODUÇÃO

No mercado competitivo, é fundamental que as empresas tenham grande versatilidade na tomada de suas decisões. Para isso, é necessário que tenham em mãos informações precisas, disponíveis e atualizadas. Em vista a esta necessidade, os executivos interagem com diversos sistemas de informação, que produzem uma quantidade volumosa de dados. Em decorrência o conjunto total de dados existentes nas empresas são produzidos por sistemas diferentes, a tarefa de integração destes dados para produzir informação gerencial é extremamente árdua. Com isso a integração de dados heterogêneos para a produção de informação gerencial, significa um passo decisivo para aumentar a eficiência em gestões e um grande desafio para as empresas.

A informação, contudo, pode não satisfazer às expectativas de seus usuários e se tornar um grande problema em tomada de decisões se não estiver acompanhada de atributos que a tornem realmente útil ao processo decisório, isto é, se não for acurada, tempestiva e adequadamente comunicada àqueles que dela necessitam.

As ferramentas de *Business Intelligence* (BI) são amplamente utilizadas em empresas de grande porte para o desenvolvimento de aplicações para a tomada de decisão e inteligência competitiva. Pequenas e médias empresas também podem se beneficiar deste tipo de ferramenta para a tomada de decisão. Porém, o alto custo desse tipo de aplicação não favorece a sua disseminação em empresas desse porte tornando-se um desafio à implementação de um ambiente de *Business Intelligence* (BI) para as mesmas.

Considerando esse contexto o objetivo geral desse trabalho é expor o melhor aproveitamento das informações de uma empresa para o alcance e sustentação de uma vantagem competitiva. Para isso, esse trabalho apresenta elementos que caracterizam uma configuração informacional que busca atingir e/ou sustentar uma vantagem competitiva empresarial, bem como um direcionamento prático da utilização de *Business Intelligence* (BI) em uma rede de ensino superior privada.

Os objetivos específicos desse trabalho são:

- Mostrar como o *Business Intelligence* (BI) pode contribuir no processo de decisão de uma empresa a partir de um estudo de caso em uma organização de ensino apoiado por uma suíte de softwares de BI.
- Demonstrar o uso de parte da suíte de softwares através de um estudo de caso.

O uso estratégico da informação tem se tornado uma necessidade cada vez mais clara para as empresas. Quando a informação é bem utilizada, ela pode agregar valor ao negócio, além de gerar outros benefícios para a empresa, como reduzir custos ou identificar novos nichos de mercado.

No entanto, nos últimos anos vem ocorrendo uma interessante mudança para as empresas, que não envolve grandes investimentos em infra-estrutura tecnológica. Esta mudança está associada à existência de excelentes softwares livres e confiáveis, que podem ser utilizados de maneira muito profissional, como bancos de dados, suítes de escritório, bem como, programas para a implantação de BI.

Portanto, a importância do *Business Intelligence* (BI) no planejamento estratégico começa a ser sentida a partir do momento em que as empresas adotam uma postura de trabalho mais voltada à gestão da informação. Somente com informação íntegra e confiável é possível criar estratégias que atendam melhor seus clientes e colocar a empresa em um patamar de competitividade mais lucrativo.

Para o presente Trabalho de Conclusão de Curso primeiramente será realizado um levantamento bibliográfico em livros e artigos, em seguida um estudo de caso para mostrar a viabilidade da implantação de um ambiente de *Business Intelligence* (BI) focando na suíte Pentaho.

## 2 DADOS, INFORMAÇÃO E CONHECIMENTO

Segundo PROBST, RAUB e ROMHARDT, (2002, p. 23) “Para que os gestores do conhecimento possam desenvolver uma abordagem integrada há a necessidade de, primeiro, distinguirem entre dados, informações e conhecimento, e segundo, reconhecer as relações entre eles”.

A exemplo da citação de PROBST, RAUB e ROMHARDT, (2002) evidencia-se a importância de se conhecer as diferenças entre dados, informações e conhecimento para que se entenda sobre qualquer tema.

De acordo com Davenport e Prusak (1998), dados, informação e conhecimento estão relacionados entre si, sendo suas diferenças uma questão de grau, conforme apresentado na Tabela 1:

Tabela 1- Dado, informação e conhecimento

<p><b>Dado:</b> conjunto de fatos distintos e objetivos, relativos a eventos;</p> <ul style="list-style-type: none"> <li>- registros estruturados de transações;</li> <li>- descrevem apenas parte daquilo que aconteceu; não fornecem julgamento nem interpretações e nem qualquer base sustentável para a tomada de ação.</li> </ul>	<p><b>Informação:</b> uma mensagem, geralmente na forma de um documento ou uma comunicação audível ou visível;</p> <ul style="list-style-type: none"> <li>- visa a modelar a pessoa que a recebe no sentido de fazer alguma diferença em sua perspectiva ou insight;</li> <li>- dados tornam-se informação quando o seu criador lhes acrescenta significado.</li> </ul>	<p><b>Conhecimento:</b> - uma mistura fluida de experiência condensada, valores, informação contextual e insight experimentado, a qual proporciona uma estrutura para a avaliação e incorporação de novas experiências e informações;</p> <ul style="list-style-type: none"> <li>- nas organizações, ele costuma estar embutido não só em documentos ou repositórios, mas também em rotinas, processos, práticas e normas organizacionais.</li> </ul>
--	---	---

FONTE: Adaptado de DAVENPORT E PRUSAK (1998).

Sobre a definição de dado e informação SETZER (1999) define dado como uma seqüência de símbolos quantificados ou quantificáveis, necessariamente uma entidade matemática e, desta forma, puramente sintática. MIRANDA (1999) fornece uma visão semelhante: dado é o conjunto de registros qualitativos ou quantitativos conhecido que organizado, agrupado, categorizado e padronizado adequadamente transforma-se em informação.

Enquanto informação, MIRANDA (1999) oferece uma visão simplista, sugerindo que informações são dados organizados de modo significativo, sendo subsídio útil à tomada de decisão.

Compreende-se então que a informação, obtida através do tratamento dos dados pelos sistemas informatizados de apoio à decisão, é matéria-prima fundamental para o entendimento, para posterior tomada de decisão e conseqüentemente, para a formulação ou adequação ágil de estratégias assertivas e inovadoras.

MIRANDA (1999) apresenta 12 categorias de informação que suportam a estratégia corporativa:

- Informação estratégica sobre clientes: estudos de tendência de comportamentos de consumo, atendimento de demandas, aferição de qualidade de atendimento, entre outros.
- Informação estratégica sobre concorrentes: preços e prazos praticados, participação no mercado, estrutura gerencial, política de investimentos etc.
- Informação estratégica cultural: acesso da população à educação, aos meios de comunicação e seus hábitos culturais.
- Informação estratégica demográfica: densidade e mobilidade populacional, distribuição da população quanto à idade, sexo, raça, entre outros, concentração de renda etc.
- Informação estratégica ecológica: ações de responsabilidade ambiental, estudos de uso de recursos naturais versus manutenção da natureza.
- Informação estratégica econômica/financeira: conjuntura econômica nacional e internacional, flutuação de câmbio, taxa de juros, poder aquisitivo da população, distribuição de renda per capita, incentivos fiscais, créditos e tributários, atuação de blocos econômicos, entre outros.
- Informação estratégica sobre fornecedores: localização, opções de fontes de fornecimento, condições de transporte, preços, prazos de pagamento, descontos, tendências quanto à formação de parcerias etc.

- Informação estratégica sobre governos/política: diretrizes do poder Executivo dentro do mercado da corporação, regulamentações e desregulamentações, políticas fiscais de exportação e importação, e mais.
- Informação estratégica legal: ações dos poderes Legislativo e Judiciário no que se refere à legislação tributária, fiscal, trabalhista, sindical, entre outros.
- Informação estratégica sindical: capacidade de mobilização, poder de arregimentação, atuação em acordos trabalhistas, integração com outros sindicatos ou outras entidades (por exemplo, partidos políticos), representação parlamentar, entre outros.
- Informação estratégica social: distribuição dos segmentos socioeconômicos, diferenças entre as classes (sistema de valores, nível cultural, poder aquisitivo, estrutura política e ideológica, influência na sociedade), atuação de Organizações Não Governamentais (ONGs), associações de bairro e entidades religiosas.
- Informação estratégica tecnológica: pesquisas realizadas e em andamento, tendências quanto à política de pesquisa e desenvolvimento nacional e internacional (investimentos, entidades patrocinadoras etc.), impactos de mudanças tecnológicas, possibilidades de transferência de tecnologia, acesso a fontes produtoras ou fornecedoras de tecnologia (universidades, centros de pesquisa, redes de inovação tecnológica e outras).

Conceber um ambiente tecnológico que extraia informação dessas 12 categorias é um grande desafio das corporações. Usualmente, as empresas limitam-se à obter e trabalhar dados de seus clientes e, em alguns casos, de seus concorrentes. Mas é muito pouco em comparação com o universo de informações que poderia ser trabalhado e convertido em conhecimento, oferecendo assim uma poderosa vantagem competitiva para as organizações em seu processo de formulação de estratégia.

LEME FILHO (2006), apresenta na Figura1 que demonstra a evolução do dado até a obtenção de vantagem competitiva:



Figura1 – Evolução do Dado  
 FONTE: LEME FILHO (2006)

Com o resumo proposto pela Figura1 acima, dado são todas as fontes utilizadas na criação de um processo de uma organização. As fontes de dados podem estar disponíveis para as empresas a partir de seus próprios sistemas computacionais (sistemas internos, ERPs, CRMs) ou a partir de fontes externas por exemplo: IBGE, Fundação Getúlio Vargas ou AC Nielsen.

Esses dados trabalhados transformam-se em informação, que oferecem às empresas um entendimento sobre sua atuação: perfil de consumo de seus clientes, produtos mais ou menos rentáveis, comparação de preços, prazos, estudos de sazonalidade, entre outros.

As informações são insumos básicos para, juntamente com experiências previamente adquiridas gerar conhecimento. Para PONCHIROLI (2005), conhecimento é informação internalizada pela pesquisa, estudo ou experiência que tem valor para a organização. NONAKA e TAKEUCHI (1997), consideram o conhecimento como um processo humano dinâmico de justificar a crença pessoal com relação à verdade. São dois entendimentos complementares, pois trata-se de uma dimensão poderosa de análise que, democratizada, se potencializa para apoiar as estratégias a se tornarem mais assertivas. Nesse momento, cruzando informações de perfis de consumo de clientes e características de participação e

rentabilidade de produtos, é possível traçar alvos específicos no mercado. Vendas cruzadas podem ser estimuladas a partir das informações de necessidades de clientes, aproveitando assim oportunidades em primeira-mão. E compreendendo a atuação da concorrência, pode-se mitigar com maior velocidades os efeitos das ameaças.

Tem-se a impressão de que, de acordo com a análise previamente feita, o conhecimento derivado da informação e dos dados, é um bem precioso para as corporações. De fato, há a máxima de que 'informação é poder'. PONCHIROLI (2005) cita:

(...) estamos no limiar de uma nova era, na qual o conhecimento é reconhecido como o principal ativo das organizações e a chave para uma vantagem competitiva sustentável. Até então fortemente caracterizada pelos bens tangíveis, como o capital financeiro e as estruturas físicas, a fonte de riqueza e competitividade passa a ser, agora, o próprio conhecimento. Sociedade do conhecimento, era do conhecimento, era do capital intelectual, sociedade pós-capitalista são algumas denominações para esta nova época.

No entanto, o autor adverte que acréscimo de informação tende a confundir em vez de esclarecer.

A quantidade e disponibilidade das informações crescem numa progressão exponencial, confundindo as pessoas e dificultando, sobretudo, a gestão do conhecimento. Sabe-se que tanto a escassez quanto o excesso podem ser prejudiciais: perder-se em uma quantidade incontrolável de informações é tão nocivo como não possuí-las. Temos de aprender a jogar fora, em vez de adorar e acumular informações. Temos de adotar a máxima "quanto menos melhor".

Sabe-se então que, buscar e tratar as informações certas (adequadas aos usuários e ao momento) é fator crítico de sucesso para gerar conhecimento e vantagem competitiva.

### 3 SISTEMAS DE INFORMAÇÃO

Um sistema de informação consiste em uma soma estruturada de elementos. Mosimann, Alves e Fisch (1993, p.52) conceituam sistema de informação como “uma rede de informações cujos fluxos alimentam o processo de tomada de decisões, não apenas da empresa como um todo, mas, também, de cada área de responsabilidade.”

Laudon e Laudon (1999), Sistemas de informação são muito mais amplos em seu escopo. Eles abrangem tecnologias, os procedimentos organizacionais, as práticas e as políticas que geram informação, assim como as pessoas que trabalham com essa informação.

Campos Filho (1994, p.34) separa o sistema de informação em quatro componentes reunidos de modo a permitir o melhor atendimento aos objetivos da organização. A saber: informação (dados formatados, imagens, sons e textos livres); os recursos humanos (que coletam, armazenam, recuperam, processam, disseminam e utilizam as informações); as tecnologias de informações (o hardware e o software); e as práticas de trabalho (métodos utilizados).

Um sistema de informação (SI) pode ser definido como um conjunto de componentes inter-relacionados trabalhando juntos para coletar, recuperar, processar, armazenar e distribuir informação com a finalidade de facilitar o planejamento, o controle, a coordenação, a análise e o processo decisório em empresas e outras organizações Laudon e Laudon (1999).

As práticas de trabalho, segundo Alter (1992, p.07), são as formas, maneiras e métodos utilizados pelas pessoas da organização para realizar os seus serviços. Elas estão intimamente ligadas às ferramentas tecnológicas e às informações utilizadas para comunicar, tomar decisão e outras tarefas necessárias ao negócio.

As tecnologias de informações, de acordo com Alter (1992, p.09), compreendem o hardware e software destinados a tarefa de processamento de dados e que capturam, manipulam, armazenam, recuperam e transmitem as

informações. A tecnologia da informação é um componente do sistema de informação e não pode ser confundido como se este elemento fosse o todo. A tecnologia de informação é o ferramental do qual se utiliza a prática de trabalho. Por sua vez, os recursos humanos, conforme Campos Filho (1994, p.35), estão incluídos no sistema de informação para, a não ser em casos totalmente automatizados, coletar, processar ou usar dados. A vinculação entre os recursos humanos e as práticas de trabalho demonstra que elas afetam os recursos humanos, ao passo que as características dos recursos humanos no sistema determinam quais práticas serão viáveis e adequadas.

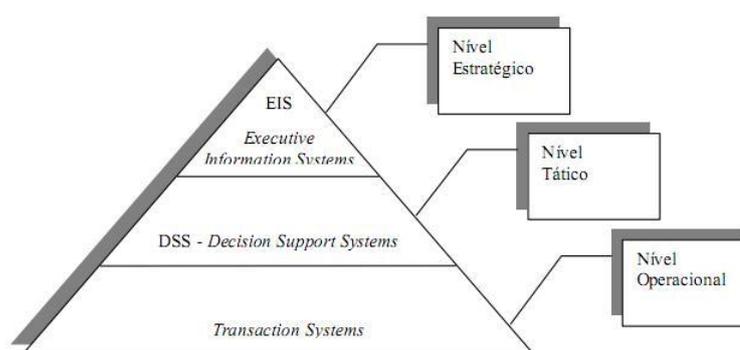


Figura2 - Pirâmide dos Sistemas de Informação

FONTE: DAMIANI (1996)

Na base da pirâmide como mostra a Figura2 estão concentrados os sistemas de processamento de transações (SPT). O segundo estágio compreende os sistemas gerenciais e divide-se em dois módulos: o SAD e o SIG e no topo da pirâmide está o sistema de informação executivas (SIE).

### 3.1 SISTEMAS DE PROCESSAMENTO DE TRANSAÇÕES (SPT)

O sistema de processamento de transações, também denominado sistema eletrônico de processamento de dados, é definido por Mcleod (1993, p.390) como a “manipulação ou transformação de símbolos tais como números e letras para o propósito de aumentar sua utilidade.”

De acordo com Alter (1992, p.127), “um sistema de processamento de transações coleta e armazena dados sobre transações e algumas vezes controla decisões que são feitas como parte de uma transação.” Uma transação é qualquer

troca relacionada com negócios, como pagamento a empregados, vendas a clientes e pagamento a fornecedores.

Na perspectiva administrativa, um SPT é destinado a desempenhar um papel de suporte às atividades da organização empresarial. De acordo com Stair (1998, p. 183), o SPT é usado para dar “suporte às atividades do pessoal não-gerencial e pelos níveis da administração operacional da organização.”

Um SPT, bem desenvolvido e implantado, torna-se uma fonte de dados valiosa, como entrada aos outros sistemas de informação. O SPT é o alicerce que sustenta a integridade e precisão da informação gerada, assegurando a confiabilidade dos sistemas de informação hierarquicamente acima dele.

### 3.2 SISTEMAS DE INFORMAÇÕES GERENCIAIS (SIG)

Sistema de informações gerenciais é descrito por Mcleod (1993, p.427) como “um sistema baseado em computador que faz avaliações das informações para usuários com necessidades similares.” Afirma, ainda, que as informações são utilizadas por administradores e não administradores para tomadas de decisão e para resolver problemas.

Para Stair (1998, p.38), “um sistema de informações gerenciais (SIG) é um agrupamento organizado de pessoas, procedimentos, banco de dados e dispositivos usados para oferecer informações de rotina aos administradores e tomadores de decisões.”

Como o foco é a saída das informações para o gerente, a geração dos relatórios em vez de mostrar as transações ocorridas, passa a apontar, por meio de resumos e filtragens, indicadores-chaves para o monitoramento e análise das informações. Dessa forma, entende-se que o sistema de informações gerenciais proporciona ao gestor condições de controlar as diversas rotinas administrativas e é fonte orientadora para o processo decisório.

### 3.3 SISTEMAS DE APOIO À DECISÃO (SAD)

Ter informações em mãos é uma poderosa ferramenta para quem precisa tomar decisões e, tendo em vista esse princípio, as empresas começaram a extrair dados dos seus sistemas operacionais e armazená-los, separados dos dados operacionais, donde surgiram os SAD.

Shim et al. (2002) dizem que os SAD são soluções computacionais desenvolvidas para apoiar a tomada de decisões complexas durante a resolução de problemas. Ferramentas clássicas de SAD compreendem componentes para gerenciamento de sofisticados bancos de dados, poderosas funções de modelagem e poderosos, embora simples, projetos de interface com o usuário, que permitem trabalhar interativamente com questões, relatórios e funções gráficas.

Turban (1995) define mais especificamente como “um interativo, flexível e adaptável sistema de informação, especialmente desenvolvido para apoiar a solução de um problema gerencial não estruturado para aperfeiçoar a tomada de decisão. Utiliza dados, provê uma interface amigável e permite ao tomador de decisão ter sua própria percepção”.

O sistema de apoio à decisão (SAD), é conceituado por Stair (1998, p.38) como “um grupo organizado de pessoas, procedimentos, banco de dados, e dispositivos usados para dar apoio à tomada de decisões referentes a problemas específicos”. Para Sprague (1991, p. 10), o SAD caracteriza-se como “um sistema computacional interativo que ajuda os responsáveis pela tomada de decisões a utilizar dados e modelos para resolver problemas não estruturados.” STAIR (1998, p.232) corrobora esta idéia e acrescenta que o foco de um SAD “está na eficácia da tomada de decisões em face de problemas comerciais não estruturados ou semi-estruturados.”

### 3.4 SISTEMAS DE INFORMAÇÕES EXECUTIVAS (SIE)

Foi mantido o termo SIE na sua forma original, mesmo que alguns autores o denominem como *Enterprise Information System*, em virtude de novas abordagens, que modificam substancialmente as bases que fundamentam o termo original.

Na literatura encontram-se diversas definições de SIE, as quais convergem para um tipo de sistema de informações que fornece suporte ao processo decisório para o alto escalão da organização.

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

Os executivos das empresas necessitam de informações para a tomada de decisão, premissa básica para o seu sucesso e das organizações para isto utilizam-se de dados transformados, que permite aos mesmos tomarem decisões assertivas.

Mcleod (1993, p. 586) ressalta que “um sistema de informações executivas é um sistema que provê informações para o executivo do desempenho global da firma.” Considera que o fornecimento destas informações ao executivo pode ser facilmente recuperado e pode ter vários níveis de detalhe. Assim, pode-se partir de um nível superficial de análise e prosseguir até um nível de detalhamento mais profundo e desejado.

Segundo Furlan (1994) o termo *Executive Information System* (SIE), foi criado no final da década de 1970, a partir dos trabalhos desenvolvidos no Massachusetts Institute of Technology (MIT) por pesquisadores como Rochart e Treacy.

A tendência da próxima fase do SIE, direcionando-o para todos dentro da organização, faz com que o limite que existe entre o SIE e o SAD (Sistema de Apoio à Decisão) seja cada vez mais estreito.

### 3.5 SISTEMAS TRANSACIONAIS (OLTP)

O processo inicial de informatização de qualquer organização é baseado fundamentalmente no desenvolvimento e na implantação de SI transacionais (também chamados de operacionais). Esses SI são também identificados pela expressão "*Electronic Data Processing*" (EDPs), e eles são necessários para o controle operacional das organizações TOM 1991.

No modelo da evolução da informática nas organizações proposto por Richard Nolan, SI transacionais se enquadram nos estágios de iniciação e contágio. São sistemas operacionais, não integrados, atendem em geral à área administrativo-financeira, controlam, na maioria das vezes, o fluxo de informações financeiras, e os usuários finais esboçam uma certa resistência a utilizá-los.

Os sistemas de folha de pagamento, contabilidade, controle de estoques, contas a pagar e a receber, faturamento, etc., são exemplos de SI transacionais.

Também conhecido como Sistema de Processamento de Transações. Estes sistemas, utilizados atualmente na maioria das organizações, monitoram, coletam, armazenam, processam e distribuem os dados das diversas transações realizadas dentro da empresa, servindo como base para os demais sistemas existentes dentro da mesma como demonstra a Figura3.

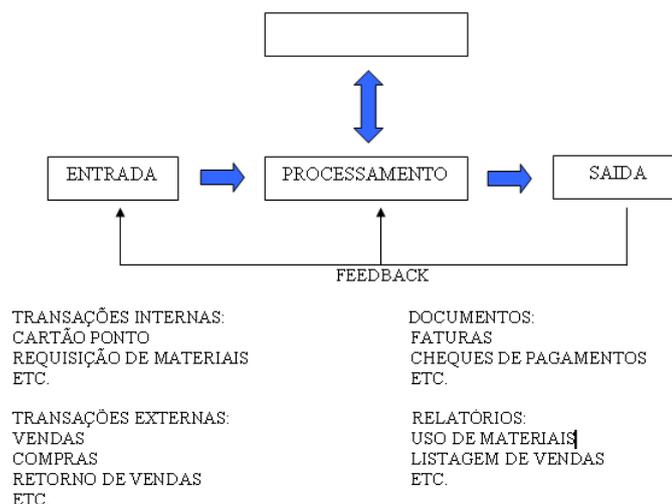


Figura3 – Sistema de informação transacional

FONTE: Adaptado de STAIR (1998)

As principais funções e características desses sistemas são:

Coletar (via digitação) os dados existentes nos documentos operacionais das organizações, validando-os; Armazenar esses dados em meio magnético; Ordenar ou indexar esses dados, de modo a facilitar o acesso a eles; Permitir consultas, on-line ou em batch, aos dados, detalhados ou agregados, que permitam retratar diferentes aspectos das operações; Gerar relatórios que possam ser distribuídos a outras pessoas que não os usuários diretos dos SI (Stair 1998).

Muito embora esses sistemas só controlem o fluxo de informações operacionais, eles também disponibilizam informações para a tomada de decisão. Um exemplo disso pode ser um sistema de controle de estoques que fornece informações sobre a movimentação do estoque para o departamento de compras. Este departamento poderá, através dessas informações, tomar decisões sobre quais produtos deverão ser comprados e em que quantidade. Um EDP pode, portanto, fornecer informações para apoio à decisão. Isso, porém, não o torna, apenas em decorrência desse fato, um SAD.

## 4 SISTEMAS DE APOIO DA DECISÃO

BISPO e CAZARINI (1998), A decisão é uma das atividades na qual todos nós estamos envolvidos diariamente e, muitas vezes, não nos damos conta de sua importância. Em consequência, o nosso processo de tomada de decisão tornou-se mais complexo. Hoje existem mais fatores que influenciam nesse processo de tomada de decisão do que na antiguidade.

Para as empresas, houve uma mudança mais radical ainda nos seus conceitos de administração, com o advento da globalização, da competitividade mais acirrada e das maiores exigências por parte dos clientes. Também o seu processo decisório se tornou muito mais complexo e tornou-se necessário o uso de ferramentas computacionais para dar suporte ao processo de tomada de decisão. Foi necessário desenvolver-se sistemas que proporcionassem o auxílio necessário aos gerentes para que pudessem enfrentar os desafios do nosso tempo.

### 4.1 A DECISÃO NA ANTIGÜIDADE

BISPO e CAZARINI (1998), Desde o início da civilização, o homem sempre procurou algo que lhe auxiliasse no seu processo decisório. Ele buscava nas divindades esse auxílio para que pudesse tomar decisões consideradas sábias. As diversas culturas antigas buscavam em pessoas com “poderes místicos” o auxílio para as suas decisões. Era considerado que essas pessoas podiam ter contato direto com as divindades e que as orientações que esses místicos forneciam eram orientações divinas. Desta forma, as decisões tomadas eram consideradas sábias e livres de erros. Quando a decisão tomada não proporcionava o resultado esperado, atribuía-se os erros às divindades que não estavam felizes com a pessoa ou com a civilização. Desta forma, as decisões tomadas nesta época eram fortemente influenciadas por entidades divinas e seus representantes.

Mais tarde surgiram divindades mais poderosas e populares, como Cristo, Maomé e Buda. Surgiram, também, líderes religiosos nas comunidades que passaram a exercer influência direta nas decisões das pessoas através dos preceitos religiosos de suas divindades. E, a exemplo do passado, quando o

resultado das decisões proporciona o sucesso, justificava-se como sendo fruto do amor pela divindade. Porém, quando o resultado era um fracasso, afirmava-se que era falta de fé. Assim, as entidades divinas continuaram a ter grande influência nas decisões.

Nos dias de hoje, a religião é um dos fatores de influência sobre as nossas decisões, porém existem alguns outros fatores que, em conjunto, nos auxiliam no processo de tomada de decisão. Esses outros fatores sempre existiram, porém, somente mais recentemente eles ganharam mais importância.

#### 4.2 A EVOLUÇÃO DA DECISÃO NESTE SÉCULO

Os critérios de tomada de decisão no começo do século eram centrados no executivo principal, em geral o proprietário, que detinha as prerrogativas de escolher o que julgasse melhor para a empresa e para os trabalhadores. Partia-se do pressuposto que os trabalhadores eram seres passivos e despreparados, sem capacitação e treinamento para atuarem nas decisões, eram avaliados quantitativamente pelo que rendiam e descartados quando não fossem mais produtivos. Como o ambiente era estável, relativamente simples, e as informações eram restritas, supunha-se que os decisores detivessem conhecimento bastante amplo de todas as alternativas e de suas conseqüências, havendo, portanto, uma única solução correta para os problemas, que deveria ser encontrada pelo decisor.

Segundo PEREIRA & FOSECA (1997) a década de 1960 foi precursora de grandes mudanças em todas as áreas, e ficou conhecida como “a década que mudou o mundo”. Na teoria administrativa, surgiu o movimento conhecido como Escola de Relações Humanas, surgido da contribuição da psicologia social, que deu um grande passo ao reconhecer o trabalhador como um ser humano capaz de pensar, de decidir e de ser motivado. Os trabalhadores deixaram de serem simples “mão-de-obra”, ou seja, aqueles que utilizam apenas as mãos para efetuarem seus trabalhos, e passaram a ser chamados de recursos humanos, ou seja, aqueles que trabalham de corpo e alma, se esforçam quando motivados e possuem habilidades que podem ser desenvolvidas e aproveitadas em benefício da empresa. Porém, a contribuição da psicologia social nas empresas foi a captação e a manipulação das

peças para manter as organizações funcionando produtivamente e aumentando cada vez mais a lucratividade da empresa.

Também foi nesta década que se passou a identificar a organização como um sistema aberto, em estreita relação com o meio ambiente. Ficou provado que os sentimentos, as atitudes e os valores do homem afetam, positiva ou negativamente, o processo de produção. Assim, as decisões passaram a ser centralizadas no cumprimento de metas. Este conceito evoluiu bastante, chegando ao seu ápice nos dias de hoje.

### 4.3 CONCEITOS BÁSICOS SOBRE DECISÃO

Segundo PEREIRA & FOSECA (1997) a palavra decisão é formada pelo prefixo *de* (prefixo latino aqui com o significado de parar, extrair, interromper) que se antepõe à palavra *caedere* (que significa cindir, cortar). Tomada ao pé da letra, a palavra decisão significa “parar de cortar” ou “deixar fluir”.

Uma decisão precisa ser tomada sempre que estamos diante de um problema que possui mais que uma alternativa para a sua solução. Mesmo quando para solucionar um problema temos uma única ação a tomar, temos as alternativas de tomar ou não esta ação.

Não costumamos nos dar conta da quantidade de decisões rotineiras que tomamos diariamente. Quando acordamos precisamos decidir se levantamos ou se permanecemos por mais alguns instantes deitados, depois precisamos decidir sobre o que queremos para o café da manhã, e diversas outras decisões precisam ser tomadas durante todo o dia. Podemos até afirmar que viver significa tomar freqüentes decisões. A maior parte das nossas decisões são rotineiras e se escolhermos a alternativa que não seja a melhor para solucionar o nosso problema, normalmente não nos acarreta maiores conseqüências, como por exemplo, se não escolhermos uma refeição balanceada para o nosso almoço, podemos sofrer as conseqüências disto, como engordar, emagrecer ou outras conseqüências que não são graves.

Existem, também, as decisões que precisamos tomar e que se não analisarmos muito bem as alternativas e escolhermos a melhor, poderá nos levar a conseqüências mais sérias, como por exemplo, decidir sobre um curso, uma viagem ou sobre onde aplicar o nosso dinheiro.

Existem, ainda, outros tipos de decisões que estão relacionadas com problemas cuja a solução deve ser uma mudança ou uma transformação. Neste caso, o grau de complexidade do problema e a escolha da melhor alternativa atingem um alto grau de responsabilidade que podem trazer a glória e o sucesso se for escolhida a alternativa mais correta ou o fracasso se for escolhida uma alternativa errada.

Segundo PEREIRA & FOSECA (1997) a racionalidade é a capacidade de usar a razão para conhecer, julgar e elaborar pensamentos e explicações e é ela que habilita o homem a escolher entre alternativas, a julgar os riscos decorrentes das suas conseqüências e efetuar uma escolha consciente da melhor alternativa.

#### 4.4 ETAPAS DO PROCESSO DECISÓRIO

URIS (1989) descreve as seguintes etapas no processo de tomada de decisão:

1. Análise e identificação da situação: A situação do ambiente onde o problema está inserido, deve ser claramente identificado, através do levantamento de informações, para que se possa chegar a uma decisão segura e precisa.
2. Desenvolvimento de alternativas: Em função do levantamento das informações, ou seja, da coleta de dados, pode se chegar a possíveis alternativas para a resolução do problema proposto.
3. Comparação entre as alternativas: Levantamento das vantagens e desvantagens de cada alternativa.

4. Classificação dos riscos de cada alternativa: As decisões sempre envolvem riscos, seja em um grau quase nulo, seja um alto grau de risco, ou sejam em um estágio intermediário de risco entre o quase nulo e o alto grau. Temos sempre que levar em consideração o grau de risco que temos em cada alternativa e escolher a alternativa que apresente comprovadamente o menor grau de risco. Porém, é necessário, muitas vezes, se combinar o grau de risco com os objetivos a serem alcançados. Às vezes, o grau de risco que se corre é muito grande, porém, o objetivo a ser alcançado, se alcançado, nos trará benefícios maiores em relação às alternativas menos arriscadas.

5. Escolher a melhor alternativa: Tendo o conhecimento das vantagens, desvantagens e riscos o decisor é capaz de identificar a alternativa que melhor solucione seu problema.

6. Execução e avaliação: A alternativa escolhida fornecerá resultados que deverão ser comparados e avaliados com as previsões anteriores.

#### 4.5 OS FATORES DE INFLUÊNCIA DO PROCESSO DECISÓRIO

Um dos maiores fatores de influência, na atualidade, é a cultura do(s) decisor(es). Um mesmo problema pode ser compreendido de forma diferente por pessoas ou povos de culturas diferentes. Mesmo numa grande cidade onde a cultura disponível é a mesma para todos os seus habitantes, o nível de aquisição desta cultura é diferente entre a população o que causa níveis de compreensão diferentes para um mesmo problema.

A quantidade de informações a respeito do problema é um fator de extrema importância para o processo decisório. Quanto mais informações conseguirmos levantar a respeito do mesmo, evidentemente, melhor será a sua compreensão e sua solução. Também, devemos nos preocupar com a qualidade das informações obtidas a respeito do problema.

Outros fatores também afetam o processo decisório, fatores como a inteligência do decisor, o seu nível social, o sexo, a religião, os costumes, as crenças, a ética, a motivação, a organização, a saúde, a família, etc.

Podemos adicionar aos fatores de influência citados acima, mais um fator de grande importância, principalmente na hora da escolha da melhor alternativa e na última fase, é o fator emocional. Muitas vezes o fator emocional é o que exerce a maior influência na hora de se tomar uma decisão nos levando a escolher a alternativa que não necessariamente é a melhor escolha.

Segundo PEREIRA & FOSECA (1997) toda decisão tem conseqüências e envolve riscos, mas uma vez processada, a escolha não tem volta. Por causa disso, o processo de tomada de decisões é sempre estressante e após a decisão acontece um relaxamento geral no organismo do decisor. Em algumas pessoas, o stress provocado pelo processo decisório é marcado por claras manifestações psicossomáticas, tais como dores de cabeça, sono, perturbações gástricas, manifestações de euforia ou depressão, etc.

Nas empresas, existem ainda, as influências externas provenientes dos clientes que solicitam produtos de boa qualidade, atendimento rápido e personalizado e preços mais competitivos; também, as condições impostas pelos fornecedores; as exigências do governo; o alerta dos concorrentes; as notícias da mídia, etc. As influências internas estão mais relacionadas com a interpretação das informações obtidas interna e externamente e englobam a cultura organizacional, as tecnologias, os recursos, as normas, os processos e as matérias-primas. Tomar decisões nesse contexto é algo complexo e administrar essa complexidade é o desafio de todo dirigente moderno.

“Informação é algo que alguém deseja saber, e está disposto a pagar por ela. A informação não é tangível e nem mensurável, mas é um produto valioso no mundo contemporâneo porque proporciona poder. O controle da informação é alvo de governos, empresas e pessoas” (GATES, 1997). A imensa quantidade de informações disponíveis atualmente nos traz grandes benefícios e grandes responsabilidades. Temos que saber muito bem como trabalhar com essa massa de informações, saber como fazer a sua integração e monitoração, como obter acesso rápido e fácil e, ainda nos preocuparmos com a sua segurança e privacidade. O processo decisório está intimamente relacionado com o tempo e com capacidade humana e/ou computacional de processar informações em tempo hábil.

#### 4.6A DECISÃO E OS SISTEMAS

Com o crescimento das organizações e a necessidade de se aumentar e se ajustar os níveis gerenciais, também faz-se necessário a divisão de tarefas e responsabilidades. E, evidentemente, a divisão do processo decisório da empresa cria vários níveis de decisões que estão relacionados com os níveis gerenciais. Desta forma, surge o conceito de decisão interdepartamental, ou seja, decisões que devem ser tomadas e as conseqüências assumidas por vários departamentos da empresa. Também surge o conceito de decisão nos níveis estratégicos, táticos e operacionais, de acordo com o nível do decisor que a toma.

A Teoria dos Sistemas afirma, em linhas gerais, que os sistemas são todos interligados e que não se pode interferir em um sistema sem afetar os demais e o sistema maior, ainda que seja de uma forma irrelevante. Portanto, não podemos nos esquecer, na hora de se tomar uma decisão, de que fazemos parte de algum subsistema, que faz parte de algum sistema e que, por sua vez, faz parte de um sistema maior. E qualquer alteração no nosso subsistema, afeta em algum grau os outros subsistemas, o sistema ao qual fazemos parte e ao sistema maior. Podemos afirmar, por exemplo, que fazemos parte de um conjunto de subsistemas: social, comercial, econômico, ecológico, político e outros, todos interligados, que fazem parte de um sistema maior que é a nossa sociedade, e fazemos parte de um sistema maior que é a nossa nação, que faz parte de um sistema maior que é a humanidade.

Segundo PEREIRA & FOSECA (1997) quando os decisores levam em conta apenas as variáveis que os afetam diretamente, sem analisar as conseqüências para os outros subsistemas ou o seu impacto no sistema maior, corre-se o risco de se resolver o problema criando outro maior em outros contextos, como por exemplo, tomar um remédio para curar uma doença e sofrer efeitos colaterais iguais ou piores que ela.

#### 4.7A EVOLUÇÃO DOS SISTEMAS DE APOIO À DECISÃO

Segundo POWER (1997) o conceito de Suporte à Decisão surgiu da evolução de duas áreas de pesquisa: os estudos teóricos sobre o Processo de Tomada de Decisão Organizacional feitos no *Carnegie Institute of Technology* durante as décadas de 50 e 60, e os trabalhos realizados com sistemas computacionais interativos no *Massachusetts Institute of Technology* nos anos sessenta.

MORTON (1971) enfocou em seu livro como os computadores e modelos analíticos poderiam ajudar os gerentes no suporte à tomada de decisões chaves. Ele criou um sistema chamado *Management Decision System* (MDS) baseado na experiência de gerentes e utilizou este sistema para coordenar o planejamento de uma linha de produção de máquinas de lavar. As pesquisas e o sistema de Morton abriram caminho para novas definições e pesquisas para se criar outros DSS (*Decision Support Systems*) específicos.

GERRITY (1971) escreveu um artigo e projetou um sistema computacional para dar suporte aos gerentes de investimentos na administração diária das aplicações de seus clientes.

LITTLE (1975) escreveu um artigo no qual ampliou as fronteiras da modelagem suportada por computador. Ele identificou critérios para projetar modelos para apoiar a tomada de decisão gerencial. Seus critérios incluíram: robustez, facilidade de controle, simplicidade, e perfeição de detalhes relevantes. Ele projetou um DSS, que chamou de “Brandaid”, para dar suporte a decisões referentes a produto, promoção, preços e decisões adversas. KEEN & MORTON (1978) descreveu em seu livro uma metodologia para análise, projeto, implementação, avaliação e desenvolvimento de DSS.

Segundo PEARSON & SHIM (1995) nos anos 70, os DSSs enfatizavam a interação dos sistemas baseados em computador que auxiliava na tomada de decisão utilizando modelos de dados que resolviam problemas semi-estruturados e não estruturados. A ênfase não era no processo decisório, mas no suporte computacional e nas ferramentas necessárias para o desenvolvimento rápido das

aplicações. Nos anos 80, surgiram uma variedade de novas tecnologias para prover a eficiência gerencial, organizacional e profissional. Um grande número de softwares foram produzidos sob o título de DSS.

Na década de 90, com os avanços tecnológicos, tanto em hardware como em software, houve grandes avanços nos Sistemas de Apoio à Decisão. Foram desenvolvidos diversos aplicativos específicos para dar suporte a categorias de sistemas mais específicos e utilizavam recursos sofisticados, incluindo algoritmos de inteligência artificial. Esta década foi marcada pelo grande avanço em Tecnologia de Informação proporcionado pela valorização das informações pelas empresas. Entre as novas ferramentas de Tecnologia da Informação desenvolvidas nesta década, está a chamada nova geração de Sistemas de Apoio à Decisão: o *data warehouse*, o OLAP. Estas novas ferramentas estão sendo muito úteis no gerenciamento dos negócios modernos e estão auxiliando os gerentes a enfrentarem os desafios dos novos tempos.

## 5 DATA WAREHOUSE

O *data warehouse* surgiu como conceito acadêmico na década de 80. Com o amadurecimento dos sistemas de informação empresariais, as necessidades de análise dos dados cresceram paralelamente. Os sistemas OLTP não conseguiam cumprir a tarefa de análise com a simples geração de relatórios. Nesse contexto, a implementação do *data warehouse* passou a se tornar realidade nas grandes corporações.

Em INMON (1997), encontramos a definição mais usual de um *data warehouse*, como sendo “uma coleção de dados integrados, orientados por assunto, não-voláteis e variáveis com relação tempo, criada para suporte a decisão”. O objetivo do *data warehouse* é fornecer uma imagem única da realidade do negócio da organização. Deve selecionar, sumarizar e organizar informações estratégicas que estão contidas nos bancos de dados operacionais de forma implícita, associando-as ao conhecimento gerencial para a construção de uma base de dados que facilite a tomada de decisão GONÇALVES (1999).

Em SINGH (2001), afirma que um *data warehouse* é definido como uma ferramenta competitiva que permite a qualquer usuário final acessar dados com uma visão uniforme para a empresa inteira. Ao armazenar dados em um ponto de armazenamento central, o *data warehouse* oferece uma representação integrada das múltiplas fontes de informação da empresa, e garante a consistência das normas de gerenciamento e das convenções aplicadas aos dados. Portanto, refletindo as necessidades da corporação e não, simplesmente as individuais.

Segundo Ralph Kimball (1998) um *Data Warehouse* é um repositório construído especificamente para a consolidação da informação da organização em um formato válido permitindo que esses mesmos dados sejam analisados de uma forma seletiva. Onde seus dados são consistentes, isso diz que a base de dado quando acessada simultânea não pode trazer resultados diferentes, os dois resultados devem está iguais para que não aja divergência de dados, se isso acontecer, os dados não foi tratado corretamente. Esses dados podem ser

separados e combinados usando-se qualquer medição possível de negócio (o requisito clássico *slice and dice*).

Para Ralph Kimball (1998) afirma que um *data warehouse* não consiste apenas em dados, mas também em um conjunto de ferramentas pra consultar, analisar e apresentar informações. Porém é um conjunto de ferramentas que serve de auxílio para consultar, analisar e apresentar informações auxiliando os gestores a tomada de decisão. Uma *data warehouse* é um local onde são armazenados dados confiáveis e não sendo apenas um acúmulo de dados para posterior liberação. Esses dados são coletados em diversas fontes de informação, limpos, tem sua qualidade assegurada para que não aja inconsistência e depois são liberados somente se forem adequados para o uso, se eles estiverem incompletos não serão liberados.

O DW não é feito apenas para simples aprisionamento de dados e sim com o intuito de se extrair e analisar informações provindas desses dados armazenados usando-se ferramentas adequadas para isso. O valor de um DW não está em colecionar dados e sim saber gerenciar aqueles dados sendo transformados em informações úteis. De acordo com Harrison (1998), um banco *data warehouse* é projetado para atender a necessidade dos executivos por informações sobre o desempenho comercial de suas organizações de maneira mais completa e rápida.

Na prática, segundo Campos (1997), *data warehouse* é um banco de dados, alimentado continuamente com dados oriundos dos sistemas operacionais, devidamente selecionados, depurados e integrados, com o propósito de gerar uma visão única e real da empresa, predispostos de maneira a agilizar o processamento de consultas e preferencialmente isentos de procedimentos transacionais.

## 5.1 ARQUITETURA DE UM DATA WAREHOUSE

A arquitetura funcional completa de um *data warehouse* inclui, além da base de dados analítica e das ferramentas de acesso, um conjunto de serviços e ferramentas que estabelecem a sua conexão com as fontes de dados, que alimentam o repositório em um processo complexo de extração, limpeza e

consolidação. Segundo GONÇALVES (1999) uma solução completa de *data warehouse* contém três componentes:

Um componente responsável pela aquisição dos dados, e que se costuma chamar de *back end*. É um conjunto de programas utilizados para prover a conectividade com as fontes de dados, realizando limpeza, conversões de formatos e correções de inconsistências, que eventualmente existam, para a adequação as necessidades dos dados do *data warehouse*.

Um repositório de dados, que é o banco de dados propriamente dito. O armazenamento dos dados pode ser feito em bancos de dados relacionais, banco de dados multidimensionais ou em uma solução híbrida. A estrutura dos dados no repositório corresponde ao modelo de dados do *data warehouse*.

Uma camada de acesso aos dados, normalmente chamada de front-end. É um conjunto de ferramentas ou aplicativos que servem para acesso e análise de dados do *data warehouse*. Permitindo ao usuário do *data warehouse* navegar entre os dados, resumi-los, compará-los, etc.

## 5.2 DADOS DO DATA WAREHOUSE

Como é colocado em SINGH (2001) o processo de transformação do dado em informação é realizado através da extração dos dados operacionais internos e dos dados disponíveis em fontes externas à empresa. Fontes externas podem ser tabelas, arquivos, ou qualquer tipo de informação que se queira agregar ao projeto, por exemplo, dados estatísticos ou financeiros fornecido por algum órgão do governo. Os dados são transformados e incluídos no *data warehouse* com base na estrutura do assunto do negócio em questão.

O *data warehouse* possui os seguintes dados:

Dados orientados por assunto: Os dados em um *data warehouse* são organizados por assunto, ou seja, pode-se dizer que são orientados por áreas da empresa – financeiro, qualidade, recursos humanos.

Dados integrados: Um dos processos mais complexos em termos de data warehouse e que envolve uma grande dose de esforço é garantir a integração dos dados. Uma vez que eles são oriundos de diversas fontes, podendo ser sistemas legados ou de fontes externas, é comum que haja diferenças de nomenclatura, de semântica, de codificação, etc.

Não voláteis: Um *data warehouse* tem como princípio básico a manutenção de grandes volumes de dados históricos, pois é sobre estas informações que podem ser formuladas as análises gerenciais de tendências de mercado e de prospecção de comportamentos evolutivos, por exemplo. Uma vez carregados não sofrem mais atualização, permanecendo estáticos e servindo para leitura somente. A atualização de informações se dá através de novas cargas de dados feitas periodicamente, em função de regras previamente estabelecidas para tanto.

Dados Sumarizados: Os dados, segundo INMON (1997), podem ser classificados genericamente em dois grupos:

Dados primitivos: Os dados primitivos são os dados de produção usados nos sistemas OLTP. São dados com alto nível de detalhe, pois são usados no dia-a-dia da organização. São dados em estado bruto. Devem estar sempre disponíveis para acesso, pois os sistemas que os utilizam os acessam repetidas vezes GONÇALVES (1999). Para estarem disponíveis instantaneamente devem sofrer constantes atualizações.

Dados derivados: Os dados derivados são dados históricos sumarizados, ou seja, sofreram uma totalização ou agregação a partir dos dados primitivos. Este tipo de dado é pouco encontrado em sistemas convencionais e são de vital importância para um *data warehouse*.

O DW oferece informações para o nível estratégico de forma agregada, ou seja, não interessa saber o valor de venda específica de um item de um produto, mas sim o total de vendas, a média ou máximo de vendas obtidas daquele produto em um determinado período de tempo.

Uma questão a ser definida em um *data warehouse* é o seu nível de detalhe, ou seja, a sua granularidade. A partir desta definição é que se estabelecem quais os tipos de consultas serão possíveis. Segundo INMON (1997) a definição da granularidade é a questão mais importante do projeto, pois se a mesma for corretamente definida os demais aspectos do projeto tendem a fluir tranquilamente. Um *data warehouse* armazena cinco tipos de dados, de acordo com a sua granularidade e agregação:

**Dados corrente detalhados:** Dados correntes detalhados são aqueles que refletem os acontecimentos mais recentes do data warehouse. A atualização se dá no momento da carga. São replicados para evitar sobrecargas de consulta nos sistema originais.

**Dados antigos detalhados:** Em função de regras a serem estabelecidas pelos projetistas, os dados passam a ser considerados antigos e podem ser deslocados para outras mídias de armazenamento, com acesso mais lento e menos freqüente.

**Dados pouco sumarizados:** Melhoram a usabilidade do sistema e o seu tempo de resposta. Tem a desvantagem de gastarem mais espaço e daí a necessidade de usarmos uma granularidade adequada.

**Dados altamente sumarizados:**Dados altamente sumarizados, são os dados utilizados pelos níveis gerenciais da organização, pois são mais adequados para a análise de comportamentos e tendências. São muito consultados. Ocupando menor volume de armazenamento na base de dados do *data warehouse*.

**Metadados:** Segundo INMON (1997) os metadados englobam o *data warehouse* e mantém as informações sobre o que está onde. Ele ainda define quais as informações os metadados mantêm: a estrutura dos dados segundo a visão do programador; a estrutura dos dados segundo a visão dos analistas de SAD; as fontes de dados que alimentam o *data warehouse*; as transformações sofridas pelos dados no momento de sua migração para o *data warehouse*; o modelo de dados; o relacionamento entre o modelo de dados e o *data warehouse*; o histórico das extrações de dados; acrescentamos ainda os dados referentes aos relatórios que são gerados pelas ferramentas OLAP assim como os que são gerados nas camadas semânticas.

A análise de viabilidade para um projeto de *data warehouse* deve ser realizada com um enfoque estratégico e deve ser encarado como um instrumento de vantagem competitiva (VILELA, 1997). Apesar de ter um retorno difícil de ser mensurado em números, os ganhos sob o ponto-de-vista estratégico podem ser percebidos rapidamente.

### 5.3 MODELAGEM DIMENSIONAL

Nos bancos de dados relacionais, usados nos sistemas transacionais tradicionais, a redundância dos dados é evitada, sendo aceita somente em determinados casos em que é realmente necessária. Esta redundância é eliminada através de processos de normalização. A normalização das tabelas traz benefícios nos casos em que muitas transações são efetuadas, pois estas se tornam mais simples e rápidas. No caso de *Data Warehouses*, ocorre o contrário, as transações operam sobre um grande volume de dados e não são simples, nem freqüentes, não sendo conveniente a normalização das tabelas, pois no ambiente de *Data Warehouse* ocorrem poucas transações concorrentes e cada transação acessa um grande número de registros (PERNAS, 2003).

Conforme Pernas (2003), outro ponto que distingue o banco de dados relacional do *Data Warehouse* está relacionado à modelagem dos dados. Enquanto que em um banco de dados relacional geralmente utiliza-se a modelagem Entidade-Relacionamento (ER), em um DW utiliza-se de uma modelagem lógica denominada de modelagem dimensional, também chamada de modelagem multidimensional. Assim, modelagem dimensional é a técnica utilizada para se ter uma visão multidimensional dos dados.

Segundo Kimball (1997), a modelagem dimensional “é uma técnica de projeto lógico que busca apresentar os dados em uma estrutura padrão e intuitiva que permite um acesso de alta performance”. Essa é uma técnica antiga usada para criar bancos de dados simples e compreensíveis.

Um modelo dimensional é composto, basicamente, pela tabela de fatos e pelas tabelas de dimensões (Figura4). A tabela de fatos traz o resultado da consulta, ou seja, os valores de medição representando transações ou eventos referentes aos negócios da organização e que podem ser passíveis de análise. Uma dimensão pode agregar sob nomes distintos, um conjunto de itens com características e posições próprias, tais como meses e trimestres em relação a um período anual FROZZA (2006).

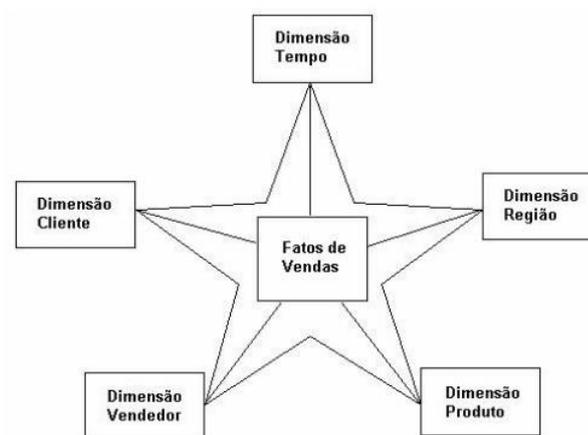


Figura4 – Representação do modelo dimensional

FONTE: SHIGUNOV (2007)

Existem várias técnicas para fazer a modelagem dimensional, sendo as principais: o modelo estrela (*Star Schema*), que se assemelha a uma estrela, e o modelo floco de neves (*Snow Flake*). Dominante no projeto de DW (KIMBALL, 1997), a modelagem dimensional no modelo estrela possui as seguintes características: Distingue bem as dimensões dos fatos medidos; Simplifica a visualização dimensional; É eficiente para a realização de consultas; Acomoda mudanças mais facilmente.

A Figura5 ilustra um esquema dimensional na forma de estrela para um DW de vendas, com as dimensões tempo, região, produto, vendedor, cliente e as medidas da tabela de fatos (valor das vendas, média das vendas).

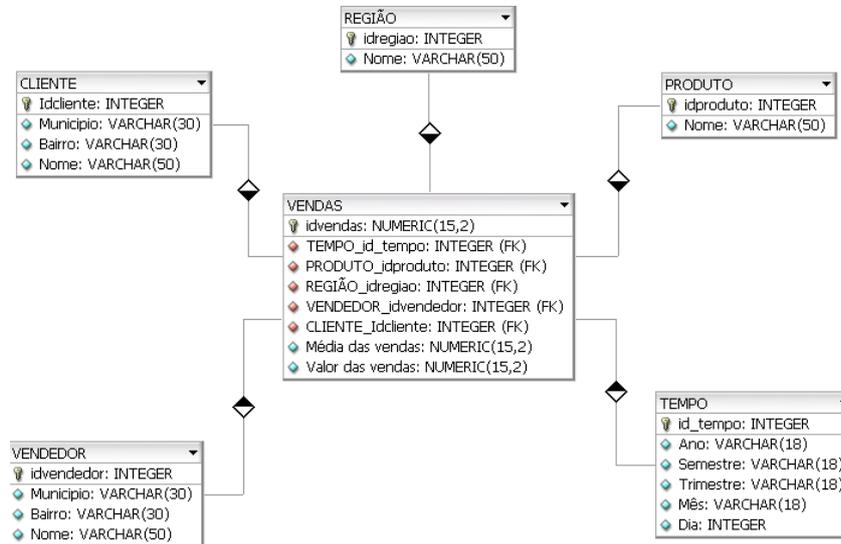


Figura5 – Representação do modelo estrela  
 FONTE: SHIGUNOV (2007)

No modelo floco de neves as tabelas de dimensão são normalizadas, evitam redundância e requerem mais junções para as consultas como ilustra a Figura6.

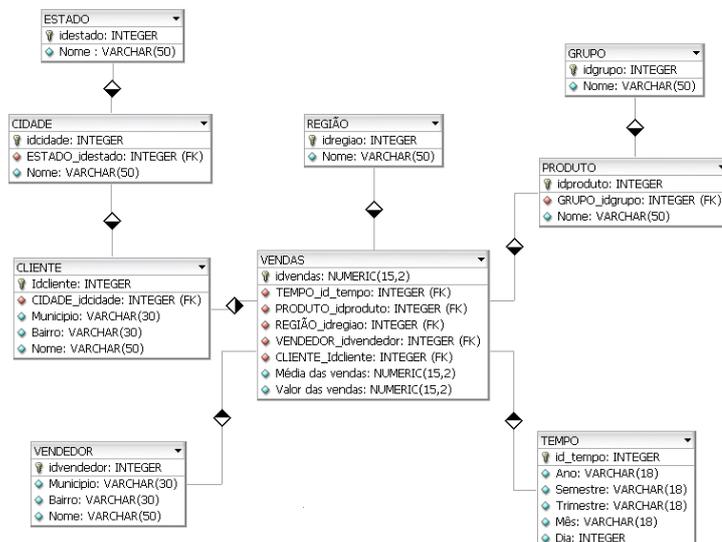


Figura6 – Representação do modelo floco de neves  
 FONTE: SHIGUNOV (2007)

De acordo com Barbieri (2001, p. 74), “a modelagem de dados é seguramente um dos fatores críticos de sucesso num projeto de *Data Warehouse*, e pode representar a fronteira entre o sucesso e o seu fracasso”. Como o modelo relacional trabalha com normalização, suas tabelas possuem menos registros e não têm redundâncias, apresentando assim uma melhor performance nas tarefas do dia-a-dia, como inclusões, alterações e exclusões de registros. Mas ele só é adequado para consultas simples de poucos registros. Para análises mais complexas, com um universo de registros maior, o modelo dimensional oferece uma melhor alternativa, economizando em junções com várias tabelas e armazenando dados que facilitam a análise das informações (HOKAMA et al., 2004, p. 32).

Para se elaborar um esquema dimensional deve-se levar em conta a granularidade desejada para a análise. Granularidade se define como o nível com que os dados estão sumarizados. O grão é o maior nível de detalhamento. O nível adequado de granularidade deve ser definido de tal forma que atenda às necessidades do usuário, tendo como limitação os recursos disponíveis (SHIGUNOV, 2007). A granularidade afeta diversas características em um DW, como o número de diferentes cruzamentos de dados que podem ser realizados, a infra-estrutura e o espaço em disco necessário.

Uma dimensão de um cubo é uma coleção de hierarquias de membros, organizados de maneira conveniente para análise. Um membro é um ponto em uma hierarquia de uma dimensão, determinado por um conjunto particular de valores de atributo. Um nível da hierarquia é uma coleção de membros que possuem a mesma distância da raiz da hierarquia. Cada hierarquia de dimensão está associada a um atributo da tabela de fatos (MACDONALD e RUBIK, 2007).

#### 5.4 DATA MART

Da mesma forma que o *data warehouse*, o *data mart* ainda não possui uma definição universalmente aceita e também está em fase de aperfeiçoamento. Os *data marts* são subconjuntos de dados, dentro de um *data warehouse*, projetados para dar suporte a negócios de unidades organizacionais específicas (NIMER, 1998). Os *data marts* são muito interessantes para resolver certos problemas, mas

não são necessariamente substitutos de um projeto de *data warehouse*. Um *data mart* não deve ser um pequeno *data warehouse*, com a finalidade de ser rápido ou possuir dados ainda não ajustados para o *data warehouse* (KIMBALL, 1997).

Os *Data Marts* são bancos de dados modelados multidimensionalmente, orientados aos departamentos de uma empresa, por exemplo, a área comercial, financeira, marketing e logística construíram vários *Data Marts* (DM's) cada uma. Ao final de dois anos, eles passaram a alimentar um DW que, depois de implementado, passou a fazer o caminho inverso, que é de alimentar os DM's. Então com essa fundamentação, podemos definir que os DM's podem surgir de duas maneiras. A primeira é *top-down* e a segunda é a *botton-up*.

*Top-down*: é quando a empresa cria um DW e depois parte para a segmentação, ou seja, divide o DW em áreas menores gerando assim pequenos bancos orientados por assuntos aos departamentos.

*Botton-up*: é quando a situação é inversa. A empresa, por estratégia sua, prefere primeiro criar um banco de dados para somente uma área. Com isso os custos são bem inferiores de um projeto de DW completo. A partir da visualização dos primeiros resultados parte para outra área e assim sucessivamente até resultar em um *Data Warehouse*.

Os *Data Marts* atendem as necessidades de departamentos específicos de negócio ao invés das da empresa inteira. Eles otimizam a entrega de informação de suporte à decisão e se focam na gerência sumarizada e/ou dados exemplificativos ao invés do histórico de níveis mais elevados, englobando a empresa inteira. Algumas empresas são atraídas aos *Data Marts* não somente pelo custo baixo e um tempo menor de implementação, mas também por causa dos correntes avanços tecnológicos.

STRANGE (1998) apresenta uma tabela comparativa entre *data mart* e *data warehouse*:

Tabela 2– Tabela comparativa entre *data mart* e *data warehouse*

Propriedades	Data Warehouse	Data Mart
Âmbito	aplicações neutras	aplicações específicas
	centralizado e compartilhado	departamento ou área de usuário
Assuntos	múltiplas áreas	área única
Fontes de dados	muitas	poucas
Tempo de implementação	<ul style="list-style-type: none"> <li>• 9 a 18 meses para 1º estágio (duas ou três áreas de assuntos);</li> <li>• múltiplos estágios de implementação;</li> </ul>	4 a 12 meses
Características	flexível	restritivo
	nível estratégico	nível tático

FONTE: STRANGE (1998)

Segundo DEPOMPA (1998), muitas empresas construíram seus *data marts* especializados para atender as necessidades imediatas de consultas e análises dos dados, sem criar primeiro uma infra-estrutura física e lógica para o *data warehouse*.

### 5.5 ETL (ETL – Extraction, Transformation, Loading)

Segundo Ralph Kimball (1999) O *Extrato-Transform-Load* (ETL) sistema é a base de dados entreposto. Um sistema concebido adequadamente, o ETL extrai os dados da fonte de sistemas, reforça a qualidade dos dados e consistência normas, conforme dados de fontes distintas para que possam ser utilizados em conjunto e, finalmente, fornece dados em uma apresentação de formato pronto, para que desenvolvedores possam construir aplicações e os utilizadores finais podem tomar decisões. O sistema ETL acrescenta um valor significativo aos dados. É muito mais do encanamento para obter dados dos sistemas fonte e os dados em armazém.

Todo sistema de ETL tem que organizar dados dentro os vários permanentes e formas de semipermanente. Quando nós dissermos organizando, nós queremos dizer dados de escritura ao disco, e por isto o sistema de ETL às vezes está chamado à área organizando. Você poderia ter notado que nós recomendamos

alguma forma de organizar depois de cada dos passos de ETL principais pelo menos (extrato limpe, conforme, e entregue). Nós discutimos as razões para várias formas de organizar (KIMBALL, 1998).

Um ETL corretamente projetado de sistema extrai dados dos sistemas de fonte, obriga qualidade de dados e padrões de consistência, conforma dados de forma que fontes separadas podem ser usadas juntas, e entrega dados em um formato de apresentação-pronta e de forma que fomentador de aplicação pode construir aplicações e usuários de fim podem tomar decisões como é apresentado no esquema ETL da Figura7.

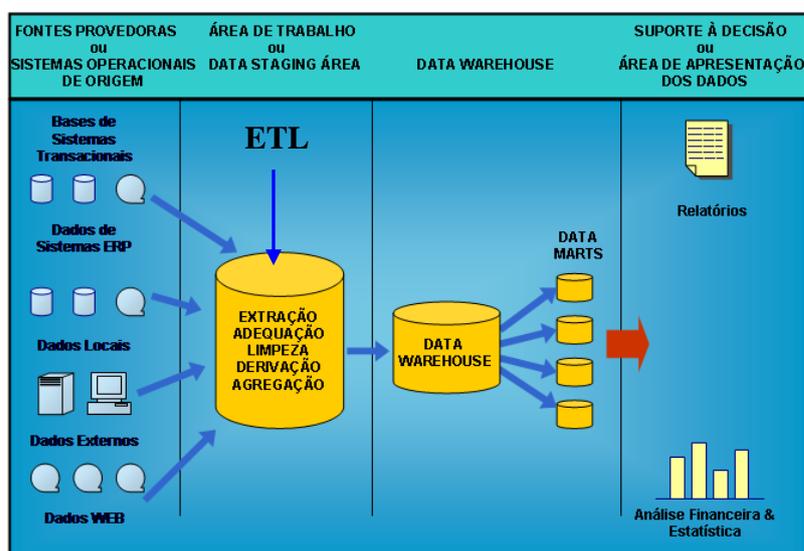


Figura7 – Esquema ETL

FONTE: Adaptado de *SUNEXPERT MAGAZINE* (1998)

### 5.5.1 EXTRAÇÃO

Segundo Ramos e Santos (2006), os requisitos do negócio que irão compor o *Data Warehouse* levantados, iniciaram então a primeira etapa do processo da ETL, a extração dos dados do ambiente operacional que realmente importam na construção e carga dos dados no DW. Estes dados a serem extraídos podem ser originados de múltiplas fontes do ambiente operacional, como arquivos TXT ou XML, planilhas ou até base de dados diferentes da do DW. O ambiente operacional é composto de vários dados / registros que contém informações que podem ser úteis ou não para a construção do *data warehouse*. Assim, esta etapa é responsável por realizar a seleção e extração dos dados das distintas fontes que realmente serão relevantes para a

montagem do *data warehouse*. Durante o processo da extração, apenas os dados que são úteis e relevantes ao negocio, são extraídos para o *Data Warehouse*.

Os dados contidos em um DW são provenientes de diversas origens. Assim, após a definição do escopo do projeto, a segunda etapa é a identificação das fontes de dados, as quais podem ser classificadas em dois grandes grupos: internas e externas.

As fontes internas são os sistemas chamados transacionais, os quais têm por objetivo o controle das rotinas operacionais da organização, tais como: almoxarifado, faturamento, vendas, etc. Esta, sem dúvida alguma, é a principal fonte para um DW, pois trata-se dos dados a respeito da própria empresa (JORGE et al., 1998).

O outro tipo de fonte de dados, a externa, refere-se a dados que não fazem parte dos sistemas da organização e que muitas vezes são comprados de outras empresas que mantêm bases de dados comerciais. Em alguns casos, os mesmos estão disponíveis na Internet e em outros, encontram-se de forma não computadorizada, como em jornais, boletins informativos, dentre outros.

Como se pode observar, as origens de dados de um DW são as mais diversas possíveis podendo portanto, ter características variadas. É dessa heterogeneidade que surgem as grandes dificuldades dessa fase. A seguir são citados alguns dos elementos dificultadores da extração de dados (CIELO, 2002):

Algumas vezes é necessário selecionar vários campos do ambiente operacional para compor um único atributo no DW;

Difícilmente existem modelos de dados e documentação dos sistemas transacionais. Isso significa que os desenvolvedores terão que realizar trabalhos de engenharia reversa para compreender as fontes dos dados;

Os dados não se encontram padronizados. Por exemplo: em um sistema transacional o campo sexo é identificado pelas letras M (masculino) e F (feminino). Em outro, esse mesmo campo utiliza as letras H (masculino) e M (feminino);

As bases de dados transacionais são desenvolvidas em plataformas e tecnologias distintas, tais como Oracle, SQL Server, DB2, sistemas de arquivos, etc.

Uma vez identificados, os dados devem ser coletados de suas fontes e armazenados em uma área intermediária, por meio de rotinas de extração. Existem diversas técnicas para realização desse processo de captura dos dados das fontes, porém todas elas são classificadas em estática ou dinâmica, conforme descrito a seguir (PEREIRA, 2000).

Técnicas de captura estática de dados baseia-se na captura de todos os dados de interesse do DW, independentemente de terem sido modificados ou não durante o período compreendido entre uma extração e outra.

Assim, não existe a necessidade de rotinas que fiquem monitorando os sistemas transacionais em busca de dados que foram alterados, incluídos ou excluídos. No lugar disso, existe um processo automatizado para extração dos mesmos, sendo a periodicidade com que essa captura é realizada dependente no nível de granularidade do DW.

Uma característica marcante da técnica estática é a aquisição apenas da versão corrente dos dados acessados. Em outras palavras, se um registro sofreu várias alterações desde sua última captura e os sistemas transacionais não tiverem sido projetados para armazenar um histórico de tais modificações, na próxima execução das rotinas de extração, somente a última versão do registro em questão será migrada para o ambiente DW.

A grande vantagem dessa técnica é o fato de poder ser executada fora dos horários mais críticos de requisição aos bancos de dados, não interferindo na performance dos sistemas transacionais.

A captura dinâmica, também conhecida como incremental, implementa uma replicação dos dados modificados entre uma extração e outra para posterior migração para o DW.

Essa técnica utiliza recursos disponíveis nos SGBD (sistema de gerenciamento de banco de dados) tais como *triggers* e arquivos de *log*, para realização da aquisição das modificações. Com isso, diferentemente do que ocorre na captura estática, as versões intermediárias dos dados podem ser adquiridas.

A desvantagem dessa abordagem é a sobrecarga dos sistemas transacionais causados pela execução das rotinas de captura das modificações dos dados, mesmo durante os horários mais críticos de trabalho.

### 5.5.2 TRANSFORMAÇÃO

Ramos e Santos (2006), pode-se dizer que esta é a etapa mais cautelosa da ETL. Esta etapa de transformação, conforme já comentado ocorre preferencialmente na área de fase inicial. É nesta etapa que tentamos garantir aos usuários do *Data Warehouse*, a integração e confiabilidade dos dados armazenados. Para isto realizamos uma série de transformações e acabamos também por filtrar apenas o que será necessário para carga do *Data Warehouse*.

Dentre as diversas funções e atribuições que podem ser executadas destacamos algumas que são descritas abaixo:

- Limpeza dos dados – Corrigir possíveis dados “sujos”, ou seja, dados que chegam nesta área de stage que não serão aproveitados, por motivos de má digitação no ambiente operacional ou duplicidade de registros.
- Codificação - devido à informação estar vir de diversas fontes de dados, pode ocorrer de a mesma ser carregada de diferentes maneiras, existindo então a necessidade de se convencionar um padrão para garantir a qualidade do dado no data warehouse. Por exemplo, ao extrair os dados de clientes das múltiplas fontes do operacional, a informação se o mesmo tem filho ou não pode ser representada de diversas formas, “S” ou “N”, “0” ou “1”, ”sim” ou “não”. No entanto ao se convencionar o uso de “S” e “N”, um trabalho de integração e codificação deve ser realizado transformando todas as outras representações para este formato único.
- Atributos de medidas – assim como data, representação de sexo as unidades de medidas também devem ser tratadas para que os dados que farão parte do data warehouse sejam armazenados em uma única unidade de medida.

No processo de transformação, os dados podem estar no ambiente operacional, de forma duplicados ou com “sujeiras”, ou em branco. Se não for tomado um cuidado especial para a transformação desses dados, o objetivo do *data*

*warehouse* pode ser perdido e as informações não poderão ser consideradas consistentes.

Na transformação, também alguns cálculos e sumarizações de valores são realizados com o objetivo de acelerar mais ainda o processamento da carga que estará por vir. Assim durante a etapa da transformação, tenta se tomar todos os cuidados para que ao realizarmos a carga no *data warehouse*, tenhamos dados em que o usuário possa ter a maior confiabilidade do que estão acessando. Após as transformações aplicadas sobre os dados para assegurar a qualidade e integração dos mesmos estes podem finalmente ser migrados para a base de dados do DW.

### 5.5.3 CARGA

Ramos e Santos (2006), é a etapa final da ETL. Os dados teoricamente já se encontram limpos e transformados. Este é o momento de ser realizada a carga dos dados da área de *stage* para o *data warehouse* em si. Depois de realizada todas as transformações, codificações, sumarizações já calculadas, os dados se encontram num ponto aceitável, assumindo que estão com todos os requisitos atendidos, para serem carregados no *data warehouse*.

Porem, antes da realização da carga, deve se realizar mais uma verificação. A verificação da integridade dos dados, realizando a validação das chaves estrangeiras com suas respectivas chaves primarias, confirmando que a informação realmente consiste.

Outro ponto a ser definido nesta etapa, é se a carga a ser realizada será incremental ou total. Definir que a carga seja incremental, quer dizer que os dados serão apenas adicionados ao *data warehouse*. Já com a carga do tipo total, os dados existentes serão excluídos e os dados são re-inseridos. Geralmente orienta-se que a carga incremental seja realizada em tabelas fatos e a carga total em dimensões.

Sugere se nesta etapa também a realização de alguns tratamentos dos dados que foram rejeitados durante as transformações, apenas para se manter um controle mais eficiente do processo. Desta forma a carga de dados depois de realizada com sucesso, completa o ciclo do processo denominado ETL (KIMBALL, 1998).

ETL é ambos um simples e um assunto complicado. Quase todo o mundo entende a missão básica do sistema de ETL: sair dados da fonte e carregar isto no *data warehouse*. E a maioria dos observadores crescentemente está apreciando a necessidade para limpar e transformar dados no caminho. Tanto para a visão simples. É um fato de vida que o próximo passo no desígnio do ETL fraturas de sistema em mil pequenas *subcases*, dependendo fontes de dados estranhas, regras empresariais, software existente, e aplicações destino - informando incomuns por conta própria. O desafio para todos nós é tolerar as mil pequenas *subcases*, mas manter perspectiva na missão global simples do sistema de ETL (KIMBALL, 1998).

## 5.6 OLAP

A tecnologia OLAP (*On Line Analytical Processing* – Processamento Analítico *On Line*) surgiu com a evolução dos sistemas de informação. Antes de surgir esta tecnologia armazenava-se grandes quantidades de dados, mas o tratamento destas informações eram muito difíceis para os usuários finais e analistas de sistema. A grande dificuldade era pelo sistema não ser relacional e tinha que procurar dados por vários arquivos, com isso tinha-se dois grandes trabalhos: primeiramente encontrar os dados e em segundo, construir o relatório no formato desejado.

Com o surgimento da tecnologia OLAP foi possível acessar, visualizar e analisar dados corporativos com alta flexibilidade e performance. Com a globalização empresas estão enfrentando maior concorrência e também estão expandindo sua área de mercado. Os executivos tem que tomar decisões precisas para que sua empresa tenha um futuro promissor, e para ajudá-los os executivos nas tomadas de decisões a OLAP apresenta informações através de um modelo de dados natural e intuitivo de simples navegação e pesquisa, gerando relatórios do tipo “*ad-hoc* (Consulta Eventual) ”.

OLAP é implementado em um modo cliente/servidor onde oferece respostas rápidas às consultas, criando um microcubo na máquina cliente ou no servidor. Estas respostas são dadas a partir de questões feitas pelo gerenciador do sistemas.

O OLAP é freqüentemente utilizado para integrar e disponibilizar informações gerenciais contidas em bases de dados operacionais, sistemas ERP (*Enterprise Resource Planning*) e CRM (*Customer Relationship Management* – Administração e Relação de Cliente), EIS (*Executive Information System* - Sistema de Informação Executivo) e *Data Warehouses* (Armazém de Dados) (DWBRASIL).

## 5.7 LINHAS DO OLAP

O OLAP é uma arquitetura de aplicação multidimensional, existem vários conceitos para se extrair as informações de maneira correta. A seguir será demonstrada algumas linhas do OLAP.

- A Visão conceitual multidimensional: enfatiza a forma como o usuário "vê" dados sem impor que os dados sejam armazenados em formato multidimensional;
- Transparência: localização da funcionalidade OLAP deve ser transparente para o usuário, assim como a localização e a forma dos dados a ser utilizados;
- Facilidade de Acesso: acesso a fontes de dados homogêneas e heterogêneas deve ser transparente;
- Desempenho de consultas consistente: não deve ser dependente do número de dimensões;
- Arquitetura cliente/servidor: produtos devem ser capazes de operar em arquiteturas cliente/servidor;
- Dimensionalidade genérica: todas as dimensões são iguais;
- Manipulação dinâmica de matrizes esparsas: produtos devem lidar com matrizes esparsas eficientemente;
- Suporte multi-usuário;
- Operações entre dimensões sem restrições;
- Manipulação de dados intuitiva;

- Relatórios/consultas flexíveis;
- Níveis de agregação e dimensões ilimitados: ferramentas devem ser capazes de acomodar 15 a 20 dimensões.

Com a arquitetura DOLAP (Desktop On Line Analytical Processing) ganha-se com pouco tráfego na rede, porque todo processamento OLAP se dá na máquina cliente, e maior agilidade de análise, além do servidor de banco de dados não ficar sobrecarregado, sem incidir em problemas de escalabilidade. A desvantagem é que o micro-cubo não pode ser muito grande, caso seja a máquina cliente não suporte a função em configuração e/ou pode se levar um tempo grande para se fazer análise (CIELO).

As ferramentas ROLAP (*Relational On Line Analytical Processing* – Processamento Analítico Relacional *On Line*) têm acesso aos dados e análise OLAP com uma arquitetura diferente. Envia-se a consulta ao Banco de dados relacional e processa no próprio servidor de banco de dados. A vantagem é que se pode analisar grandes quantidades de dados e a desvantagem é que se várias pessoas utilizarem ao mesmo tempo pode comprometer muito da performance do servidor ou até mesmo ocorrer seu travamento (CIELO).

A ferramenta MOLAP (*Multidimensional On Line Analytical Processing* - Processamento Analítico Multidimensional *On Line*), tem uma arquitetura que processa com um servidor multidimensional, com isso o acesso aos dados ocorre diretamente no banco, o usuário trabalha, monta e manipula os dados do cubo diretamente do servidor. Há um melhoramento na performance, mas tem problemas com escalabilidade e tem alto custo (CIELO).

A ferramenta HOLAP (*Hybrid On Line Analytical Processing* – Processamento Analítico Híbrido *On Line*) tem a arquitetura desta ferramenta a união de tecnologias que há entre o MOLAP e o ROLAP, a grande vantagem é que pode pegar o que há de melhor destas duas ferramentas que é a alta performance do MOLAP com a escalabilidade do ROLAP (CIELO).

## 6 BUSINESS INTELLIGENCE

As ferramentas de BI podem fornecer uma visão sistêmica do negócio e ajudar na distribuição uniforme dos dados entre os usuários, sendo seu objetivo principal transformar grandes quantidades de dados em informações de qualidade para a tomada de decisões. Através delas, é possível cruzar dados, visualizar informações em várias dimensões e analisar os principais indicadores de desempenho empresarial (BATISTA, 2004). Essa facilidade, considerando-se as características dessas ferramentas, pode contribuir diretamente para as funções da área de controladoria na obtenção, análise e comunicação do recurso informação aos gestores, além de permitir a essa área o estreito monitoramento das atividades da empresa como um todo.

Define-se BI como guarda-chuva conceitual, visto que se dedica à captura de dados, informações e conhecimentos que permitam às empresas competirem com maior eficiência em uma abordagem evolutiva de modelagem de dados, capazes de promover a estruturação de informações em depósitos retrospectivos e históricos, permitindo sua modelagem por ferramentas analíticas. Seu conceito é abrangente e envolve todos os recursos necessários para o processamento e a disponibilização da informação ao usuário (BARBIERI, 2001).

A história do Business Intelligence, segundo (Serra, 2002), teve início na década de 70, quando alguns produtos de BI foram fornecidos para os analistas de negócios. O problema destes produtos era que exigiam muito esforço de programação, além de não disponibilizar informação em tempo hábil nem de forma flexível, além de um custo elevado de implantação.

Segundo Ramos e Santos (2006) Business Inteligência são o resultado de um processo que começa com a coleta de dados. Esses dados são organizados e transformados em informações, que depois de analisadas e contextualizada se transforma em inteligência. Essa, por sua vez, quando aplicada a processos de decisão geram vantagem competitiva para a organização. Essas tecnologias tornam-se um componente essencial de grande importância para as organizações, ela tem como objetivo coletar dados tanto interno quanto externo que seja favorável e que é

de suma importância para que os gerentes possa agir e criar condições que facilite a construção necessária para planejamento, obtendo soluções para problemas e desafios, que isso venha garantir uma tomada de decisão satisfatória. Um BI tem como objetivo em uma organização aumentar a sua competitividade direcionando a empresa para um no mercado futuro.

Segundo Geiger (2001), BI é todo conjunto de processos e estrutura de dados utilizados para compreender o ambiente do negócio da empresa, objetivando apoiar a análise estratégica e a tomada de decisão. Os componentes principais do BI são o DW e *Data Mart*, interfaces para apoio à decisão e processos para coletar dados, inseri-los em um DW e entregá-los à comunidade do negócio. O maior benefício do ambiente de BI, além da razão técnica representada pela simplificação e visão global dos dados consolidados em um DW, é a possibilidade de se pensar sobre as questões estratégicas da organização e a projeção do futuro por meio de cenários fundamentados em dados precisos do presente e do passado.

Para Ramos e Santos (2006) O conhecimento organizacional é o suporte da inteligência coletiva, a qual permite à organização manipular o conhecimento de que dispões para resolver problemas, tirar partido de oportunidades, e influenciar o ambiente em seu próprio benefício. Benefícios de um Sistema formal de Business Intelligence: antecipar mudanças no mercado, antecipar ações dos competidores, descobrir novos ou potenciais competidores, aprender com os sucessos e as falhas dos outros, conhecer melhor suas possíveis aquisições ou parceiros, conhecer novas tecnologias, produtos ou processos que tenham impacto no seu negócio, entrar em novos negócios, rever suas próprias práticas de negócio, auxiliar na implementação de novas ferramentas gerenciais.

Ramos e Santos (2006), todo este conhecimento é fundamental para apoiar o processo de tomada de decisão. A informação gerada pelas aplicações informáticas disponibilizada aos gestores um conjunto de indicadores sobre o negócio, que lhes dão indicações do que aconteceu no passado e lhe permitem traçar cenários para o futuro. Business Intelligence é a utilização de uma série de ferramentas para coletar, analisar e extrair Informações, que serão utilizadas no auxílio ao processo de gestão e tomadas de decisão.

Os sistemas de Business Intelligence estão tradicionalmente associados a várias tecnologias. Fazem parte dos pacotes de Business Intelligence existente: o *Data warehouse* (DW), Sistemas de gestão integrados (ERP), ferramentas OLAP, *Data Mining*, *Data Mart*, CRM, Metadados, dentre outros. Como sabemos é o principal armazenamento de dados, um repositório que interage com diversos outros repositórios permitindo e facilitando informações relevantes auxiliando gerentes para a tomada de decisão. Este repositório tem seus sistemas integrados que podem ser analisadas com ferramentas específicas para obter resultados relevantes como OLAP (*On-line Analytical Processing*) e *Data Mining*, essas ferramentas tem como objetivo a extração, transformação e manipulação e carregamento de dados. As ferramentas OLAP (*On-Line Analytical Processing*), apresenta diversas funções de manipulação dos cubos.

Ramos e Santos (2006) Os sistemas Business Intelligence contribuem para aumentar, a inteligência coletiva da organização na medida em que facilitam a construção do conhecimento necessário para planejar e implementação soluções para problemas e desafios percebidos como podendo ameaçar a sobrevivência de bem-estar da organização. É o resultado desse esforço humano para reduzir a incerteza, criar significado e controlar a sua própria existência. As organizações definem o significado da ação dos seus membros através de regras, procedimentos e conhecimento compartilhado. Esta ação é organizada em torno de metas e propósitos comuns, os quais são definidos de acordo com as exigências e condicionamentos do ambiente externo. Ao longo do período de existência das organizações, grande parte do esforço humano é direcionada para a manutenção do ajuste entre as metas organizacionais e as transformações ambientais, e para a aquisição dos recursos necessários para prever e influenciar essas transformações.

Uma empresa busca, através das ferramentas de BI, acessar e integrar indicadores de performance e tendência com diferentes graus de sintetização, capazes de auxiliá-la na condução do negócio. Nessa linha, McGeever (2000) apresenta uma outra visão do conceito de BI, referente ao tipo de informação segmentada que o gestor de um ramo de negócios procura.

Fazendo uso de BI, o usuário pode formatar suas próprias informações, assim como pode conectá-las a outras a fim de obter uma melhor análise e um melhor

resultado com o seu uso, ou seja, ele pode tornar-se mais independente na busca de informações adequadas, não precisando de relatórios distintos.

## 6.1 VISUALIZAÇÃO DE DADOS

A Visualização é a apresentação de dados numéricos em formato gráfico. Apresentando os dados de uma forma visualmente apelativa, o utilizador é confrontado com uma compreensão rápida e qualitativa da informação.

Os dados (e, por consequência, a informação) desempenham um papel importante nas operações diárias de uma organização. Os utilizadores empresariais encontram dados numéricos informativos, mas difíceis de interpretar sem uma análise mais aprofundada. Os mesmos dados, quando apresentados em formato de gráfico (gráfico de barras, queijo, etc...), comunicam a informação de forma mais eficaz. É ainda importante salientar que os elementos numéricos originais mantêm-se inalteráveis na visualização de dados.

Uma visualização de dados bem-sucedida tem as seguintes três características: Exatidão – apresentação dos dados (com a seleção apropriada de escalas e tipo de gráficos); Elementos estéticos – experiência de utilização agradável; Flexível – permite alteração de estilos visuais.

O uso de ferramentas de visualização de dados em ferramentas de análise de dados tradicionais, centenas de linhas de dados podem ser representadas em uma única tela, através de gráficos, mapas e alertas, fazendo com que muitos problemas não sejam percebidos, pois questões incomuns costumam ficar automaticamente de fora. O trabalho de uma ferramenta de visualização de dados é apontar o que existe de anormal. Os gráficos de visualização de dados podem identificar singularidades em cada minúcia, detalhe ou particularidade do dado, isolá-lo e investigar sua fonte de origem mais rapidamente do que em métodos analíticos tradicionais.

Sendo assim para usuários com perfil de Executivos ou Gerentes se forem apresentados a um painel com indicadores de desempenho, alertas, cores e gráficos, esses usuários terão muito mais facilidade em visualizar a informação e, conseqüentemente, aderir e reconhecer os benefícios de um projeto de BI.

## 7 ESTUDO DE CASO

O estudo de caso apresenta a aplicação de BI na empresa FTC EAD, através da utilização das ferramentas disponíveis no Pentaho agregadas da ferramenta DTS do *Microsoft SQL Server*. Neste capítulo é apresentada uma descrição geral da plataforma Pentaho, juntamente com algumas definições e processos que se faz necessário para o entendimento do estudo de caso. Logo após, é abordado como pode ser obtida e sua licença de uso. Ainda, descreve-se a arquitetura do Pentaho e demonstra-se, através de um estudo de caso, como funciona o processo de desenvolvimento de um ambiente de BI a partir da plataforma Pentaho. O processo de criação do DW inicia pela análise de um sistema OLTP – (*On-line Transaction Processing*) de uma instituição de ensino superior privada. É feita a identificação de alguns dados atualmente utilizados para tomada de decisão na empresa e, a partir desta análise, é realizada a modelagem dimensional do DW. Logo após, é programado o processo de extração, transformação para posterior carga no DW. Ao final, com a modelagem dimensional criada e o banco de dados do DW devidamente preenchido, é feita a configuração do cubo para possibilitar a análise dos dados na plataforma Pentaho e em seguida a visualização dessas informações geradas para consultas dos gestores.

### 7.1 PLATAFORMA PENTAHO DE BUSINESS INTELLIGENCE

A plataforma Pentaho é um conjunto de software open source para criação de soluções de BI (*Business Intelligence*). Ela possui ferramentas para atender ao processo de criação de soluções de BI de ponta-a-ponta, integrada à uma gama de opções para banco de dados e outras ferramentas, conforme ilustra a Figura8 (TEMATEC, 2008).

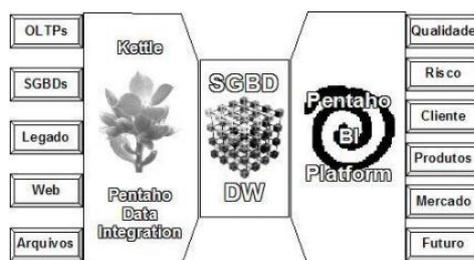


Figura8 – Diagrama de uma solução de BI

FONTE: TEMATEC (2008)

Quando uma organização precisa tomar uma decisão é indispensável ter dados corretos e disponíveis para consulta. Para conseguir isso, ela deve tratar e consolidar as informações armazenadas nos sistemas e fontes de dados que apóiam seu negócio em um repositório centralizado, criando “uma única versão da verdade”, limpa e confiável (processo ETC). Depois, pessoas que entendem do negócio da empresa devem ter acesso a esse repositório e, usando ferramentas de visualização e exploração de dados, interpretá-los para finalmente tomar uma decisão (TEMATEC, 2008).

O conjunto destes componentes de softwares, dados, operações e processos usados para atender uma necessidade específica, para tomar uma decisão, é chamado de solução.

A Pentaho Inc. integrou e promoveu o desenvolvimento de várias ferramentas open source que fornecem os recursos necessários para criação de soluções de BI. Esse conjunto é conhecido por Pentaho Open BI Suite e inclui ferramentas para consolidar dados de fontes diversas, criar interfaces visuais para exploração desses dados e montar soluções para necessidades específicas (TEMATEC, 2008).

A plataforma Pentaho BI é distribuída como código aberto, através da Licença Pública da Pentaho (PPL - *Pentaho Public License*). A PPL é uma licença para software livre de código aberto, similar à Licença Pública do Mozilla (versão 1.1) (PENTAHO, 2007).

A plataforma Pentaho BI pode ser baixada gratuitamente, através do endereço <http://www.sourceforge.net/projects/pentaho>. Existem várias versões disponíveis. Através do endereço <http://www.pentaho.org> pode-se ter acesso a um wizard que ajuda a determinar a versão correta para cada caso.

Para iniciantes, é recomendável fazer o download de uma versão de demonstração (por exemplo, a *pentaho\_demo\_hsqldb-1.6.0.GA.863*). Estas versões incluem uma aplicação pré-configurada em um servidor Jboss, juntamente com alguns exemplos e dados em um servidor *Sun Microsystems* JRE.

## 7.1.2 ARQUITETURA

O projeto Pentaho BI é constituído de um conjunto completo de ferramentas de BI e bibliotecas que fornecem funcionalidades de BI aos desenvolvedores. É uma solução com suporte a relatórios, análises, *data mining* e *workflow*, através de uma série de componentes que podem ser distribuídos juntos ou separados (PENTAHO, 2007).

O servidor roda de acordo com o padrão de servidores Java, tais como Apache TomCat e JBoss. Ele utiliza um método de desenvolvimento, distribuição e suporte que torna possível o modelo de negócios open source (PENTAHO, 2007).

A Figura9 apresenta a arquitetura do Pentaho, a qual é composta por componentes de integração de dados, infra-estrutura e apresentação dos dados e a origem destes dados.

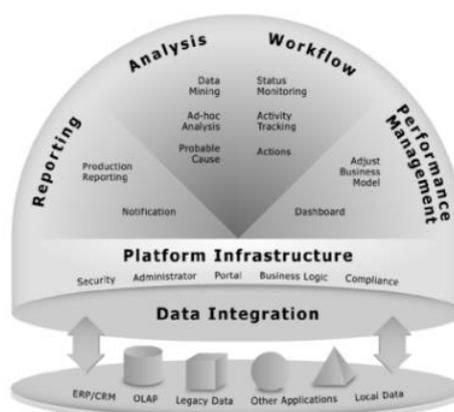


Figura9 – Arquitetura do Pentaho BI

FONTE: PENTAHO (2007)

O Pentaho BI abrange, principalmente, as seguintes áreas de aplicação (PENTAHO, 2007):

- Relatórios (Reporting): Fornece desde simples relatórios em uma página web, até relatórios de alta qualidade, tais como relatórios de indicações financeiras e relatórios ricos em conteúdos, como tabelas, gráficos, entre outros;

- Análises (Analysis): Permite consultas, exploração interativa com operações slice-and-dice, drill-down e pivoting. Inclui front-end gráfico para exploração dos cubos OLAP;
- Painéis (Dashboards): Reúne relatórios, análises e outras exposições em um único local para simplificar o acesso, podendo ser customizado por usuário ou assunto;
- Data mining: Descobre relacionamentos ocultos nos dados, que podem ser utilizados para otimizar os processos de negócio e prever resultados futuros. Permite que os resultados sejam exibidos em um formato de fácil entendimento ao usuário;
- Workflow: Liga diretamente as medidas de desempenho de negócio aos processos, promovendo um ciclo contínuo de melhorias.

Todos os componentes da plataforma são de código aberto, o projeto Pentaho BI oferece uma solução que pode ser utilizada por desenvolvedores Java, os quais podem utilizar os componentes do projeto para montar rapidamente soluções BI sob medida. O Pentaho pode ser utilizado por empresas desenvolvedoras de software para adicionarem as funcionalidades de BI em seus produtos. O projeto pretende, ainda, permitir que usuários finais tenham acesso a soluções de BI com a qualidade dos softwares comerciais tradicionais, porém com um custo bem mais acessível.

## 7.2 AMBIENTE OPERACIONAL DA EMPRESA

O sistema transacional deste estudo de caso é da empresa FTC EAD uma instituição de ensino superior privada onde seus cursos de graduação e pós-graduação são ministrados a distancia. Os cursos contemplam a mediação pedagógica por meio de: *Vídeostreaming* com retorno por voz, realizadas pelos docentes, geradas pela FTC EAD Salvador para as salas de recepção localizadas nas Unidades Pedagógicas espalhadas por todo o Brasil, propiciando a interatividade entre o docente e os aproximadamente 50 estudantes de cada turma. As aulas são monitoradas remotamente por professores assistentes, por meio de plataforma de aprendizagem, da Internet e através das salas de atividades *online*,

gerando a interatividade necessária para a troca de conhecimento entre os alunos e professores.

O sistema OLTP, que faz o controle de alunos é de grande importância para a empresa, mas apresenta carência de informações estratégicas e não responde eficientemente a questões que requerem cruzamento de dados com uma visualização. A exemplo desses cruzamentos temos: qual a quantidade de alunos por período do ano, total de alunos inadimplentes por mês, total de alunos por Estado, Região. No levantamento das informações, fica claro a necessidade de uma aplicação que apresente as informações estratégicas aos dirigentes da empresa, diretamente da origem dos dados, para que possam visualizá-las e utilizá-las na tomada de decisão.

Os dados do sistema OLTP estão disponíveis no SGBD (Sistema Gerenciador de Banco de Dados) SQL Server, versão 2000. Para facilitar a análise, os dados de interesse deste estudo de caso estão representados no diagrama apresentado na Figura10.

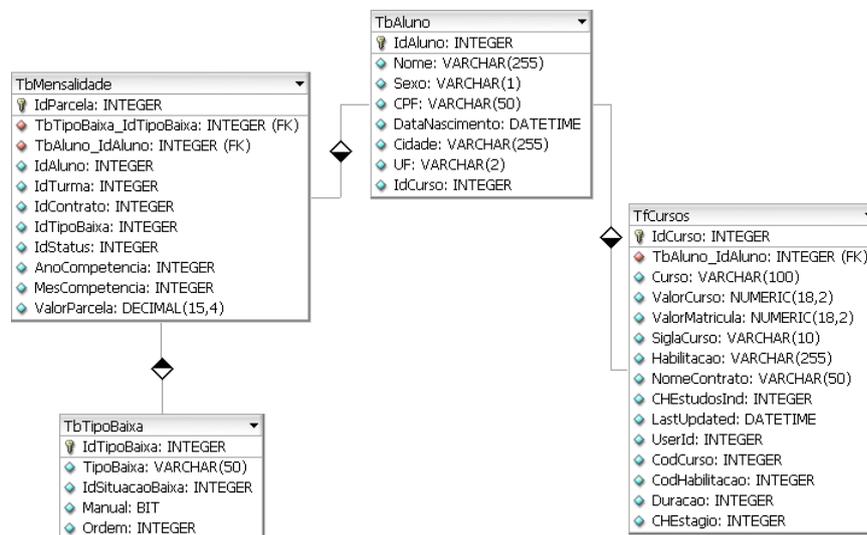


Figura10 – Diagrama físico do banco de dados do estudo de caso.

### 7.3 CRIAÇÃO DO DW (*DATA WAREHOUSE*)

No levantamento das informações, fica claro a necessidade de uma aplicação que apresente as informações estratégicas aos dirigentes da empresa, diretamente da origem dos dados, para que possam visualizá-las e utilizá-las na tomada de decisão.

A arquitetura proposta para esse estudo de caso é apresentada na Figura11. Ela ilustra o sistema transacional ligado em um servidor OLAP, com o acesso local (*localhost*). O servidor OLAP é utilizado para construir e visualizar o *Data Warehouse*. Esta arquitetura é ideal para a demonstração deste estudo de caso, porque utiliza o sistema, bem como o servidor de banco de dados já existente, e requer um investimento mínimo em hardware e software.

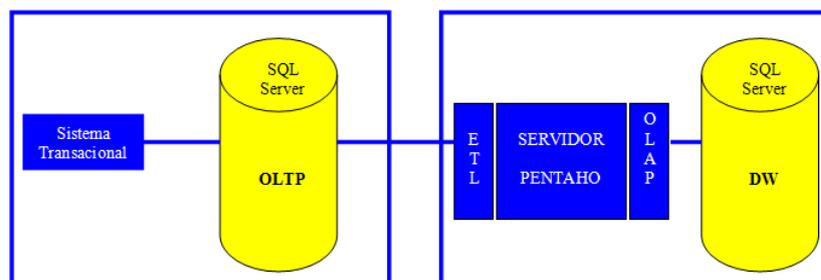


Figura11 - Arquitetura sistema transacional

Com isso é apresentada a fase de projeto para a construção do *Data Warehouse*, segundo as nove etapas propostas por Kimball (2002):

- Primeira etapa - seleção do processo do negócio a ser modelado:  
Após reunião com os gerentes da empresa, entre os processos de negócio existentes, escolheu-se os alunos por serem os clientes da empresa e de fundamental importância para os negócios e principal fonte geradora de receitas. Assim sendo, o *Data Warehouse* proposto compreende um cubo OLAP com dados sobre os alunos e parcelas;
- Segunda etapa – declaração do grão do processo de negócio:

Escolheu-se que a granularidade, ou nível de detalhes, forneça informações sobre parcelas por período (ano, mês, região, estado, cidade). Alunos (Adimplentes, Inadimplentes, por curso, por sexo, por faixa etária. Pode-se observar o dicionário de indicadores no APÊNDICE A deste trabalho.

- Terceira etapa – escolha das dimensões:

Com base na definição da granularidade, as dimensões definidas foram:

Aluno, Tempo, Mensalidade e Geografia. Nesta etapa, também são definidos os atributos das dimensões, os quais correspondem à descrição dos fatos e estabelecem os níveis de detalhe que a informação pode ser apresentada.

- Quarta etapa – identificação dos Fatos:

Com a finalidade de analisar as medidas de desempenho, escolheu-se como fatos: a quantidade de alunos e o valor da parcela do curso.

- Quinta etapa – armazenar os dados na tabela de Fatos:

Esta etapa, descrita no tópico 8.4, é suportada pelo processo de Extração, Transformação e Carga (ETC), que extrai os dados da base transacional, passa pelo processo transformação e carrega para a base dimensional.

- Sexta etapa – preenchendo as tabelas de dimensão:

Esta etapa também está descrita no tópico 8.4, que trata sobre o processo de ETC;

- Sétima etapa – preparar dimensões para suportar mudanças:

Até o presente momento, a única necessidade de alteração prevista para as tabelas de dimensões é a inclusão de novos valores provenientes das tabelas de cadastro do sistema transacional. Cabe ressaltar que não há exclusão de registros em um banco de dados de um *Data Warehouse*;

- Oitava etapa – escolha da duração do banco de dados:

Nesta etapa definiu-se que o período de duração do banco de dados do *Data Warehouse* é de 4 anos. A definição deste tempo leva em consideração o período

de formação dos alunos, principal foco pela empresa deste estudo de caso. Desta forma após 4 anos, os dados no banco de dados do DW podem ser retirados da base de dados e armazenados em dispositivos de armazenamento secundário, possibilitando sua recuperação caso seja necessário;

- Nona etapa – definir espaço de tempo de extração e carga de dados:

O espaço de tempo em que os dados devem ser extraídos do banco de dados do sistema transacional e carregados no *Data Warehouse* é mensal. O processo de extração e a programação do mesmo são detalhados no próximo tópico.

Seguindo as nove etapas propostas por Kimball (2002), aqui é apresentado o Modelo Estrela do banco de dados para o DW proposto conforme Figura12, composto por uma tabela de Fatos e quatro tabelas de Dimensões:

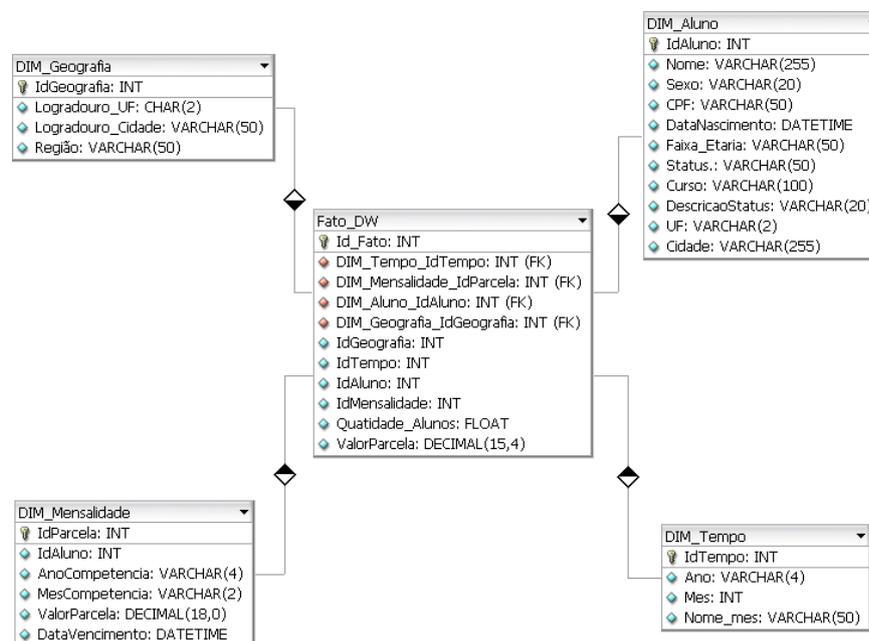


Figura12 – Diagrama físico do banco de dados para o *Data Warehouse*

## 7.4 PROCESSO ETL

Os detalhes do povoamento do *Data Warehouse* através do processo de ETL—*Extraction, Transformation and Load* (Extração, Transformação e Carga) são explicados neste tópico. Para tanto, foi utilizado no projeto a ferramenta DTS Designer que é distribuída junto com o *SQL Server Enterprise 2000* e auxilia a integração de dados e a construção do DW, como ilustra a Figura13.

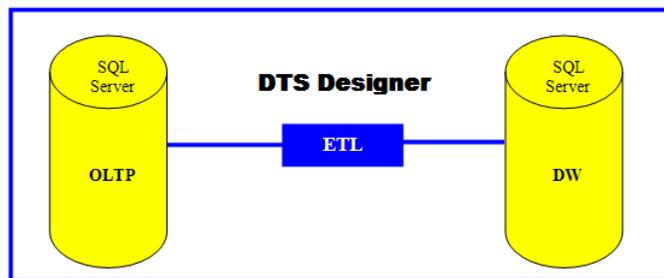


Figura13 - Arquitetura do processo de ETL

A seguir, são exemplificados os passos necessários para passar os dados contidos no banco de dados OLTP para o banco de dados do DW, através do processo de ETL.

A Figura14 mostra a interface da ferramenta DTS Designer, responsável pelo processo de ETL. A base do processo de ETL no DTS Designer é a criação de transformações, ou seja, para cada tabela do DW, é criado um processo de transformação que lê os dados da fonte (OLTP), faz um pré-processamento e grava na tabela do DW.

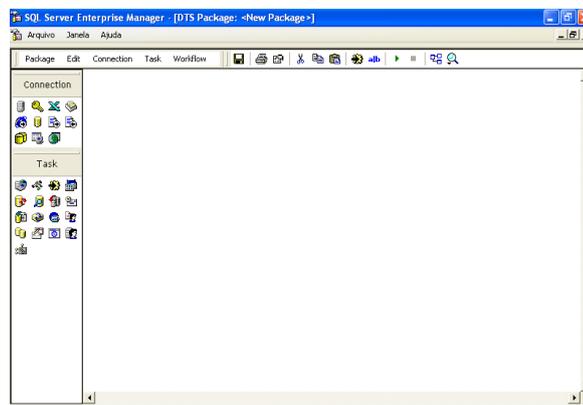


Figura14 - DTS Designer

O *Data Transformation Services* no Microsoft SQL Server 2000 é uma injeção de dados de alto desempenho. É uma ferramenta para copiar, mover, consolidar, limpar e validar dados. A injeção de dados carrega cada linha da origem de dados, manipula os valores nessa linha e insere a linha no destino de dados. O DTS também é uma ferramenta de Rápido Desenvolvimento de Aplicações (*Rapid Application Development – RAD*) para programação orientada a dados. Ele fornece um conjunto abrangente de ferramentas de manipulação de dados organizado em um ambiente de desenvolvimento que é tanto conveniente como poderoso (PETERSON, 2001).

A ferramenta DTS Design foi escolhida para esse trabalho devido a seu fácil manuseio e a mesma já se encontrar no SGBD do OLTP da empresa deste estudo de caso. Com o DTS Designer podemos utilizar qualquer tipo de conexão mostrada na primeira parte do painel localizada à esquerda a interface (*Connection*). Na segunda parte do painel (*Task*) podemos escolher qual a tarefa será realizada. Com estas opções podemos montar transformações simples como uma cópia de linhas entre duas tabelas de um mesmo SQL Server ou transformações complexas como a obtenção de um arquivo .txt armazenado em um servidor de ftp, a pré-seleção de somente algumas linhas deste arquivo, o mapeamento de algumas colunas do arquivo .txt para uma base localizada em um banco DB2 e outras colunas para um Oracle, a notificação de qualquer tipo de problema que ocorra na execução dos passos e ainda um fluxo gráfico de todo este processo.

Tudo deve começar pelo entendimento do processo: quais são as origens dos dados, quais as transformações, validações, verificações, checagens, etc que estes dados provenientes da origem podem sofrer e qual o destino destes dados.

A realização do ETL desse estudo de caso usando a ferramenta DTS Designer se deu da seguinte forma:

Inicialmente vamos criar um pacote através do DTS Designer como visto na Figura15. Feito isso, devemos clicar e arrastar o primeiro ícone da parte superior do painel à esquerda da interface (*Connections*). O nome desta opção é “*Microsoft OLE DB Provider for SQL Server*”, já que a origem dos nossos dados se encontra em uma tabela no SQL Server.

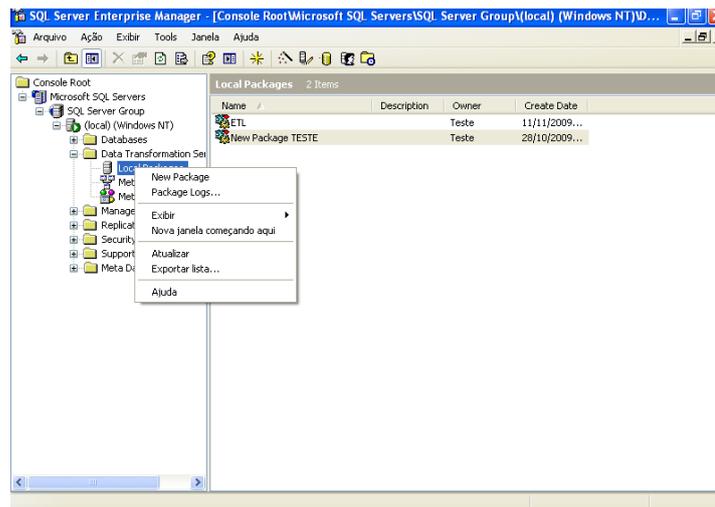


Figura15 – Novo Pacote

Isso fará com que a tela de edição de conexão seja mostrada. A configuração foi feita como mostrado na Figura16. Outros tipos de conexão podem solicitar outros tipos de informação, como o nome do DSN caso seja feita uma conexão através do ODBC.

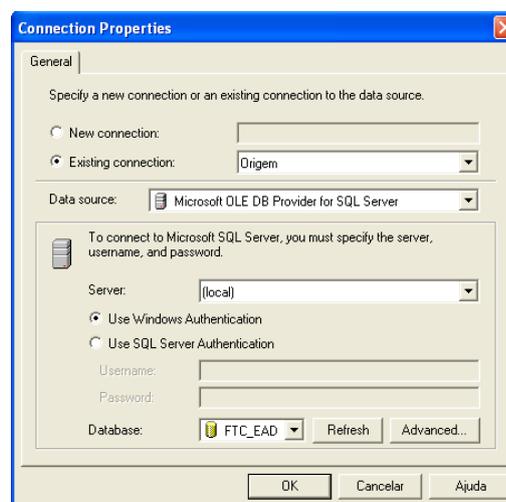


Figura16 – Propriedades de Conexão

Para o tratamento de cada dimensão foi usada *procedures* que podem ser visualizadas no APÊNDICE B deste trabalho criadas no próprio SQL Server 2000 e executadas a partir da ferramenta *Execute SQL Task*, a caixa de diálogo *Execute SQL Task Properties* do *Package Designer* é mostrada na Figura17.

Ela oferece três maneiras de configurar a instrução de SQL:

- Escrevendo-a na caixa de instrução de SQL
- Utilizando o botão *Browse* para carregar a consulta de um arquivo.
- Utilizando o botão *Build Query* para criar uma consulta utilizando uma interface visual.

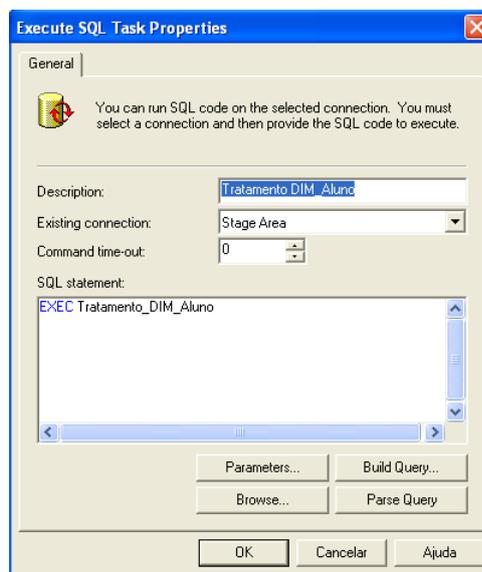


Figura17 – Executar Propriedades da Tarefa

Deve-se selecionar uma conexão de DTS que foi definida previamente no pacote. Pode-se fornecer uma descrição de um valor para o tempo limite de comando, também possui botões para analisar sintaticamente a consulta e fornecer parâmetros para ela.

As *Procedures* criadas fazem o tratamento dos dados oriundos do OLTP e inserem os mesmos em tabelas temporárias que agem como *Stage Area* para o DW com mostrado no esquema de ETL na Figura18.

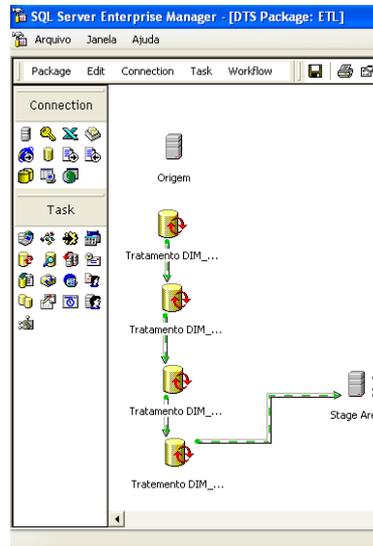


Figura18 - Esquema de ETL

Após todos os dados estarem tratados e inseridos na *Stage Area* os dados são carregados nas dimensões do DW e posteriormente são carregados na Tabela Fato como é mostrado na Figura19 o esquema de ETL completo feito no DTS Design.

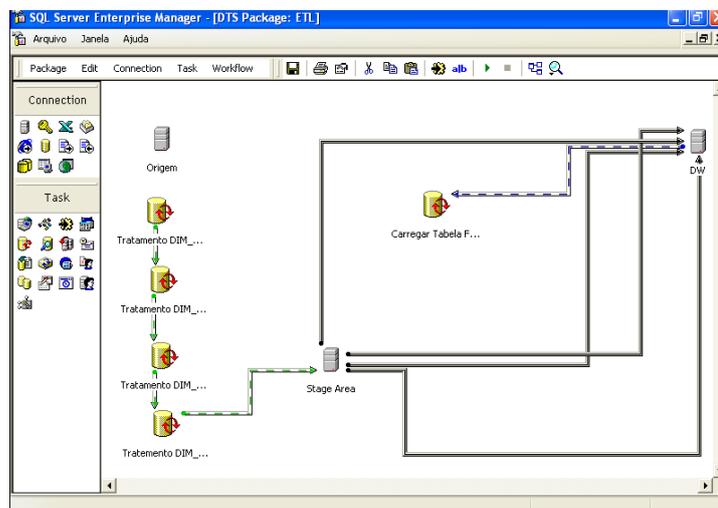


Figura19 - Esquema de ETL completo

## 7.5 CRIAÇÃO DO CUBO

O próximo passo é gerar o cubo com os dados que devem aparecer no módulo Pentaho *Analysis*. O Pentaho *Analysis* é a interface final do usuário, desenvolvida sobre o *engine Mondrian* OLAP, que é um servidor OLAP, ou seja, o *software* responsável por prover os recursos de gerenciamento de dados de um DW em um ambiente de BI. O Mondrian é configurado através de esquemas, que são arquivos XML criados em uma estrutura específica utilizada pelo *engine*. Neste trabalho, descreve-se a configuração do cubo com a utilização da ferramenta *Schema Workbench*.

O *Schema Workbench* é uma interface de *design* que permite a criação e o teste de esquemas de cubos OLAP (arquivos XML) do Mondrian. Esses modelos XML utilizam as tabelas de Fatos e Dimensões encontradas no banco de dados do *Data Warehouse* criado.

O *Schema Workbench* fornece as seguintes funcionalidades:

- Editor de esquema integrado, apresentando a fonte dos dados no rodapé, para validação;
- Teste das consultas MDX nos esquemas e nas bases de dados;
- Visualização da estrutura da base de dados.

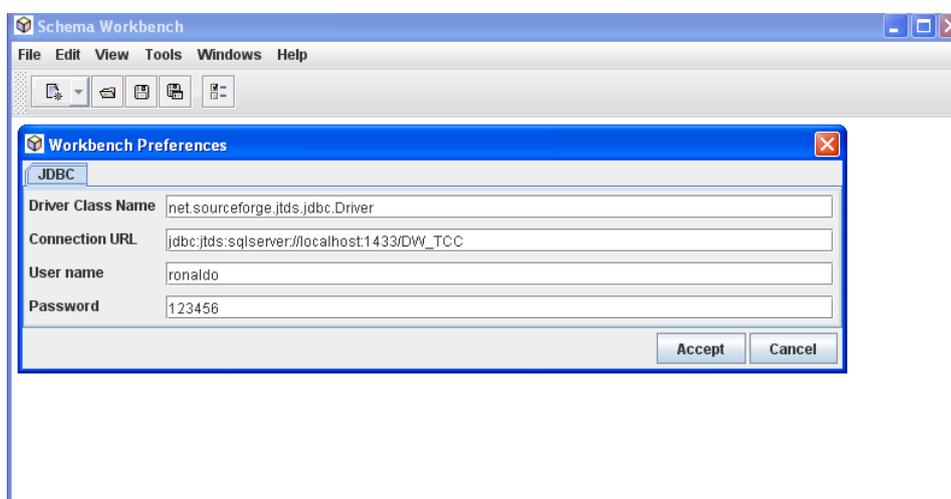


Figura20 - Conexão com o cubo de dados no *Schema Workbench*

As propriedades de conexão com o cubo de dados são configuradas através do menu *Tools > Preferences* como ilustra a Figura20. Devem ser informados:

- O nome da classe de conexão do *driver* do banco de dados;
- A URL da conexão com o banco de dados;
- O usuário e a senha;

Ao criar um novo esquema ou abrir um já existente, o *Schema Workbench* verifica se as tabelas e colunas definidas no cubo existem realmente na base de dados. Para criar ou editar elementos nos esquemas, o *Workbench* valida as modificações nas tabelas e colunas do cubo de dados. O procedimento para configuração das estruturas da *engine* Mondrian são detalhadas através do passo-a-passo seguinte:

- a) Clica-se em *File>New>Schema* para criar um novo esquema;
- b) Com o botão direito do mouse clica-se em *Schema>Add cube* para adicionar um novo cubo como ilustra a Figura21;

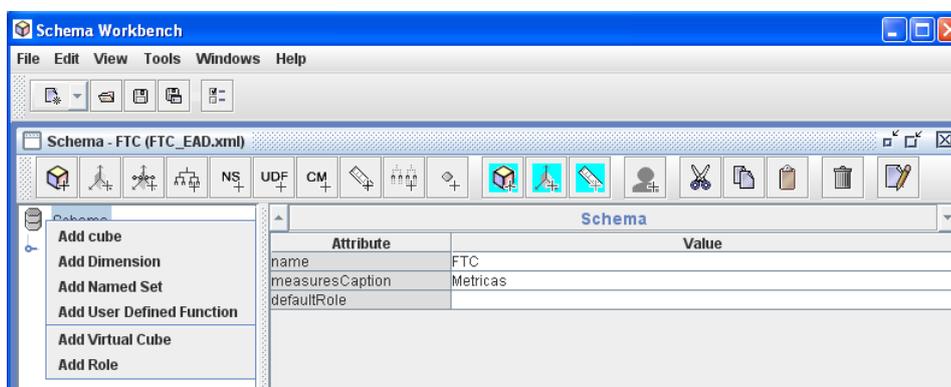


Figura21 – Novo Cubo

- c) Seleciona-se *Table* e, se for caso, seleciona-se a tabela *Fatos*;
- d) Clica-se com o botão direito do mouse no novo cubo criado e seleciona-se *Add Measure*, para criar uma nova medida;

e) Na nova medida criada, escolhe-se um agregador, por exemplo, soma, e seleciona-se a coluna e o tipo de dado como ilustra a Figura22;

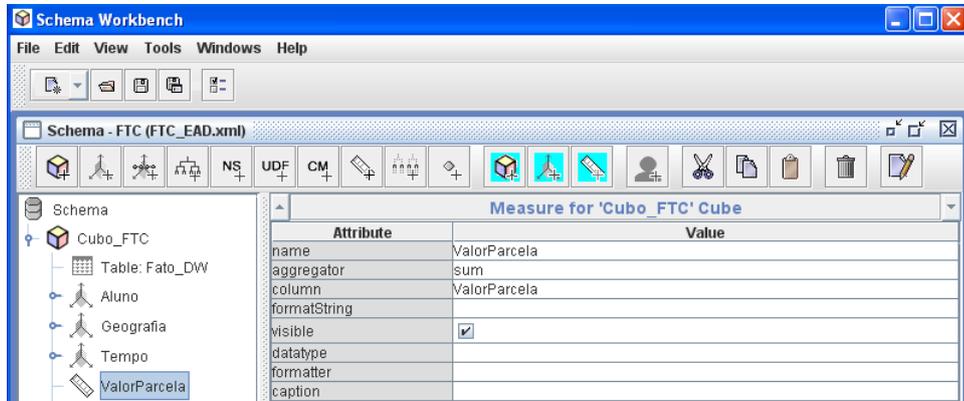


Figura22 - Nova Medida

f) Para criar uma nova dimensão, clica-se com o botão direito e seleciona-se *Add Dimension*;

g) Seleciona-se a chave estrangeira adequada para a dimensão e atribui-se um nome como ilustra a Figura23;

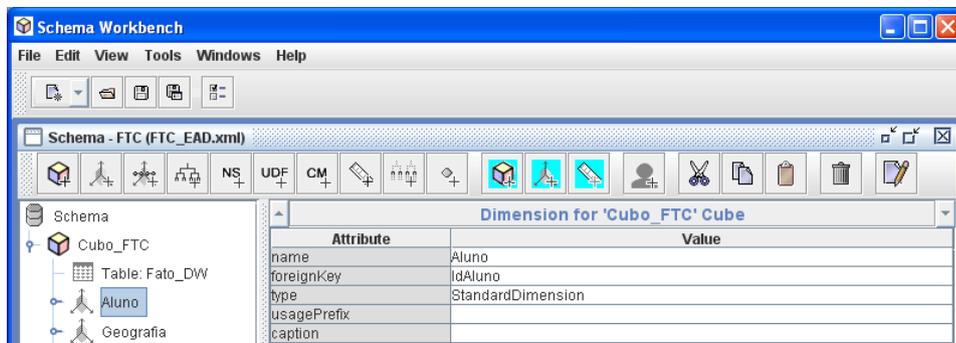


Figura23 - Nova Dimensão

h) Para expandir uma hierarquia nesta dimensão, seleciona-se *Add Hierarchy*;

- i) No nome da hierarquia, seleciona-se *Primary Key* para chave primária e atribui-se um nome a tabela como ilustra a Figura24;

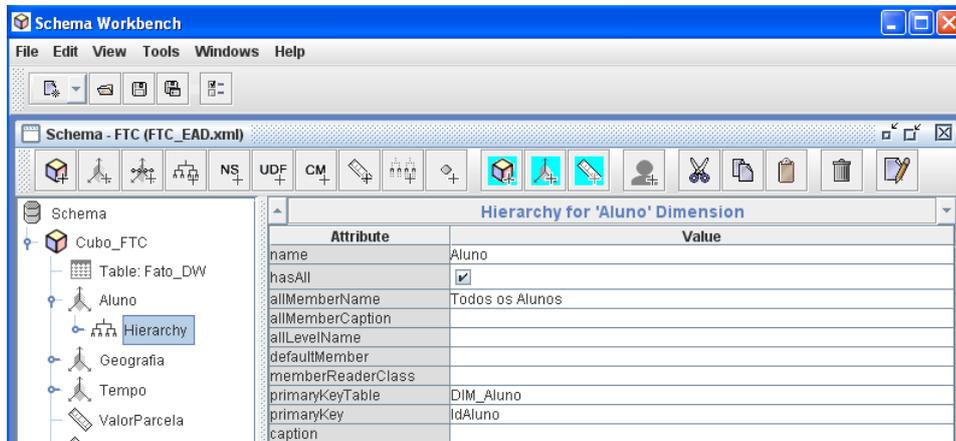


Figura24 - Nova Hierarquia

- j) Clica-se com o botão direito em *Hierarchy* e adiciona-se o nível;
- k) Seleciona-se o nome do nível, a tabela e as colunas como ilustra a Figura25;

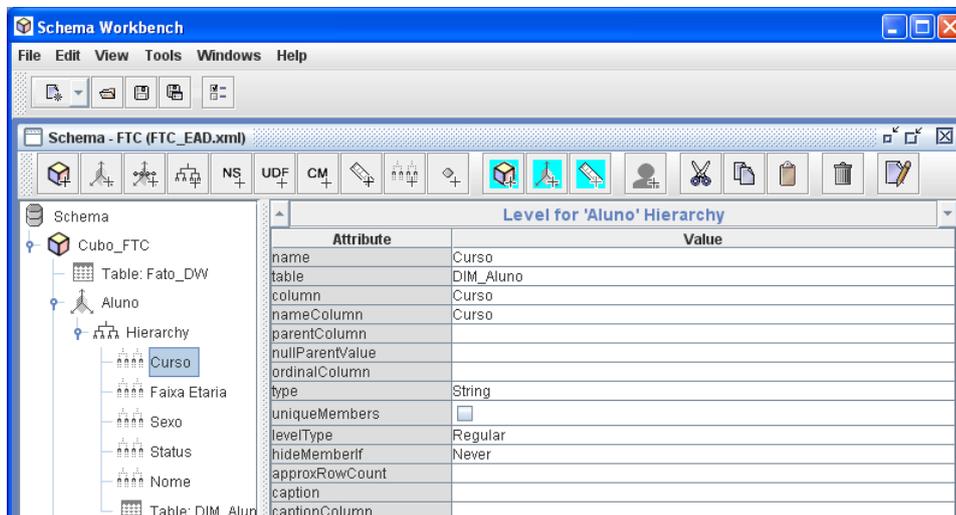


Figura25 – Novo Nível

Ao final, repetir os mesmos procedimentos para atribuir mais dimensões e medidas. A Figura26 mostra o cubo finalizado, com a opção de visualizar o XML do cubo em *View>View XML*.

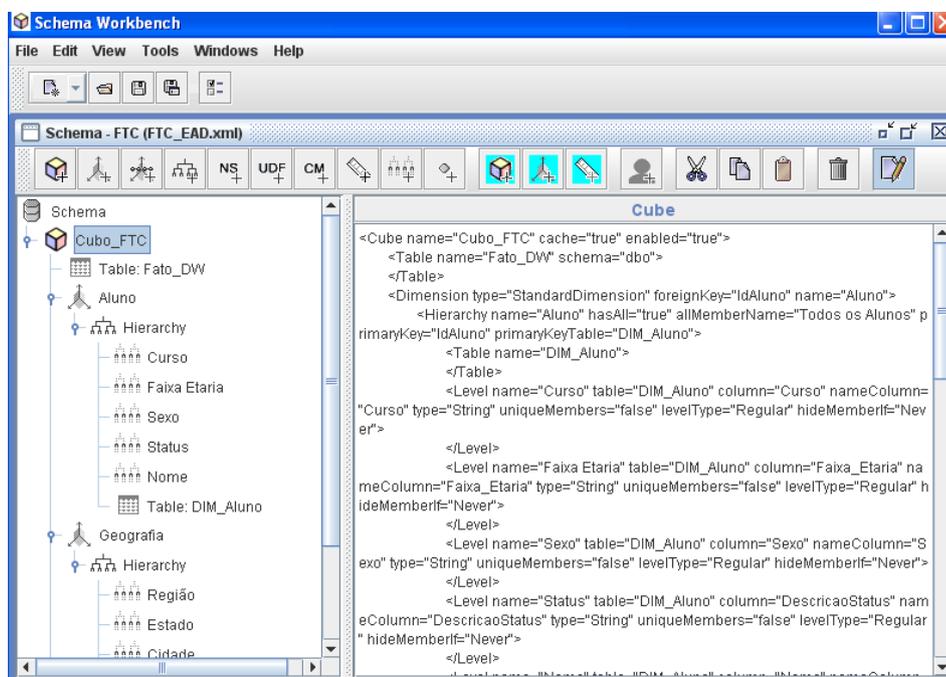


Figura26 - Visualização do cubo criado

## 7.6 VISUALIZAÇÃO

O Pentaho Analysis apresenta as funcionalidades gerais para visualização do cubo configurado. Nas ferramentas disponíveis, destaca-se a opção que exporta os dados extraídos para Excel e a opção *OLAP Navigator*, que oferece filtros altamente customizáveis para relacionar informações do cubo. Outras funcionalidades relevantes são as operações de *Drill Up* e *Drill Down*, que consistem, respectivamente, em agregação de métricas e detalhamento das métricas. Esta opção é indicada pelo sinal de + e - ao lado das colunas.

Os principais recursos da barra de ferramentas são:

a) *OLAP Navigator*. Define o layout geral da *query*, como as colunas e linhas a serem exibidas e os campos para os filtros. Pode-se, através desse ícone, modificar

as estruturas das pesquisas OLAP, visualizando e adicionando ou removendo dimensões ao cubo propriamente dito como ilustra a Figura27;

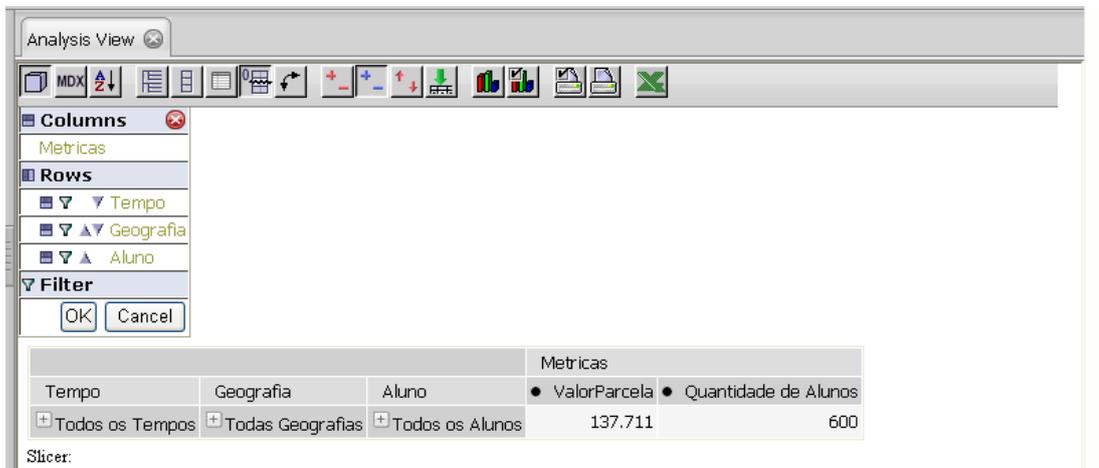


Figura27 – Navegador OLAP

b) MDX: o Analysis utiliza a MDX *Query Language* para definir *queries* multidimensionais como ilustra a Figura28;

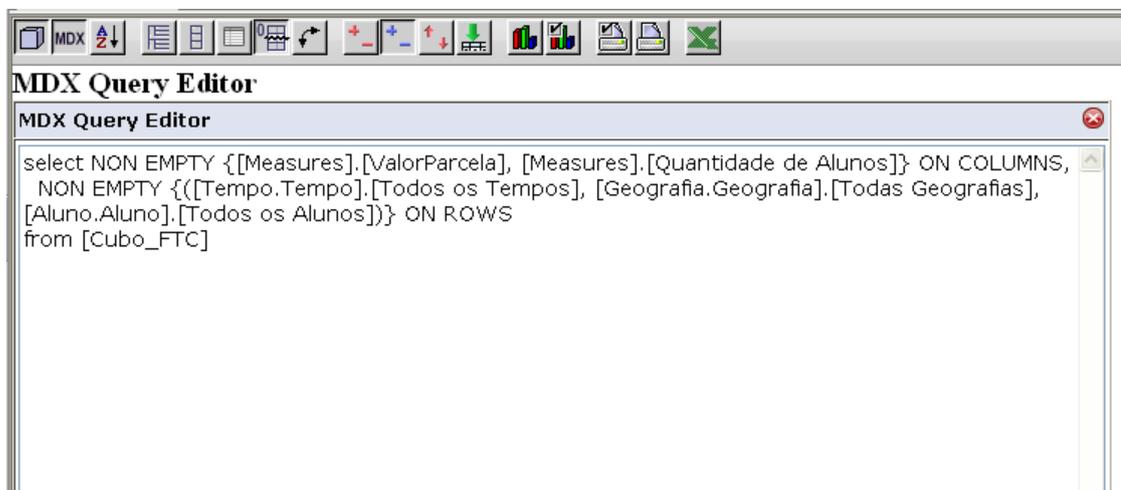


Figura28 – Editor de consulta MDX

c) Config OLAP Table: permite configurar a tabela gerada pelo cubo. Oferece recursos como manter a hierarquia ascendente ou descendente, importantes para uma melhor visualização das pesquisas como ilustra a Figura29.

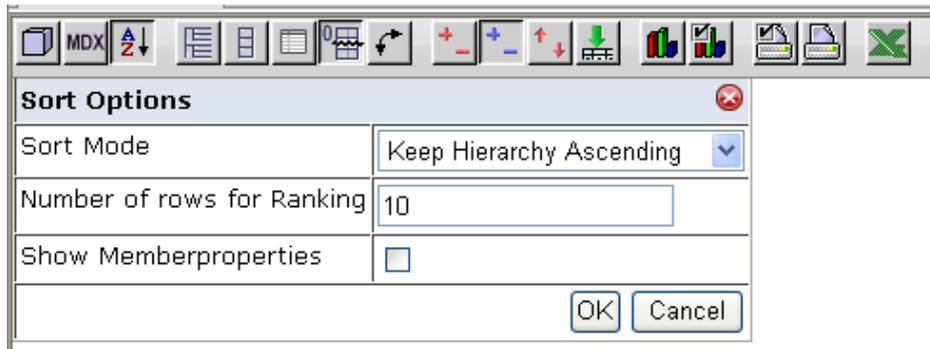


Figura29 – Configuração da tabela OLAP

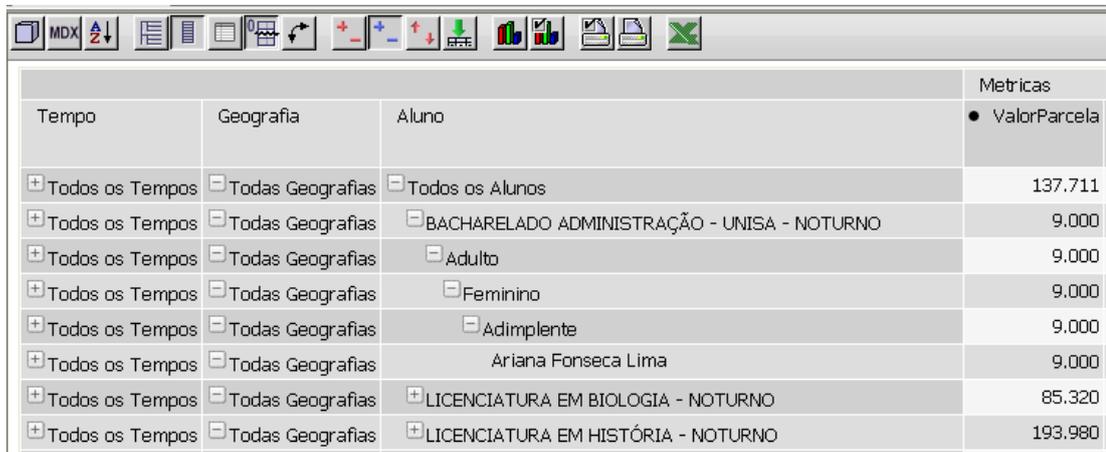
d) *Show Parent Members*: mostra os *parent members* para as linhas ou colunas, ou seja, preenche os espaços vazios na tabela com os membros pais para cada nível nas suas respectivas linhas como ilustra a Figura30;

Tempo		Geografia		Aluno	Metricas				
(All)	Ano	Mes	(All)	Região	(All)	• ValorParcela	• Quantidade de Alunos		
<input type="checkbox"/>	Todos os Tempos		<input type="checkbox"/>	Todas Geografias	<input type="checkbox"/>	Todos os Alunos	137.711	600	
<input type="checkbox"/>	Todos os Tempos		<input type="checkbox"/>	Todas Geografias	<input type="checkbox"/>	Todos os Alunos	45.581	198	
	2005	Dezembro	<input type="checkbox"/>	Todas Geografias	<input type="checkbox"/>	Todos os Alunos	22.701	99	
			<input type="checkbox"/>	Todas Geografias	<input type="checkbox"/>	Nordeste	Todos os Alunos	18.701	83
			<input type="checkbox"/>	Todas Geografias	<input type="checkbox"/>	Norte	Todos os Alunos	4.000	16
		Novembro	<input type="checkbox"/>	Todas Geografias	<input type="checkbox"/>	Todos os Alunos	22.880	99	
	<input type="checkbox"/>		2006	<input type="checkbox"/>	Todas Geografias	<input type="checkbox"/>	Todos os Alunos	91.230	398
	<input type="checkbox"/>		2007	<input type="checkbox"/>	Todas Geografias	<input type="checkbox"/>	Todos os Alunos	900	4

Slicer:

Figura30 – Mostra membros com parentescos

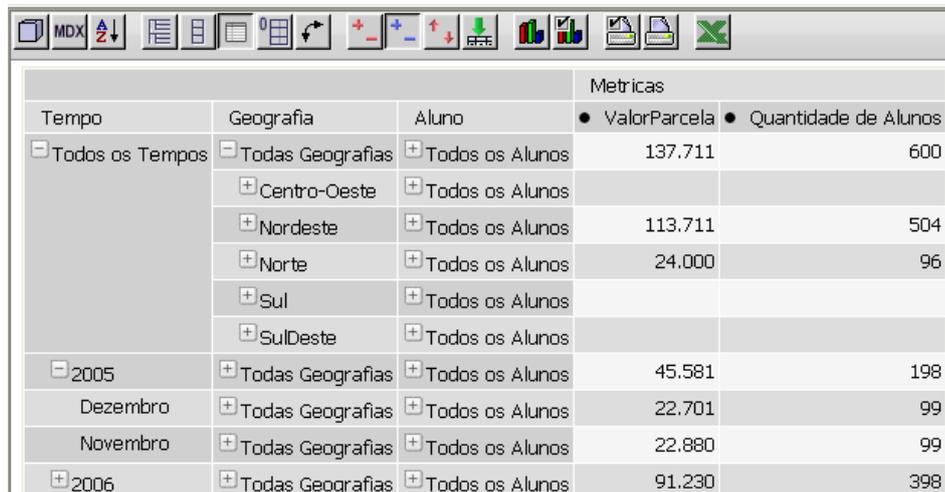
e) *Hide Spans*: retira cabeçalhos repetidos em resultados comuns. Esta opção realiza uma tarefa parecida com a ferramenta anterior, porém com relação às colunas como ilustra a Figura31;



Tempo	Geografia	Aluno	Metricas
+ Todos os Tempos	- Todas Geografias	- Todos os Alunos	137.711
+ Todos os Tempos	- Todas Geografias	- BACHARELADO ADMINISTRAÇÃO - UNISA - NOTURNO	9.000
+ Todos os Tempos	- Todas Geografias	- Adulto	9.000
+ Todos os Tempos	- Todas Geografias	- Feminino	9.000
+ Todos os Tempos	- Todas Geografias	- Adimplente	9.000
+ Todos os Tempos	- Todas Geografias	Ariana Fonseca Lima	9.000
+ Todos os Tempos	- Todas Geografias	+ LICENCIATURA EM BIOLOGIA - NOTURNO	85.320
+ Todos os Tempos	- Todas Geografias	+ LICENCIATURA EM HISTÓRIA - NOTURNO	193.980

Figura31 – Ocultar informações repetidas

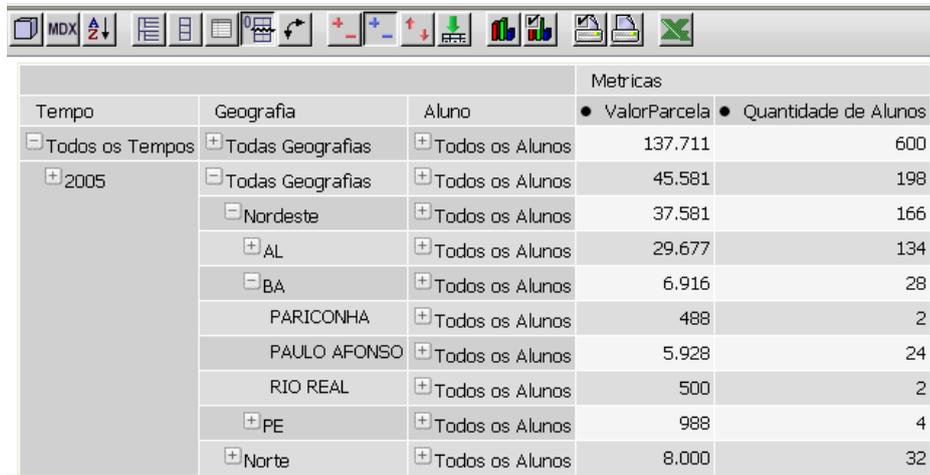
f) *Show Properties*: através da ativação dessa opção tornam-se visíveis as propriedades das dimensões expandidas como ilustra a Figura32;



Tempo	Geografia	Aluno	Metricas
- Todos os Tempos	- Todas Geografias	+ Todos os Alunos	137.711
	+ Centro-Oeste	+ Todos os Alunos	
	+ Nordeste	+ Todos os Alunos	113.711
	+ Norte	+ Todos os Alunos	24.000
	+ Sul	+ Todos os Alunos	
	+ SulDeste	+ Todos os Alunos	
- 2005	+ Todas Geografias	+ Todos os Alunos	45.581
Dezembro	+ Todas Geografias	+ Todos os Alunos	22.701
Novembro	+ Todas Geografias	+ Todos os Alunos	22.880
+ 2006	+ Todas Geografias	+ Todos os Alunos	91.230

Figura32 - Mostrar Propriedades

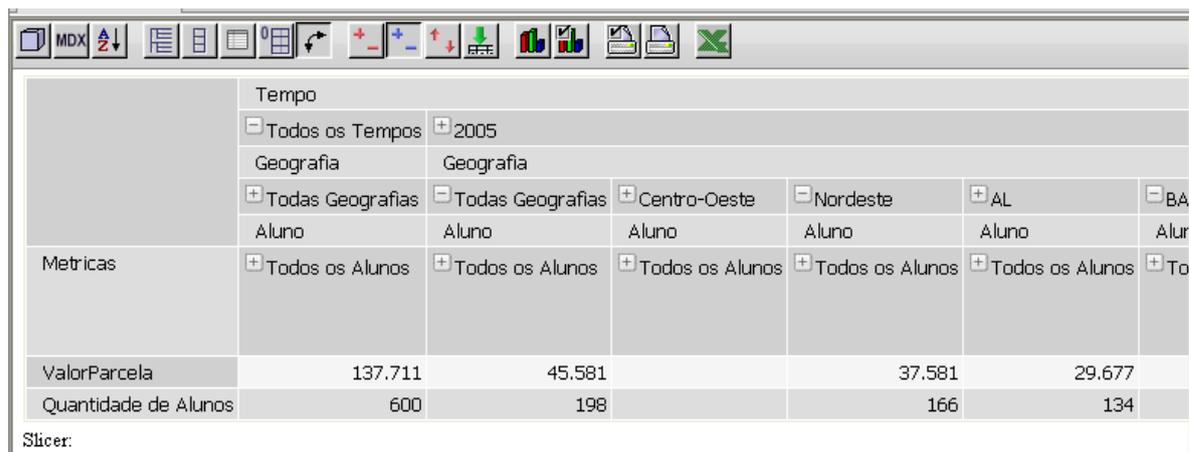
g) *Suppress Empty Rows / Columns*: mais uma função importante na boa visualização da tabela. Espaços em branco são ocultados por esta ferramenta para uma melhor visualização como ilustra a Figura33;



Tempo	Geografia	Aluno	Metricas	
			• ValorParcela	• Quantidade de Alunos
[-] Todos os Tempos	[+] Todas Geografias	[+] Todos os Alunos	137.711	600
[+] 2005	[-] Todas Geografias	[+] Todos os Alunos	45.581	198
	[-] Nordeste	[+] Todos os Alunos	37.581	166
	[+] AL	[+] Todos os Alunos	29.677	134
	[-] BA	[+] Todos os Alunos	6.916	28
	PARICONHA	[+] Todos os Alunos	488	2
	PAULO AFONSO	[+] Todos os Alunos	5.928	24
	RIO REAL	[+] Todos os Alunos	500	2
	[+] PE	[+] Todos os Alunos	988	4
	[+] Norte	[+] Todos os Alunos	8.000	32

Figura33 – Reprimir linhas e colunas vazias

h) *Swap Axes*: realiza a inversão entre linhas/colunas e colunas/linhas; É equivalente à operação de Dice – rotação para visualizar o cubo sobre outra perspectiva como ilustra a Figura34;



Metricas	Tempo					
	[-] Todos os Tempos		[+] 2005			
ValorParcela	Geografia			Geografia		
	[+] Todas Geografias	[-] Todas Geografias	[+] Centro-Oeste	[-] Nordeste	[+] AL	[-] BA
Quantidade de Alunos	Aluno			Aluno		
	[+] Todos os Alunos	[+] Todos os Alunos	[+] Todos os Alunos	[+] Todos os Alunos	[+] Todos os Alunos	[+] To
ValorParcela	137.711	45.581		37.581	29.677	
Quantidade de Alunos	600	198		166	134	

Slicer:

Figura34 – Inversão de eixos

i) *Drill buttons*: pode-se escolher entre os itens j, k, l e m para refinar uma operação de *Drill down* ou *Drill up*;

j) *Drill Member*: Em *Drill Member*, quando expande-se uma dimensão, todas as instâncias do membro são expandidas como ilustra a Figura35;

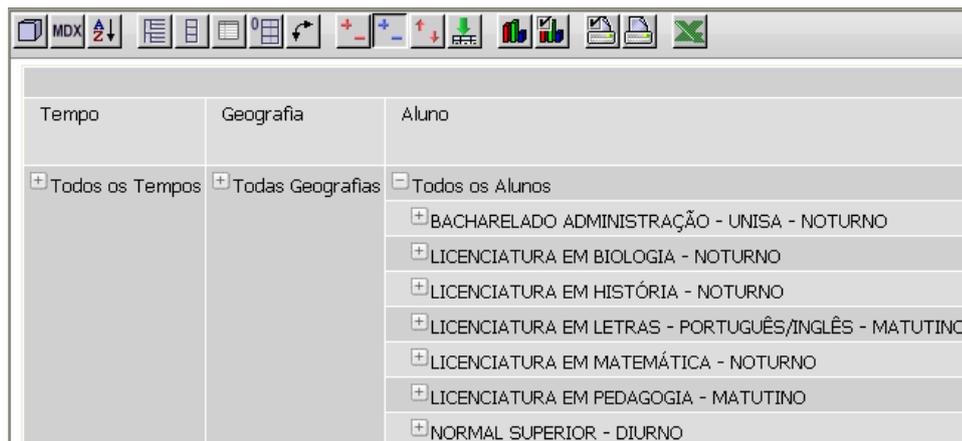


			Metricas	
Tempo	Geografia	Aluno	● ValorParcela	● Quantidade de Alunos
- Todos os Tempos	+ Todas Geografias	+ Todos os Alunos	137.711	600
+ 2005	+ Todas Geografias	+ Todos os Alunos	45.581	198
+ 2006	+ Todas Geografias	+ Todos os Alunos	91.230	398
+ 2007	+ Todas Geografias	+ Todos os Alunos	900	4

Slicer:

Figura35 – Expandir membros

k) *Drill Position*: seguindo o *Drill Member*, apenas o membro selecionado é expandido, mesmo que existam outras instâncias do mesmo membro como ilustra a Figura36;



Tempo	Geografia	Aluno
+ Todos os Tempos	+ Todas Geografias	- Todos os Alunos
		+ BACHARELADO ADMINISTRAÇÃO - UNISA - NOTURNO
		+ LICENCIATURA EM BIOLOGIA - NOTURNO
		+ LICENCIATURA EM HISTÓRIA - NOTURNO
		+ LICENCIATURA EM LETRAS - PORTUGUÊS/INGLÊS - MATUTINO
		+ LICENCIATURA EM MATEMÁTICA - NOTURNO
		+ LICENCIATURA EM PEDAGOGIA - MATUTINO
		+ NORMAL SUPERIOR - DIURNO

Slicer:

Figura36 – Expandir posição

l) *Drill Replace*: essa opção permite agregar linhas ou colunas durante o *Drill down* como ilustra a Figura37;

Tempo	Geografia	Aluno	Metricas	
			● ValorParcela	● Quantidade de Alunos
2005	Centro-Oeste	BACHARELADO ADMINISTRAÇÃO - UNISA - NOTURNO		
		LICENCIATURA EM BIOLOGIA - NOTURNO		
		LICENCIATURA EM HISTÓRIA - NOTURNO		
		LICENCIATURA EM LETRAS - PORTUGUÊS/INGLÊS - MATUTINO		
		LICENCIATURA EM MATEMÁTICA - NOTURNO		
		LICENCIATURA EM PEDAGOGIA - MATUTINO		
		NORMAL SUPERIOR - DIURNO		
	Nordeste	BACHARELADO ADMINISTRAÇÃO - UNISA - NOTURNO	3.000	12
		LICENCIATURA EM BIOLOGIA - NOTURNO	19.536	84

Figura37 – Expandir substituindo

m) *Drill Through*: adiciona uma seta indicativa nos Fatos apresentados para possível visualização mais específica do usuário como ilustra a Figura38;

Tempo	Geografia	Aluno	Metricas	
			● ValorParcela	● Quantidade de Alunos
2005	Centro-Oeste	BACHARELADO ADMINISTRAÇÃO - UNISA - NOTURNO		
		LICENCIATURA EM BIOLOGIA - NOTURNO		
		LICENCIATURA EM HISTÓRIA - NOTURNO		
		LICENCIATURA EM LETRAS - PORTUGUÊS/INGLÊS - MATUTINO		
		LICENCIATURA EM MATEMÁTICA - NOTURNO		
		LICENCIATURA EM PEDAGOGIA - MATUTINO		
		NORMAL SUPERIOR - DIURNO		
	Nordeste	BACHARELADO ADMINISTRAÇÃO - UNISA - NOTURNO	↕3.000	↕12
		LICENCIATURA EM BIOLOGIA - NOTURNO	↕19.536	↕84
		LICENCIATURA EM HISTÓRIA - NOTURNO	↕47.556	↕192

Figura38 – Expandir com detalhe

n) *Show Chart*: permite a construção de gráficos para análise dos dados em outra visão como ilustra a Figura39;

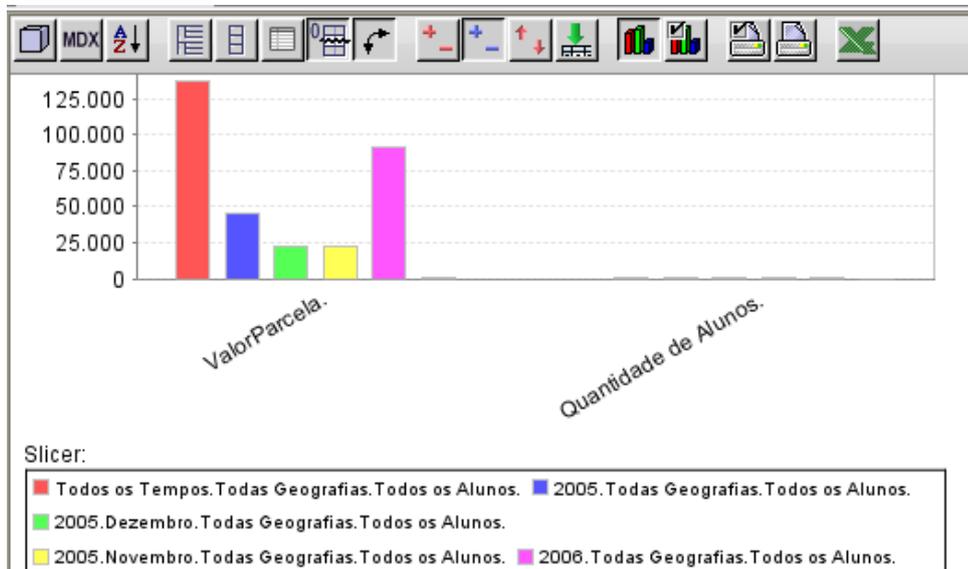


Figura39 - Visualizar Gráfico

o) *Chart Config*: como alguns gráficos ficam desconfigurados pelo grande número de colunas, ou pelo tamanho dos nomes das mesmas, essa ferramenta permite configurar fatores que possam tornar o gráfico mais apresentável como ilustra a Figura40;

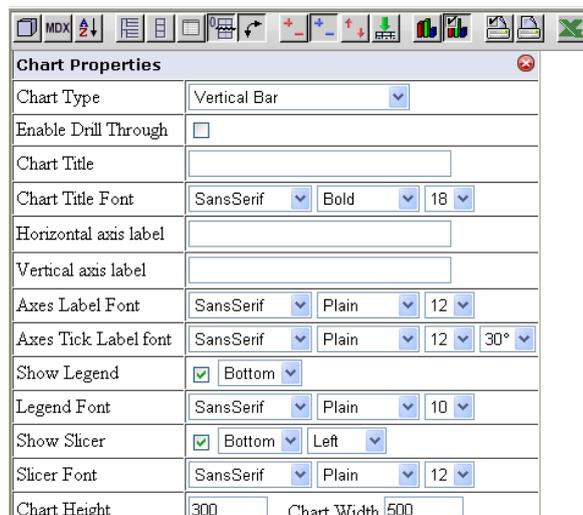


Figura40 – Configuração do gráfico

p) *Configure Print Settings*: permite a configuração de opções de impressão como ilustra a Figura41;

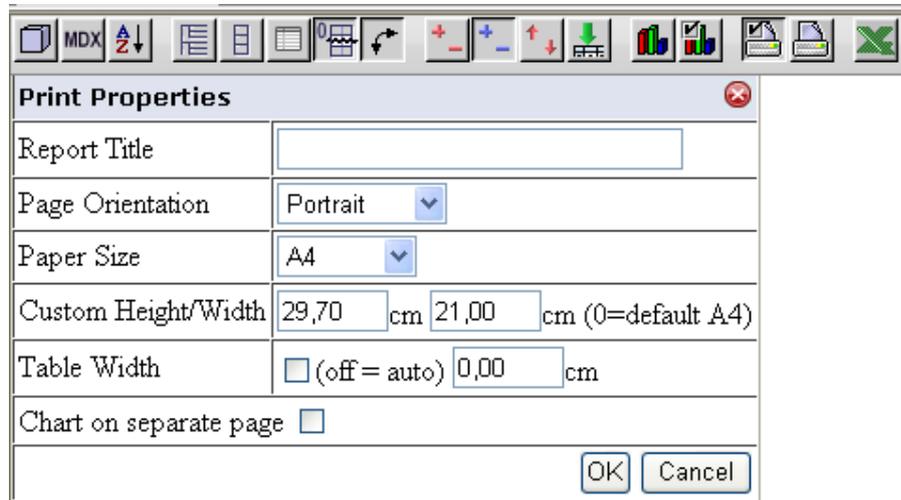


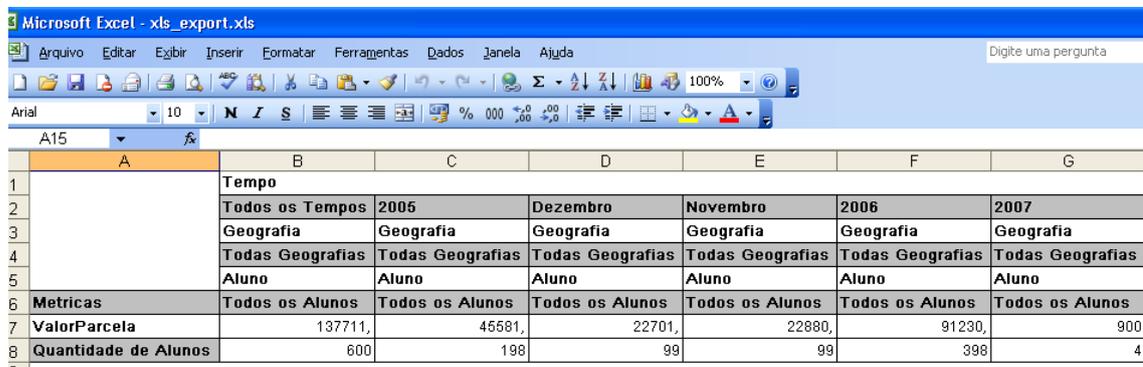
Figura41 - Configurar Definições de impressão

q) *Print This Page Via PDF*: gera um documento em formato PDF com a tabela em questão como ilustra a Figura42;

	Tempo					
	Todos os Tempos	2005	Dezembro	Novembro	2006	2007
Geografia	Geografia	Geografia	Geografia	Geografia	Geografia	Geografia
Todas Geografias	Todas Geografias	Todas Geografias	Todas Geografias	Todas Geografias	Todas Geografias	Todas Geografias
Aluno	Aluno	Aluno	Aluno	Aluno	Aluno	Aluno
Metricas	Todos os Alunos					
ValorParcela	137.711	45.581	22.701	22.880	91.230	900
Quantidade de Alunos	600	198	99	99	398	4

Figura42 - Imprimir esta página através de PDF

r) *Start Excel*: gera um documento de planilha eletrônica, no formato XLS (MS Excel), com a tabela em questão como ilustra a Figura43.



	A	B	C	D	E	F	G
1		<b>Tempo</b>					
2		<b>Todos os Tempos</b>	<b>2005</b>	<b>Dezembro</b>	<b>Novembro</b>	<b>2006</b>	<b>2007</b>
3		<b>Geografia</b>	<b>Geografia</b>	<b>Geografia</b>	<b>Geografia</b>	<b>Geografia</b>	<b>Geografia</b>
4		<b>Todas Geografias</b>					
5		<b>Aluno</b>	<b>Aluno</b>	<b>Aluno</b>	<b>Aluno</b>	<b>Aluno</b>	<b>Aluno</b>
6	<b>Métricas</b>	<b>Todos os Alunos</b>					
7	<b>ValorParcela</b>	137711,	45581,	22701,	22880,	91230,	900
8	<b>Quantidade de Alunos</b>	600	198	99	99	398	4

Figura43 - Iniciar o Excel

O Pentaho é um pacote completo para criar aplicações de Business Intelligence. O JPivot e o Mondrian são as principais ferramentas constantes no pacote do Pentaho e alvo do trabalho aqui apresentado. O Mondrian é responsável por implementar um servidor OLAP, enquanto o JPivot é usado para implementar a interface final do usuário. Este capítulo foi muito importante para demonstrar o uso da plataforma Pentaho, através do estudo de caso pode-se mostrar todo o potencial de desenvolvimento com soluções de *Business Intelligence*.

Com o conhecimento obtido no desenvolvimento do trabalho e descritos nos capítulos anteriores viu-se a importância desta ferramenta por ser completa para desenvolvimento de *Business Intelligence*. Devida a falta de documentação, algumas dificuldades foram encontradas durante o desenvolvimento do trabalho, mas muitas delas resolvidas através da comunidade Pentaho-br, e com esta interação possibilitou ainda contribuir com outros desenvolvedores com as mesmas dificuldades.

## 8 CONCLUSÃO

As contribuições deste trabalho acadêmico consistem no estudo de uma tecnologia existente para o auxílio às pequenas empresas que têm limitações no desenvolvimento de aplicações para a tomada de decisão, devido ao seu alto custo.

Conclui-se que com a suíte de ferramentas Pentaho, o desenvolvedor pode criar aplicações completas para *Business Intelligence*. As contribuições de cunho social, baseiam-se no fato de que este estudo serve como base para a disseminação de tecnologias de software livre e Ambientes de BI em pequenas empresas no Brasil e para o domínio da mesma, além da redução no seu custo.

Diante deste cenário, foi descrito um estudo de caso de desenvolvimento de um ambiente de BI em uma empresa de ensino superior privada, de forma a coletar, consolidar e analisar informações referentes às práticas institucionais que possam futuramente servir de base para uma melhor compreensão e um melhor planejamento da empresa.

Os objetivos específicos propostos para o trabalho foram alcançados. Foi alcançada através da plataforma Pentaho BI, com a descrição de sua arquitetura e dos principais softwares que compõe esta suíte. Além disso, foi elaborado um estudo de caso que demonstra passo-a-passo o uso do Pentaho para criação de um ambiente de BI simples.

Como alternativas de trabalhos futuros, sugere-se aplicar os conhecimentos aqui apresentados para implantar o ambiente de BI no ambiente de produção da empresa abordada nesse trabalho sendo estudado futuramente e avaliados os impactos organizacionais.

## REFERÊNCIAS

PEREIRA, Ana Altina Cambuí. **Manual Institucional do Centro Universitário da Bahia – FIB**. Centro Universitário da Bahia – FIB, Salvador, 2008.

ALTER, Steven. **Information system: a management perspective**. United States of America: Addison Wesley Publishing Company, 1992.

BARBIERI, C. **Business Intelligence: modelagem e tecnologia**. Rio de Janeiro: Axcel Books, 2001.

BATISTA E. O. **Sistemas de informação**. São Paulo: Saraiva, 2004.

BISPO, C.A.F., CAZARINI, E.W. **A evolução do processo decisório**. Niterói: Anais Niterói, TEP – UFF, artigo 94.doc, 1998

CAMPOS FILHO, M. P. **Os sistemas de informação e as modernas tendências da tecnologia e dos negócios**. RAE. São Paulo, v.34, n. 6, 2001.

CAMPOS FILHO, Maurício Prates. **Os sistemas de informação e as modernas tendências da tecnologia e dos negócios**. Revista de Administração de Empresas, São Paulo, v.34, n.6, nov/dez 1994.

CAMPOS, Maria L., ROCHA, Arnaldo V. Filho. **Data Warehouse**, in: Congresso da Sociedade Brasileira de Computação, 17, 1997, Brasília. Anais do XVII Congresso da SBC. Brasília: UNB, 1997.

CIELO, I. **Arquiteturas OLAP**. Disponível em: <<http://www.dcc.ufmg.br/pos/html/spg98/anais/anapat/anapat.html>>. Acessado em: 26 OUT. 2009.

DALFOVO, Oscar. **Desenho de um modelo de sistemas de informação**. Blumenau, 1998. Dissertação (mestrado em Administração de Negócios) Centro de Ciências Sociais e Aplicadas, FURB.

DALFOVO, Oscar. **Metodologia sistema de informação estratégico para o gerenciamento operacional (SIEGO): um modelo para a universidade com aplicação na gestão ambiental baseado em Data Warehouse**. Florianópolis, 2001. 308 f. Tese de doutorado (curso de pós-graduação em Ciência de Computação)

Universidade Federal de Santa Catarina, UFSC.

DALFOVO, Oscar; AMORIM, Sammy Newton. **Quem tem informação é mais competitivo**. Blumenau: Acadêmica, 2000.

DAMIANI, Wagner B. - **Pesquisa sobre o uso de EIS nas 500 maiores empresas americanas e brasileiras** - EAESP/FGV - Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas Endereço eletrônico: São Paulo, 1996 Disponível em: <<http://www.fgvsp.br/eis/eisbr/indexbr.html>> Acessado em: 19 OUT. 2009.

DAVENPORT, T.; PRUSAK, L. **Conhecimento empresarial**: como as organizações gerenciam o seu capital intelectual. Rio de Janeiro: Campus 1998.

DWBRASIL. **Data Mart**. Disponível em: <<http://www.dwbrasil.com.br/html/dmart.html>>. Acessado em: 26 OUT. 2009.

FROZZA, A. A. **Data warehouse**: modelagem dimensional. Notas de Aula da disciplina *Data Warehouse*. Lages: UNIPLAC, 2008.

FURLAN, J. D. **Modelagem de negócio**. São Paulo, Makron Books, 1997.

FURLAN, José Davi; IVO, Ivonildo da Motta; AMARAL, Francisco Piedade. **Sistemas de informação executiva - EIS**. São Paulo: Makron Books, 1994.

GATES, B. **A Estrada do Futuro**. São Paulo, Cia das Letras, 1994.

GATES, B. **A estrada do futuro**. São Paulo, Companhia das Letras, 1995.

GERRITY Jr., T. P. **The Design of Man-Machine Decision Systems**. Sloan Management Review, vol. 12, n. 2, 1971.

GONÇALVES, J. L. M. **Data warehouse é necessariamente um megaprojeto?** *Developers' Magazine*, n. 6, 1997.

HARRISON, Thomas H. **Intranet Data Warehouse**. 1.ed. São Paulo: Berkeley, 1998.

HOKAMA, D. D. B. et al. **A modelagem de dados no ambiente Data Warehouse**. 2004. 121 f. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Faculdade de Computação e Informática, Universidade Presbiteriana Mackenzie. Disponível em:  
<<http://meusite.mackenzie.com.br/rogerio/tgi/2004ModelagemDW.pdf>>. Acessado em: 20 SET. 2009.

INMON, William H. **Como construir o data warehouse**. Rio de Janeiro: Campus, 1997.

INMON, William H.; HACKATHORN, Richard D. **Como usar o Data Warehouse**. Trad. Olavo Faria. Rio de Janeiro: Infobook, 1997.

KIMBALL, R. **Data Warehouse Toolkit**. São Paulo: Makron Books, 1997.

KIMBALL, Ralph. **Data Warehouse Toolkit: Técnicas para Construção de Data Warehouses Dimensionais**. São Paulo: Makron Books, 1998.

LAUDON, Kenneth C.; LAUDON, Jane Price. **Sistemas de informação com Internet**. Tradução Dalton Conde de Alencar. Rio de Janeiro: LTC, 1999.

LEME FILHO, Trajano. **Aulas ministradas na Fasp da disciplina Sistemas de Suporte à Decisão**. Transparências em PowerPoint, 2006.

LITTLE, J. D. C. **Models and Managers: The Concept of a Decision Calculus**. Management Science, vol. 16, n. 8, 1975.

MACDONALD, G. C.; RUBIK, J. R. **Pesquisa e seleção de ferramentas livres e baseadas em padrões de sistemas abertos para a elaboração de interfaces OLAP sobre a web**. 2007. 114 p. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) - Departamento de Informática e Estatística, Universidade Federal de Santa Catarina, Florianópolis, 2007.

MCGEEVER, C. **Business Intelligence**. Computerworld. Arizona, p.54, jul-2000.

MCLEOD JR., Raymond. **Management information system: a study of computer-base information system**. United States of America: Macmillan Publishing Company, 1993.

MIRANDA, Roberto Campos da Rocha. O uso da informação na formulação de ações estratégicas pelas empresas. **Ci. Inf., Brasília**, v. 28, n. 3, 1999. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19651999000300006&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651999000300006&lng=pt&nrm=iso)>. Acesso em: 11.Outubro.2009.

MORTON, M. S. S. **Management Decision Systems**: Computer-Based Support for Decision Making. Boston, Division of Research, Graduate School of Business Administration, Harvard University, 1971.

MOSIMANN, Clara Pellegrinello, ALVES, Osmar de C., FISCH, Silvio. **Controladoria**: seu papel na administração de empresas. Florianópolis: ESAG, 1993.

NONAKA, Ikujiro; TAKEUCHI, Hirotaka. **Criação de conhecimento na empresa**: como as empresas japonesas geram a dinâmica da inovação. 4. ed. Rio de Janeiro : Campus, 1997.

OLIVEIRA, Adelise G. **Data warehouse conceitos e soluções**. Florianópolis - SC: Advanced, 1998.

OLIVEIRA, Djalma de Pinho Rebouças. **Sistemas de informações gerenciais**. São Paulo: Atlas, 1992.

PEARSON, J. M.; SHIM, J. P. **An Empirical Investigation into DSS Structures and Environments**. Decision Support Systems, n. 13, 1995.

PENTAHO. **Pentaho Open Source Business Intelligence**. Disponível em: <<http://www.pentaho.com>>. Acessado em: 12 out. 2009.

PEREIRA, M. J. L. B.; FONSECA, J. G. M. **Faces da decisão**: as mudanças de paradigmas e o poder da decisão. São Paulo, Makron Books, 1997.

PERNAS, A. M. da R. **Modelagem de um Data Webhouse voltado a produção e comercialização de sementes**. 2003. 70 f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Instituto de Física e Matemática, Universidade Federal de Pelotas, Pelotas. Disponível em: <[http://www.ufpel.edu.br/prg/sisbi/bibct/acervo/info/2003/mono\\_ana\\_ pernas.pdf](http://www.ufpel.edu.br/prg/sisbi/bibct/acervo/info/2003/mono_ana_ pernas.pdf)>. Acessado em: 20 out. 2009.

PETERSON, Timothy. Microsoft SQL Server 2000 (DTS). Rio de Janeiro: Campus, 2001.

PONCHIROLLI, O., FIALHO, Francisco Antonio P. **Gestão estratégica do conhecimento como parte da estratégia empresarial**. Revista FAE, Curitiba, v. 8, n. 1, pp. 127-138, 2005. Disponível em: <[http://www.fae.edu/publicacoes/pdf/revista\\_da\\_fae/rev\\_fae\\_v8\\_n1/rev\\_fae\\_v8\\_n1\\_11.pdf](http://www.fae.edu/publicacoes/pdf/revista_da_fae/rev_fae_v8_n1/rev_fae_v8_n1_11.pdf)>. Acesso em: 22.Outubro.2009.

POWER, D. **A brief history of Decision Support Systems**, 1997. Disponível em: <<http://power.cba.uni.edu/isworld/dsshhistory.html>> Acessado em: 19 OUT. 2009.

POZZEBON, M. & FREITAS, H. **Construindo um EIS (Enterprise Information System) da (e para a) Empresa**. RAUSP - Revista de Administração da USP, vol.31, No. 4, 1996.

PROBST, G.; RAUB, S.; ROMHARDT, K. **Gestão do conhecimento: os elementos construtivos do sucesso**. Porto Alegre: Bookman 2002.

RAMOS, I., SANTOS, M.Y. **Business Intelligence – Tecnologia da Informação na Gestão de Conhecimento**, Lisboa: FCA – Editora de Informática, 2006.

SANTOS, Maribel Yasmina ; RAMOS, Isabel - "**Business Intelligence : tecnologias da informação na gestão de conhecimento**". Lisboa: FCA Editora de Informática, 2006.

SETZER, Valdemar W. Dado, Informação, Conhecimento e Competência. **DataGramaZero - Revista de Ciência da Informação**, n. 0, 1999. Disponível em: <<http://www.inf.pucrs.br/~gilberto/Sistemadeinformacao/dado-competencia.pdf>>. Acesso em: 10. Outubro. 2009.

SHIGUNOV, F. **Uma Aplicação OLAP sobre a Web para Análise dos Dados do Vestibular da UFSC e Diretrizes para a sua Integração com GIS**. 2007. 88 f. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação)- Departamento de Informática e Estatística, Universidade Federal de Santa Catarina-UFSC, Florianópolis.

SINGH, H. S. **Data Warehouse – Conceitos, tecnologias, implementação e Gerenciamento**. São Paulo. 2001.

SPRAGUE JR., Ralph H. e WATSON, Hugh J. (Org.). **Sistemas de apoio a decisão**: colocando a teoria em prática. Rio de Janeiro: Campus, 1991.

STAIR, Ralph M. **Princípios de sistemas de informação**: uma abordagem gerencial. Rio de Janeiro: LTC - Livros Técnicos e Científicos, 1998.

STRANG, K. **Can data marts grow?**. CIO Magazine, 1 Jul. 1997. Disponível em: <[http://www.cio.com/archive/070197\\_gartner\\_content.html](http://www.cio.com/archive/070197_gartner_content.html)> Acessado em: 20 SET. 2009.

TEMATEC. **Por dentro da Pentaho Open BI Suite**: Conceitos, Arquitetura e Componentes. Disponível em: <<http://br.groups.yahoo.com/group/pentahobr/>>. Acessado em: 26 OUT. 2009.

TOM, Paul L. **Managing Information as a Corporate Resource**. Harper Collings Publishers, 1991.

TURBAN, E & WALLS, J.G. **Executive Information Systems - a Special Issue**. Decision Support Systems, vol. 14, 1995.

URIS, A. **O livro de mesa do executivo**. São Paulo, Pioneira, 1989.

VILELA, E. **Computerworld technology breakfast**: prévia *data warehouse*. Computerworld, 1997. Suplemento Especial. Disponível em: <<http://www.computerworld.com.br/polemica/data.htm>> Acessado em: 20 OUT. 2009.

## APÊNDICE A – Dicionário de Indicadores

FTC-EAD	<b>DICIONÁRIO DE INDICADORES</b>		rev. 1.0
<b>Posicionamento no mapa</b>			
NOME DO INDICADOR:	<i>Quantidade de alunos por Curso</i>		
<b>Descrição do Indicador</b>			
Compreende na quantidade de alunos por Cursos ministrados pela instituição			
FÓRMULA DE CÁLCULO:		UNIDADE DE MEDIDA:	Númeral
PERIODICIDADE DE CARGA	Semestral	CASAS DECIMAIS:	6
<b>Disponibilização</b>			
FONTE DE DADO:	Base de Dados SQLServer do SGE		

FTC-EAD	<b>DICIONÁRIO DE INDICADORES</b>		rev. 1.0
<b>Posicionamento no mapa</b>			
NOME DO INDICADOR:	<i>Quantidade de alunos por Região</i>		
<b>Descrição do Indicador</b>			
Compreende na quantidade de alunos por Região do Brasil (Norte, Nordeste, Sul, Suldeste, Centro-Oeste)			
FÓRMULA DE CÁLCULO:		UNIDADE DE MEDIDA:	Númeral
PERIODICIDADE DE CARGA	Semestral	CASAS DECIMAIS:	6
<b>Disponibilização</b>			
FONTE DE DADO:	Base de Dados SQLServer do SGE		

FTC-EAD	<b>DICIONÁRIO DE INDICADORES</b>		rev. 1.0
<b>Posicionamento no mapa</b>			
NOME DO INDICADOR:	<i>Quantidade de alunos por Estado</i>		
<b>Descrição do Indicador</b>			
Compreende na quantidade de alunos por Estados do Brasil onde houver alunos matriculados			
FÓRMULA DE CÁLCULO:		UNIDADE DE MEDIDA:	Númeral
PERIODICIDADE DE CARGA	Semestral	CASAS DECIMAIS:	6
<b>Disponibilização</b>			
FONTE DE DADO:	Base de Dados SQLServer do SGE		

FTC-EAD	<b>DICIONÁRIO DE INDICADORES</b>		rev. 1.0
<b>Posicionamento no mapa</b>			
NOME DO INDICADOR:	<i>Quantidade de alunos por Cidade</i>		
<b>Descrição do Indicador</b>			
Compreende na quantidade de alunos por Cidades do Brasil onde houver alunos matriculados			
FÓRMULA DE CÁLCULO:		UNIDADE DE MEDIDA:	Númeral
PERIODICIDADE DE CARGA	Semestral	CASAS DECIMAIS:	6
<b>Disponibilização</b>			
FONTE DE DADO:	Base de Dados SQLServer do SGE		

FTC-EAD	<b>DICIONÁRIO DE INDICADORES</b>		rev. 1.0
<b>Posicionamento no mapa</b>			
NOME DO INDICADOR:	<i>Quantidade de alunos por Sexo</i>		
<b>Descrição do Indicador</b>			
Compreende na quantidade de alunos por Sexo			
FÓRMULA DE CÁLCULO:		UNIDADE DE MEDIDA:	Numeral
PERIODICIDADE DE CARGA	Semestral	CASAS DECIMAIS:	6
<b>Disponibilização</b>			
FONTE DE DADO:	Base de Dados SQLServer do SGE		

FTC-EAD	<b>DICIONÁRIO DE INDICADORES</b>		rev. 1.0
<b>Posicionamento no mapa</b>			
NOME DO INDICADOR:	<i>Quantidade de alunos por Faixa Etaria</i>		
<b>Descrição do Indicador</b>			
Compreende na quantidade de alunos por faixa etaria (Velhice, Adulto, Jovem)			
FÓRMULA DE CÁLCULO:		UNIDADE DE MEDIDA:	Numeral
PERIODICIDADE DE CARGA	Semestral	CASAS DECIMAIS:	6
<b>Disponibilização</b>			
FONTE DE DADO:	Base de Dados SQLServer do SGE		

FTC-EAD	<b>DICIONÁRIO DE INDICADORES</b>		rev. 1.0
<b>Posicionamento no mapa</b>			
NOME DO INDICADOR:	<i>Quantidade de alunos Adimplentes</i>		
<b>Descrição do Indicador</b>			
Compreende na quantidade de alunos Adimplentes em determinado periodo			
FÓRMULA DE CÁLCULO:		UNIDADE DE MEDIDA:	Numeral
PERIODICIDADE DE CARGA	Semestral	CASAS DECIMAIS:	6
<b>Disponibilização</b>			
FONTE DE DADO:	Base de Dados SQLServer do SGE		

FTC-EAD	<b>DICIONÁRIO DE INDICADORES</b>		rev. 1.0
<b>Posicionamento no mapa</b>			
NOME DO INDICADOR:	<i>Quantidade de alunos Inadimplentes</i>		
<b>Descrição do Indicador</b>			
Compreende na quantidade de alunos Inadimplentes por periodo.			
FÓRMULA DE CÁLCULO:		UNIDADE DE MEDIDA:	Numeral
PERIODICIDADE DE CARGA	Semestral	CASAS DECIMAIS:	6
<b>Disponibilização</b>			
FONTE DE DADO:	Base de Dados SQLServer do SGE		

FTC-EAD	<b>DICIONÁRIO DE INDICADORES</b>		rev. 1.0
<b>Posicionamento no mapa</b>			
NOME DO INDICADOR:	<i>Total Mensalidade por Ano</i>		
<b>Descrição do Indicador</b>			
Compreende no total em reais de mensalidades arrecadadas em determinado ano			
FÓRMULA DE CÁLCULO:	=soma()	UNIDADE DE MEDIDA:	R\$
PERIODICIDADE DE CARGA	Semestral	CASAS DECIMAIS:	6
<b>Disponibilização</b>			
FONTE DE DADO:	Base de Dados SQLServer do SGE		

FTC-EAD	<b>DICIONÁRIO DE INDICADORES</b>		rev. 1.0
<b>Posicionamento no mapa</b>			
NOME DO INDICADOR:	<i>Total Mensalidade por Mês</i>		
<b>Descrição do Indicador</b>			
Compreende no total em reais de mensalidades arrecadadas em determinado mês			
FÓRMULA DE CÁLCULO:	=sima()	UNIDADE DE MEDIDA:	R\$
PERIODICIDADE DE CARGA	Semestral	CASAS DECIMAIS:	6
<b>Disponibilização</b>			
FONTE DE DADO:	Base de Dados SQLServer do SGE		

FTC-EAD	<b>DICIONÁRIO DE INDICADORES</b>		rev. 1.0
<b>Posicionamento no mapa</b>			
NOME DO INDICADOR:	<i>Total Mensalidade por Região</i>		
<b>Descrição do Indicador</b>			
Compreende no total em reais de mensalidades arrecadadas por região			
FÓRMULA DE CÁLCULO:	=soma()	UNIDADE DE MEDIDA:	R\$
PERIODICIDADE DE CARGA	Semestral	CASAS DECIMAIS:	6
<b>Disponibilização</b>			
FONTE DE DADO:	Base de Dados SQLServer do SGE		

## APÊNDICE B - Procedures para tratamento das dimensões

```

CREATE PROCEDURE Tratamento_DIM_Aluno
AS
BEGIN

insert into [DW_TCC].[dbo].[DIM_TEMP_Aluno]
SELECT distinct [FTC_EAD].[dbo].[TbAluno].IdAluno,
Nome,
Sexo,
CPF,
DataNascimento,
Faixa_Etaria = ",
[FTC_EAD].[dbo].[TbTipoBaixa].IdSituacaoBaixa,
Curso,
DescricaoStatus = ",
[FTC_EAD].[dbo].[TbAluno].UF,
[FTC_EAD].[dbo].[TbAluno].Cidade
FROM [FTC_EAD].[dbo].[TbAluno] inner join
[FTC_EAD].[dbo].[TbMensalidade] on
[FTC_EAD].[dbo].[TbAluno].IdAluno =
[FTC_EAD].[dbo].[TbMensalidade].IdAluno
inner join
[FTC_EAD].[dbo].[TbTipoBaixa]
on
[FTC_EAD].[dbo].[TbMensalidade].IdTipoBaixa =
[FTC_EAD].[dbo].[TbTipoBaixa].IdTipoBaixa
inner join [FTC_EAD].[dbo].[TfCursos] on
[FTC_EAD].[dbo].[TbAluno].IdCurso =
[FTC_EAD].[dbo].[TfCursos].IdCurso

update [DW_TCC].[dbo].[DIM_TEMP_Aluno] set Faixa_Etaria = 'Juventude' where
(Datepart(yy,GETDATE()) - Datepart(yy,DataNascimento)) <=24
update [DW_TCC].[dbo].[DIM_TEMP_Aluno] set Faixa_Etaria = 'Adulto' where
(Datepart(yy,GETDATE()) - Datepart(yy,DataNascimento)) > 24
and (Datepart(yy,GETDATE()) - Datepart(yy,DataNascimento)) <= 59
update [DW_TCC].[dbo].[DIM_TEMP_Aluno] set Faixa_Etaria = 'Velhice' where
(Datepart(yy,GETDATE()) - Datepart(yy,DataNascimento)) >=60

update [DW_TCC].[dbo].[DIM_TEMP_Aluno] set Status = 1 where Status >= 1
update [DW_TCC].[dbo].[DIM_TEMP_Aluno] set Status = 2 where Status = 0

update [DW_TCC].[dbo].[DIM_TEMP_Aluno] set Sexo = 'Masculino' where Sexo = 'M'
update [DW_TCC].[dbo].[DIM_TEMP_Aluno] set Sexo = 'Feminino' where Sexo = 'F'

update [DW_TCC].[dbo].[DIM_TEMP_Aluno] set DescricaoStatus = 'Adimplente' where Status = 1
update [DW_TCC].[dbo].[DIM_TEMP_Aluno] set DescricaoStatus = 'Inadimplente' where Status = 2

END

```

```

CREATE PROCEDURE Tratamento_DIM_Geografia
AS
BEGIN

insert into [DW_TCC].[dbo].[DIM_TEMP_Geografia]
SELECT distinct UF,
Cidade,
Região = ''
FROM [FTC_EAD].[dbo].[TbAluno]

update [DW_TCC].[dbo].[DIM_TEMP_Geografia] set Região = 'Norte' where Logradouro_UF in
('AC','AM','RO','RR','PA','AP','TO')

update [DW_TCC].[dbo].[DIM_TEMP_Geografia] set Região = 'Centro-Oeste' where Logradouro_UF
in ('MT','MS','GO','DF')

update [DW_TCC].[dbo].[DIM_TEMP_Geografia] set Região = 'Sul' where Logradouro_UF in
('PR','SC','RS')

update [DW_TCC].[dbo].[DIM_TEMP_Geografia] set Região = 'Nordeste' where Logradouro_UF in
('MA','PI','CE','RN','PB','PE','AL','SE','BA')

update [DW_TCC].[dbo].[DIM_TEMP_Geografia] set Região = 'SulDeste' where Logradouro_UF in
('MG','ES','RJ','SP')

END

```

```

CREATE PROCEDURE Tratamento_DIM_Mensalidade
AS
BEGIN

insert into [DW_TCC].[dbo].[DIM_TEMP_Mensalidade] (IdParcela, IdAluno, AnoCompetencia,
MesCompetencia, ValorParcela, DataVencimento)
SELECT IdParcela, IdAluno, AnoCompetencia,
MesCompetencia, ValorParcela, DataVencimento
FROM [FTC_EAD].[dbo].[TbMensalidade]

END

```

```

CREATE PROCEDURE Tratamento_DIM_Tempo
AS
BEGIN

insert into [DW_TCC].[dbo].[DIM_TEMP_Tempo] (Ano, Mes, Nome_mes)
SELECT distinct AnoCompetencia, MesCompetencia, Nome_mes=""
FROM [FTC_EAD].[dbo].[TbMensalidade]
GROUP BY AnoCompetencia, MesCompetencia
ORDER BY AnoCompetencia, MesCompetencia

update [DW_TCC].[dbo].[DIM_TEMP_Tempo] set Nome_mes = 'Janeiro' where Mes = 1
update [DW_TCC].[dbo].[DIM_TEMP_Tempo] set Nome_mes = 'Fevereiro' where Mes = 2
update [DW_TCC].[dbo].[DIM_TEMP_Tempo] set Nome_mes = 'Março' where Mes = 3
update [DW_TCC].[dbo].[DIM_TEMP_Tempo] set Nome_mes = 'Abril' where Mes = 4
update [DW_TCC].[dbo].[DIM_TEMP_Tempo] set Nome_mes = 'Maio' where Mes = 5
update [DW_TCC].[dbo].[DIM_TEMP_Tempo] set Nome_mes = 'Junho' where Mes = 6
update [DW_TCC].[dbo].[DIM_TEMP_Tempo] set Nome_mes = 'Julho' where Mes = 7
update [DW_TCC].[dbo].[DIM_TEMP_Tempo] set Nome_mes = 'Agosto' where Mes = 8

```

```

update [DW_TCC].[dbo].[DIM_TEMP_Tempo] set Nome_mes = 'Setembro' where Mes = 9
update [DW_TCC].[dbo].[DIM_TEMP_Tempo] set Nome_mes = 'Outubro' where Mes = 10
update [DW_TCC].[dbo].[DIM_TEMP_Tempo] set Nome_mes = 'Novembro' where Mes = 11
update [DW_TCC].[dbo].[DIM_TEMP_Tempo] set Nome_mes = 'Dezembro' where Mes = 12

```

```
END
```

```

CREATE PROCEDURE Inserir_DW_Fato
AS
BEGIN

```

```

insert into [DW_TCC].[dbo].[Fato_DW] (IdGeografia, IdTempo, IdAluno, IdMensalidade, Quantidade,
ValorParcela)

```

```

SELECT [DW_TCC].[dbo].[DIM_Geografia].IdGeografia,
[DW_TCC].[dbo].[DIM_Tempo].IdTempo,
[DW_TCC].[dbo].[DIM_Aluno].IdAluno,
[DW_TCC].[dbo].[DIM_Mensalidade].IdParcela,
1,
[DW_TCC].[dbo].[DIM_Mensalidade].ValorParcela

```

```

FROM [DW_TCC].[dbo].[DIM_Geografia],
[DW_TCC].[dbo].[DIM_Tempo],
[DW_TCC].[dbo].[DIM_Aluno],
[DW_TCC].[dbo].[DIM_Mensalidade]
where

```

```

[DW_TCC].[dbo].[DIM_Aluno].UF =
[DW_TCC].[dbo].[DIM_Geografia].Logradouro_UF
and
[DW_TCC].[dbo].[DIM_Aluno].Cidade =
[DW_TCC].[dbo].[DIM_Geografia].Logradouro_Cidade
and
[DW_TCC].[dbo].[DIM_Aluno].IdAluno =
[DW_TCC].[dbo].[DIM_Mensalidade].IdAluno
and
[DW_TCC].[dbo].[DIM_Mensalidade].MesCompetencia =
[DW_TCC].[dbo].[DIM_Tempo].Mes
and
[DW_TCC].[dbo].[DIM_Mensalidade].AnoCompetencia =
[DW_TCC].[dbo].[DIM_Tempo].Ano

```

```

group by IdGeografia, IdParcela, IdTempo, [DW_TCC].[dbo].[DIM_Aluno].IdAluno,
[DW_TCC].[dbo].[DIM_Mensalidade].ValorParcela
order by IdGeografia, IdParcela, IdTempo, [DW_TCC].[dbo].[DIM_Aluno].IdAluno

```

```
END
```