

UNIVERSIDADE FEDERAL DO PAMPA

RAFAEL DOS SANTOS FERREIRA

CONSTRUINDO UM DATA WAREHOUSE PARA O AGRONEGÓCIO

**BAGÉ
2013**

RAFAEL DOS SANTOS FERREIRA

CONSTRUINDO UM DATA WAREHOUSE PARA O AGRONEGÓCIO

Trabalho de Conclusão de Curso
apresentado ao Curso de Especialização
em Sistemas Distribuídos com Ênfase em
Banco de Dados da Universidade Federal
do Pampa, como requisito parcial para
obtenção do Título de Especialista
Orientador: Dr. Sandro da Silva Camargo

**BAGÉ
2013**

RAFAEL DOS SANTOS FERREIRA

CONSTRUINDO UM DATA WAREHOUSE PARA O AGRONEGÓCIO

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Sistemas Distribuídos com Ênfase em Banco de Dados da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Especialista.

Trabalho de Conclusão de Curso defendido e aprovado em: 08/08/2013

Banca examinadora:

Prof. Dr. Sandro da Silva Camargo
Orientador
UNIPAMPA

Prof^a Dr^a Ana Paula Lüdtkke Ferreira
UNIPAMPA

Prof^a Msc. Sandra Dutra Piovesan
UNIPAMPA

Dedico esse trabalho primeiramente a Deus pela inspiração, à minha família- Priscila, Mariana , meus pais e meu irmão – pelo incansável apoio emocional .Agradeço á minhas amigas, Ariane e Vanessa, pela grande paciência. Aos demais amigos pelo incentivo, e à todos que direta ou indiretamente. Meus sinceros agradecimentos ao professor Sandro pela orientação e ajuda, assim como , aos demais mestres que encontrei durante esse período que passei na UNIPAMPA

" Agradeço todas as dificuldades que enfrentei;
não fosse por elas, eu não teria saído do lugar.
As facilidades nos impedem de caminhar.
Mesmo as críticas nos auxiliam muito.."

Chico Xavier

RESUMO

O agronegócio é um setor de grande importância na economia do Brasil, representando em torno de um terço do PIB do país. O aumento das atividades nesse setor vem se dando de forma ampla e rápida exigindo o uso de novas tecnologias para seu gerenciamento e otimização. Dentro deste escopo, este trabalho tem por objetivo mostrar as capacidades da aplicação da tecnologia de *Data Warehouse* no setor do agronegócio. Foi utilizado como referência um estudo de caso efetuado em uma cooperativa da Região da Campanha, no estado do Rio Grande do Sul, conceituando e destacando as potencialidades do *Data Warehouse*. Neste estudo foi utilizada uma ferramenta chamada SPAGO BI.

Palavras-chave: SPAGOBI, *Data Warehouse*, *Data Mart*, BI.

ABSTRACT

Agribusiness has a large importance in Brazil economy, representing a third of Brazilian GNP. The growth of activities in this area is happening in a broad and fast way, demanding the use of new technologies for its management and optimization. In this scope, this paper has as main objective to present capabilities of Data Warehouse technology applied to agribusiness. It was used, as reference, a study case done in a cooperative of Campanha Region, in Rio Grande do Sul state, where it were defined concepts and emphasizing Data Warehouse potencialities. In this study, it was used a software tool called SPAGO BI.

Keywords: SPAGOBI, Data Warehouse ,*Data Mart*, BI.

LISTA DE FIGURAS

Figura 1 Exemplo de <i>Data Mart</i>	18
Figura 2 Esquema Estrela.....	23
Figura 3 Esquema da estrutura do DW	28
Figura 4 Exemplo de Cubo	33
Figura 5 Exemplo de hierarquia de uma dimensão	34
Figura 6 Resultado sem Aplicação de <i>Ranking</i>	36
Figura 7 Resultado com Aplicação de <i>Ranking</i>	36
Figura 8 Exemplo de MOLAP	36
Figura 9 Exemplo de Modelo ROLAP.....	37
Figura 10 Exemplo de HOLAP.....	37
Figura 11 Gráfico da divisão das cooperativas no RS.....	41
Figura 12 Representação Gráfica Receitas 2012.....	41
Figura 13 Diagrama do <i>Data Warehouse</i>	47
Figura 14 Diagrama Modelo E/R banco de dados Relacional	47
Figura 15 Aplicativo de Exportação.....	48
Figura 16 Arquivo gerado pela aplicação	48
Figura 17 Aplicativo Delphi de importação	49
Figura 18 Menu <i>Data Source</i>	49
Figura 19 Visualização Conexões Existentes.....	50
Figura 20 Configuração da Conexão.....	50
Figura 21 Criar Consulta.....	50
Figura 22 Seleção de Fields para a consulta	51
Figura 23 Resultado dos <i>Fields</i>	51
Figura 24 Seleção do Gráfico.....	51
Figura 25 Resultado da Consulta	52
Figura 26 Fluxo de um Engenho.....	54
Figura 27 Gráfico de Recebimento Mensal.....	55
Figura 28 Gráfico Tempo Médio de Descarga e Quantidade Kg.....	56
Figura 29 Peso Médio Placa.....	56

LISTA DE TABELAS

Tabela 1- Resultado da consulta Maio/2010	34
Tabela 2- Resultado <i>DrillDown</i>	34
Tabela 3- Resultado <i>Drill Across</i>	35
Tabela 4- Resultado <i>Drill UP</i>	35
Tabela 5 - Entidade Recebimento	45
Tabela 6 - Entidade Data	46
Tabela 7 - Entidade Pessoas	46
Tabela 8 - Entidade Produtos	46

LISTA DE SIGLAS

BI	<i>Business intelligence</i> (Inteligencia do Negocio)
CAMAL	Coopertiava Agricola Mista Acegua LTDA
CPF	Cadastro Pessoa Fisica
DOLAP	<i>Desktop OLAP</i>
DW	<i>Data Warehouse</i> (Armazem de Dados)
ETL	Extract Transform and Load (Extração Tranformação e Migração)
HOLAP	<i>Hybrid OLAP</i>
MDB	<i>Model Data Base</i>
MOLPA	Multidimensional OLPA
OCERGS	Organização das Cooperativas do Rio Grande do Sul
ODS	<i>Operation Data Store</i>
OLPA	On-Line Analytical Processing (Processamento Analítico ON-LINE)
PIB	Produto Interno Bruto
ROLAP	Relacionamento OLAP
SAD	Sistemas de Apoio a Decisão
SGDBS	Sistema Gerenciados de Banco de Dados
SQL	<i>Structured Query Language</i> (Linguagem de Consulta Estruturada)
WOLAP	Web OLAP

SUMÁRIO

1	INTRODUÇÃO.....	13
1.1	Justificativa.....	14
1.2	Objetivos	15
1.2.1	Objetivo Geral	15
1.2.2	Objetivos Específicos	15
1.3	Metodologia.....	15
1.4	Organização do Texto	16
2	<i>DATA WAREHOUSE</i>	17
2.1	Introdução.....	17
2.2	Objetivo do DW	17
2.3	Arquitetura do DW.....	18
2.4	Etapas do Desenvolvimento do DW.....	19
2.5	Busca pelos dados Operacionais	20
2.6	Modelagem.....	23
2.6.1	Normalização/ Desnormalização.....	24
2.7	Acomodação dos diferentes níveis de granularidade	26
2.7.1	Fusão (<i>merge</i>) de tabelas	26
2.7.2	Criação de <i>Array</i> de dados	26
2.7.3	Organização dos dados de acordo com suas características de estabilidade	26
2.8	Migração de Dados.....	27
2.9	Granularidade	28
2.10	Particionamento.....	29
2.10.1	Visões Particionadas.....	29
2.10.2	Tabela e índices Particionados	29
2.10.3	Vantagens do Particionamento.....	30
2.11	Povoando o Data Warehouse.....	30

	12
2.12	Data Warehouse X BD Operacional..... 31
2.13	Processamento Analítico On-Line (OLAP)..... 32
2.13.1	Tipos de Ferramentas 38
3	ESTUDO DE CASO 40
3.1	Origem das Cooperativas 40
3.2	Desenvolvimento do Protótipo 42
3.3	Requisitos Trabalhados 42
3.4	Tecnologias e Ferramentas Utilizadas..... 43
3.4.1	<i>SQL SERVER</i> 44
3.4.2	<i>Delphi</i> 44
3.5	Modelagem..... 45
3.5.1	Dicionário de Dados..... 45
3.6	Aplicativo <i>Delphi</i> 48
3.7	Spago BI..... 49
3.7.1	Conexão com o Banco de Dados..... 49
3.7.2	Criando Consultas 50
3.8	Problemas Enfrentados..... 52
4	RESULTADOS 54
4.1	Gráfico com a Produção de Cereis por ano 54
4.2	Gráfico: Tempo Médio de Descarga e Quantidade de KG..... 55
4.3	Peso médio dos caminhões..... 56
5	CONCLUSÕES..... 58
6	REFERÊNCIAS 59

1 INTRODUÇÃO

O Agronegócio é um setor vital na economia brasileira representando mais de 22% do Produto Interno Bruto (PIB) e mais de 30% dos empregos do País. Ministério da Agricultura, Pecuária e Abastecimento (2012). Em face de suas características e diversidades, o Brasil é um país com perspectiva satisfatória para o agronegócio. Para autores como Bacha (2000) o aumento da demografia mundial e sua consequente demandam por alimentos, nos conduzem a previsão de que o país alcançará o patamar de líder mundial no fornecimento de alimentos e *commodities* ligadas ao agronegócio, solidificando sua economia e alavancando seu crescimento.

Paralelamente, conforme dados levantados pela Organização das Nações Unidas (2012), as cooperativas possuem um papel fundamental para a redução da fome e da pobreza em comunidades rurais pobres, impulsionando as economias nacionais. Além disso, em um escopo regional, o Rio Grande do Sul contabiliza o maior número de cooperativas no país.

Ultimamente, no mundo dos agronegócios já existem *softwares* e metodologias aplicáveis para acolher os anseios desse tipo de mercado, tais como: identificação de padrões e tendências de negócios para destacar empresas no mercado. Para que estas ferramentas ou técnicas funcionem corretamente é necessário que os dados utilizados sejam corretos e completos; caso contrário, as informações geradas⁵¹ não serão informações satisfatórias ou confiáveis.

De acordo com Sorathia (2005), a negociação de produtos agrícolas é uma tarefa complexa, que além de envolver diferentes produtos e cenários, envolve a análise de grande quantidade de dados gerados em toda a cadeia produtiva. Apesar da importância crucial deste setor, há poucas soluções no Brasil para gerência de informação que auxiliam no processo de tomada de decisão e que tomam como base dados operacionais gerados pela cadeia produtiva do agronegócio como explicitado por Elias Correa (2009). Exemplos como o da Índia, conforme Rai (2005), que vem implementando uma *Data Warehouse* (DW) governamental para armazenar os dados gerados pelo agronegócio e deveriam ser seguidos pelo Brasil.

Assim, o presente trabalho tem como objetivo apresentar um estudo de caso sobre a implementação de um *Data Warehouse* em uma cooperativa localizada no sul do Rio Grande do Sul. O trabalho visa demonstrar os benefícios que esta tecnologia traz a este ramo de atividade que é fundamental na economia do estado e do país.

1.1 Justificativa

O acesso às técnicas, bem como, o acesso aos dados objetivando a procura de informações, é de suma importância, principalmente tratando-se de agronegócio. Desta forma, neste processo, a tomada de decisões rápidas e precisas têm sido imprescindível para as organizações a fim de torná-las mais competitivas e rentáveis nos dias de hoje. Segundo Carlos José Bacha (2000), estas adequam-se as modificações que ocorrem em seu meio ambiente, buscando informações corretas e facilmente acessíveis para tomada de decisões em um tempo bastante hábil. Com este objetivo, desde a década de 70, um conjunto de tecnologias vem sendo desenvolvido, dentre elas, os Sistemas de Apoio à Decisão (SAD), que através de uma evolução proporcionou o surgimento do *Data Warehouse*. Tais sistemas analíticos servem para os usuários identificarem, de forma geral, as ações e modelos que devem ser seguidos pela organização a qual pertencem, tornando a tarefa de acesso à informação mais dinâmica, rápida e confiável, uma vez que as informações encontram-se resumidas.

A investigação por sistemas computacionais mais eficientes, rápidos e com baixos custos vem sendo feita juntamente com o armazenamento de informações através de sistemas transacionais. Embora seja ideal começar o armazenamento em repositórios de dados totalmente do zero; dos pontos de visto financeiro, humano e do fator tempo, seria muito dispendioso para qualquer corporação que já armazena informações há anos em banco de dados existentes, conceber isto.

Assim, o presente trabalho justifica-se pela necessidade de se fazer um estudo da aplicação desta nova tecnologia, o *Data Warehouse*, e também, de identificar de que forma estas podem auxiliar na gestão de cooperativas do agronegócio.

Diante das justificativas adotadas, destaca-se, ao concluir este trabalho, que poder-se-á ter uma visão mais precisa do agronegócio da cooperativa escolhida para o estudo. Tal estudo mostra-se também, de grande valia para outras cooperativas, já que auxilia e maximiza os processos de tomada de decisões.

1.2 Objetivos

1.2.1 Objetivo Geral

Identificar e apresentar os benefícios trazidos pelo *Data Warehouse* aplicado ao agronegócio, em uma cooperativa localizada no sul do Rio Grande do Sul, visando uma melhora no processo de recebimento e gerência de cereais.

1.2.2 Objetivos Específicos

- Levantar os requisitos para a construção do DW;
- Construir um DW;
- Migrar dados do banco de dados relacional para o *Data Warehouse*;
- Fornecer recursos que auxiliem a cooperativa na tomada de decisão;
- Buscar e aplicar a ferramenta *Business intelligence* (BI).

1.3 Metodologia

O trabalho de pesquisa foi desenvolvido da seguinte maneira: em um primeiro momento, foi feita uma pesquisa do referencial bibliográfico sobre o tema aqui tratado. Logo em seguida, foi realizada a análise do projeto do Data Warehouse (levantamento de dados) e a migração dos dados do banco relacional para DW, permitindo uma implementação bem sucedida da ferramenta de BI.

Após, foi feita uma abordagem geral na tecnologia do *Data Warehouse*, consistindo no embasamento teórico e na compreensão dos conceitos básicos relacionados ao *Data Warehouse* utilizados no decorrer deste trabalho. Em seguida, foram realizadas a análise e o levantamento de requisitos do *Data Warehouse* para aplicação da metodologia.

A segunda fase consiste na apresentação do estudo de caso, onde a metodologia de organização de dados DW foi aplicada. Localizada na região da campanha do Estado do Rio Grande do Sul, a empresa escolhida para tal estudo atua no setor do agronegócio, no recebimento de cereais, tais como: arroz, soja, milho e sorgo. Por fim, com base nas informações obtidas na fase de estudo de caso foram apresentados benefícios no uso do *Data Warehouse*.

1.4 Organização do Texto

No capítulo 2 deste trabalho tratar-se-á sobre o *Data Warehouse*. Para tanto, serão apresentadas suas características, desenvolvimento, granularidade, dentre outros. No capítulo posterior será apresentado o estudo de caso e os benefícios que a migração de sistemas trouxe para a cooperativa aqui analisada.

Na parte final do trabalho, têm-se os resultados obtidos até o momento, seguidos pela conclusão e pela bibliografia utilizada na presente pesquisa.

2 DATA WAREHOUSE

Neste capítulo, tópicos referentes ao *Data Warehouse* serão tratados. Após uma breve explanação sobre aspectos históricos do DW, serão apresentados sua arquitetura e objetivos, suas etapas de desenvolvimento, além dos passos para extração dos dados de um sistema e sua migração para o DW.

2.1 Introdução

Ao longo de muitas décadas, as empresas armazenaram as informações em Sistemas de Gerenciamento de Banco de Dados (SGDBS) sem a obtenção de vantagens com os dados armazenados. Mas, com o passar do tempo, o que causava transtorno e gasto de recurso passa a ser visto como um benefício, de acordo com Machado (2011). Há, assim, o surgimento de muitas técnicas para extração de informações dos sistemas gerenciadores de banco de dados (SGDB), de forma a apoiar o processo de tomada de decisão. Hoje, as empresas que conseguirem extrair informações para o apoio à decisão ganham um maior destaque em relação às outras. Para tanto, o *Data Warehouse* é uma das tecnologias que ganham espaço.

Segundo Júnior (2004), DW é um banco de dados histórico, separado, lógica e fisicamente, do ambiente de produção da organização, concebido para armazenar dados extraídos deste ambiente. Já para Barbalho (2002), complementa-se o conceito:

“ O *Data Warehouse* é um sistema de suporte à decisão, composto por um conjunto de ferramentas que centralizam, armazenam, gerenciam e extraem informações históricas da empresa, em um formato “mastigado” para o tomador de decisão”

Os dados históricos agora podem gerar conhecimento estratégico, trazendo a ideia de centralização da informação, visualização multidimensional dos dados e a descoberta de padrões de comportamento, podendo, assim, os administradores tomarem decisões mais precisas, já que os resultados exploram dados legados.

2.2 Objetivo do DW

O principal objetivo do DW é disponibilizar informações de apoio à decisões para empresa, propiciando uma sólida e concisa integração dos dados, para a realização de análises gerenciais estratégicas de seus principais processos de negócio. A crescente utilização do *Data Warehouse* em empresas está relacionada à necessidade do domínio de

informações estratégicas para a garantia de respostas e ações rápidas, assegurando a competitividade de um mercado altamente competitivo e mutável como destacado por Barbalho (2002).

2.3 Arquitetura do DW

Para Júnior (2004), um *Data Warehouse* em sua arquitetura pode ser dividido em partes menores chamadas *Data Mart*. Tais partes são descritas pelo referido autor como um subconjunto de dados contidos em um *Data Warehouse* extraído para um ambiente separado. Esse subconjunto traz alguns benefícios, como:

Melhor desempenho;

Segurança dos dados, já que o *Data Mart* é um DW menor, dirigido a usuários específicos;

Manutenção mais facilitada.

Figura 1- Exemplo de *Data Mart*



Fonte 1- Junior (2004)

Para Machado (2011), existem três diferentes classificações nas quais os *Data Mart* e o *Data Warehouse* podem ser inseridos:

Global;

Independente;

Integrada.

A classificação global é mais generalista e é a que suportará parte dos requerimentos ou necessidades de um *Data Warehouse*, integrando as informações com grande grau de acesso e as utilizando em todos os departamentos de uma empresa.

Já a arquitetura independente implica em DW controlados por um grupo específico de usuários que atendem somente as suas necessidades específicas e departamentais sem foco corporativo. Este fato faz com que não exista nenhuma conectividade de um *Data Warehouse* com os DW de outros departamentos ou áreas de negócio.

Por fim, a arquitetura de *Data Warehouse* integrados é basicamente uma distribuição de implementação. Apesar dos DW serem implementados separadamente por grupos de trabalho ou departamentos, eles são integrados ou interconectados, provendo uma visão corporativa maior dos dados e informações. De fato o alto nível de integração é similar ao da arquitetura global. Por outro lado, os usuários de um departamento podem acessar e utilizar os dados de um *Data Mart* de outro departamento.

2.4 Etapas do Desenvolvimento do DW

A construção de um DW passa por quatro fases, segundo Barbalho (2002). O primeiro deles é o levantamento onde avaliam-se, junto aos tomadores de decisão, os conhecimentos que desejam ser adquiridos. Logo em seguida, tem-se a modelagem multidimensional. Nesta forma de modelagem representamos a ideia central e suas dimensões e definimos como os dados serão armazenados. A terceira fase desta construção é o ETL (*Extract, Transform and Load*), que é a exportação dos dados nos sistemas corporativos e transformação para carga no DW. Por fim, tem-se a visualização do resultado através de ferramentas para interação com o usuário com interfaces amigáveis.

Já para INNON (1997), as fases para a elaboração de um *Data Warehouse* são:

Busca pelos dados operacionais;

Modelagem dos dados;

Normalização/ desnormalização;

Migração dos dados.

Pode-se verificar ainda que para Innon (1997), a fase mais importante é a de levantamento de requisitos, pois nela definimos a real necessidade do usuário. Levando-se

isso em conta, dá-se início a modelagem do DW ou *Data Smart* utilizada no presente trabalho.

2.5 Busca pelos dados Operacionais

A extração dos dados operacionais nem sempre é uma tarefa fácil. A falta de integração entre os sistemas existentes é comum, já que na época da criação das aplicações de sistemas legados, a possibilidade de integração desta não era considerada. Desta forma, a falta de integração passa a ser o grande desafio da extração dos dados operacionais.

Um exemplo comum que pode ser destacado é o atributo sexo em uma tabela de clientes. Muitos sistemas podem tratar este atributo como masculino e feminino, ou homem e mulher, ou 0 ou 1 e até mesmo, erros de digitação por parte dos usuários.

Há três tipos de cargas que podem ser feitas do ambiente operacional para o *Data Warehouse*, segundo INNON (1997):

- carregamento de dados históricos;
- carregamento de dados de valor corrente no ambiente operacional;
- carregamento de alterações do *Data Warehouse* a partir de alterações (atualizações) que tenham ocorrido no ambiente operacional desde a última atualização do DW.

O carregamento de dados de sistemas legados pode ser considerado como o menor desafio, pois será feito de uma única vez. Já a extração do sistema relacional corrente poderá ser feita através de um arquivo sequencial já tratado, dando carga assim, no *Data Warehouse*.

Pode-se considerar a atualização do *Data Warehouse* a maior dificuldade, pois é difícil prever o que foi alterado, incluído ou deletado a partir da última carga. Quanto menor o intervalo de atualização mais difícil é a previsão de atualizações. Um exemplo que pode ser demonstrado é a venda de calçados em site. Se a carga de vendas for semanal, o cliente tem a possibilidade de trocar o produto em sete dias, o evento da venda, pode ter acionado muitos eventos como: baixa no estoque, contas a receber e/ou pedidos ao fornecedor. Se for gerada uma carga neste intervalo, provavelmente, o estoque ou contas a receber terão informações desatualizadas. A varredura de arquivos existentes é, portanto, uma importante questão a ser enfrentada pelo arquiteto do *Data Warehouse*.

Conforme descrito por Innon (1997), em seu livro *Como construir o Data Warehouse*, há cinco técnicas comumente usadas para limitar a quantidade de dados pesquisados. A primeira delas consiste em buscar dados que têm como domínio marcas do tempo, quando uma aplicação registra uma alteração ou atualização de um registro.

A segunda técnica é a criação de um arquivo “*delta*”, no qual o *software* gerador dos dados operacionais é responsável por manter um *log* do que foi alterado. Quando é possível contar com este tipo de arquivo, o processo de carregamento se torna mais eficiente, já que os dados que não forem candidatos à varredura não serão salvos. Contudo, poucas aplicações são capazes de gerar este tipo de arquivo.

Com a terceira técnica, percorre-se o arquivo de *log* gerado pelo banco de dados. Este contém basicamente dados do mesmo tipo dos de um arquivo *delta*. Um obstáculo encontrado consiste no fato de o formato interno ser gerado para atender os objetivos de um sistema e não aos de uma determinada aplicação. Outra falha dos arquivos *log* é que, geralmente, eles contêm muitas informações além daquelas procuradas pelo desenvolvedor do *Data Warehouse*.

A quarta técnica usa a modificação da aplicação em produção para a mesma gerar o arquivo *delta* ou um *log* dos registros modificados. Porém, esta técnica geralmente não é empregada, já que muitas vezes o *software* em produção não pertence a mesma empresa fornecedora da solução de DW.

Na sequência, a quinta técnica visa criar um arquivo de imagem do banco de dados com a posição da última extração e comparar com o arquivo de imagem atual. Os dois arquivos serão comparados serialmente entre si, detectando o que tiver de diferente entre eles. Essa técnica é pesada, complexa e demanda alocação muito grande de recursos.

Junior (2004) faz uma observação importante de que é preciso agrupar os dados do ambiente relacional para o multidimensional. Nesta forma, o volume de dados contidos no DW será muito grande para ser controlado. A condensação deverá ser feita no momento da extração dos dados. Para ele, todas as informações obtidas de um DW devem ser originadas de sistemas externos. A colocação destes dados e a preparação completa, de modo a poder utilizá-los, é uma das principais etapas de um DW. Junior (2004) ainda enumera algumas etapas as quais são necessárias para transformar dados não processados, de origem externa, em informações prontas para serem acessadas pelo *Data Warehouse*.

Para o autor os dados precisam ser convertidos para o formato interno do banco de dados escolhido para o DW, a partir de uma variedade de representações externas, incluindo registros de comprimento variável e fixo, formatos de caracteres e binários,

dentre outros. Estes dados precisam ser filtrados de modo a rejeitarem valores inválidos, chaves repetidas ou registros com outros tipos de erros.

Já os registros precisam ser reorganizados a partir de representações de *flat files*¹, de modo a ficarem compatíveis com o esquema do DW. Quando o sistema fonte utiliza um banco de dados diferente ou incompatível ou, até mesmo, utiliza outra estrutura de arquivos diferente de um SGBD, o mesmo deverá gerar *flat files* com os dados a serem carregados no DW. Estes precisam ser gravados fisicamente, observando as exigências de configuração para particionamento de dados, balanceamento de disco, entre outros. Ainda, precisam ser verificados em relação ao banco de dados existente de modo a assegurar consistências, mantendo uma integridade referencial completa. É importante ressaltar que os registros precisam estar corretamente indexados.

O sistema deverá possuir algumas tabelas, as quais serão responsáveis pelo gerenciamento do processo de carga. A tabela de carga contém os arquivos gerados *flat file* para o DW, marcando a periodicidade e a data da última carga, enquanto que a tabela de sistemas e seus respectivos *flat files* gerados contém os *flat files* que cada sistema gera, bem como as rotinas de carga de cada um, nas diversas fases do processo. As rotinas de carga são geralmente escritas na linguagem do banco de dados escolhido e/ou empregando utilitários de conversão de banco de dados.

A tabela de tarefas contém os arquivos de cada sistema, que serão carregados diariamente, e a data do movimento. Deverá existir, também, um campo para controlar as diversas situações do arquivo ou tabela, tais como: em espera, carregando na *staging* área, carregando no *Operational Data Store* (ODS), carregando no DW, carregado com erro e invalidado. A tabela de dependência de arquivos deve conter as dependências entre arquivos para a sequência correta de carga, porém, alguns arquivos não podem ser carregados antes de outros.

Tem-se ainda a tabela de controle de fim de carga, que contém os nomes de todos os arquivos já carregados. Assim, para carregar os dados nas diversas camadas, o programa deverá exibir a relação de sistemas com suas respectivas datas de processamento. O operador poderá fazer alguma alteração nas datas antes do processo começar, montando assim a tabela de carga. A partir das tabelas de carga e de sistemas, é montada a tabela de tarefas, ou seja, a lista de sistemas que serão carregados no dia.

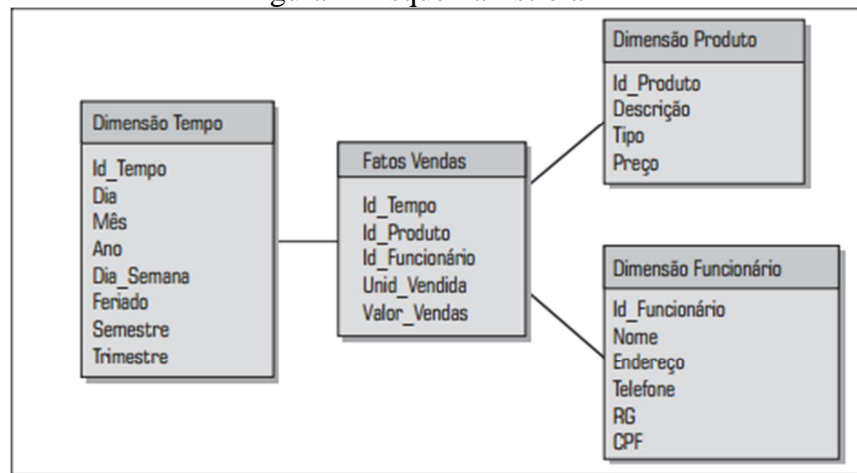
¹ É um arquivo plano geralmente do tipo texto, formatado para conversão de outro banco de dados. O *flat file* possui registros de tamanho único, onde cada registro de detalhe gera uma inclusão no banco de dados.

2.6 Modelagem

A modelagem de um DW consiste em um modelo chamado *Star Schema* (Esquema Estrela). O conceito estrela foi criado pelo Dr. Ralph Kimball, segundo Júnior (2004), e tem como características básicas a presença de dados altamente redundantes para se obter um melhor desempenho. Essa desnormalização ocorre devido a junções de tabelas para que não haja a necessidade de ocorrerem junções de execução.

O nome estrela foi adotado pela semelhança com uma estrela. Este esquema é composto de uma tabela dominante, chamada de tabela de fatos, no centro, rodeada por tabelas auxiliares, chamadas de tabelas dimensão. A tabela de fatos conecta-se as demais por múltiplas junções e as tabelas de dimensões se conectam com apenas uma junção à tabela de fatos.

Figura 2- Esquema Estrela



Fonte 2- Junior (2004)

Neste esquema, a consulta ocorre inicialmente nas tabelas de dimensão e depois na tabela de fatos, assegurando assim a precisão dos dados através de uma estrutura completa de chaves onde não é preciso percorrer todas as entidades. Isso garante um acesso mais eficiente e o mais alto desempenho possível.

As tabelas dimensionais são desnormalizadas para aumentar o desempenho e armazenam as informações a respeito das dimensões dos dados. Cada coluna deverá conter descrições textuais significativas para os usuários e deverão ser apropriadas para relatórios.

2.6.1 Normalização/ Desnormalização

Conforme Isabel Italiano (2010), a modelagem relacional desnormalizada tem, normalmente, como ponto de partida, um modelo de dados da empresa já construído. Estes modelos de dados são construídos pela utilização de técnicas clássicas de modelagem com objetivo de agrupar e sintetizar as necessidades dos dados da organização, geralmente não fazendo distinção entre ambientes transacionais e de suporte a decisão.

Italiano (2010) resume as oito regras de desnormalização de dados de INNON (1997), sejam elas:

- Remoção de dados puramente operacional;
- Adição de um elemento de tempo a estrutura da chave do *Data Warehouse*;
- Adição de dados derivados apropriados;
- Criação de artefatos de relacionamentos;
- Acomodação dos diferentes níveis de granularidade encontrados no DW;
- Fusão (*merge*) de dados semelhantes provenientes de diferentes tabelas;
- Criação de *array* de dados;
- Separação dos atributos de dados de acordo com suas características de estabilidade.

A sequência de aplicação dessas etapas não deve ser considerada como imutável, mas como um critério geral a ser seguido.

Remoção de dados puramente operacionais

A remoção de dados puramente operacionais possibilita apenas selecionar os elementos importantes para o *Data Warehouse*, excluindo assim, aqueles que não terão relevância na tomada de decisão. Porém, é importante ressaltar que um atributo que primeiramente é classificado como dispensável, mais tarde poderá ser visto como uma qualidade importante na tomada de decisão.

“Infelizmente, podemos imaginar circunstâncias onde quase qualquer dado possa ser usado em DSS (Decision Support Systems). Com imaginação fértil quase qualquer dado se qualifica como sendo aplicável a sistemas de suporte à decisão”

(ITALIANO e ESTEVES, 2010).

Podem se classificar como dados não operacionais o número de identidade, CPF (Cadastro de Pessoa Física), do cartão de crédito, entre outros, ressaltando que tudo dependerá do objetivo do *Data Warehouse*.

Adição de um elemento de tempo à estrutura da chave do *Data Warehouse*

Um elemento tempo, que não está presente no sistema, pode ser classificado como data da operação, mês que foi efetuado, bimestre ou semestre.

Adição de dados derivados apropriados

Devem ser adicionados dados derivados quando estes são muito acessados e calculados uma única vez. A adição de dados derivados faz sentido porque reduz o processamento necessário quando se acessa os dados do DW.

Criação de Artefatos de relacionamentos

Os relacionamentos de dados utilizados na modelagem de dados clássica são para o ambiente operacional. Estes pressupõem a existência de apenas uma única regra de negócio subordinada ao relacionado. Para a premissa de que os dados são precisos no momento de seu acesso (dados operacionais), a representação clássica de um relacionamento é correta. Entretanto, para um DW, existirão geralmente várias regras de negócio entre tabelas de dados. Isto se deve ao fato de que dados em um DW representam situações que abrangem um grande aspecto temporal. O relacionamento entre tabelas no DW é obtido por intermédio da criação de “artefatos”.

Artefatos são itens tangíveis, criados e usados para se desenvolver uma tarefa. Eles incluem dados, documentos e conteúdo armazenados em um formato eletrônico, que podem ser gerenciados e utilizados em benefício das empresas, seus clientes, fornecedores entre outros.

O relacionamento de dados no ambiente transacional é um relacionamento ativo e contínuo, preciso no momento que o usuário acessa os dados sobre o relacionamento de dados. Já o relacionamento de dados no DW é preciso em um determinado instante de tempo, geralmente no passado.

2.7 Acomodação dos diferentes níveis de granularidade

Uma das características do *Data Warehouse* é a existência de diferentes níveis de granularidade das informações. Em algumas situações o nível de granularidade não se altera na medida em que os dados passam do ambiente transacional para o DW. Em outras situações, o nível de granularidade do modelo de dados *Data Warehouse* precisará refletir essas alterações.

Alguns pontos relativos à mudança da granularidade - considerados pelo modelador do DW - podem refletir em perguntas referentes ao período de tempo para a sumarização dos dados (isto é, por dia, semana, mês e etc), aos elementos de dados que deverão estar contidos na tabela sumarizada e ao ambiente operacional que suportará os elementos de dados sumarizados. Desta forma, a partir de tais questionamentos, seria o arquiteto do DW possível de ser calculado a partir das fontes de dados transacionais?

2.7.1 Fusão (*merge*) de tabelas

Nesta fase define-se a transformação de tabelas corporativas em uma tabela do DW. O *merge* economiza espaço e melhora a performance. Assim, as condições nas quais uma fusão (*merge*) faz sentido são onde as tabelas compartilham uma chave em comum e ainda, onde os dados de diferentes tabelas são usados em conjunto, sem esquecer o necessário padrão de inserção de dados. De acordo com Innon (1997), se uma dessas condições não for atendida, não se recomenda efetuar a fusão (*merge*) das tabelas.

2.7.2 Criação de *Array* de dados

Os dados no modelo de dados de uma empresa estão geralmente normalizados, significando que grupos de repetição não são mostrados como parte do modelo de dados. Entretanto, em condições apropriadas, o DW pode e deve conter grupos de repetição.

A criação de *arrays* de dados não é uma opção genérica. Desta maneira, somente sob as circunstâncias adequadas serão obtidos benefícios pela criação de *arrays* de dados no modelo do DW.

2.7.3 Organização dos dados de acordo com suas características de estabilidade

O modelo de dados da empresa faz pouca ou nenhuma menção à taxa de alterações realizadas em uma tabela, mas o DW é bastante sensível a mudanças em seus dados. Com

a organização otimizada em um DW é possível dividir em tabelas distintas, dados com alta e baixa taxa de atualização. A estrutura do DW é, assim, composta por estruturas que são compatíveis em volatilidade.

As oito regras de Inno, aqui mencionadas, mostram uma possível solução de arquitetura de dados para o modelo de nível atômico e sumarizado do DW. Estas regras relatam a arquitetura utilizada na construção do *Data Warehouse*, mas que nem sempre consistem na solução final que poderá estar contida nestas regras.

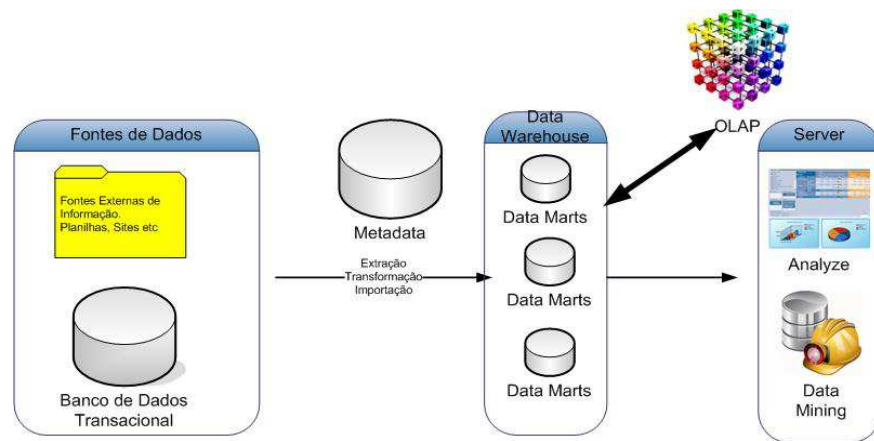
2.8 Migração de Dados

Um *Data Warehouse* é povoado com dados a partir do ambiente operacional que pode possuir diversas fontes de informação espalhadas, tanto geograficamente, quanto em diferentes plataformas de *software* e *hardware*. O povoamento do *Data Warehouse* se dará através da aquisição de dados que envolvem os processos de identificação, captura e transformação de dados originados de sistemas operacionais, bem como, o processo da carga dos dados que pode ocorrer em um *Data Warehouse* ou *Data Mart*.

A migração de dados pode ser uma das partes mais complexas, demoradas, e de altos custos para a construção e gerenciamento do *Data Warehouse*. O motivo principal de tal processo migratório é que empresas descobrem frequentemente que os dados que elas querem são, lamentavelmente, inexatos e incompatíveis. As funções principais executadas em migração de dados são comumente definidas como extração, transformação, transporte e carga de dados.

Segundo Laura Sarkis (2001), o processo de extração dos dados envolve adquirir dados relevantes ao negócio nos bancos de dados operacionais. Os dados extraídos são transformados, sofrem sumarizações e agregações, para que então possam estar mais adequados para o suporte à decisão. Em seguida, os dados passam por processos de limpeza e de consistência das informações obtidas pelos processos anteriores. Ao término destes processos, é feita a integridade dos dados, onde ocorre a padronização de definições, taxonomia de atributos e o registro das variações que possam ser encontradas nas múltiplas fontes de dados, tanto no ambiente operacional, como nas tendências externas. É necessário que os atributos e dados relevantes sejam cuidadosamente tratados para então povoarem o DW, principalmente, para tornar os processos de acesso aos dados uma solução mais eficiente.

Figura 3- Esquema da estrutura do DW



Fonte 1 -Própria

Como podemos ver na figura 2 acima, o processo de migração de dados começa com a extração dos dados fontes, geralmente sistemas legados, para uma área de estágio comum, chamada esquema intermediário, para que então sejam executados os demais processos necessários para se chegar ao povoamento do *Data Warehouse*.

2.9 Granularidade

Conforme Machado (2011), a granularidade de dados refere-se ao nível de sumarização dos elementos e de detalhes disponíveis nos dados sendo considerado, o mais importante aspecto do projeto de um *Data Warehouse*.

A razão pela qual a granularidade é a principal questão de projeto consiste no fato de que ela afeta profundamente o volume de dados que reside no *Data Warehouse*, e ao mesmo tempo, o tipo de consulta que pode ser atendida. Desta forma, quando se tem um nível de granularidade muito alto, o espaço em disco e o número de índices necessários se tornam bem menores, mesmo que haja uma correspondente diminuição da possibilidade de utilização dos dados para atender a consultas detalhadas.

O mais importante da granularidade em um projeto é entendermos que ela não se limita ao tempo, mas que ela abrange todos os fatores de classificação da informação que estiverem sendo utilizados. À medida que o nível de granularidade se eleva, há uma correspondente diminuição da possibilidade de utilização dos dados para atendimento a consultas. Já com um nível mais baixo de granularidade é possível responder a qualquer consulta necessária.

O bom-senso é o método mais indicado para a agregação e sumarização dos dados sobre granularidade, assim como a análise detalhada das necessidades de informação levantadas para o projeto. Ouvir atentamente o usuário, discutir e propor alternativas que levam a uma correta granularidade do projeto.

2.10 Particionamento

Um DW, por definição, representa um banco de dados histórico com grande volume de dados. Tal característica traz consigo problemas relacionados ao gerenciamento de estruturas com centenas de *gigabytes* e até mesmo *terabytes* de dados (JÚNIOR, 2004).

Para solucionar os problemas de desempenho, os SGBDs passaram a oferecer mecanismos de suporte a grandes estruturas de dados para sistemas de missão crítica e DW. Embora os SGDBS apresentem diferentes implementações, todos compartilham da mesma ideia: decompor uma estrutura em estruturas menores que possam ser mais facilmente gerenciadas e que possam oferecer melhor desempenho. Assim, o principal mecanismo oferecido atualmente é o particionamento.

2.10.1 Visões Particionadas

Também conhecido como pseudoparticionamento ou particionamento manual, esse mecanismo, ainda presente como o único em muitos SGDBS, representou a primeira solução para o problema de frágeis estruturas de dados. Esse tipo de particionamento é alcançado através de duas atividades: a primeira consiste em dividir uma tabela em diversas outras com estruturas idênticas e, através de restrições, separar os dados pertencentes a cada uma delas, já a segunda atividade é a criação de uma visão de banco de dados que faça a união dos dados de todas as estruturas idênticas. Uma vez definida a visão, consultas direcionadas a ela que contenham restrições no atributo de divisão, farão referências apenas às estruturas efetivamente necessárias.

2.10.2 Tabela e índices Particionados

O particionamento de tabelas e índices, muitas vezes referenciado com particionamento nativo, representa o mecanismo mais moderno de particionamento oferecido pelos SGDBS. Nesses modelos, os SGDBS disponibilizam objetos, como tabelas e índices, que podem ser particionados. A vantagem desse tipo de particionamento

é que ele é inerte ao objeto, deixando transparente para o usuário todo o controle da separação de dados.

2.10.3 Vantagens do Particionamento

O particionamento de estruturas, sejam elas tabelas ou índices, apresenta as seguintes vantagens: redução do tempo de execução de consultas, ao eliminar partições que não atendem ao critério da consulta e a redução do tempo de indisponibilidade em função de tarefas de manutenção, sendo algumas tarefas de criação, a reconstrução de índices e a eliminação de dados referentes a determinado período altamente favorecidos pelo particionamento. Outro benefício trazido é o da redução do tempo para execução de *backup* e restauração, podendo as partições serem armazenadas individualmente (espaços de tabela). Pode-se, por exemplo, distribuir partições mais recentes em dispositivos mais rápidos e mover partições com dados históricos poucos acessados para dispositivos mais lentos.

2.11 Povoando o Data Warehouse

A extração, a limpeza, a transformação e a migração de dados dos sistemas existentes na empresa para o *Data Warehouse*. Segundo Innon (1997), constituem tarefas críticas para o seu funcionamento efetivo e eficiente. Diversas técnicas e abordagens têm sido propostas, algumas bastante genéricas e outras especialmente voltadas para a manutenção de integridade dos dados num ambiente caracterizado pela derivação e a replicação de informações.

O grande desafio por trás da alimentação de dados das fontes para o *Data Warehouse* não é técnico, mas gerencial. Muitos dos processos envolvidos, como mapeamento, integração e avaliação de qualidade, ocorrem de fato durante a fase de análise e projeto do *Data Warehouse*. Especialistas afirmam que identificar fontes, definir regras de transformação e detectar e resolver questões de qualidade e integração consomem cerca de 80% do tempo de projeto. Infelizmente, não é fácil automatizar estas tarefas.

Embora algumas ferramentas possam ajudar a detectar problemas na qualidade dos dados e gerar programas de extração, a maioria das informações necessárias para desenvolver regras de mapeamento e transformação existem apenas para os analistas e usuários. Fatores que certamente influenciam na estimativa de tempo para estas tarefas são o número de fontes e a qualidade dos metadados mantidos sobre estas fontes. As

regras de negócio associadas a cada fonte, tais como validação de domínios, regras de derivação e dependências entre elementos de dados, são outra fonte de preocupações. Se estas regras tiverem de ser extraídas do código fonte das aplicações, o tempo para mapeamento e integração pode dobrar.

Os processos associados à migração de dados das fontes para o *Data Warehouse* incluem extração de dados, limpeza, transformação e carga no *Data Warehouse*.

2.12 Data Warehouse X BD Operacional

A principal diferença entre o Data Warehouse e o BD operacional está no fato de *Data Warehouse* ser um ambiente informacional, enquanto que os bancos de dados convencionais das organizações, como os localizados em suas filiais, são ambientes operacionais.

De acordo com Innon (1997), em um ambiente operacional, manipula-se um volume grande de transações que geralmente são simples, pequenas e acessam poucos registros por vez. Já no ambiente informacional, manipula-se um baixo volume de transações longas e complexas e ainda, acessam-se muitos registros, necessitando muitas vezes, realizar funções de junção e agregação.

Uma diferença entre DW e Banco de Dados relacional segundo Jacobson (2007) reside na característica de que um banco de dados transacional ajuda pessoas a executar atividades, enquanto o DW ajuda a planejar. Por exemplo, um banco de dados de transação pode informar quais assentos estão disponíveis em um voô para que o agente possa fazer uma nova reserva. Um DW, por outro lado, pode informar o padrão histórico de assentos vagos por voô para que um gerente da companhia aérea decida se fara ajustes nas agendas de voô futuramente.

Outra diferença está no fato de que um banco de dados transacional se concentra nos detalhes, enquanto um DW se concentra em agregados de alto nível. Ainda tem-se que um banco de dados de transação geralmente é projetado para um aplicativo, tendo que propiciar recuperação e atualização rápidas de firmações detalhadas enquanto um DW integra dados de fontes diferentes, devendo proporcionar uma recuperação rápida de informações altamente resumidas. Por exemplo, seu aplicativo de processamento de pedidos e seu respectivo banco de dados provavelmente incluem informações detalhadas sobre descontos para cada pedido, mas nada sobre excessos de custos de fabricação, ao contrário, provavelmente incluem informações detalhadas sobre custos, mas nada sobre descontos em vendas. Assim, combinando as duas fontes de um DW, seria possível

calcular a rentabilidade real das vendas de um produto, revelando, possivelmente, que o preço final com desconto é menor que o custo real de fabricação.

Um banco de dados transação se preocupa com o agora, sendo volátil. Suas informações mudam constantemente conforme novos pedidos chegam ou são cancelados, conforme novos produtos são criados ou enviados ou ainda, conforme novas reservas são feitas. Já um DW se preocupa com a atividade todo tempo, sendo estáveis, suas informações são atualizadas em intervalos padrão.

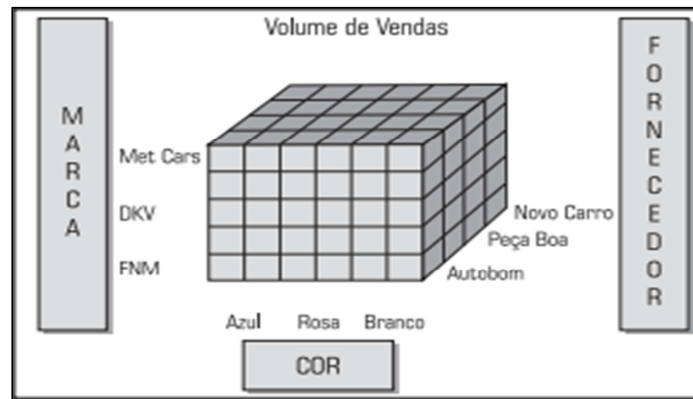
2.13 Processamento Analítico On-Line (OLAP)

O termo *On line Analytical Processing* (OLAP) refere-se a um conjunto de ferramentas voltadas para acesso e análise *ad-hoc* de dados. Assim sendo, como nos mostra Carvalho (2010), o objetivo final de uma ferramenta OLAP é transformar dados em informações capazes de dar suporte a decisões gerenciais de forma amigável e flexível e em tempo hábil ao usuário. Tal termo foi criado em 1993 por E.F.Codd, que instituiu 12 regras básicas, que serviriam para definir se uma ferramenta seria ou não considerada OLAP.

Em meados de 90 surgia no mercado uma nova classe de ferramentas que traziam tecnologia OLAP embutida. Tais ferramentas trouxeram uma grande capacidade de efetuar cálculos complexos como previsões, percentuais de crescimento e médias diversas, considerando-se a variável tempo, muito importante no contexto gerencial. Assim, OLAP ajuda a analisar de forma mais eficiente à quantidade de dados crescentes armazenados pelas organizações, transformando-os em informação, possibilitando o acesso mais rápido e permitindo uma análise dinâmica e multidimensional

A multidimensionalidade é o conceito chave da análise feita através de ferramentas OLAP. Neste tipo de análise os dados são modelados em uma estrutura conhecida como cubo que permite a observação de vários assuntos (dimensões) para uma mesma massa de dados. As dimensões desse cubo representam os componentes dos negócios da empresa como cliente, produtos, fornecedores e tempo, que na prática, são os diversos ângulos ou óticas sob as quais uma informação pode ser analisada. Desta forma, a célula resultante da interseção das dimensões do cubo é chamada de média, representando dados numéricos como vendas, lucros e quantidade de unidades vendidas.

Figura 4 - Exemplo de Cubo



Fonte 2 - (Júnior, 2004)

Atualmente as ferramentas OLAP fornecem suporte para funções de derivação de dados complexos, que recebem o nome de *slice and dice*. Para Júnior (2004) O suporte *slice and dice* é uma das principais características de uma ferramenta OLAP, servindo para a modificação da ordem das dimensões e alteração de linhas por colunas de maneira a facilitar a compreensão dos usuários. Essa característica é de extrema importância, pois com ela é possível analisar as informações de diferentes prismas, permitindo ao usuário investigar diferentes inter-relacionamentos. O *slice and dice* compreendem quatro operações: *ranking*, *drilling*, *rotation* e *ranking*.

Ranking

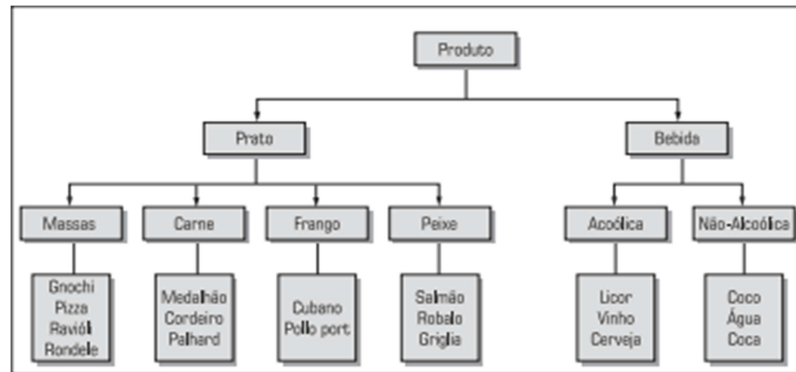
É a operação responsável por alterar o resultado das consultas a qualquer momento, inserindo novas posições ou removendo as que estão em foco. Para que isto ocorra é preciso que o usuário informe o que está modificando e o que, então, será feito, o que pode ser, por exemplo, a inserção de um novo produto em uma consulta. O resultado desse *Ranking* será considerado para todas as demais operações, ou seja, pode-se encarar o resultado como um novo cubo gerado a partir do cubo original.

Drilling

Após escolher o que deseja analisar, o analista ainda pode mudar o escopo do que está analisando, porém essas informações podem encontrar-se agregadas em diversos níveis. O *drilling* permite a navegação por entre estes níveis.

Existem três operações OLAP que permitem ao analista mudar o escopo dos dados (MACHADO, 2011): *Drill Down*, *Drill Up* e *Drill Across*.

Figura 5- Exemplo de hierarquia de uma dimensão



Fonte 3- (JÚNIOR, 2004)

Drill Down

Esta operação navega verticalmente na hierarquia, no sentido em que os dados são mais atômicos. Por exemplo, ao totalizar as vendas de um restaurante, poderíamos obter os seguintes resultados: total de vendas por diferentes categorias de alimentos e outro, por tipos de pratos, analisando somente aqueles feitos com massa.

O primeiro poderia ser do tipo:

	MASSAS	CARNES	FRANGO	PEIXE
TOTAL	563	278	286	198

Tabela 1- Resultado da consulta Maio/2010

Com esses resultados o analista não pode tomar decisão alguma, assim, para que os dados sejam mais específicos será preciso utilizar o conceito de *drilling*, alterando, desta forma, o escopo da análise. A fim de conseguir informações mais detalhadas sobre os pratos de massas e carnes, o analista precisará navegar ao longo das hierarquias de cada dimensão, até chegar no nível mais específico de cada prato.

Executado o *DrillDown* em massas obtém-se o seguinte:

	GNOCHI	PIZZA	RAVIOLI	RONDELE
TOTAL	96	274	84	103

Tabela 2- Resultado *DrillDown*

O resultado, no gráfico acima, mostra que o motivo da elevação das vendas de massas foi gerado pelo produto Pizzas.

Drill Across

Esta operação permite a navegação transversal no meio da árvore hierárquica. O *Drill Across* é uma operação de grande utilidade, pois permite inserir e retirar posições do cenário corrente.

Se for necessário comparar as vendas de pizza com os demais pratos de carne, seria preciso navegar no mesmo nível de detalhe das posições da dimensão. Neste caso, o resultado seria o seguinte:

	PIZZA	MEDALHÃO	CORDEIRO	PALHARD
TOTAL	96	274	84	103

Tabela 3- Resultado *Drill Across*

Drill Up

A última operação do *driling* realiza a função inversa do *Drill Down*. Ela permite ao usuário uma visão mais agregada das informações. Nesta técnica podemos navegar nos diversos níveis.

	PRATOS	BEBIDAS
TOTAL	1325	2529

Tabela 4- Resultado *Drill UP*

Rotation

Além de mudar as posições de fato, o analista também tem a flexibilidade de alterar a forma de visualização das informações. Vale salientar que o *rotation* não adiciona nem retira posições do cenário, sendo a inexistência de operações de disco no servidor sua grande vantagem. O que muda é como os dados serão dispostos para o usuário.

Ranking

Com esta operação o analista pode filtrar as informações que ele deseja ver, sendo possível fazer uma classificação dos dados obtidos e operar diretamente sobre os valores das células. Vale frisar que as operações anteriores atuavam apenas sobre as posições ou dimensões dos dados. Através do Ranking, o analista pode executar diversos tipos de filtros, eliminando assim informações desnecessárias.

Figura 6 - Resultado sem Aplicação de *Ranking*

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
GNOCCHI	86	89	86	87	83	90	90	89	86	78	75	91
PIZZA	115	110	111	103	106	98	119	112	109	109	115	113
RAVIOLLI	75	74	76	78	74	77	80	78	79	77	73	78
RONDELE	93	96	92	94	91	93	98	97	93	95	96	98

Fonte 3 - Junior (2004)

Utilizando a planilha acima, é possível aplicar o ranking de quais os 4 pratos mais vendidos.

Figura 7 - Resultado com Aplicação de Ranking

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
GNOCCHI												
PIZZA	115						119				115	113
RAVIOLLI												
RONDELE												

Fonte 4 - Junior (2004)

TIPOS DE PROCESSAMENTO

Segundo (CARVALHO, 2010) existem três tipos básicos de processamento:

Mecanismo do servidor Model Data Base MDB em MOLAP (Multidimensional OLAP) e HOLAP (Hybrid OLAP).

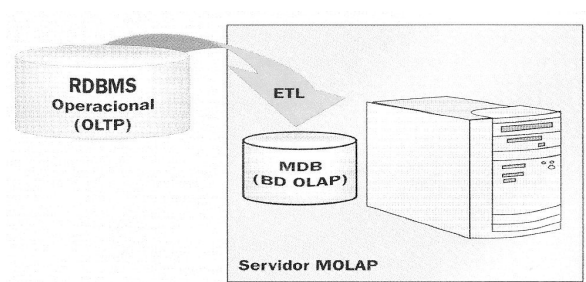
A Linguagem SQL que utiliza em ROLAP (OLAP Relacional) e HOLAP (Hybrid OLPA);

Mecanismos multidimensionais nas estações clientes, em caso de DOLAP (Desktop OLPA) e WOLAP (Web OLPA).

MOLPA (Multidimensional OLPA)

Na arquitetura MOLAP os dados ficam armazenados em um banco de dados multidimensional, onde o servidor MOLAP atua. O usuário trabalha, monta e manipula os dados diretamente no servidor.

Figura 8- Exemplo de MOLAP



Fonte 5 - Carvalho 2010

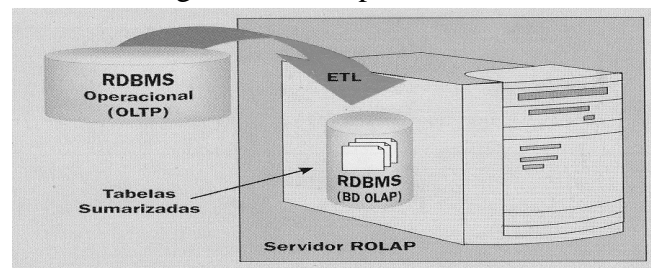
O armazenamento em um banco de dados multidimensional utiliza um espaço menor que o utilizado para a armazenagem dos mesmos dados em um banco de dados relacional, não utilizando chaves estrangeiras para os relacionamentos.

ROLAP (OLAP Relacional)

ROLAP é uma simulação das ferramentas *front-end* que permitem efetuar requisições multidimensionais, transformando estas consultas em rotinas SQL da tecnologia OLAP, feita em banco de dados relacional e que também, possuem a vantagem de não restringir o volume de armazenamento de dados.

Suas principais características são a possibilidade de fazer qualquer consulta, visto que não se está limitado ao conteúdo de um “Cubo” e a capacidade de navegar nos dados até atingir o nível de detalhe mais baixo, ou seja, de menor granularidade. Como limitações têm-se o pobre conjunto de funções para análise dimensionais e o baixo desempenho da linguagem SQL na execução de consultas pesadas. Assim, as ferramentas ROLAP atendem melhor os usuários que não tem um escopo de análise bem definido.

Figura 9 - Exemplo de Modelo ROLAP

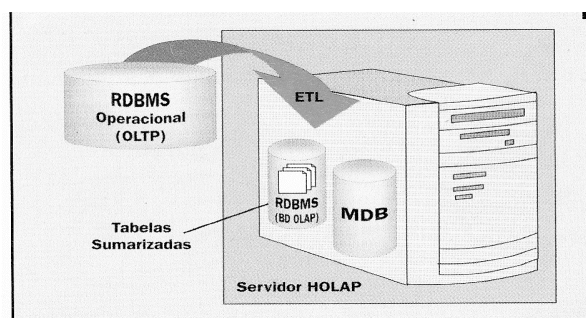


Fonte 6 - Carvalho 2010

HOLAP (Hybrid OLAP)

Uma forte tendência de arquitetura OLAP atualmente é o HOLAP ou processamento híbrido que combina a capacidade e escalabilidade das ferramentas ROLAP com o desempenho superior dos bancos de dados multidimensionais.

Figura 10 - Exemplo de HOLAP



Fonte 7- Carvalho 2010

DOLAP (Desktop OLAP)

O DOLAP é uma arquitetura desktop do OLAP, ou seja, é uma ferramenta para usuários que possuem uma cópia de base multidimensional ou de um subconjunto dela ou ainda, que queiram acessar, localmente, o repositório de dados central. Neste último caso, o usuário, através do disparo de uma instrução SQL, acessa os cubos já existentes no banco de dados multidimensional, residente no servidor OLAP, obtendo de volta um micro-cubo para ser analisado em sua estação de trabalho.

O ganho desta arquitetura é a redução da sobrecarga no servidor de banco de dados uma vez que todo o processamento OLAP acontece na máquina cliente. A desvantagem está no fato do tamanho do micro-cubo não ser muito grande, caso contrário, a análise pode ser demorada e a máquina cliente não suportar.

WOLAP (Web OLAP)

Resumidamente, WOLAP é a utilização de uma ferramenta OLAP a partir de um browser. Esta solução traz os benefícios de duas tecnologias que estão em constante evolução: a web e as ferramentas OLAP. O grande diferencial entre WOLAP e as demais arquiteturas está, justamente, na utilização web já que esta facilita, principalmente, a distribuição da ferramenta, o acesso remoto aos dados a serem analisados e a utilização da aplicação independente de plataforma.

Estas ferramentas seguem os mesmos princípios de um sistema web qualquer. As solicitações são enviadas via HTTP ao servidor que então, as processa, enviando, o resultado da solicitação, ao cliente.

2.13.1 Tipos de Ferramentas

Os processos de um DW consistem na extração dos dados dos sistemas operacionais, na organização e integração desses dados de forma consistente e, no acesso aos dados integrados de forma simples, fácil, eficiente e flexível para consultas. Contudo, a extração, organização e integração dos dados devem ser realizadas com o propósito de garantir a consistência e integridade das informações, construindo, desta forma, uma base de dados de alta qualidade e confiabilidade, que retrate efetivamente a realidade de negócios da empresa.

Essas ferramentas de DW são responsáveis pela filtragem, limpeza, sumarização e concentração dos dados espalhados pelas fontes externas e nos sistemas operativos. Em

sua maioria esses sistemas/aplicações possuem interfaces gráficas e interativas que facilitam a realização do mapeamento de dados e a automação do processo de extração, limpeza e carga. Fortemente baseadas no conhecimento da linguagem SQL, esses sistemas também possuem funções predeterminadas para sua utilização e permitem um acesso intuitivo aos dados, possibilitando a análise daqueles mais significativos.

Segundo Machado (2011), o sucesso de um DW pode depender da disponibilidade da ferramenta certa para as necessidades dos usuários e, para garantir essa flexibilidade, normalmente, as ferramentas para pesquisa e relatórios são empregadas, sendo estas simples e com uma interface gráfica para a geração de relatórios e análise de dados históricos. No decorrer do processo de análise de dados as ferramentas OLAP (“*On-Line Analytical Processing*”) também permitem avaliar o que aconteceu e o porquê dos resultados obtidos.

Atualmente, existe disponível no mercado uma variedade dessas ferramentas com diferentes abordagens. Uma dessas ferramentas são os Sistemas de Informações Executivas, que apresentam uma visualização mais simplificada e consolidada dos dados, não requerendo do usuário experiência e tempo para a execução de uma análise. As ferramentas OLAP e o *Data Mining* representam uma categoria de ferramentas de análise denominada *open-em*, que permite ao usuário analisar tendências e padrões não conhecidos entre os dados. Esse tipo de ferramenta utiliza-se das mais modernas técnicas de computação, como redes neurais, algoritmos genéricos e lógica nebulosa.

3 ESTUDO DE CASO

Neste capítulo será abordado a aplicação prática do DW, método utilizado para extração dos dados bem como a aplicação da ferramenta de BI SPAGO.

3.1 Origem das Cooperativas

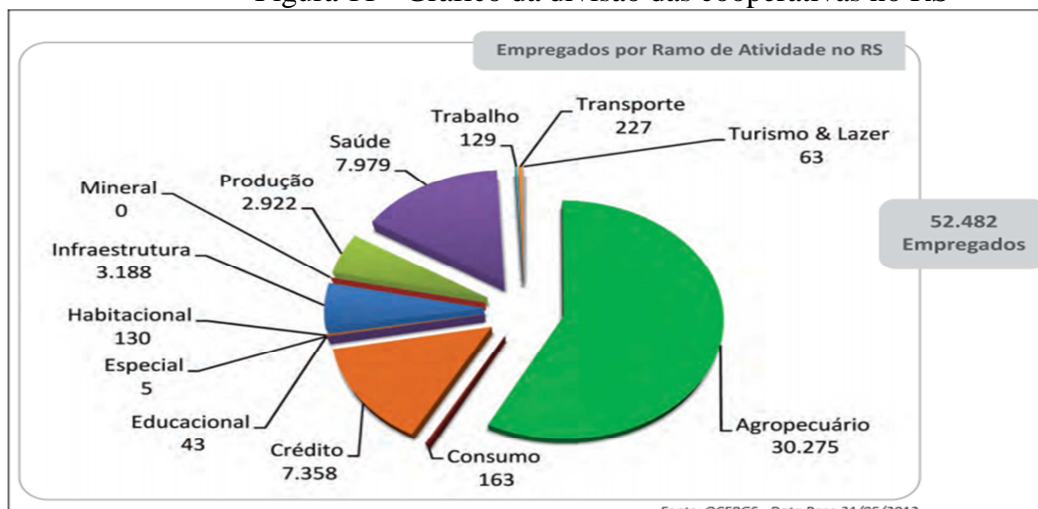
No Brasil o cooperativismo surge no século XIX com as cooperativas de consumo no estado do Rio de Janeiro, apresentando os mesmos propósitos das europeias. Entretanto, foi só na década de 1970 que a Política Nacional de Cooperativismo foi instituída, definida através da Lei nº 5.764/71, como elencado por Maraschin (2004), estabelecendo-se o regime jurídico das sociedades cooperativas.

No início do século XX esta forma de organização chega ao Rio Grande do Sul, surgindo no estado, nos anos seguintes, o cooperativismo agrícola que buscava excluir os intermediários e facilitar a venda dos produtos dos colonos, permitindo uma maior lucratividade por parte dos agricultores.

Nas décadas de 1960 e 1970, as cooperativas possuíam papel fundamental na agricultura brasileira sob a ótica do Estado. Estas tinham a incumbência de difundir, entre os produtores, o pacote da “Revolução Verde”, tendo como principais elementos, a difusão de relações de trabalho capitalistas no meio rural e a incorporação de insumos industriais à tecnologia de produção. Este processo de modernização da agricultura promovida pelo Estado através das cooperativas reduziria os custos operacionais, de circulação e produção, de acordo com Benetti (1982), facilitaria a compra de grãos, oportunizaria a difusão e incorporação de tecnologia avançada e garantiria maior produtividade física e econômica da lavoura.

Segundo relatórios do OCERGS (2012), cerca de 1.026 cooperativas estão registradas no Rio Grande do Sul, colocando o nosso Estado em 1ª posição no ranking brasileiro. O Rio Grande do Sul representa cerca de 14,5% das cooperativas do Brasil, estando 499 destas cooperativas com o cadastro inativo.

Figura 11 - Gráfico da divisão das cooperativas no RS



Fonte 8 - OCERGS 2012

A cooperativa aqui analisada foi fundada em 1959 por colonos europeus, com a finalidade de congregiar os produtores familiares da região de Bagé/RS e municípios vizinhos e ainda, facilitar comercialização dos seus produtos. Em sua fundação, o leite era o produto com maior destaque.

A partir da década de 1990, as mudanças políticas e econômicas ocorridas na agropecuária brasileira tornaram a produção de grãos mais atrativa. Com este novo cenário a cooperativa CAMAL mudou seu foco para a produção de grãos e manteve a estrutura ligada ao beneficiamento de leite em segundo plano. Segundo relatórios contábeis, o faturamento da cooperativa está distribuído conforme figura 12, em anexo:

Figura 12 - Representação Gráfica Receitas 2012



Fonte 4 - Relatório Contábil CAMAL

Mesmo o leite ainda tendo grande representatividade dentro da cooperativa, pode-se verificar que cerca de 44% do faturamento apresentado no gráfico acima se refere a venda de cereais: 31% de cereais em geral e somente 13% de arroz.

3.2 Desenvolvimento do Protótipo

A cooperativa encontra-se hoje, em estado migratório de um sistema legado COBOL 4.5, para um sistema ERP 100% integrado e armazenado em um Sistema Gerenciador de Banco de Dados (Microsoft SQL Server, 2005). O fato de o sistema ERP atual não apresentar uma ferramenta de apoio à decisão, propiciou a possibilidade e a oportunidade de aplicar os conhecimentos adquiridos em aula sobre o DW.

Levando em consideração os objetivos propostos por este trabalho, procurou-se construir um DW que apoiasse a tomada de decisão do setor de recebimento de cereais da cooperativa, tomando por base informações do recebimento de grãos. No item subsequente, serão apresentadas as principais tecnologias e ferramentas utilizadas no trabalho, também far-se-á a descrição da especificação e implementação do sistema proposto.

3.3 Requisitos Trabalhados

Para a obtenção de informações foram feitas entrevistas e consultas aos setores responsáveis pelo recebimento de cereais, podendo ser, assim, feito o levantamento de todas as variáveis necessárias ao protótipo e a identificação de procedimentos executados na entrada de grãos. Outros fatores como os campos e funções para o controle destes processos foram levantados para a implementação do sistema proposto neste trabalho.

Um dos fatores que impulsionaram a realização do presente trabalho foi a possibilidade de acrescentar ao setor responsável pelo recebimento de cereais, informações específicas referentes às safras de anos anteriores e também, ao recebimento da safra do ano corrente. Com isso, buscaram-se maneiras de medir a qualidade do grão recebido e os sócios ativos da cooperativa. Verificou-se, também, a necessidade de disponibilização de relatórios e gráficos que mostrassem mais claramente a situação da safra.

A cooperativa atualmente conta com aproximadamente 1.190 associados entregando leite e/ou cereais. Cada associado produz em média 7.314 litros de leite e cerca de 30,5 toneladas de cereais, distribuídos entre soja, trigo, sorgo, arroz, milho e cevada. Hoje, as tabelas que armazenam o recebimento de cereais têm aproximadamente

800.000 registros, com dados desde 2008, possibilitando uma análise mais completa do recebimento de cereais.

Além disso, no DW, alguns dados foram sumarizados e, tabelas referentes a datas, semanas, meses, bimestres, semestres e anos foram criadas. Desta maneira, o DW poderá alcançar os objetivos, ao fornecer acesso imediato a um conjunto de dados confiáveis reunidos da cooperativa. O sistema disponibiliza consultas e gráficos que possibilitam a visualização dos resultados.

Utilizando-se da subtécnica de particionamento, considerou-se a divisão do banco de dados criado, já que futuramente outros *Data Marts* serão agregados ao DW. Os próximos *Data Mart* criados serão para o supermercado, recebimento de leite e posto de combustível. O banco de dados *My SQL* foi utilizado neste trabalho pela facilidade de conexão com *Spago BI*.

A técnica OLAP foi utilizada para a construção dos gráficos do sistema. A visualização pelos gráficos apresenta informações condensadas das tabelas existentes no sistema, permitindo uma consulta flexível, uma vez que se pode analisar o recebimento dos cereais sobre vários aspectos, como por período, por centro e por *ranking* quantidade, cada um subdividido conforme os parâmetros escolhidos pelo usuário.

Outra característica de OLAP observada nas análises dos gráficos é a origem dos dados para a confecção dos mesmos. As informações são obtidas no *Data Warehouse*, que por sua vez é alimentado a partir do ambiente operacional da cooperativa.

3.4 Tecnologias e Ferramentas Utilizadas

Para a especificação do sistema foram utilizadas algumas ferramentas. A cooperativa aqui analisada utiliza o banco de dados SQL Server 2005 e como ferramenta de desenvolvimento, o *Delphi 7* ambos licenciados. Já para o trabalho proposto foi utilizado banco de dados *My SQL 5.4* pela facilidade de conexão com o BI, ferramenta de migração *Delphi 7* já que a mesma é licenciada pela cooperativa e a ferramenta de *BI Spago BI*.

3.4.1 SQL SERVER

O *MS SQL Server* é um sistema gerenciador de banco de dados criado pela Microsoft e Sybase. Inicialmente o *SQL Server* foi criado para a plataforma OS/2. Com o fim da parceria com a Sybase, em 1998, a Microsoft assume o projeto do SGDB. Hoje, o *SQL Server* encontra-se na versão 2012, possuindo vários tipos de distribuição, inclusive, uma distribuição gratuita com grandes limitações.

3.4.2 Delphi

Embarcadero Delphi, anteriormente conhecido como *CodeGear Delphi*, *Inprise Delphi* e *Borland Delphi*, também conhecido como *Delphi*, é um compilador, uma IDE e uma linguagem de programação, produzido anteriormente pela *Borland Software Corporation* e atualmente produzido pela *Embarcadero*. O Delphi foi originalmente direcionado para a plataforma *Windows*.

3.3.1 SPAGO BI

O *Spago BI* é um suíte desenvolvido para apoio à decisão mantida e desenvolvido pela empresa *Engineering Informática* (www.eng.it). De fácil customização, ajuda na diminuição de custos, adequando-se a complexidade do negócio, ao tamanho e à quantidade de funcionários da empresa que o utiliza. Tal suíte suporta diversos bancos de dados, porém, não suporta diversos idiomas, inclusive português. Outro diferencial importante do *Spago BI* é o seu modelo de segurança, que acompanha a forma natural de pensar em soluções BI e a separação dos dados com a seleção e a apresentação.

Este modelo de segurança, chamado de modelo comportamental, representa o pilar fundamental para parametrização do acesso e visibilidade dos dados. A camada de acesso aos dados é preparada com a definição de fontes de dados que inclui banco de dados SQL até as *webservices*, planilhas e arquivos de textos. Desta maneira, o modelo de segurança acaba por definir os drivers analíticos, aqui vistos como um conjunto de definições de associação de roles, validação de valores, de forma que este conjunto possa responder a necessidade do negócio.

Segundo Lacy (2010), o *Spago BI* é o único sistema BI completo em software livre que pode ser implementado e usado corporativamente sem pagamento de licenças. Outras

empresas mantêm versões diferentes do suíte “community” com funcionalidades reduzidas.

3.5 Modelagem

A modelagem do banco de dados do DW deu-se através do esquema em estrela. Na qual a tabela fato, armazena as principais informações sobre o recebimento dos cereais. As tabelas dimensões armazenam dados referentes ao produto, associado e tempo.

3.5.1 Dicionário de Dados

O dicionário de dados é composto por três entidades, relacionadas através da modelagem em estrela. São elas recebimento, pessoas, datas e produtos

A tabela recebimento é responsável pelo armazenamento de dados referente a pesagem e armazenamento dos cereais. Nela contem atributos que identificam a qualidade do grão, procedência e tipo de grão.

Tabela 5 - Entidade Recebimento

Tabela: Recebimento		Tipo: Fato
Nome	Tipo	Descrição
SocioNSocio	Alfanumérico (6)	Tipo de pessoa sócio e não sócio
DataRecebimento	Data	Data/Hora da entrada do caminhão no engenho
DataEntrada	Data Hora	Data/Hora descarga do caminhão
DataSaida	Data Hora	Data/Hora saída da descarga
CodProduto	Numérico	Código Produto
Bairro	Alfanumérico (30)	Nome do Bairro
Cidade	Alfanumérico (30)	Nome Cidade
PesoBruto	Numérico	Peso da Entrada do Caminhão
pesoSaida	Numérico	Peso saída Caminhão
Limpo_Seco	Numérico	Grão limpo e seco
KGCapital	Numérico	Quantidade de capital
PesoAssociado	Numérico	Peso liquido do associado
KGUmidade	Numérico	Peso Umidade
KGImpureza	Numérico	Peso da Impureza
KGDespesa	Numérico	Peso da despesa de secagem
Placa	Alfanumérico (7)	Placa do caminhão
Idade	Numérico	Idade do associado
CodigoPessoa	Numérico	Codigo da pessoa

Na entidade recebimento pode-se também verificar a existência de atributos curiosos como os de *SocioNSocio* e *Idade*. É comum que não sócios, que já entregavam suas safras na cooperativa, venham a associar-se. A classificação *SocioNSocio* é assim, utilizada como um meio importante de medição da fidelização na entrega de cereais. Já a classificação *Idade*,

mostra o perfil do associado no decorrer dos anos, apresentando ainda, a ocorrência de apostas em outras culturas.

A entidade dada armazena as datas referentes ao recebimento de cereais. Dividindo as datas em dia, semana do ano, mês , bimestre ,semestre e ano. Facilitando assim a sumarização dos dados referente ao recebimento dos cereais.

Tabela 6 - Entidade Data

Tabela: Data		Tipo :Dimensão
Nome	Tipo	Descrição
Data	Data	Armazena a Data com dia , mês e ano
Dia	Numérico	Dia do mês
Semana	Numérico	Semana do Ano
Mes	Numérico	Mês do ano
bimestre	Numérico	Bimestre do Ano
semestre	Numérico	Semestre do Ano
Ano	Numérico	Ano

Já a entidade Pessoa refere-se as informações dos associados, como o nome ,data de nascimento nome, tipo de pessoa e sexo.

Tabela 7 - Entidade Pessoas

Tabela: Pessoas		Tipo :Dimensão
Nome	Tipo	Descrição
CodigoPessoa	Numérico	Código do associado
DataNascimento	Data	Data de nascimento
Nome	Alfanumérico(10)	Nome do Associado
Fisjur	Alfanumérico(1)	Tipo físico ou jurídico
Sexo	Alfanumérico(1)	Sexo
DataAssociacao	Data	Data associação

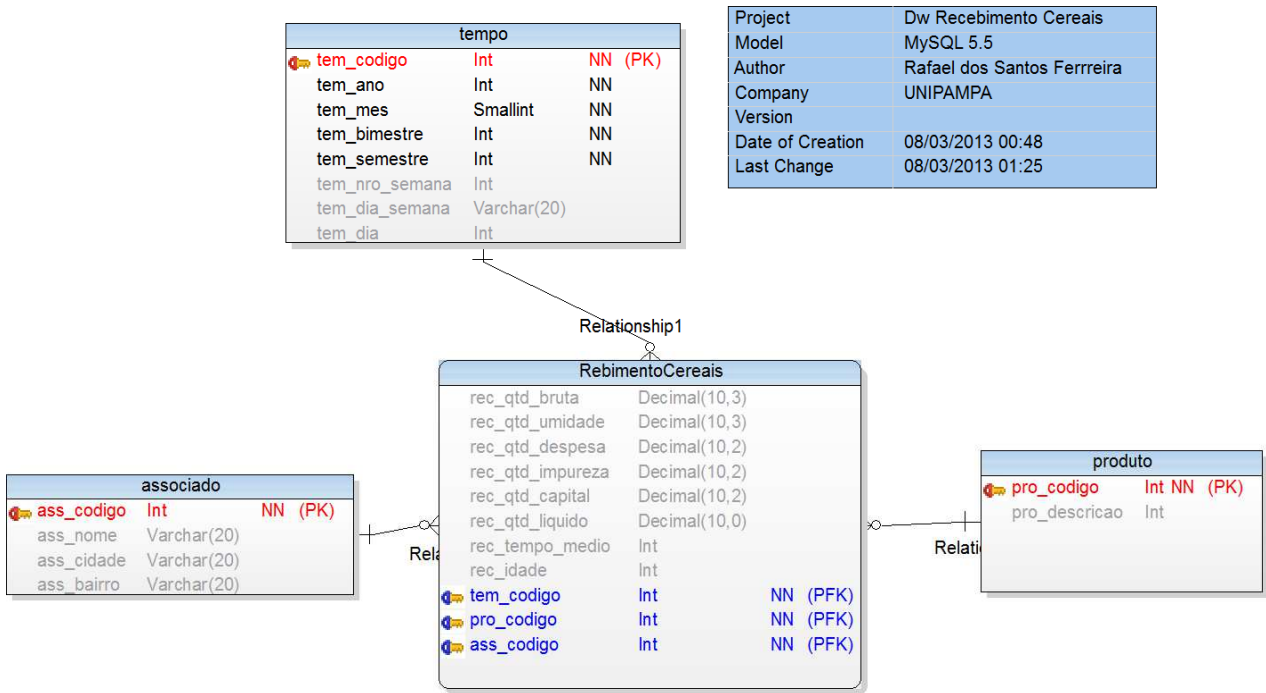
A entidade produto a armazena informações referente ao produto recebido do associado.

Tabela 8 - Entidade Produtos

Tabela: produtos		Tipo :Dimensão
Nome	Tipo	Descrição
CodProduto	Numérico	Código do Produto
Produto	Alfanumérico(20)	Nome do Produto

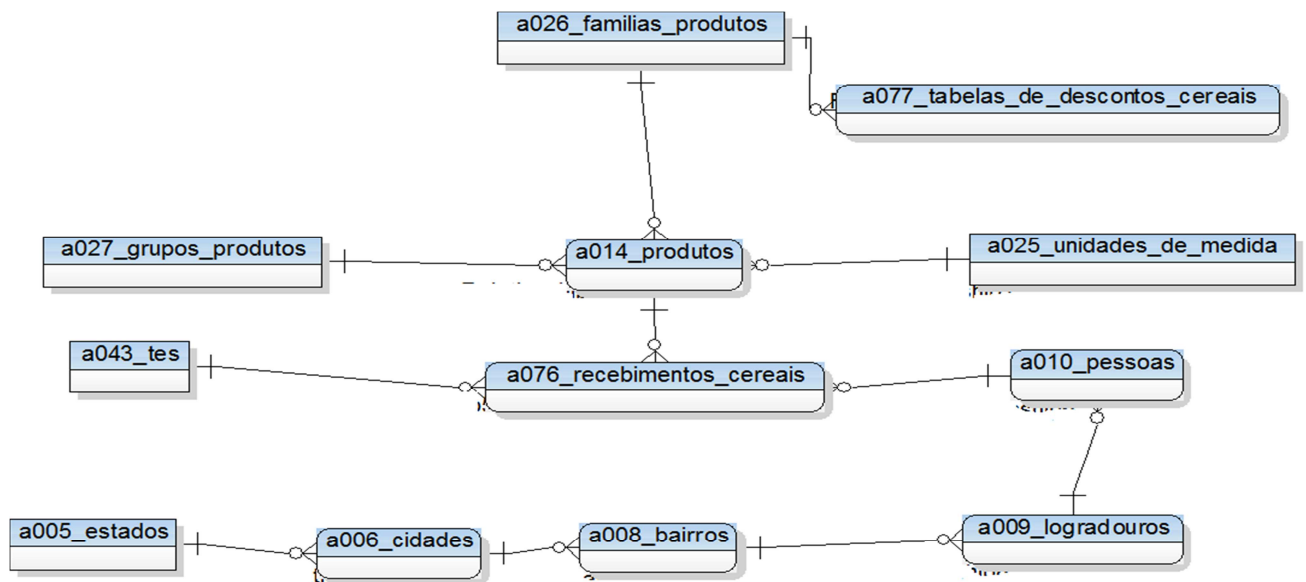
A figura 13 mostra o diagrama E/R do modelo do DW adotado. Ao centro a tabela fato/recebimento e ao lado estão representadas as dimensões, produto, pessoas e data. Hoje, o banco de dados DW armazena aproximadamente 900.000 (Novecentos mil) registros, contendo atributos importantes para a tomada de decisão.

Figura 13 - Diagrama do *Data Warehouse*



Abaixo está o diagrama ER do banco de dados relacional, com entidades que são envolvidas no recebimento dos cereais.

Figura 14 - Diagrama Modelo E/R banco de dados Relacional



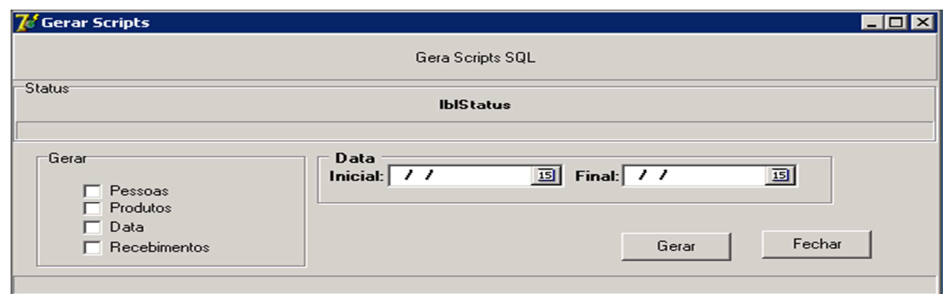
O banco relacional que está em produção encontra-se com aproximadamente 250 entidades distribuídas nos mais diversos setores (cereais, supermercado, setor pessoal,

entre outros), com aproximadamente 100 *Gigabytes* de informações. Como se pode notar no diagrama figura 14, foi ocultado os atributos das tabelas. Pois muitas delas possuem um grande numero de atributos, ficando ,assim, impossível de mostrá-los no modelo E/R. Pode-se notar que algumas entidades foram absorvidas pelo modelo do DW e transformadas em atributos, outras foram eliminadas do modelo alvo, ocorrendo a desnormalização do modelo adotado.

3.6 Aplicativo Delphi

Para facilitar a importação dos dados para o banco DW, foi implementado um aplicativo simples em *Delphi* o qual lê os dados da base relacional e gera um arquivo texto. Uma informação importante que podemos destacar é que produtos, datas e pessoas já geradas em cargas passadas não são geradas novamente, assim, torna este processo mais rápido e ágil. A figura 15 mostra a tela do aplicativo Delphi para exportação dos arquivos.

Figura 15 - Aplicativo de Exportação



O usuário poderá gerar somente os arquivos do seu interesse, como: pessoas, produtos, datas e/ou recebimento. Existe, ainda, a possibilidade de gerar todos arquivos, marcando todos os *check box* e em seguida o botão de Gerar. A figura 16 mostra de forma resumida o arquivo gerado pela aplicação.

Figura 16 Arquivo gerado pela aplicação

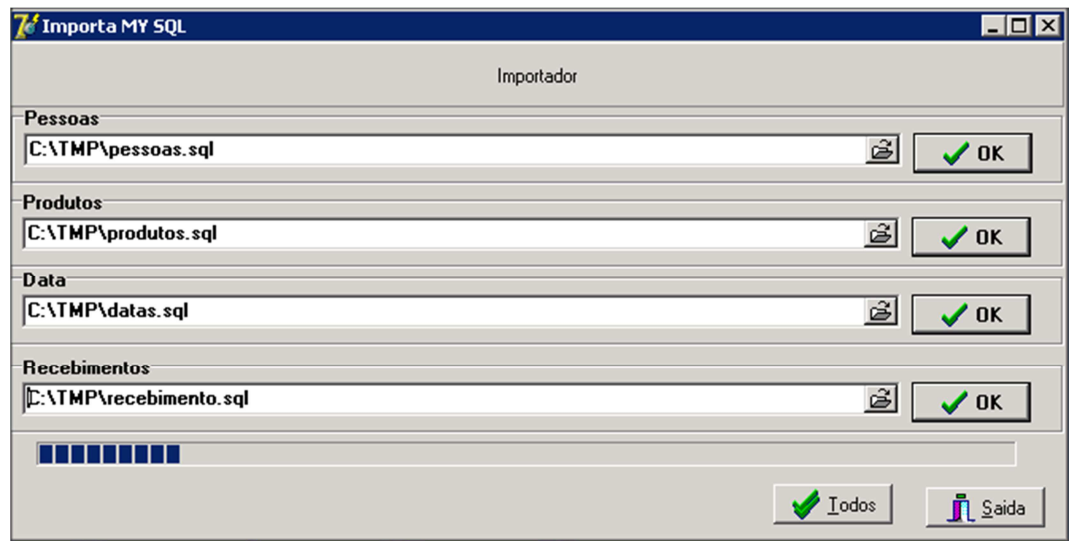
```

1  -- TABELA DE PRODUTOS
2  Insert into produtos(CodProduto,
3  Produto) values ('48','SOJA EM GRAO (RR)');
4
5  -- TABELA DE DATAS
6  INSERT INTO DATAS (data,dia,semana,mes,bimestre,semestre,ano)
7  values ('2008-11-10','10','Segunda','11','6','2','2008');
8
9  -- TABELA PESSOAS
10 Insert into pessoas (CodigoPessoa,
11 DataNascimento,
12 Associado,
13 fisjur) values ('1717','2008-01-01','ABIMAEI SILVEIRA TEIXEIRA ','Fisj
14
15 --TABELA RECEBIMENTO
16 Insert into recebimento (SocioNSocio,
17 DataRecebimento,DataEntrada,DataSaida,CodProduto,
18 Bairro,Cidade,PesoBruto,Limpo_Seco,
19 KGCapital,PesoAssociado,KGUmidade,
20 KGImpureza,KGDespesa,Placa,
21 pesoSaida,perSecagem,perUmidade,
22 perCapital,Idade,
23 fisjur,CodigoPessoa)
24 values ('Socio','2009-03-27 00:00:00','2009-03-26 18:48:40'
25 ,'2009-03-27 09:17:28','2','COLONIA NOVA'
26 ,'Aceguá',16700,12637.8,212,12630,21.4,3,1303
27 ,'ifa-7023',7460,7.8,12.3,1.5,52,'Fisica','');

```


Já a carga dos registos no DW é feita através de um outro aplicativo Delphi, o qual lê os arquivos textos gerados e os insere na seguinte sequência: pessoas, produtos, datas e após recebimento. Caso ocorra algum erro na inserção dos comandos, o aplicativo gera um arquivo de *log*, no qual poderá ser consultado no final do processo. A figura 17 mostra o aplicativo que importa o arquivo texto gerado pelo exportador

Figura 17 - Aplicativo Delphi de importação



3.7 Spago BI

A implementação dentro do Spago BI deu-se através de três passos:

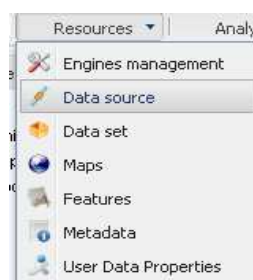
- Conexão com o Banco de Dados;
- Criação de Consultas através do Spago Web;
- Criação de Consultas através do Spago *Studio*.

Porém para demonstrar as potencialidades da ferramenta vamos exemplificar a conexão com o banco de dados e a criação de uma consulta simples através do Spago Web.

3.7.1 Conexão com o Banco de Dados

A conexão com o banco de dados MySQL deu-se através de 3 passos:

Figura 18 - Menu *Data Source*



No menu *resources* seleciona-se a opção *Data Source*, a partir dela podemos verificar as conexões exemplos do SpagoBi.

Figura 19- Visualização Conexões Existentes

LABEL	DESCRIPTION
Bam	Bam
FoodMart	Foodmart
FOODMART_MYSQL	Foodmart MySQL
SpagoBI	SpagoBI
dwCereaisBi	dwCereaisBi

Pode-se criar outras conexões com o *Spago* clicando-se no ícone da lupa no canto direito superior.

Figura 20 - Configuração da Conexão

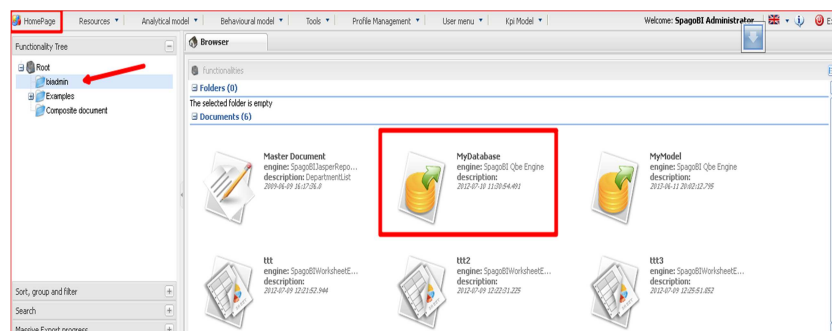
LABEL	dwCereaisBi
DESCRIPTION	dwCereaisBi
DIALECT	MySql
MULTISCHEMA	<input type="radio"/> Yes <input checked="" type="radio"/> No
SCHEMA	
ATTRIBUT	
TYPE	<input type="radio"/> Jndi <input checked="" type="radio"/> Jdbc
JNDI NAME	
URL	jdbc:mysql://localhost:3306/dwcereais
USER	root
PASSWORD	****
DRIVER	com.mysql.jdbc.Driver

Na opção *label* insere-se o nome da conexão na opção *description*, ou seja, a descrição da conexão, enquanto que em *dialect* insere-se o tipo de conexão e as informações para o drive JDBC para conectar ao banco de dados. Terminada esta etapa, a conexão já está pronta para ser efetuada e testada no *Spago*.

3.7.2 Criando Consultas

Após a conexão ser efetuada, podemos realizar algumas consultas e gráficos, de forma dinâmica, de dentro da própria ferramenta *SpagoWeb*.

Figura 21- Criar Consulta



Primeiramente, deve-se selecionar a opção *HomePage*, apresentada no canto superior esquerdo da página, logo, seleciona-se a pasta *bdadmin* e o nome do *Data Source* criado anteriormente, no nosso caso *MyDatabase*.

Na tela que se segue, o usuário poderá criar uma consulta dinâmica com os *fields* necessários para a geração do gráfico pretendido. No exemplo abaixo, precisa-se verificar a evolução do recebimento do arroz por ano.



Figura 22 - Seleção de Fields para a consulta

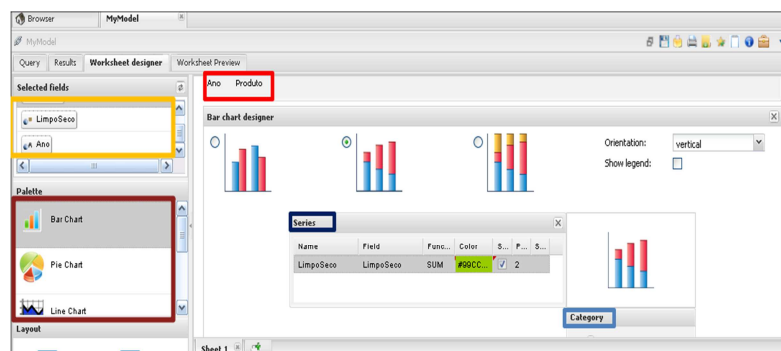
Para tanto se seleciona somente os campos necessários para o resultado do nosso gráfico. Como é possível notar foram agrupados os campos produtos e ano, já o campo limpo seco foi somado, obtendo-se, assim, o seguinte resultado:

Figura 23 - Resultado dos *Fields*

	Produto	LimpoSeco	Ano
1	ARROZ EM CAS...	14,601,736.20	2,009
2	ARROZ EM CAS...	9,287,492.00	2,010
3	ARROZ EM CAS...	14,162,022.50	2,011
4	ARROZ EM CAS...	6,249,366.10	2,012
5	ARROZ EM CAS...	3,795,967.90	2,013
6	AVEIA AMARELA	2,512.97	2,008
7	AVEIA AMARELA	89,864.84	2,009
8	AVEIA AMARELA	94,431.00	2,010
9	AVEIA AMARELA	74,207.51	2,011
10	AVEIA AMARELA	12,575.00	2,012

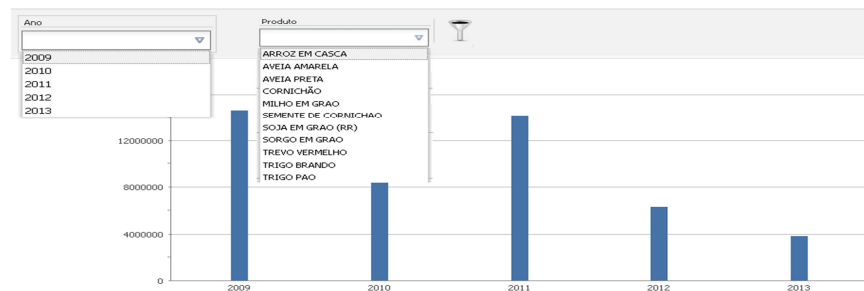
A partir deste resultado o usuário poderá ter uma visão melhor da consulta executada. Então, o próximo passo será escolher o gráfico pretendido.

Figura 24- Seleção do Gráfico



No *select Fields* encontramos os campos que fazem parte da consulta, em *palette*, e são os tipos de gráficos que podem ser gerados pela ferramenta. Na parte superior, ao centro, em um quadrado, são os filtros, que poderão ser executados durante a execução do gráfico. Neste caso, podemos fazer o filtro por ano e produto.

Figura 25 - Resultado da Consulta



De forma simples e fácil e com um conhecimento básico em SQL, se pode gerar um gráfico na ferramenta. O *Spago BI* é uma ferramenta muito completa e fácil de ser programada, podendo gerar gráficos, consulta e relatórios de forma muito fácil.

3.8 Problemas Enfrentados

Foram encontrados alguns problemas na execução do projeto: dados inconsistentes, escolha da ferramenta BI e modelo em estrela.

O banco de dados onde se extraíram os dados continham alguns erros. Um dos que podemos destacar é a relação dos endereços nos quais existiam erros de grafia e associações com as localidades erradas. Podemos citar o exemplo da cidade da Hulha Negra, no qual a grafia foi escrita de diversas maneiras tais como: ulha negra, ulhanegra, hulha, entre outros tipos de grafia. A correção foi executada no banco de dados relacional através de uma busca fonética e *update*. Algumas localidades e/ou bairros estavam associados a cidades erradas, como a localidade de Colônia Nova que estava associada à cidade de Bagé. A correção foi executada pela inscrição estadual do produtor e corrigida por *update* no banco de dados de produção.

A ferramenta escolhida para o projeto do BI iria ser o *Pentaho BI*, porém a distribuição gratuita da ferramenta possui muitas limitações, as quais poderiam prejudicar o resultado do projeto. Com isto adotou-se o *Spago BI* por ser uma ferramenta livre, gratuita e sem limitações de uso.

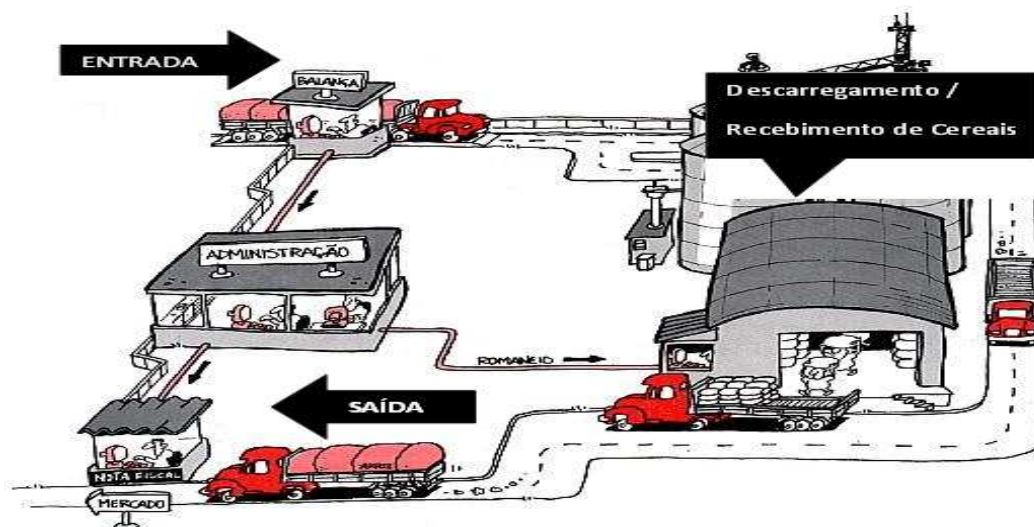
A modelagem estrela é o padrão adotado para os projetos de DW, porém ao se aplicar este modelo ao *Spago BI*, o mesmo travava ou não respondia as consultas. Como

solução criou-se uma visão no banco de dados com as junções necessárias para obter-se os resultados pretendidos, resolvendo-se, assim, os problemas.

4 RESULTADOS

Alguns resultados foram obtidos com a adoção do *Data Warehouse* junto com o *Spago BI*. Podendo ser melhor entendidos através da figura abaixo na qual mostra o fluxo do funcionamento de um engenho.

Figura 26 Fluxo de um Engenho

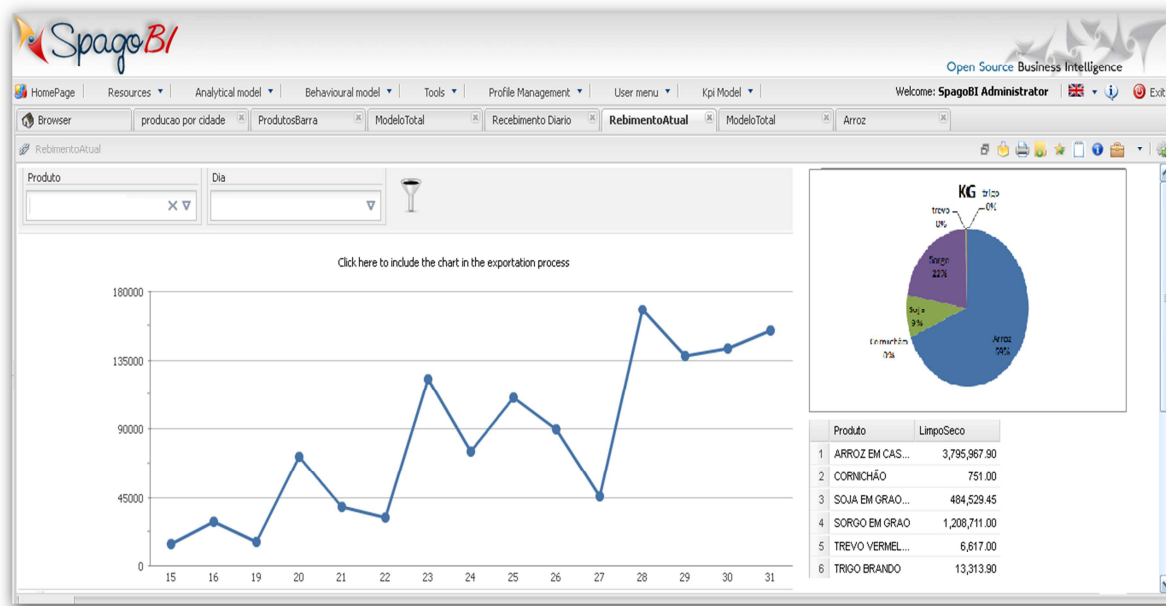


Primeiramente o caminhão antes de descarregar faz a pesagem na entrada do engenho, após este processo o caminhão é descarregado e novamente passa por uma nova pesagem. A diferença entre a pesagem de entrada e saída é a quantidade descarregada no engenho. Em alguns engenhos como é o caso da cooperativa estudada o descarregamento é feito de forma manual, ou seja, os funcionários sobem com pás e enxadas e descarregam o cereal de forma manual. O processo em si leva em torno de 40 minutos dependendo da quantidade descarregada e o número de pessoas envolvidas no processo.

4.1 Gráfico com a Produção de Cereis por ano

O gráfico a seguir foi construído utilizando-se os dados do recebimento do mês de março de 2013. Os atributos *data*, *limposeco* e *produto recebido* foram utilizados na consulta. Para a elaboração do gráfico em linha foi utilizado a sumarização do atributo *limposeco* junto com o *group by* da *data*. Já o gráfico em formato pizza foi sumarizado o atributo *limposeco* junto com o *group by* do *produto* e finalmente o grid é o resultado da consulta do gráfico em formato pizza.

Figura 27 - Gráfico de Recebimento Mensal

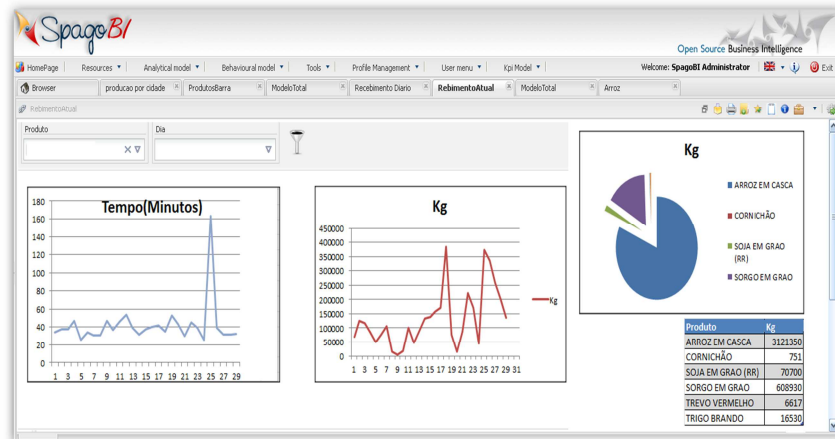


O gráfico principal, em linha, mostra o total de cargas recebidas no mês de março de 2013, o que pode se notar é um crescimento no total recebido durante o mês. O gráfico em pizza mostra o total de cereal recebido no mês de março a área em azul mostra o produto arroz e roxo o sorgo e o verde a soja. Com base nestes dados o gestor poderá ter uma previsão do que esta sendo armazenado e o que poderá ser vendido futuramente

4.2 Gráfico: Tempo Médio de Descarga e Quantidade de KG

Como atributos utilizados na criação destes gráficos foram utilizados o limposeco, dataentrada, datasaida e produto. A consulta do tempo foi utilizada a subtração da datasaida com dataentrada, sumarizando o limposeco e agrupando por dataentrada. Para o gráfico em quilos foi utilizada a sumarização do limposeco agrupando-se pela dataentrada. Já o gráfico de pizza e o grid utilizam a sumarização do limposeco agrupando-se pelo produto.

Figura 28- Gráfico Tempo Médio de Descarga e Quantidade Kg

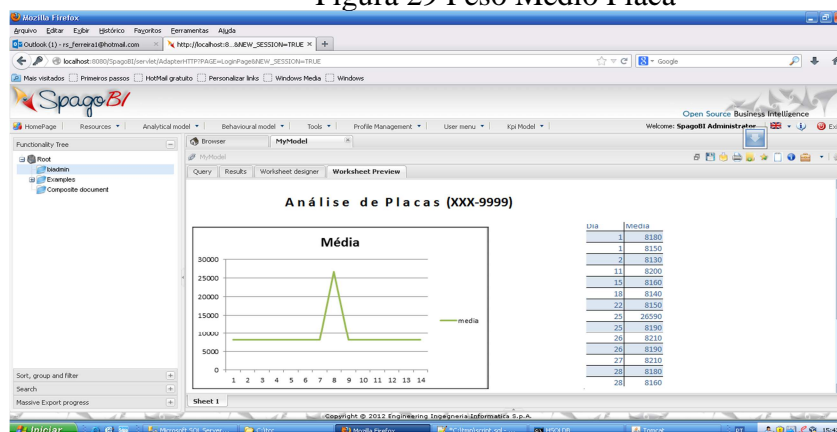


O gráfico apresentado mostra o tempo médio das descargas diárias e a quantidade de quilos recebidos. Já o gráfico em pizza e o *grid* mostram os cereais recebidos no período consultado. Com o cruzamento dos dois gráficos em linha o gestor poderá ter a informação do tempo médio de descargas por quilo recebido. Podendo, assim, deslocar mais ou menos recursos para as áreas de descarga. Um fato que destaca-se no gráfico é que o tempo médio não variava em relação ao total de cargas recebidos, porém em uma determinada data, o tempo aumentou muito e o total de cargas recebidos não acompanhou esta variação, e a mesma quantidade de recursos, funcionários, reservados para a tarefa sempre se manteve a mesma.

4.3 Peso médio dos caminhões

Como atributos deste gráfico foi utilizado a placa do caminhão e as pesoSaida. Foi utilizada a pesagem media de saída (pesosaida) agrupando-se pela placa. A análise deste gráfico mostra uma maior variação da pesagem de saída do caminhão em relação aos outros dias. O que na verdade surpreendeu o gestor, já que tecnicamente os caminhões teriam que sair vazios do engenho, ou seja, descarregar totalmente a sua carga.

Figura 29 Peso Médio Placa



O gráfico em linha e o grid utilizam a mesma consulta, no grid pode-se ver peso médio de saída do caminhão já o em linha pode ser notado mais facilmente alguma variação que poderá existir na pesagem do caminhão

5 CONCLUSÕES

O uso de sistemas de *Data Warehouse* na área agrícola para subsidiar o funcionamento de sistemas de apoio à tomada decisão é de crucial importância, principalmente devido ao grande volume de informação gerada na cadeia produtiva e dos altos valores envolvidos em cada processo decisório. As funcionalidades de navegação e cruzamento de dados fornecidos pela *Data Warehouse* permite ao tomador de decisão explorar facilmente, e de maneira flexível, grandes quantidades de dados sobre diferentes perspectivas. Apesar da *Data Warehouse*, implementada neste estudo de caso, estar funcional e já apresentar claros benefícios no processo decisório, diversos aspectos ainda estão sendo aprimorados a fim de criar um modelo de implementação sistemático.

Futuras extensões deste trabalho incluem criação do DW para o supermercado da cooperativa e, ainda, a utilização de técnicas de mineração de dados para detectar padrões dentro dos processos decisórios, a fim de buscar a minimização de erros e maximização de lucro.

6 REFERÊNCIAS

BACHA, C. J. C. **Economia e Política Agrícola no Brasil**. Agricultura na virada do milênio: velhos e novos desafios. Viçosa: [s.n.]. 2000. p. 93-116.

BARBALHO, P. Descubra o Data Warehouse Produtividade e rapidez. **SQL Magazine**, Rio de Janeiro, v. 01, n. 03, p. 34-38, Março 2002. ISSN ISSN 16779185.

CAMAL, Copertiava Agrícola Mista Acegua LTDA. **Relatorio 2012**. Cooperativa Agrícola Mista Acegua LTDA. Bagé, p. 19. 2012.

CARVALHO, B. F. Arquiteturas de Ferramenta OLAP. **SQL Magazine**, Rio de Janeiro, v. 1, n. 9, p. 12-16, ago. 2010. ISSN 1677918-5.

ELIAS CORREA, F. et al. **Data warehouse for soybeans and corn market on Brazil**. [S.l.]: [s.n.]. 2009. p. 6.

INNON, W. H. **Como Construir o Data Warehouse**. Rio de Janeiro: Campus, 1997.

ITALIANO, C.; ESTEVES, L. A. A Modelagem de Data Warehouse e Data Marts. **Sql Magazine**, Rio de Janeiro, v. 2, n. 16, p. 59-66, ago. 2010. ISSN 1677918-5.

ITALIANO, ISABEL CRISTINA; ESTEVES, ANTONIO LUIZ. Modelagem de Data Warehouse e Data Marts - Parte 01. **Sql Magazine**, Rio de Janeiro, v. 1, n. 13, p. 42-45, Junho 2006. ISSN 1677918-5.

JACOBSON, R.; MISNER, S.; COSULTING, H. **Microsoft Sql Server 2005 Analysis Services**. 1. ed. Porto Alegre: Bookman, v. 1, 2007. 352 p. Tradução Claudio Belleza Dias.

JÚNIOR, M. C. **Projetando Sistemas de Apoio à Decisão Baseados em Data Warehouse**. 1. ed. Rio de Janeiro: Axcel Books, v. 1, 2004.

LACY, M. K. Sistemas Gerenciadores de Conteúdo. **Revista Espirito Livre**, 2010. Disponível em: <<http://revista.espiritolivre.org>>. Acesso em: 30 fev. 2013.

MACHADO, F. N. R. **Tecnologia e Projeto de Data warehouse**. São Paulo: Erica, v. 5, 2011.

MAPA. Política Agrícola. **Ministério da Agricultura**, 2012. Disponível em: <<http://www.agricultura.gov.br/>>. Acesso em: 20 maio 2013.

MARASCHIN, A. F. As relações entre produtores de leite e cooperativas. **Um Estudo de caso de bacia leiteira de Santa Rosa**, Porto Alegre, 2004. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/6407/000485056.pdf?sequence=1>>. Acesso em: 20 mar. 2013.

MY SQL. Disponível em: <<http://www.mysql.com/why-mysql/white-papers/>>. Acesso em: 20 maio 2013

.

OCERGS. Expressão do Cooperativismo Gaúcho. **OCERGS**, jan. 2013. Disponível em: <<http://intranet.sescooprs.coop.br/arquivos/arqs/20120719102955.pdf>>. Acesso em: 29 maio 2013.

ONU. Organização das Nações Unidas para Alimentação e Agricultura. **ONU**, 2012. Disponível em: <<http://www.onu.org.br/>>. Acesso em: 20 maio 2013.

RAI, Anil; NILAKANTA, S.; SCHEIBE, Indiano Agrícola Design Data Warehouse. **Information Resources Management Association**, 2005. Disponível em: <<http://www.irma-international.org/proceeding-paper/indian-agricultural-data-warehouse-design/33130/>>. Acesso em: 20 maio 2012.

SARKIS, Vikram. **DATA WAREHOUSE: O PROCESSO DE MIGRAÇÃO DE DADOS**. Florianópolis: [s.n.]. 2001. p. 146.

SORATHIA, . Semantic Web - Software Architecture - Distributed Computing - Social Network Analysis. **University of Southern California**, 2005. Disponível em: <http://scholar.google.com/citations?view_op=view_citation&hl=en&user=CefuLl4AAAAJ&citation_for_view=CefuLl4AAAAJ:u-x6o8ySG0sC>. Acesso em: 23 maio 2013.

SORATHIA, V.; LALIWALA, Z.; CHAUDHARY, S. **Towards Agricultural Marketing Reforms: Web Services Orchestration Approach**. Proceedings of the 2005 IEEE International Conference on Services Computing (SCC'05). [S.l.]: [s.n.].