FERNANDO ELIAS CORREA

**REPRESENTAÇÃO DE COMERCIALIZAÇÃO AGROPECUÁRIA
ATRAVÉS DE MODELO DE DATA WAREHOUSE**

São Paulo

2010

FERNANDO ELIAS CORREA

**REPRESENTAÇÃO DE COMERCIALIZAÇÃO AGROPÉCUARIA
ATRAVÉS DE MODELO DE DATA WAREHOUSE**

Dissertação apresentada à Escola
Politécnica da Universidade de São Paulo
para obtenção do título de Mestre em
Engenharia.

São Paulo

2010

FERNANDO ELIAS CORREA

**REPRESENTAÇÃO DE COMERCIALIZAÇÃO AGROPÉCUARIA
ATRAVÉS DE MODELO DE DATA WAREHOUSE**

Dissertação apresentada à Escola
Politécnica da Universidade de São Paulo
para obtenção do título de Mestre em
Engenharia.

Área de Concentração:
Sistemas Digitais

Orientador:
Prof. Dr. Pedro Luiz Pizzigatti Corrêa

São Paulo
2010

Este exemplar foi revisado e alterado em relação à versão original, sob a responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 28 de Janeiro de 2010.

Assinatura do autor

Assinatura do orientador

FICHA CATALOGRÁFICA

Correa, Fernando Elias

Representação de comercialização agropecuária através de modelo de Data Warehouse / F.E. Correa. - ed. rev. - São Paulo, 2010.

70 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1. Banco de dados 2. Agribusiness 3. Modelagem de dados 4. Sistemas de informação I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II. t.

Dedico

Aos meus pais Pedro Correa e Inair Pascoina Tabai Correa por fornecerem toda a estrutura para a minha vida, me ensinando os valores da honestidade, humildade, justiça, amor e respeito ao próximo, tudo que sou hoje se deve a eles.

Aos meus irmãos Sandra, Celso, Silvana, Eliana e Vande por me apoiarem e incentivarem nos caminhos da minha vida.

Aos meus sobrinhos e sobrinhas, Camila, Gabriel e Mariana que os amos muito e desejo poder auxiliá-los no caminhar dessa vida.

A Tâmara, minha amada namorada que sempre me incentivou e apoiou, e pela dedicação e ajuda em todos os momentos em que passamos juntos.

AGRADECIMENTOS

Primeiramente agradeço a Deus por me conceder a vida e por mais uma conquista.

Ao meu orientador, Prof. Dr. Pedro Luiz Pizzigatti Corrêa, pela oportunidade, valiosas orientações, dedicação e confiança durante todo o trabalho.

Aos Professores e amigos de trabalho do Cepea, pela permissão, incentivo e compreensão para que esse trabalho pudesse ser concluído.

Ao Professor Dr. Antônio Mauro Saraiva pelos ensinamentos, confiança e convivência.

Ao Professor Dr. Jorge Rady pelas valiosas colaborações e ensinamentos dedicados.

Aos Professores do LAA, Prof. Dr. André Hirakawa e Prof. Dr. Carlos Eduardo Cugnasca pela confiança.

Aos meus amigos Pablo Salvanha, Leandro Halle e Edson Souza pela amizade, apoio e pelas viagens que certamente irei sempre me lembrar.

Aos meus amigos, Antonio Bianchi e Gustavo Bertola que me auxiliaram em minha caminhada desde meu ingresso até a conclusão de meu trabalho.

Aos meus amigos, Lucien, Ana Julia e Fabiana por todo o apoio e que apesar da distância nunca deixaram de estar próximo e me incentivar.

E por fim agradeço a todos que me ajudaram de forma direta ou indireta para a realização deste trabalho.

RESUMO

A comercialização de produtos agropecuários tem grande importância para o Brasil, sendo responsável por altos índices de produção e financeiro. E como para todo grande segmento, o acesso à informação tem se tornado essencial para manter a competitividade no ambiente globalizado. Apesar de o agronegócio possuir um grande volume de informações, as mesmas muitas vezes estão inconsistentes ou inacessíveis. Nesse contexto, ainda se observam a necessidade de ampliação de pesquisas que busquem minimizar essa falta de dados que podem auxiliar produtores, agentes de mercado e principalmente pesquisadores para que desenvolvam análises para o mercado agropecuário para servir de apoio para as tomadas de decisões sobre a comercialização diária desses produtos. A técnica usada por diversos segmentos que visam a melhorar e a permitir acesso a dados para tomada de decisão é conhecida como *Data Warehouse*, pois provê um ambiente propício para análises e acesso a informações consolidadas e históricas. O objetivo do estudo foi demonstrar o processo de análise efetuado por pesquisadores para obtenção de dados, e aplicar dois estudos de casos usando técnicas de *Data Warehouse*, sendo para dados de comercialização de pecuária e para dados de grãos (soja e milho), gerando assim o modelo dimensional, a matriz de barramento e a aplicação da ferramenta de processamento analítico online. O resultado do estudo de pecuária permitiu gerar os modelos iniciais para a cadeia, e podem ser expandidos com a inclusão de novos produtos, bem como auxiliar o pesquisador na geração de análises com a utilização das ferramentas para acesso ao *Data Mart*. Assim como foi possível gerar análises de comercialização de grãos a partir de dados armazenados no *Data Mart*. Portanto, pode-se afirmar que a aplicação e expansão da técnica de *Data Warehouse* é viável para outras cadeias do agronegócio possibilitando a ampliação e a melhoria da oferta de informações.

Palavras-chave: *Data Warehouse*, modelagem dimensional, sistema de apoio a decisão, agronegócio.

ABSTRACT

The agricultural and livestock markets in Brazil play important roles both for the production and the financial sectors. However, like other business, information is essential to maintain the competitiveness in the global market. Although the agribusiness sector processes high amount of information on a daily basis, it is almost always inconsistent or inaccessible. Moreover, there is little research aimed at minimizing the lack of agribusiness information available for the producers, market agents and researchers in order to develop data analyses on agribusiness, to help in the decision process. The technique commonly used in other segments, such as industries, is called *Data Warehouse*, which provides an environment to provide access to data and consistent information. The goals of this research were to present what researchers what they can do to perform agribusiness analysis, how to apply *Data Warehouse* modeling and the application of the tool to this process. To reach these goals, we evaluated two case studies. The first study used data from the livestock market, as it allowed to generate initial data for the chain and because it can be applied to other products, as well as to help the researcher to analyze using the OLAP tool. At the same line, the study about grain market was developed, and models and OLAP tools were developed to help with the particular points for this chain. Concluding, the research shows that it is viable to use *Data Warehouse* techniques to create an agribusiness data environment, consistent and organized, which can be expanded to new agricultural products.

Keywords: Data Warehouse, dimensional model, decision-making system, agribusiness.

LISTA DE ILUSTRAÇÕES

Figura 1: Descrição dos níveis para tomada de decisão. (IBM, 2006).....	23
Figura 2: Estrutura <i>Bottom-up</i> de <i>Data Warehouse</i> . (DWHINFO, 2009).....	26
Figura 3: Ambiente <i>Data Warehouse</i> orientado por assunto	27
Figura 4: Conversão de dados para o <i>Data Warehouse</i> . (HANSON, 2009).....	28
Figura 5: Componentes existentes no <i>Data Warehouse</i>	29
Figura 6: Matriz de barramento de exemplo. (DWHINFO, 2009)	31
Figura 7: Modelo estrela. Adaptada de (INMON, 1999).....	32
Figura 8: Modelo constelação (ELMASRI, 2000)	33
Figura 9: Exemplo de granularidade, (HANSON, 2009)	34
Figura 10: Fluxograma de análise de dados do agronegócio do Cepea.....	40
Figura 11: Macro processo definição de tema e fonte de dados para análise.....	41
Figura 12: Macro processo de organização dos dados.	42
Figura 13: Macro processo de análise primária dos dados.	42
Figura 14: Macro processo de aplicação estatística e análise final.	43
Figura 15: Matriz de barramento: <i>Data Mart</i> Pecuária	46
Figura 16: Modelagem dimensional aplicada à comercialização de pecuária.	47
Figura 17: Tela do <i>Data Warehouse</i> para pecuária, software <i>Mondrian</i>	49
Figura 18: Matriz de barramento do <i>Data Mart</i> Grãos.....	54
Figura 19: Modelagem dimensional para comercialização de grãos.....	55
Figura 20: Tela de aplicação do <i>Data Warehouse</i> , software <i>Mondrian</i>	56
Figura 21: Relação de preços entre soja e milho.....	57
Figura 22: Modelo de arquivo xml para o <i>Mondrian</i>	68
Figura 23: Tela do <i>mondrian</i> com consulta inicial.	69
Figura 24: Cubo com as dimensões e medidas do <i>Data Mart</i>	70
Figura 25: Configuração e gráfico do <i>Mondrian</i>	70

LISTA DE TABELAS

<i>Tabela 1: Dados provenientes do Data warehouse de comercialização de Pecuária.</i>	
.....	50

LISTA DE ABREVIATURAS E SIGLAS

CEPEA – Centro de Estudos Avançados em Economia Aplicada.

CNA – Confederação de Agricultura e Pecuária do Brasil.

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária.

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

IBGE – Instituto Brasileiro de Geografia e Estatística.

JIAC – *Joint International Agricultural Conference*.

OLAP – Processamento Analítico Online (*Online Analytical Processing*).

PIB – Produto Interno Bruto.

SUMÁRIO

1	INTRODUÇÃO.....	13
1.1	INFORMAÇÕES SOBRE O AGRONEGÓCIO	13
1.2	MOTIVAÇÃO E JUSTIFICATIVAS	15
1.3	OBJETIVOS	16
1.3.1	Objetivo principal	16
1.3.2	Objetivos específicos	16
1.4	MATERIAIS E MÉTODOS.....	17
1.4.1	Ferramentas OLAP para Data Warehouse	18
1.5	ORGANIZAÇÃO DA DISSERTAÇÃO	18
2	REVISÃO BIBLIOGRÁFICA	20
2.1	OFERTA DE INFORMAÇÃO NO AGRONEGÓCIO	20
2.2	COMERCIALIZAÇÃO DE PRODUTOS AGROPECUÁRIOS	21
2.3	PROCESSO DE AQUISIÇÃO DO CONHECIMENTO	22
2.4	DATA WAREHOUSING	24
2.5	ARQUITETURA DE DATA WAREHOUSE	29
2.6	MATRIZ DE BARRAMENTO	30
2.7	MODELO LÓGICO	31
2.7.1	Modelo Estrela.....	31
2.7.2	Modelo Constelação	33
2.7.3	Granularidade de dados	33
2.8	MODELAGEM DIMENSIONAL	35
2.9	OLAP.....	35
2.10	TRABALHOS RELACIONADOS	36
2.11	CONSIDERAÇÕES PARCIAIS.....	38
3	PROCESSO DE ANÁLISE AGROPÉCUARIA	39
3.1	CONSIDERAÇÕES PARCIAIS.....	44
4	DATA MART DE COMERCIALIZAÇÃO DE PECUÁRIA	45
4.1	MATRIZ DE BARRAMENTO: DATA MART PECUÁRIA.....	46
4.2	MODELO DIMENSIONAL	47
4.3	APLICAÇÃO DE FERRAMENTA OLAP	48
4.4	GERAÇÃO DE ANÁLISES USANDO ESTRUTURA DE DATA MART	49

4.5	CONSIDERAÇÕES PARCIAIS	50
5	<i>DATA MART</i> DE COMERCIALIZAÇÃO DE GRÃOS	53
5.1	MATRIZ DE BARRAMENTO	53
5.2	MODELO DIMENSIONAL	54
5.3	APLICACÃO DE FERRAMENTA OLAP	56
5.5	CONSIDERAÇÕES PARCIAIS	58
6	CONCLUSÃO	60
6.1	CONSIDERAÇÕES FINAIS	60
6.2	TRABALHOS FUTUROS	61
	REFERÊNCIAS	63
7	APLICAÇÃO DO SOFTWARE MONDRIAN	66
7.1	REQUISITO DE INSTALAÇÃO	66
7.2	PROCESSO DE INSTALAÇÃO DO MONDRIAN	66
7.3	ARQUIVO XML DE EXEMPLO	68
7.4	EXECUÇÃO DO MONDRIAN	69

1 INTRODUÇÃO

1.1 INFORMAÇÕES SOBRE O AGRONEGÓCIO

O agronegócio brasileiro enfrenta desafios na organização de informação estratégica em seus diferentes setores (agricultura, pecuária e floresta). A competitividade e a busca por redução de custos forçam a revisão e o aprimoramento dos processos de produção e comercialização. No entanto, somente esses pontos não são suficientes para garantir o sucesso das operações comerciais no mercado (BATALHA, 2007).

No Brasil, em especial, o agronegócio tem forte influência sobre a economia, sendo que no ano de 2008 a agricultura e pecuária tiveram evolução no Produto Interno Bruto (PIB) nacional na ordem de 5,81% e 8,56% respectivamente (CNA, 2008).

Além dos fatores financeiros, o agronegócio tem um dos maiores índices de produção, o que permite a criação de muitos empregos e favorece o desenvolvimento interno.

Considerando a dimensão do agronegócio para economia brasileira, torna-se fundamental a gestão integrada das informações da cadeia de comercialização de seus produtos, como: os preços comercializados, volumes exportados, políticas internacionais, altas e baixas de produtos e insumos, negociações financeiras, entre outros (CEPEA, 2007).

A gestão da informação só é possível se a mesma estiver acessível para as partes interessadas do agronegócio. O acesso à informação é importante para a comunidade agrícola, sejam agricultores, corretores, empresas agrícolas, órgãos públicos, entre outros, pois melhora a competitividade do Brasil no mercado internacional, possibilitando que especialistas econômicos observem tendências e proponham novos rumos nacionais e internacionais.

Contudo, a organização de informação de comércio é complexa para todas as áreas de negócio, devido à descentralização e carência de dados consistentes, séries históricas, comparativos entre outros setores e áreas agregadas (CHAUDHARY, 2004).

Essa organização tem sido o grande desafio, uma vez que estas não se encontram agrupadas e prontas para uso, pelo contrário, a informação geralmente está dispersa, muitas vezes incompletas, separadas geograficamente e em sistemas computacionais legados.

Outro ponto a ser considerado são as diferentes terminologias utilizadas nas cadeias, o que gera redundância e inconsistências, como exemplo o uso de nomes diferentes para um mesmo produto, ou uso de unidades distintas como quilo ou litro (HAVERKORT, 2006).

Agrupar e organizar a informação em um modelo de dados integrado e homogêneo é fundamental para que sejam obtidas análises que apóiem um entendimento da dinâmica do agronegócio, seja para os produtores adaptarem-se às mudanças no mercado, seja para órgãos governamentais instituírem políticas públicas para regular o mercado.

Órgãos públicos como o Centro de Estudos Avançados em Economia Aplicada – Cepea (CEPEA, 2007) e o Instituto Brasileiro de Geografia e Estatística – IBGE (IBGE, 2008), buscam diminuir essa escassez de informação, coletando dados relevantes ao mercado agropecuário.

Essas bases de dados têm crescimento exponencial, constituindo-se em gigantescos bancos de dados, os quais passam por um processo de análises estatísticas. Os resultados gerados são armazenados para estudos e disponibilizados para a sociedade.

Assim, *Data Warehousing* é um processo que permite agrupar e organizar dados de diferentes fontes (banco de dados, planilhas e textos) em um ambiente de dados integrado, orientado por assunto, com dados históricos e consolidados. O mesmo tem por objetivo dar suporte ao processo de tomada de decisão (INMON, 1999).

Uma abordagem reduzida do escopo de *Data Warehouse* é conhecida como *Data Mart*, o qual dá um nível de detalhamento maior de dados sobre um determinado assunto. Assim conjuntos de *Data Mart* agrupados são um *Data Warehouse* (INMON, 1999).

Foram relatadas pesquisas sobre a gestão de informação gerencial do agronegócio. Toni (2000) destaca a importância da utilização de ferramentas para apoio à tomada de decisões para o agronegócio, através de *Data Marts* em

cooperativas agropecuárias, enquanto Marques (2002) descreveu a aplicação do *Data Warehouse* para análises de dados de melhoramento genético da raça nelore.

Porém, observa-se a necessidade de aplicar ferramentas de apoio à decisão voltadas, especificamente, ao agronegócio com uma visão da cadeia produtiva, tornando-se assim necessário o desenvolvimento de trabalhos que mostrem os benefícios da aplicação dessas ferramentas e divulguem experiências em modelagem de dados voltadas para sistemas de apoio à decisão no setor.

Como objetivo de atender a essa demanda e propor modelos de dados integrados que apoiem a tomada de decisões gerenciais relacionadas às cadeias de comercialização de produtos do agronegócio, este trabalho apresenta um estudo sobre modelos de dados dimensionais, aplicados na definição de *Data Warehouse* e *Data Marts*.

1.2 MOTIVAÇÃO E JUSTIFICATIVAS

A aplicação de *Data Warehouse* sobre os dados de comercialização agropecuária que tem maior representatividade se aplicada sobre cadeias agrícolas de produtos para fins alimentares de origem animal e vegetal.

Entre os produtos para fins alimentares de origem animal, a pecuária tem sido destacada. Os principais produtos são a comercialização de boi gordo e bezerro, mas inclui-se também a vaca gorda, a comercialização de cortes de carnes, comércio de frango e suínos.

As pesquisas sobre informação do mercado brasileiro de pecuária ganha importância quando se analisa o consumo interno desses produtos, que é de 80% de toda a produção nacional, e o Brasil possui o maior rebanho comercial do mundo (ABIEC, 2009).

Outra consideração são os contratos negociados em operações de compra, vendas de animais e mercado futuro, na Bolsa de Mercadorias e Futuros – BM&F que atinge totais de negociações superiores a 14 mil em contratos abertos no mês de outubro desse ano (BM&F, 2009).

O levantamento de dados e acompanhamento de informação de produtos como soja também é destacado em diversas pesquisas econômicas do agronegócio. O Brasil tem grande competitividade no mercado internacional para tais produtos, como para a produção e exportação de soja do Brasil, que é o segundo maior do mundo, sendo responsável por 26% de toda produção mundial nos últimos três anos (USDA, 2009).

A análise dessas cadeias agrícolas juntas permitem uma visão expandida dos pontos relevantes para os estudos e proposta de modelos para criação e implementação de *Data Warehouse* que irá auxiliar no processo de análises para o agronegócio.

1.3 OBJETIVOS

1.3.1 Objetivo principal

Esta dissertação tem por objetivo principal estudar modelos de dados dimensionais aplicados à criação de *Data Warehouse* de comercialização de produtos agropecuários brasileiro.

1.3.2 Objetivos específicos

- Estudo do processo para geração de análise de informações para o agronegócio.
- Aplicação de dois estudos de caso de cadeias de comercialização agrícola, sendo pecuária e grãos, que contemplam geração de modelos dimensionais, criação de *Data Marts* e implementação de ferramentas *OLAP* (*Online Processing Analytical* – Processamento analítico online).

1.4 MATERIAIS E MÉTODOS

Inicialmente foi necessário estudar os processos para tomada de decisão junto à cadeia de comercialização de produtos. Para isso foram estudados o processo que o pesquisador efetua para obter dados que são utilizados sobre uma determinada cadeia agropecuária e fazer análises do agronegócio.

Esse levantamento permitiu obter os seguintes resultados:

- Mapeamento do processo de Análise.
- Identificação das fontes primárias e tratamentos dos dados obtidos.
- Volume de dados obtidos, variáveis e tempo para agrupamentos dos dados.
- Identificação das ferramentas utilizadas para o processo de análise.

Os pontos relevantes obtidos desse levantamento forneceram suporte para apresentar os modelos para geração de um *Data Warehouse* e em quais estágios do processo de análise o *Data Warehouse* será empregado.

O levantamento teve como base um processo para geração de análise típica para o agronegócio. A análise está baseada em fontes de dados primários do Cepea e para geração das informações diárias que são disponibilizadas para a comunidade do agronegócio.

Com base na identificação dos processos, foram propostos modelos e ferramentas de *Data Warehouse* com objetivo de apoiar o processo de análise.

As cadeias de comercialização do agronegócio analisadas foram a de pecuária e a de grãos (soja e milho). Os motivos para o uso dessas cadeias foram sua importância no agronegócio brasileiro como descrito anteriormente e pelo tempo de dados históricos disponível pelo Cepea para o processo de *Data Warehouse*.

Junto com essa identificação foi necessário obter informações sobre a comercialização dos produtos do agronegócio e suas relações para que possa se gerar modelos para criação do *Data Warehouse*.

Como forma de analisar a eficiência dos modelos e uso de ferramenta foram feitos dois estudos de caso com validação por parte dos pesquisadores.

1.4.1 Ferramentas OLAP para Data Warehouse

As ferramentas utilizadas neste trabalho são aquelas definidas como *OLAP*, empregadas tipicamente nos sistemas de *Data Warehouse* (NILAKANTA, 2008). As ferramentas *OLAP* selecionadas para serem usadas neste trabalho seguiram o critério de serem *opensource*. Existem outras implementações com ferramentas *OLAP* proprietárias, como a descrita por Meira (2003), em que a implementação foi realizada utilizando o software de banco de dados proprietário.

Entretanto, neste estudo foram utilizadas tais ferramentas por serem *opensource* e possuírem um bom grau de estabilidade e portabilidade.

As ferramentas utilizadas foram:

- *APACHE-TOMCAT* (2007): Software para servidores de páginas *web*, que trabalha integrado com o *software Tomcat*, sendo que ambos permitem a publicação de sites com estrutura baseada em sistemas Java.
- *MONDRIAN* (2007): Software para aplicações *OLAP*, criando todo o ambiente para *Data Warehouse*, contendo consultas *ad hoc* e pré definidas, geração dinâmica de gráficos, navegação e criação dos cubos, *drill up* e *drill down* sobre os cubos, exibição de dados fonte, entre outros.
- *MYSQL* (2007): software gerenciador de banco de dados responsável por armazenar as relações do *Data Warehouse*.
- *Framework* de desenvolvimento em linguagem *JAVA* (2007): A aplicação acima descrita utiliza a plataforma *Java*, por ser independente do sistema operacional, executável em qualquer sistema.

1.5 ORGANIZAÇÃO DA DISSERTAÇÃO

O presente trabalho está organizado da seguinte forma:

- No capítulo 1, são descritos os problemas para aquisição de informação no setor de agronegócio, a importância da informação para o setor, bem

como as técnicas e trabalhos existentes para buscar soluções em *Data Warehouse*, os objetivos e métodos utilizados nesta pesquisa.

- No capítulo 2, são apresentados os conceitos para dar suporte à pesquisa do trabalho, sendo descritos os conceitos de *Data Warehouse*, a modelagem dimensional e matriz de barramento e suas técnicas para expandi-las na criação do modelo de referência, as ferramentas disponíveis para aplicação ao *Data Warehouse*, os conceitos existentes para o agronegócio e, por fim, como esses conceitos se adequam ao estudo do trabalho.
- No capítulo 3 são apresentados o processo de análise de dados de comercialização do agronegócio, descritas as técnicas usadas, as ferramentas e as aplicações de *Data Warehouse* nesse modelo.
- No capítulo 4, apresenta-se o estudo de caso realizado aplicado à cadeia produtiva de pecuária, com modelos e resultado da aplicação.
- No capítulo 5, é apresentado o estudo de caso da cadeia produtiva de grãos, onde são demonstrados os modelos e seus resultados obtidos.
- No capítulo 6 é apresentada a conclusão do trabalho.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo inicialmente discute conceitos de agronegócio e fontes de informação. Em seguida, são discutidos os processos envolvidos na obtenção de dados até se tornarem informação para apoio a tomada de decisão; Apresentam-se conceitos sobre *Data Warehouse* e técnicas de modelos que norteiam a aplicação de *Data Warehouse*, que é matriz de barramento e a modelagem dimensional e, então, é finalizado com a revisão dos trabalhos relacionados.

2.1 OFERTA DE INFORMAÇÃO NO AGRONEGÓCIO

Na década de 50 nos Estados Unidos, nos trabalhos de Davis e Goldberg foi definido agronegócio como sendo:

“a soma total das operações de produção e distribuição de suprimentos agrícolas, das operações de produção nas unidades agrícolas, do armazenamento, processamento e distribuição dos produtos agrícolas e itens produzidos a partir deles” (DAVIS; GOLDBERG, 1957).

A definição mostra a necessidade de trabalhar com informação de diversas partes da agropecuária, ou seja, utilizando informação em forma de visão da cadeia produtiva de produtos. Para isso é feita a busca de diferentes fontes de dados em diferentes partes da cadeia produtiva.

Muitas fontes de informação sobre agronegócio estão disponíveis em jornais, revistas, rádios, telejornais e Internet. No entanto, o nível de conteúdo obtido pelas mídias é variável, bem como a precisão e qualidade da informação fornecida em tais meios. Não só as fontes de informação são heterogêneas, mas também o perfil dos agentes que a demandam, ou seja, diferentes níveis de necessidades por informação são gerados por fazendeiros, agricultores, corretores, agentes de mercados, pesquisadores entre outros (BOTEON, 2004).

A pesquisa de Silva (2005) mostra a dinâmica e os diferentes meios de comunicação utilizados para obter informação que apóiem e/ou subsidiem as

pesquisas e atividades do agronegócio, em que se destaca o uso eficiente dos diferentes meios que podem melhorar o desenvolvimento das pesquisas no setor.

As fontes de informação do agronegócio no Brasil são providas por setores públicos como: IBGE e secretaria da agricultura, por setores privados (exportadores, corretores, cooperativas). Além de órgãos de pesquisa como Cepea e a Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA (EMBRAPA, 2008).

Boteon (2004) apresenta um fluxo de informação agro econômicas em suas pesquisas gerando três grupos que ofertam informações. Sendo fontes primárias aquelas que produzem dados e informação direta. Fazem parte desse grupo empresas, cooperativas, órgãos públicos e privados. Outro grupo são os intermediários, nos quais a informação é sumarizada, processada e distribuída em agências de notícias e sites de conteúdo específico. Por fim, o grupo dos usuários finais são compostos por tomadores de decisões e/ou pesquisadores no agronegócio.

Compor análises sobre o agronegócio é sempre um desafio, dado o número de fontes de dados heterogêneas e de diferentes granularidades de informação. Mas, se os conteúdos estiverem agrupados por assunto e tratados para gerar informação consolidada, é possível agilizar a obtenção de variáveis para apoiarem as pesquisas e análises para o agronegócio.

2.2 COMERCIALIZAÇÃO DE PRODUTOS AGROPECUÁRIOS

Conforme descrito por Barros (2006), a comercialização pode ser vista como uma interação entre agentes econômicos através de instituições apropriadas e dentre essas a que se destaca é a de mercado.

Nesse contexto, mercado pode ser definido como o local em que se há operações de oferta e demanda de produtos agropecuários e tais operações ocorrem entre compradores e vendedores.

A transferência de um produto do vendedor para o comprador é conhecida como operação de venda e compra. Quando o produto é entregue pelo vendedor ao comprador simultaneamente à transferência de posse do mesmo, o processo é conhecido como mercado à vista. Já a transferência de posse do produto se dá

através de contrato de compromisso de entrega no futuro, o mesmo é conhecido como mercado a termo.

Barros (2006) ainda define que “local” se trata de um termo amplo que pode caracterizar desde um local existente, como o caso de mercado atacadista de São Paulo, ou a um produto consolidado como exemplo mercado de milho.

A comercialização dos produtos agrícolas também engloba a integração das regiões de produção e venda dos mesmos, ou seja, o mercado de um produto ocorrerá em diferentes regiões se os preços pagos pelas mesmas forem superiores aos custos de fretes praticados para o transporte dos produtos (Barros, 2006).

Dessa forma, as variáveis sobre comercialização de produtos agropecuários devem conter dados de regiões, produtos, insumos, fretes, tipos de negociações, entre outras para auxiliar no entendimento do mercado agropecuário.

2.3 PROCESSO DE AQUISIÇÃO DO CONHECIMENTO

Com o crescimento das organizações e empresas, gerou-se a necessidade de criar e organizar banco de dados, que foram aprimorando seus processos e formaram sistemas complexos de banco de dados.

Esses processos resultam da coleta e armazenamento de dados baseado no ambiente operacional ao qual a instituição é organizada, como se fizesse uma reprodução desse ambiente.

No entanto, para o processo de tomadas de decisões é necessário que os bancos de dados tenham seu histórico de informação disponível e ajustadas, fazendo novamente a analogia para o mundo real; o gerente de uma empresa precisa de informação de balanço no mínimo anual para que possa tomar decisões para futuros investimentos.

Figura 1: Descrição dos níveis para tomada de decisão. Extraído de (IBM, 2006).

A Figura 1 apresenta uma pirâmide que descreve em níveis os ambientes dos quais a informação são tratadas em uma organização. O ambiente operacional, representado na Figura 1 como nível 0 e parte do 1, possui informação diária da organização, todos os dados coletados, transações e comercialização, ou seja, um volume de informação ainda não tratada. Uma analogia com o ambiente diário de uma empresa seriam os papéis nas mesas, os recados, telefonemas (IBM, 2006).

Os próximos níveis descrevem o ambiente “tático”, ou seja, a fase na qual a informação coletada é tratada e armazenada seguindo modelos descritos para bancos de dados, como organização e normalização (criação de índices e eliminação de dados inconsistentes).

Dado o volume de informação adquirido nos níveis inferiores e o prévio tratamento feito à informação, iniciam-se as etapas desenvolvidas no ambiente estratégico. Tal ambiente tem início em parte do nível 2, e finaliza no nível 5 da pirâmide. No terceiro nível a informação é resumida e tratada para eliminação de informação redundante, também são efetuadas extrações das camadas inferiores para alimentar com informação sumarizada nível 4.

No nível 4, a informação é segmentada, dando início ao conceito de *Data Warehouse*, que é descrito segundo Inmon (1999): “[...] *Data warehouse* integra um conjunto de dados consolidados, organizados por assuntos e que evoluem no tempo, utilizado para dar suporte ao processo de tomada de decisão[...]”.

Por fim, no topo da pirâmide está a informação retornada ao público de gerência que fará uso em forma de relatórios e planilhas, providas de diversas bases de dados e diferentes anos, para a tomada de decisões.

Com a finalidade de auxiliar e aperfeiçoar as consultas e a geração de relatórios e planilhas criados para análise de dados históricos no ambiente estratégico e auxílio à tomada de decisão, foram utilizadas técnicas conhecidas como sistema *OLAP* (SCHULZE, 2007).

A sofisticação das ferramentas analíticas gerou o que conhecemos como sistemas *OLAP*, que fornecem funcionalidades para atender às consultas para análise de dados históricos concentrados por assuntos, o que os diferencia dos dados operacionais mantidos em bancos de dados relacionais (ELMASRI, 2000).

Um meio destinado a dar apoio à aplicação das técnicas de *OLAP* é o *Data Warehouse*, cujas fases consistem em projetar, construir, acessar e gerenciar o *Data Warehouse* (KIMBALL; ROSS, 2002).

2.4 DATA WAREHOUSING

O *Data Warehouse* em sua aplicação, ou seja, com o agrupamento de dados, armazena um volume de dados na ordem dos mega bytes a giga bytes, diferentemente dos dados de um banco de dados transacional, onde os volumes são bem menores. Daí a necessidade de se trabalhar com *Data Warehouse* separada das bases de dados operacionais (KIMBALL; ROSS, 2002).

Outro ponto importante é que, enquanto o banco de dados transacional tem um volume muito grande de usuários com pequenas transações, o *Data Warehouse* recebe poucos usuários, porém com um volume grande de consultas, pois o mesmo é designado para efetuar consultas *ad-hoc* à base de dados históricos (KIMBALL; ROSS, 2002).

Como mencionado para *Data Warehouse*, os dados são armazenados em uma estrutura separada da base de dados operacional, para isso, é feito o processo de *ETL* (*Extract, Transform and Load* - Extração, Transformação e Carga) que fará a extração dos dados primários do banco de dados operacionais, faz-se a padronização dos dados obtidos e a inserção dos dados pertinentes à área de negócio a ser armazenado no *Data Warehouse*(VASSILIADIS, 2002).

Mas antes de se iniciar o processo de *ETL* ou definição de dados, alguns pontos devem ser considerados como: quais modelos devem ser usados, a arquitetura que será adotada e a granularidade dos dados.

A granularidade dos dados é descrita como o nível de detalhamento dos dados que são armazenados. Esse fator é importante, pois sua definição tem impactos diretos sobre o volume de dados armazenados fisicamente, ao desempenho das consultas executadas e na flexibilidade de informação contida no *Data Warehouse*.

O modelo comumente utilizado é o de Modelagem Dimensional. Ela é utilizada para definir de forma intuitiva o modelo de negócio, dividindo as estruturas de banco de dados em formas não normalizadas. A Modelagem Dimensional divide as relações em fatos e dimensões, onde os fatos são geralmente medidas numéricas e as dimensões são as descrições dos dados existentes (KIMBALL;ROSS, 2002).

A arquitetura do *Data Warehouse* pode ser subdividida em *Data Marts*, sendo que o conjunto de *Data Marts* compatíveis e integrados em um ambiente único formam um barramento de *Data Warehouse*.

A redução de seu escopo é o diferencial dessa estrutura, pois o intuito é gerar os *Data Marts* focados em um determinado assunto e com uma granularidade maior de dados, permitindo pesquisas mais aprofundadas sobre o assunto, sem comprometer o desempenho do *Data Warehouse*.

Essa estrutura contendo diversos *Data Mart* individuais compatíveis uns com os outros visa criar o esqueleto para um *Data Warehouse* integrado.

A Figura 2 apresenta a visão de Kimball e Ross (2002), que trata da aplicação do *Data Warehouse* seguindo a abordagem conhecida como *bottom-up*, ou seja, sugere que a primeira etapa da criação do *Data Warehouse* seja constituída pela criação dos modelos para os *Data Marts* e, após constituir dois ou mais deles, pode se generalizar para o *Data Warehouse*.

Figura 2: Exemplo de estrutura *Bottom-up* de *Data Warehouse*. Adaptada de (DWHINFO, 2009)

Outro ponto a ser avaliado conforme Kimball e Ross (2002) são as características do *Data Warehouse* que podem ser divididas em quatro grupos:

- Orientado por Temas:

O *Data Warehouse* tem como premissa ser orientado por temas ou assuntos, o que o difere dos ambientes operacionais que são modelados para processos. Essa abordagem de temas apresentada na Figura 3 mostra a estrutura de *Data Warehouse* dividido em três *Data Mart*: vendas, engenharia e financeiro. O acesso por temas específicos dentro do *Data Mart* permite uma busca mais eficiente aos dados para análises.

Figura 3: Ambiente *Data Warehouse* orientado por assunto

- Integrado:

A integração dos dados que serão armazenados nas bases de dados do *Data Warehouse* é proveniente de diferentes fontes, às vezes com dados incompatíveis entre si, como exemplo datas em padrões diferentes, medidas como peso em unidades diferentes entre outros. No entanto, trata-se de dados correlatos, que serão agregados para um mesmo assunto.

Assim, é necessário, no *Data Warehouse*, executar o processo de transformação desses dados para uma base integrada, como apresentado na Figura 4, em que as fontes de dados distintas possuem unidades de medidas distintas (libra, quilo, oz) e após a conversão são armazenadas em uma unidade padrão (gramas).

Figura 4: Conversão de dados para o *Data Warehouse*. Adaptada de (HANSON, 2009)

- Não-Volátil:

Um das principais premissas de um ambiente de *Data Warehouse* é que serão armazenados somente dados consolidados; assim, uma vez que os dados são inseridos os mesmos não serão mais alterados com o passar do tempo. Isso é importante para manter a integridade e consistência das análises providas através do *Data Warehouse*.

Essa propriedade justifica ainda mais a não aplicação desses tipos de análises sobre bases operacionais, pois a volatilidade de dados dessas bases é intensa.

- Variante no Tempo:

O agrupamento dos dados contidos no ambiente de *Data Warehouse* é tipicamente modelado para que seja relativo ao tempo, ou seja, são dados históricos agrupados, para que possam ser comparados entre diferentes meses e períodos durante o ano.

2.5 ARQUITETURA DE DATA WAREHOUSE

Uma estrutura comumente definida para *Data Warehouse* é apresentada na Figura 5, onde os dados operacionais, que podem ser encontrados em bases externas e descentralizados, passam pelo processo de *ETL*.

Figura 5: Componentes existentes no *Data Warehouse*.

Nesse processo são extraídos os dados definidos para a modelagem dimensional, feito o tratamento sobre os dados, como por exemplo, eliminar

redundância de dados. No processo final do *ETL*, os dados são armazenados nas relações de fatos e dimensões do *Data Warehouse*.

Para a carga dos *Data Marts*, os dados são inseridos sobre o assunto ao qual foi designado. Assim, tanto o *Data Warehouse* quanto os *Data Marts* estarão atualizados e aptos para uso no processo de obtenção de informação para tomada de decisão.

Esse uso de informação sobre o *Data Warehouse*, ou seja, a navegação em um ambiente que permite consultas, geração de relatórios, análises e gráficos, é feita através de ferramentas OLAP.

2.6 MATRIZ DE BARRAMENTO

A Matriz de barramento descreve a modelagem dimensional organizada em estrutura de dados de matriz. Isso auxilia a selecionar os pontos de aplicação e deixa mais intuitivo quais são os domínios de aplicação do *Data Warehouse*.

Na representação da Matriz de barramento, como mostra a Figura 6, as linhas da matriz são as áreas de negócios contendo valores numéricos (relações de Fatos) e as colunas representam as dimensões que são informação descritiva sobre as áreas de negócio. As células da matriz são marcadas caso uma determinada dimensão faça parte de uma determinada área de negócios (KIMBALL; ROSS, 2002).

A estrutura de Matriz permite adicionar novas áreas de negócio e dimensões e uni-las ao restante da Matriz, sem grande complexidade, ainda assegurando a facilidade da representação.

Figura 6: Exemplo de atriz de barramento de exemplo. Adaptada de (DWHINFO, 2009)

2.7 MODELO LÓGICO

O modelo lógico é parte importante da definição do *Data Warehouse*, pois a partir dele é que serão organizados os dados no banco de dados que o manterá.

2.7.1 Modelo Estrela

O modelo estrela é o mais utilizado entre os modelos disponíveis para *Data Warehouse*, sendo que dentre suas características está o fato de não seguir o padrão de normalização conhecido nos modelos de OLTP (Online transaction

Processing) e criar uma maior redundância de dados para pesquisas no banco de dados, o que melhora a eficiência no desempenho das mesmas.

O modelo é dividido basicamente por dois tipos de entidade: as dimensões e os fatos. Cada dimensão possui uma ligação única com uma entidade fatos, o que permite uma ligação consistente entre as entidades.

O nome de modelo estrela é proveniente de sua aparência, como apresentado na Figura 7. O centro do modelo é a relação de fatos, onde serão armazenados dados numéricos, medidas e chaves de relacionamento. Já as relações de dimensões armazenam dados descritivos sobre a área de negócios definida como tema do *Data Warehouse*.

Figura 7: Modelo estrela. Adaptada de (INMON, 1999)

Entre as vantagens do modelo estrela está o número reduzido de relacionamentos existente entre as relações, o que facilita as operações de OLAP e as consultas a grande volume de dados que exigem um alto poder de processamento dos bancos de dados.

A desvantagem fica por conta da não normalização dos dados contidos nas relações, o que inviabiliza alterações nessas bases. Contudo conforme citado, uma das premissas de um ambiente de *Data Warehouse* é possuir apenas dados consolidados.

2.7.2 Modelo Constelação

Conforme descrito por ELMASRI (2000), o modelo constelação é uma derivação do modelo estrela, sendo que a diferença é que uma ou mais relações de fatos podem compartilhar a mesma dimensão.

Isso permite a compatibilidade entre as relações de fatos que estão armazenadas no *Data Mart*, bem como a integração entre outros *Data Marts* para compartilhar as mesmas dimensões.

A Figura 8 mostra um exemplo de aplicação das relações de fatos de “vendas varejo” e de “promoções” compartilhando a mesma dimensão produtos.

Figura 8: Modelo constelação, adaptada de (ELMASRI, 2000)

2.7.3 Granularidade de dados

A coleta de dados pelas organizações atualmente é muito grande, a todo momento gestores recebem relatórios, consultas, dados de todo tipo de sistemas. Exemplo disso são os sistemas como ERP que armazenam todas as operações diárias da empresa.

Mas esses dados em sua grande maioria não são aproveitados e são armazenados em sistemas de backup para futuras consultas.

Em sistemas de apoio à decisão, os dados históricos armazenados são como um baú de peças pronto para serem analisados em busca informações importantes.

Mesmo assim, transpor toda a informação armazenada em *backups* para armazená-la em um sistema de informação como o *Data Warehouse* não será a solução, pois, como descrito acima, é necessário analisar quais conjuntos de dados são relevantes para a área de negócio a ser analisada. Esse processo está associado à granularidade dos dados.

Como descrito por Inmon (1999) a granularidade dos dados é definida pelo volume de dados requerida em um projeto de *Data Warehouse* que dará o nível de detalhamento que poderá ser obtido pelo *Data Warehouse*. O inverso também será verdadeiro, ou seja, quanto menos dados houver em um *Data Warehouse*, menor será seu nível de detalhes de informações obtidas.

A Figura 9 mostra como a granularidade pode ser definida num projeto de *Data Warehouse*, sendo que quanto maior a granularidade, mais sumarizados serão os dados e menor será o nível de detalhamento obtido; e quanto menor a granularidade, maior será o nível de detalhamento possível de ser extraído do *Data Warehouse*.

É importante manter o equilíbrio em relação a granularidade de dados, pois o maior volume de dados também terá impacto sobre o desempenho do sistema como um todo, sendo necessário maiores investimentos para não gerar mais informações.

E por outro lado, como o *Data Warehouse* irá trabalhar com dados históricos, a carga dos dados que serão inicialmente recuperados de sistemas de *backup* é interessante que seja feita com a granularidade correta para que tal processo não tenha que ser refeito por não atender ao detalhamento desejado.

Figura 9: Exemplo de granularidade, adaptada de (HANSON, 2009)

2.8 MODELAGEM DIMENSIONAL

A modelagem dimensional caracteriza-se por apresentar os dados de uma forma intuitiva e padronizada, permitindo um acesso em alto desempenho (KIMBALL; ROSS, 2002).

Um esquema comumente utilizado para representar a modelagem dimensional é o modelo estrela, no qual as medidas para gerenciamentos dos negócios podem estar em uma única relação (Relação de Fato), e os atributos relevantes que descrevem cada dimensão do negócio são armazenados em relações auxiliares, chamadas de Relações de Dimensões (INMON, 1999).

A divisão dos dados pertinentes ao negócio no modelo dimensional é feita em Cubos, sendo as informações mensuráveis para determinado ambiente chamadas de Células. Os limites do Cubo, ou seja, a informação proveniente de dados externos aos Cubos são denominadas dimensões. (KIMBALL; ROSS, 2002).

2.9 OLAP

As ferramentas OLAP são utilizadas para navegar sobre o repositório de dados do *Data Warehouse*. Dentre suas características esta a capacidade de executar consultas com grande volume de dados e retornar os resultados para o usuário final agregados e/ou transformar em relatórios e gráficos (THOMSEN, 2002).

A navegação sobre o *Data Warehouse* usando ferramentas OLAP pode ser feita por níveis de detalhamento desejado e escolha de variáveis que irão compor o resultado final, assim é possível executar operações como *drill down*, que significa aumentar o nível de detalhe desejado dos dados, ou *drill up*, que generaliza dos dados acessados, e ainda o *drill Throught*, que permite visualizar os dados agrupados que formaram o valor que foi obtido (THOMSEN, 2002).

Existem ainda várias outras formas de navegação, todas voltadas para a extração mais eficiente e rápida dos dados contido no *Data Warehouse*.

2.10 TRABALHOS RELACIONADOS

2.10.1 Modelos dimensionais para agropecuária

No artigo de Nilakanta et. al. (2008) são apresentados os desafios e problemas para a criação de modelo dimensional e criação de um *Data Warehouse* para a agropecuária na Índia baseado em dados operacionais providos por órgãos governamentais. No total são propostos modelos para 13 cadeias.

Entre os aspectos relevantes apontados na pesquisa está a apresentação de problemas sobre a modelagem, tais como relacionar dados de diferentes fontes com diferentes granularidades, dimensões com diferentes níveis de hierarquias, ou seja, dados fornecidos em nível nacional e outros regionais.

Por outro lado, o trabalho não deixa claro como estão divididas as cadeias produtivas do modelo, mostrando somente a cadeia de pecuária, e esse fator é importante ser exposto, pois a criação de modelos para cadeias produtivas similares entre si poderá gerar no próprio *Data Warehouse* redundância de informação.

Outro trabalho analisado é o de Rai et. al. (2008) que faz parte da mesma pesquisa apresentada no trabalho de Nilakanta et. al. (2008). Entretanto, nesse trabalho é exposto o modelo para cadeia agropecuária de animais. São modelos para pecuária e frango.

No entanto, integrar em um mesmo modelo de fatos essas duas cadeias (pecuária e frango), no Brasil, não representa um conjunto de informação correlatas. Informação sobre comercialização e produção de frango tem forte semelhança com sistemas de suinocultura, e pecuária contém informação distinta, o que justifica criar modelo para fato e dimensões voltadas especificamente para essa cadeia produtiva.

Os fatos e dimensões expostos no trabalho de Rai et. al. (2008) contribuem para analisar determinadas dimensões que poderão fazer parte também do proposto trabalho.

2.10.2 *Data warehouse* aplicados à agropecuária no Brasil

A dissertação de Marques (2002) apresenta uma aplicação de *Data Warehouse* para analisar dados sobre melhoramento genético da raça nelore. O trabalho tem uma ampla exposição dos conceitos de *Data Warehouse* e *Data Mart*.

Porém, por ter um foco específico no caso, a análise de dados de um programa específico de melhoramento genético não permite muito a generalização do modelo para outras áreas agrícolas. Além disso a aplicação é efetuada com ferramentas proprietárias, o que pode inviabilizar financeiramente essa aplicação.

Outro trabalho é o de Toni (2000) que descreve uma aplicação de *Data Marts* em cooperativas agrícolas. A dissertação coloca em foco a necessidade de melhorar o acesso à informação e organização das cooperativas.

O foco de aplicação é sobre a gestão dessas cooperativas, o que por um lado é interessante por mostrar um escopo que é possível ser aplicado a outras cooperativas de diferentes áreas. Porém, não há uma aplicação clara para a área agropecuária.

2.10.3 Informação no agronegócio

O trabalho de Boteon (2004) expõe as ofertas e demandas de informação sobre o agronegócio. A pesquisa mostra quais as fontes e níveis de detalhes de informação existente pela ótica de agronegócio e que podem auxiliar os agentes de mercado (agricultores, corretores, órgãos governamentais).

Outro ponto relevante é sobre o perfil dos usuários dessa informação, como níveis de escolaridade, uso da informação para tomada de decisão ou pesquisas

acadêmicas. O trabalho também apresenta os diferentes tipos de mídias usadas e se são públicas ou privadas.

2.11 CONSIDERAÇÕES PARCIAIS

As técnicas de *Data Warehouse*, como por exemplo o modelo dimensional e matriz de barramento, serão usadas como base inicial para a aplicação e definição do ambiente de *Data Marts* modelado para comercialização agropecuária.

O processo de *Data Warehousing* descrito acima também é importante para se ter a percepção integrada de todo o processo de *Data Warehouse*, desde os modelos aplicados até as ferramentas de OLAP que podem ser utilizadas para a navegação no *Data Warehouse*.

3 PROCESSO DE ANÁLISE AGROPÉCUARIA

Com o intuito de facilitar e permitir uma melhora nas análises efetuadas para o agronegócio, foi mapeado todo o processo, envolvendo desde a escolha do tipo de análise, depois a parte de aquisição dos dados, tratamentos dessas informações e meios de divulgação.

O mapeamento dos processos se faz necessário para que se possa identificar os pontos no qual este trabalho poderá ser incluído e, assim, melhorar ou possibilitar novas opções de informações e otimizar o tempo para a obtenção das informações para a produção das análises.

Para esse mapeamento foram consultados pesquisadores do Cepea, de duas áreas de estudos econômicos de comercialização distintas: Grãos e Pecuária.

O levantamento feito com os pesquisadores teve como resultado o fluxograma apresentado na Figura 10.

Figura 10: Fluxograma de análise de dados do agronegócio do Cepea.

O fluxograma está em um formato acordado com os pesquisadores para facilitar o entendimento, sendo necessária assim uma breve explicação do mesmo.

As caixas iniciadas com a letra “E” e seguidas de números são referentes às entradas dos processos, entradas essas que podem ser pesquisas externas, normas, solicitações de associações para estudos ou mesmo geração de dados do próprio processo, e são enumeradas para mostrar sua sequência no fluxo.

As caixas que iniciam em “P” e seguidas de números indicam os processos, ou seja, são as ações feitas em cada fase. Serão sobre esses processos que poderão ser aplicados os recursos propostos nesse trabalho. Os processos também seguem uma ordem de numeração sequencial.

As caixas que iniciam com “S” e seguidas de números são as saídas dos processos; são os resultados necessários a serem obtidos para a continuação do processo. Essas saídas poderão ser relatórios, bases de dados, consultas, entre outros e são seqüência no processo.

Os conectores interligam as entradas, processos e saídas mostrando o fluxo linear para obtenção da análise.

O fluxo apresentado na Figura 10 foi subdividido em cinco partes. Cada uma representa um macro processo, sendo que o macro processo da Figura 11 mostra como o pesquisador inicia o processo.

Figura 11: Macro processo definição de tema e fonte de dados para análise.

Assim, após definido o tema da análise que envolve “P1” e “S1”, é necessário buscar as fontes de informações. O processo “P2” e tem como entrada “E2”. Nesse processo “P2”, é possível ter como umas das opções em “E2” o *Data Warehouse* com bases de dados e sites que irão facilitar ao pesquisador a busca de dados.

Já a Figura 12 apresenta o macro processo de organização dos dados obtidos, macro processo esse que se inicia com a entrada “E3” e processo “P3” que, de posse das fontes primárias, os pesquisadores precisam trabalhar sobre essas bases de dados, ou seja, definir o período de todos os dados, analisar se todas ocorrem com mesma frequência, retirar dados redundantes, entre outras operações.

O processo “P4”, após esse nivelamento, segue o agrupamento desses dados. Como o foco é agronegócio, nesse ponto é necessário definir os produtos agrupando-os em cadeias de comercialização, organizar por unidades de medidas, pois produtos da mesma cadeia podem ser apresentados como quilo ou arroba, entre outros.

Figura 12: Macro processo de organização dos dados.

Esse é um macro processo chave para a qualidade dos dados obtidos inicialmente, e é todo feito de forma operacional, despendendo boa parte do tempo do pesquisador e que dificulta exponencialmente, dependendo do período e quantidade de produtos escolhidos para a análise.

Como exposto na revisão bibliográfica, após definido o modelo dimensional, que definirá o conjunto de informações que irá compor o *Data Warehouse*, é possível utilizar-se de técnicas automatizada para executar o processo de ETL.

Assim, as operações feitas no macro processo da Figura 6 podem ser adquiridas automaticamente de bases de dados operacionais e armazenadas no *Data Warehouse*, eliminando duplicidades, agrupando por períodos, unidades e produtos, organizando por assuntos, ou no caso cadeias de comercialização, e assim serem disponibilizadas para navegação para os pesquisadores. Outro fator relevante é o volume de informações, pois o *Data Warehouse* é projetado para armazenar dados históricos de muitos anos.

Já o macro processo de análise primária dos dados, representado pela Figura 13, que envolvem o processo “P5” em que depois de obtidos os dados consistentes o pesquisador fará o entendimento e gráficos sobre os dados desejados, período que melhor representa a hipótese que o mesmo definiu. Esse processo tem a sua criticidade em analisar se o conjunto e volume de dados obtidos corresponderam a necessidades da pesquisa.

Figura 13: Macro processo de análise primária dos dados.

Então ocorre o processo “P6” que será uma decisão sobre a utilização dos dados obtidos ou terão de ser refeitas as consultas e aquisição de dados conforme apresentado pelo processo “P7”. Nesse caso, como visto acima sem uma técnica de *Data Warehouse* será um longo processo novamente até o ponto de análise.

Existem ferramentas conhecidas como *OLAP* que permitem a navegação sobre os dados armazenados no *Data Warehouse*; tal navegação, como descrito na revisão, segue como se fosse em cubo, em que é possível obter cruzamentos de muitas variáveis numa mesma operação. Assim, ferramentas *OLAP* podem ser utilizadas para a aquisição primária dos dados já agrupados por assunto do *Data Warehouse*, permitindo adquirir os dados de diversas formas distintas e de periodicidade diferentes.

Por fim, o macro processo apresentado pela Figura 14 mostra o processo “P8” para a aplicação de modelos estatísticos por parte do pesquisador para que possa confirmar hipóteses feitas sobre os dados e acrescer a qualidade das análises. Para isso é utilizada a entrada “E6” em que foram obtidas dos dados já consolidados.

Figura 14: Macro processo de aplicação estatística e análise final.

Depois é executado o processo “P9” para a análise final dos dados consolidados em conjunto com os modelos estatísticos. E segue para a produção do relatório final através do processo “P11”.

Nesse último macro processo exposto pela Figura 14, as informações da entrada “E6” podem ser providas pelos dados existentes no *Data Warehouse* e a própria base do *Data Warehouse* pode ser realimentada com os dados da saída “S6”, pois se trata de dados consolidados.

Essa metodologia foi utilizada para a aplicação dos estudos de casos utilizando o processo de análise apresentado com as técnicas de *Data Warehouse* e modelo dimensional definidos para o agronegócio.

3.1 CONSIDERAÇÕES PARCIAIS

O processo de análise descrito traz uma síntese de como são geradas as informações para o agronegócio feito pelos pesquisadores do Cepea, sendo que esse processo pode ser dinâmico, ou seja, adaptado e otimizado.

Como proposto, o uso de técnicas de *Data Warehouse* nos macro processos descrito auxilia e minimiza os problemas que existentes no tratamento manual desses dados.

Outro ponto, é a otimização de aquisição de dados já consolidados através dos *Data Marts* para uso nos macros processos para a geração de análises.

4 DATA MART DE COMERCIALIZAÇÃO DE PECUÁRIA

O intuito da pesquisa voltada para a cadeia de produtos alimentares de origem animal, especificamente para produtos de pecuária, se deu pela sua representatividade no cenário agronegócio brasileiro.

Basta lembrar que a pecuária no Brasil possui cerca de 170 milhões de cabeças de gado (FAPRI, 2009). Outro ponto é o montante financeiro negociado nos últimos meses de contrato futuro na BM&F, superiores a R\$ 350 milhões (BM&F, 2009).

Assim como descrito, por Kimball e Ross (2002) sobre a criação do *Data Warehouse* baseado na estrutura *bottom-up*, sendo então criado os *Data Marts* específicos para cada área de negócio modelada.

Com o estudo do processo de definição de fontes de dados para análise com o pesquisador da área de pecuária, conforme apresentada na Figura 11, foram identificadas quais são as fontes primárias de dados e a forma como são extraídos os dados das mesmas.

Nesse macro processo são utilizados dados da base primárias do Cepea (2007), sendo dividido em diversos bancos de dados anuais e com alguns dados em backup.

Como essa divisão de dados está em um banco de dados operacional, as consultas têm que ser solicitadas à área de tecnologia e o tempo proporcional para a solicitação varia com a demanda existente, mas, em geral, isso ocorre com, no mínimo, dois dias úteis para consultas básicas, como valores no ano corrente.

Para auxiliar no processo descrito na Figura 11, primeiro se identificou qual tipo de informações poderia ser extraída da base de dados para a pesquisa.

A pesquisa foi desenvolvida para abordar dados relacionados a preços de produtos pecuários do Estado de São Paulo e demais regiões, relacionando preços regionais e indicadores, apresentando valores diferenciais dos mesmos.

Os preços dos produtos retirados do *Data Mart* utilizados no trabalho aborda a diferença entre o valor comercializado do produto na região em relação ao preço definido pelo indicador (sejam esses: Indicador de Boi Gordo – Estado de São, Indicador de Bezerro - Estado Mato Grosso do Sul e São Paulo).

Este diferencial é importante para os agentes do mercado futuro, uma vez que esses diferenciais representam uma forma de risco que foge ao controle dos agentes que fazem operações na Bolsa de Mercadorias e Futuros - BM&F e possuam produtos ou necessitam deles fora do Estado de São Paulo.

4.1 MATRIZ DE BARRAMENTO: *DATA MART* PECUÁRIA

A matriz de barramento do *Data Mart* pecuário apresentada na Figura 15 mostra à esquerda da matriz os fatos “Indicador” e “regional”, e, na parte inferior, as dimensões que contêm as descrições que dão suporte aos fatos. Nessa matriz, o fato “Regionais” não possui dados perminentes à dimensão indicadores.

Outro ponto é que, a cada relação de fatos ou dimensão nova que for inclusa ao *Data Mart*, deverá ser colocada na matriz para que possa ampliar as possibilidades de relações e consultas que o mesmo poderá atender.

Figura 15: Matriz de barramento: *Data Mart* Pecuária

A importância dessa representação através de matriz é que possibilita apresentar as dimensões e fatos que compartilham informações para os gestores das áreas de negócio de forma intuitiva, gráfica, que facilita a compreensão de consultas que poderão ser atendidas pelo *Data Mart*.

4.2 MODELO DIMENSIONAL

O *Data Mart* de pecuária foi modelado sobre dois *Star Schemas* (ELMASRI, 2000), sendo o primeiro modelo estrela definido para preços regionais e o segundo definido para preços de indicadores.

As relações modeladas são “Fato: Regionais_Pecu” e a “Fato: Indicador_Pecu”, a primeira delas possui dados referentes a preço médio, máximo e mínimo, divididos em valores a vista e a prazo.

A segunda possui os valores de comercialização a vista e a prazo, coeficientes de variações da amostra e os valores de taxas, dos quais os indicadores são ponderados.

A representação da Figura 16 demonstra um *Constellation* (ELMASRI, 2000) agregadas por dois *Stars Schemas*:

Figura 16: Modelagem dimensional aplicada à comercialização de pecuária.

As regiões onde se localizam as regiões em que são coletados preços de comercialização da pecuária, estão representas na relação denominada “Dimensão: Regiões”. A relação “Dimensão_Série” contém a descrição dos produtos. As demais dimensões descrevem os indicadores, datas e unidades utilizados.

4.3 APLICAÇÃO DE FERRAMENTA OLAP

A facilidade de navegação e de cruzamentos proporcionada pelo sistema de *Data Warehouse* permitiu a obtenção dinâmica dos dados de séries históricas de comercialização de pecuária.

A Figura 17 mostra a *interface* do sistema para *Data Warehouse*, como descrito no processo de análise, a otimização do macro processo de coleta de dados consolidados oferecida ao pesquisador através da ferramenta de navegação no *Data Mart* de pecuária, permite ao mesmo buscar intervalos de tempo maior que poderia se organizar em planilhas.

Outro detalhe é a possibilidade de inclusão e exclusão de variáveis de forma rápida pelo próprio pesquisador; a ferramenta ainda oferece recursos de navegação interna sobre os registros, exportação para outros formatos como “Excel” e “pdf” e geração dinâmica de gráficos (para maiores detalhes ver Apêndice A).

Data Mart Pecuaria

		Measures		
Series	Regioes	ValorVista	Min_Vista	Max_Vista
Boi Gordo	+MT	44,926	44,521	45,331
	+SP	51,775	51,237	52,308
Vaca	+MT	37,533	37,067	38,04
	+SP	46,699	46,114	47,31

Slicer: [Year=2006]

Figura 17: Tela de aplicação do *Data Warehouse* para pecuária, utilizando software Mondrian.

4.4 GERAÇÃO DE ANÁLISES USANDO ESTRUTURA DE *DATA MART*

Após a obtenção de dados, conforme a Figura 17, para efetuar o processo de análise, o pesquisador pode dar continuidade no macro processo de análise de dados primária, obtendo as planilhas de dados desejadas.

Conforme a Tabela 1, os dados permitiram verificar que os diferenciais seguem uma tendência distinta no decorrer do ano. No primeiro semestre, os diferenciais tendem a ser maiores, isto ocorre pela maior oferta de animais de pasto na época do verão.

Tabela 1: Dados provenientes do *Data Warehouse* de comercialização de Pecuária.

Boi Gordo	2002		2003		2004		2005		2006	
Valores a Prazo	1º S	2º S	1º S	2º S	1º S	2º S	1º S	2º S	1º S	2º S
Média Prazo SP	44.56	52.78	56.27	59.52	60.18	61.81	57.41	53.80	50.47	56.83
Tri. Mineiro - MG	43.19	51.52	54.17	58.46	57.84	60.00	55.61	52.48	50.98	56.28
Dourados - MS	41.90	50.74	53.52	57.50	56.95	59.45	53.14	49.97	46.70	54.28
Diferencial										
Tri. Mineiro - MG	1.37	1.26	2.09	1.06	2.34	1.81	1.80	1.32	-0.51	0.55
Dourados - MS	2.66	2.04	2.74	2.02	3.23	2.37	4.27	3.69	3.77	2.55

Outro ponto relevante que foi possível observar está relacionado ao comportamento histórico da série. As diferenças entre os preços tendem a seguir a lógica dos mercados de preços únicos, descrito por Barros (2006). Por essa regra, as diferenças entre os preços das regiões tendem a ser o preço, mais os custos de transferência e os impostos. Essa análise só foi possível porque foram utilizados dados históricos, não disponíveis nas bases operacionais.

4.5 CONSIDERAÇÕES PARCIAIS

A pesquisa tem apresentado que a área de agronegócio possui carência de informações que auxiliem nas pesquisas com dados consistentes e disponíveis para pesquisadores e outros interessados como, produtores, agentes de mercado que também possa fazer uso dessas informações para apoio a tomada de decisões diárias nas comercializações efetuadas.

Apesar de o mercado de pecuária ter uma grande importância no mercado brasileiro, o volume de oferta de informações consistentes não acompanha essa demanda.

O presente trabalho, como destacado em seus objetivos, buscou com essa pesquisa analisar a viabilidade da aplicação da técnica de obtenção e/ou otimização de dados, conhecida como técnicas de *Data Warehouse*, que já tem uma grande utilização em outros setores como financeiro e indústrias para gerar propostas de

modelos que venham a diminuir essa defasagem de acesso e/ou consistência de informações.

E reconhecido que a agropecuária em si possui uma grande extensão de outras variáveis e informações que, nesse trabalho não foram abordadas por motivos claros de sua grande quantidade de informações.

O que se trata de dados de comercialização, buscando atender a demandas de dados e ambiente para que pesquisadores do Cepea pudessem minimizar o processo desgastante de busca, nivelamento, armazenamento e duplicação de dados, como apresentado no processo de análise do capítulo 3, o trabalho atendeu às expectativas.

Vale lembrar que a criação de um *Data Mart* com dados específicos da cadeia de comercialização de pecuária possui cerca de 15 anos de dados históricos armazenados no mesmo. Divididos por preços comercializados em diversas regiões do Brasil, separados por diferentes produtos, sejam eles: Boi Gordo, Vaca Gorda, Bezerro, carnes e Indicadores que formam preços que são usados como base diária de negociações entre frigoríficos e produtores.

O *Data Mart*, em volume de dados, têm armazenado algo próximo a 200 mil registros em suas relações fatos, como apresentado nos modelos, e algo próximo a 200 registros em suas dimensões.

Se as operações para recuperação de dados diários fossem feitas sobre as bases de dados operacionais, haveria uma sobrecarga de processamento ocasionada por consultas históricas realizadas, prejudicando assim as coletas e geração de informações sobre essas bases. Outro fator é que como conhecido, as bases de dados operacionais apresentam pouco tempo de retenção de dados históricos e alta volatilidade, o que inviabiliza a realização de comparativos e análises de eventos passados.

Contudo, não só a questão computacional deve ser considerada para os benefícios do uso das técnicas aplicadas a esse estudo de caso, pois os ganhos obtidos pelos pesquisadores em relação ao tempo para obtenção e a consistência dos dados obtidos otimiza e facilita o trabalho do mesmo. Para exemplificar isso, basta analisar o processo descrito no capítulo 3, onde o pesquisador, ao definir o tema, terá que buscar pelas bases de dados e períodos definidos para efetuar sua análise.

Assim, após esse processo, o mesmo deverá solicitar a equipes de tecnologia a recuperação dos dados nas bases de dados do Cepea. Mas a análise de longos períodos de informações é demorada, tendo em vista os problemas já passados acima e o volume de dados grande sem ferramentas adequadas fica impraticável. Ainda para complementar o cenário, se após a solicitação o pesquisador descobrir que errou no período, produtos ou nas unidades desejadas, o processo de solicitação terá que ser refeito.

Já com o uso de ferramentas OLAP, nesse caso a ferramenta *Mondrian* (ver apêndice A), o pesquisador tem à sua disposição uma facilidade de navegação e obtenção de dados organizados e nivelados por assunto e permitindo trabalhar com tabelas e gráficos dinâmicos e sem prejudicar o processo operacional do Cepea.

Outro ponto é a aplicação da técnica para a criação do *Data Mart*. É interessante ver que a matriz de barramento pode ser aplicada para auxiliar na geração do modelo dimensional, o que permitiu organizar os dados por assunto na base de dados do *Data Mart* e, com isso, poder atender a aplicações *OLAP* que seriam feitas.

Ainda nesse contexto, a matriz de barramento de pecuária e o modelo dimensional são propostas iniciais para a cadeia de pecuária, sendo que o intuito é que o trabalho seja expandido com criação de novos modelos e agregado ao modelo já existente, já que o destaque desses modelos é que eles possam ser ampliados.

A aplicação dos modelos e os resultados obtidos nesse estudo de caso foram publicados também no congresso brasileiro de agroinformática, para que possa ser usado como auxílio para futuros trabalhos de aplicação de *Data Mart* para pecuária (CORREA, 2007).

5 DATA MART DE COMERCIALIZAÇÃO DE GRÃOS

A comercialização de grãos no mercado brasileiro é destaque nos índices de negociação em bolsa de valores, principalmente porque esses tipos de produtos são utilizados como base para a produção de outros produtos no mundo todo. (BM&F, 2009)

Como prova disso, nos últimos meses os contratos em aberto de Milho na BM&F têm ocupado a segunda posição dentre todos os outros produtos agropecuários comercializados (BM&F, 2009).

Isso faz com que cada vez mais as informações sobre essa cadeia, tais como preços negociados nas regiões, produtos ofertados, entre outras, sejam desejadas. Assim, o Cepea possui bases de dados de vários anos contemplando tais cadeias.

Contudo, o acesso a essas informações e a de outras fontes, nem sempre são otimizadas ou estão em ambientes operacionais, o que inviabiliza aplicações de consultas com grande volume de informações.

Com objetivo de atender a essa demanda e otimizar os processos de análises de comércio de Grãos, sem prejudicar o ambiente operacional e gerar consistência de informações, foram desenvolvidos modelos e aplicadas ferramentas de *Data Warehouse* gerando o *Data Mart orientado ao assunto*.

5.1 MATRIZ DE BARRAMENTO

A modelagem dimensional pode ser apresentada na forma de matriz de barramento, como mostrado na Figura 18, onde apresentam-se as áreas de negócios nas linhas e as colunas como dimensões. Essa representação permite a visão geral e intuitiva da área de negócio.

Figura 18: Matriz de barramento do *Data Mart Grãos*

5.2 MODELO DIMENSIONAL

O modelo foi criado para permitir análises nas cadeias de preços de grãos, que são soja e milho, nas diferentes praças de comercialização do Brasil, que contém informação onde os preços são coletados. As praças de comercialização relevantes para o mercado de grãos são apresentadas pela dimensão chamada “dimensão: Regional”.

A dimensão “Produto” contém a descrição dos principais produtos da cadeia. E a dimensão “Indicador” contém a descrição dos principais indicadores de preços da cadeia de grãos.

As outras dimensões descritas como datas, unidades e taxas também foram usadas.

A relação fato “Comércio Grãos” contém valores como datas, média, mínimos, máximos de preços da cadeia agrícola de grãos.

Essa estrutura é mostrada na Figura 19, que exhibe as dimensões e suas ligações e campos contidos na relação fato. Tais como na relação “dimensão: Regional” que tem informação de cidades e estados e permitem obter os preços médios na relação “fato: grãos”.

Figura 19: Modelagem dimensional para comercialização de grãos.

5.3 APLICAÇÃO DE FERRAMENTA OLAP

O *software mondrian* mostrado na Figura 20 é utilizado como ferramenta OLAP que permitiu ao pesquisador ter um ambiente que permitiu a criação de análise dos dados históricos obtidos. É possível ver as médias de regiões como Rio Verde e o gráfico obtido dos dados.

Data Mart Graos

	Measures		
	ValorVista		
	Regioes		
Series	Rio Verde	Norte PR	Passo Fundo
-All Series	30,498	32,042	33,913
Milho (Atacado)	20,368	21,479	24,04
Milho (Produtor)	19,191	19,002	21,87
Soja (Atacado)	42,658	45,299	46,393
Soja (Produtor)	39,782	42,443	43,361

Slicer: [Year=2008]

Slicer: Year=2008

— ValorVista.GO.Rio Verde. — ValorVista.PR.Norte PR. — ValorVista.RS.Passo Fundo.

Figura 20: Tela de aplicação do *Data Warehouse*, baseado em software Mondrian.

5.4 GERAÇÃO DE ANÁLISES USANDO ESTRUTURA DE *DATA MART*

A produção agrícola de grãos no Brasil tem pelo menos duas características próprias. Primeiro, há a dispersão da produção, com destaque para as regiões Sul e Centro-Oeste, em que o período de plantio e colheita se diferencia, assim como os momentos de tomada de decisão do produtor. Segundo, que no Brasil há pelo menos duas safras distintas, verão e inverno, em que as culturas concorrem em área entre si.

No verão, a principal concorrência ocorre entre soja e milho. A opção do produtor entre outra cultura depende, entre outros fatores, da relação de preços entre os produtos. Se o preço da soja estiver mais atrativo, opta-se pelo seu plantio, caso contrário, pelo milho.

Na Figura 21 observa-se esta relação. Em 2007 a relação ficou praticamente abaixo de 2, com os menores valores sendo observado no início e no final do ano. esta relação apontava atratividade do plantio de milho em detrimento da de soja, situação que resultou em crescimento expressivo da produção de milho na safra 2007/08 no Brasil. No ano seguinte, porém, os preços do milho decresceram e o da soja aumentou, fazendo crescer a relação, que mostrava atratividade para a soja. Como resultado, na safra 2008/09, a área plantada com milho no Brasil diminuiu e a de soja aumentou.

Figura 21: Relação de preços entre soja e milho.

5.5 CONSIDERAÇÕES PARCIAIS

Como definido nos objetivos, foi aplicado às técnicas de *Data Warehouse* sobre dados de comercialização de Grãos e, assim, gerando o *Data Mart* de assunto específico sobre essa cadeia.

Similar às considerações descritas no capítulo 4, a cadeia de comércio de Grãos possui grande destaque no agronegócio brasileiro, sendo assim, objeto desse estudo.

Os dados trabalhados nesse estudo de caso visaram consolidar a aplicação dos modelos definidos, gerando-se assim os novos modelos para a cadeia de grãos.

A comercialização dos produtos de grãos desse *Data Mart* abrangeram dados de produtos de Soja e seus derivados (Farelo, óleo) e Milho, contemplando dados históricos de diversas regiões, índices, entre outros.

Outro fator foi que, apesar de já ter sido modelado uma cadeia anteriormente (capítulo 4), por se tratar de mercados distintos, se fez necessário o entendimento junto ao pesquisador responsável pela área desse tipo de mercado no Cepea para mapeamento e definição dos modelos criados.

A produção desse *Data Mart* envolveu inicialmente o mapeamento das informações solicitadas para as análises feitas pelos pesquisadores, seguindo pela observação quanto à compatibilidade em relação aos modelos criados para o *Data Mart* de pecuária e, por conseqüente, as criações dos modelos.

Após as definições e carregamento dos dados contido nas bases operacionais, foi possível aos pesquisadores usufruírem da contribuição do *Data Mart* em suas análises para obtenção e navegação sobre os dados históricos dos produtos mencionados acima.

A pesquisa nesse estudo de caso aplicou a estrutura de criação da matriz de barramento, em que foram definidos os fatos e dimensões para o *Data Mart*. A criação do modelo dimensional de comércio de Grãos, e por fim a sua implementação em ferramenta OLAP, mesma ferramenta usada anteriormente.

O trabalho atendeu às necessidades de consultas para análises, mas também é um trabalho em evolução, sendo estes modelos iniciais que permitem serem ampliados para inclusão de novos modelos. Também permitiu gerar artigo que foi

publicado em congresso internacional (Correa et. al. 2009) da área agrícola e tecnológica, o que reforça ainda mais a importância de ampliação e continuidade desse tipo de pesquisa.

6 CONCLUSÃO

6.1 CONSIDERAÇÕES FINAIS

A decisão da criação dos *Data Marts* (Pecuária e Grãos) independentes focados por assuntos foi baseada na literatura de Kimbal (2002), como mencionada anteriormente, em conformidade com um processo *bottom-up*, em que descreve a estrutura de vários *Data Marts* para criar um barramento de *Data Warehouse*.

A sequência definida para os objetivos do trabalho visou à criação inicial dos *Data Marts* para que, a partir do conhecimento adquirido e dos modelos gerados, fosse possível criar um barramento de *Data Warehouse* que permita expandir essa aplicação para outras áreas de comercialização agropecuária.

Assim, a criação do *Data Mart* de pecuária se torna uma contribuição relevante, pois, como apresentado no capítulo 3, as etapas necessárias para obtenção dos dados, sendo que com o uso de ferramentas OLAP e com os dados organizados por assunto no *Data Mart* facilitou-se a produção das análises desse mercado.

Como mostrado no capítulo 4, foram extraídas informações de diversas regiões e produtos desde o ano de 2000 até 2007, sendo apresentada para o pesquisador uma ferramenta de fácil navegação e extração de dados.

O modelo dimensional criado para o estudo de caso de pecuária apresenta as dimensões, como Séries, regiões e indicadores, que são as formas de cruzamento de informações feitas para o mercado de pecuária para auxílio a tomada de decisão sobre compra e venda de produtos (Boi Gordo e Bezerro).

A criação do *Data Mart* de grãos permitiu aos pesquisadores ter uma agregação de dados de outras fontes, o que, como mostrado no macro processo de análise do capítulo 3 é demorado pois precisa nivelar esses dados e agrupar sobre a mesma unidade, agregando produtos que possuem relações diretas de comercialização e produção. O acesso aos dados (soja e milho) igualados ao tempo,

unidades e tipo de comércio (atacado e produtor) possibilitou avaliar os impactos das variações de preços em diversas regiões simultaneamente.

Outro fator importante é que, como apresentado no modelo dimensional de grãos, foi possível agregar os produtos de soja e milho utilizando as mesmas dimensões, que permitiu uma análise conjunta desses produtos.

A criação dos dois *Data Mart* permitiu dar início a esse processo e foi criada a compatibilidade entre as dimensões, fatos e modelos para poder assim gerar um *Data Warehouse* integrado para cruzamento de dados entre as cadeias de comercialização agropecuária.

Contudo, é importante observar que esses ainda são passos iniciais para que seja permitido dizer que as informações disponíveis para o agronegócio estão próximas das formas de obtenção e acesso de informações existentes em outras áreas, como exemplo a área financeira.

Entretanto, a área agropecuária se destaca pelas suas peculiaridades pertinentes ao negócio, pois a mesma possui muitas diversidades e variáveis que se alteram com frequência. Como exemplo pode se destacar o processo de custo de produção no agronegócio, que envolve uma gama extensa de variáveis a se considerar, o que já não ocorre em uma indústria em que as variáveis existentes podem ser mapeadas mais facilmente.

6.2 TRABALHOS FUTUROS

A continuidade da aplicação de técnicas de *Data Warehouse* em comercialização agropecuária, ou seja, a geração de modelos e aplicação de ferramentas OLAP deverá ser intensificada e expandida para outros produtos e também sobre outros segmentos, por exemplo, custo de produção.

A importância de serem ampliados esses trabalhos é para que possa futuramente se propor um processo de referencia para a criação de um ambiente de *Data Warehouse* integrado para o agronegócio que permitirá agregar outras

instituições e ter informações completas sobre diversos elos do produto, como: produção, clima, frete.

Outra técnica que deve ser implantada para otimização de análises é a de *Data Mining*, que nesse caso seriam para identificação e busca por padrões de dados ou sequência nas bases de dados do *Data Warehouse*.

A aplicação de técnicas de *Data Mining* é mais bem empregada como descrito na revisão sobre ambientes já consolidados e livre de inconsistência ou redundância de variáveis; assim o *Data Warehouse* já atende a essas restrições e se torna uma ótima alternativa.

O uso de *Data Mining* para o agronegócio também precisa ser mais explorado por pesquisas, e principalmente aplicações estatísticas dessa técnica, o que possibilitará melhorar a qualidade dos dados disponíveis, pois permitirá encontrar dados que deveriam seguir uma mesma tendência de volume de dados ou preços e estarem divergentes por erros de coleta, e ampliar o número de variáveis analisadas simultaneamente.

REFERÊNCIAS

ABIEC. **Associação Brasileira das Indústrias Exportadoras de Carnes.** Disponível em: <<http://www.abiec.com.br/>>. Acessado em: 10 Jun. 2009.

APACHE-TOMCAT. **Software para servidor web.** Disponível em: <<http://tomcat.apache.org/>>. Acessado em: 20 Abr. 2007.

BARROS, G. S. C., **Economia da Comercialização Agrícola**, FEALQ, 2006. 221 p.

BATALHA, M. O. **Gestão Agroindustrial**, Atlas, 2007. 420 p.

BM&F. **Bolsa de Mercadorias e Futuros.** Disponível em: <<http://www.bmf.com.br/>>. Acessado em: 10 Set. 2009.

BOTEON, M., **Mercado de informação digital agroecômica**, 2004, 90 p. Tese de Doutorado. Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo. Piracicaba, São Paulo, 2004.

CEPEA. **Centro de Estudos Avançados em Economia Aplicada.** Disponível em: <<http://www.cepea.esalq.usp.br/>>. Acessado em: 05 Nov. 2007.

CHAUDHARY, S., SORATHIA, V.; LALIWALA, Z., **Architecture of sensor based agricultural information system for effective planning of farm activities.** Services Computing. IEEE International Conference. p. 93-100. Set de 2004.

CNA. **Confederação da Agricultura e Pecuária do Brasil.** Disponível em: <<http://www.cna.org.br/>>. Acessado em: 20 Abr. 2008.

CORREA, F. et al. **Data warehouse de comercialização de Pecuária para análise de diferenciais entre regiões do Brasil.** 6º Congresso Brasileiro de Agroinformática, SBIAgro, v.1, p.1-5. 2007.

CORREA, F. et al. **Data warehouse for soybean and corn market on Brazil.** Joint International Agricultural Conference. Wageningen, p. 1-8. 2009.

DAVIS, J. H., GOLDBERG, R. A. **A concept of agribusiness. Division of research. Graduate School of Business Administration.** Boston, Harvard University, 1957.

DWHINFO. **Data Warehouse Information Site.** Disponível em: <<http://www.dwhinfo.com>>. Acessado em: 15 Set 2009.

ELMASRI, R.; NAVATHE, S.B. **Fundamentals of Database Systems.** 3ª Ed. Menlo Park: Addison-Wesley, 2000. p. 841-872.

EMBRAPA. **Empresa Brasileira de Pesquisa Agropecuária.** Disponível em: <<http://www.embrapa.br/>>. Acessado em: 02 Dez. 2008.

FAPRI. **Food and Agricultural Policy Research Institute.** Disponível em: <<http://www.fapri.iastate.edu>>. Acessado em: 15 Abr. 2009.

HANSON, G.S.B. **Aplicação da técnica de Associação para realizar Mineração de Dados em um repositório na Área da Saúde Pública,** 2009. Monografia, Escola de Engenharia de Piracicaba, Piracicaba, 2009.

HAVERKORT, A.J., TOP, J. L., VERDENIUS, F., **Organizing Data in Arable Farming: Towards an Ontology of Processing Potato,** V. 49, I. 3, p. 177-201. Out de 2006.

IBGE – **Instituto Brasileiro de Geografia e Estatística.** Disponível em: <<http://www.ibge.gov.br/>>. Acessado em: 09 Nov. 2008.

IBM, **Dimensional Modeling: In a Business Intelligence Environment.** p. 670. 2006. Acessado em: 20 Mai. 2008.

INMON, W. H. **Management Data Warehouse,** Makron Books, 1999. 375 p.

JAVA. **Plataforma para programação e ambiente.** Disponível em: <<http://www.java.com/>>. Acessado em: 20 Fev. 2007.

KIMBALL, R.;ROSS, M. **The Data Warehouse Toolkit.** 2º Ed. Rio de Janeiro: Editora Campus, 2002. 494 p.

MARQUES, V.F. **Analisando os dados do programa de melhoramento genético da raça de nelore com data warehousing e data mining**, 2002, 134 p. Dissertação (Mestrado). ICMC, Universidade de São Paulo. São Carlos. 2002.

MEIRA, C. A. A., **Projeto Físico de Data Warehouse no Oracle: um Estudo de Caso**. p. 1-21. 2003, Embrapa (Estudo técnico).

MONDRIAN. **Software OLAP para Data Warehouse**. Disponível em: <<http://mondrian.pentaho.org/>>. Acessado em: 24 Mar. 2007.

MYSQL. **Software de gerenciamento de banco de dados**. Disponível em: <<http://www.mysql.org/>>. Acessado em: 25 Mar. 2007.

NILAKANTA, S., SCHEIBE K., RAI A., **Dimensional issues in agricultural data warehouse designs**, Computers and Electronics in Agriculture. V. 60, I. 2, p. 263-278. Mar 2008.

RAI, A. et al. **Original papers: Design and development of data mart for animal resources**. Computers and Electronics in Agriculture. V. 64, I. 2, p. 111-119. Dez de 2008.

SCHULZE, C., SPILKE J., LEHNER W., **Data modeling for Precision Dairy Farming within the competitive field of operational and analytical tasks**, Computers and Electronics in Agriculture. V. 59, I. 1-2, p. 39-55, Nov 2007.

SILVA, A. P., **Da conversa na praça ao via satélite, a busca por informação agropecuária**. 2005, 113 p. Dissertação (Mestrado). Escola Comunicação e Artes da Universidade de São Paulo. São Paulo, São Paulo, 2005.

TONI, J. A. **Definição de um Data Mart em Cooperativas Agropecuárias**, 2000, 172 p. Dissertação (Mestrado). Universidade Federal de Santa Catarina, Florianópolis. 2000.

USDA. **United State Department of Agriculture**. Disponível em: <<http://www.usda.gov>>. Acessado em: 10 Abr. 2009.

THOMSEN, E. **Olap Solutions: Building Multidimensional Information Systems**, John Wiley & Sons, Inc, 2nd. 2002. 608 p.

VASSILIADIS, P., SIMITSIS, A., AND SKIADOPOULOS, S. **Conceptual modeling for ETL processes**. ACM. p. 14-21. Ago. 2002.

APÊNDICE A

7 APLICAÇÃO DO SOFTWARE MONDRIAN

A ferramenta usada no trabalho é o Mondrian, que faz parte de um pacote de ferramentas para BI chamada Pentaho.

O software Mondrian segue a licença Open-Source, ou seja, são software de uso gratuito. Pelo fato do software ser desenvolvido em linguagem Java e ter seu código fonte disponível é permitido fazer as alterações que sejam conveniente para adaptação ao ambiente que será instalado.

A seguir serão apresentados os requisitos de software necessários para a utilização do mondrian, quais as principais alterações precisam ser feitas e os principais recursos disponíveis.

7.1 REQUISITO DE INSTALAÇÃO

Por ser baseado em plataforma Java, o software pode ser executado na maioria dos sistemas operacionais disponíveis (Microsoft Windows, Linux, Mac). Mas é necessário alguns do software instalados para sua execução, sendo:

- Ambiente JRE (Java Runtime Environment).
- Ambiente Web Apache-Tomcat versão 5.5 ou superior.

Ambos os software acima mencionados precisam ser instalados antes de iniciar o processo de utilização do Mondrian e podem ser obtidos no site: <http://tomcat.apache.org> para o apache-tomcat e <http://www.java.com> para o JRE.

7.2 PROCESSO DE INSTALAÇÃO DO MONDRIAN

Após a instalação dos software é necessário somente o download do software Mondrian, disponível em: <http://mondrian.pentaho.org>.

Para a instalação do Mondrian é necessário extrair o arquivo que foi feito download do mondrian e na pasta procurar dentro dos arquivos de fonte o arquivo chamado "mondrian.war".

O próximo passo é colocar o arquivo “mondrian.war” na pasta do apache-tomcat, dentro da pasta “webapp” e iniciar o software apache-tomcat.

Quando o software apache-tomcat for iniciado, automaticamente será criada uma pasta dentro do “webapp” chamada mondrian.

Para a configuração do mondrian, podem ser alterados os seguintes arquivos:

- Pasta “webapps/mondrian/”:
 - index.jsp: A configuração principal da página, definição de páginas que serão solicitada entre outras configurações. Esse arquivo pode ser mantido como está por padrão.
 - testpage.jsp: Alteração de alguns textos também da página principal, o mesmo pode ser mantido sem alteração.
- Pasta “webapps/mondrian/WEB-INF/”:
 - datasource.xml: Esse arquivo é necessário ser alterado, pois nele irá conter dados sobre a conexão com o Banco de Dados de repositório, ou seja, o Banco de dados que irá armazenar os dados do *Data Mart*.
 - mondrian.properties: Esse arquivo também tem informações sobre o Banco de dados e qual usuário do Banco de Dados fará a conexão. Nesse arquivo também é definido o nome do arquivo XML do cubo possui as definições de dimensões e fatos do *Data Mart*, como exemplo iremos usar o nome FoodMart.xml (que é o padrão que vem no mondrian).
- Pasta “webapps/mondrian/WEB-INF/queries”:
 - mondrian.jsp: Esse arquivo tem informações sobre a consulta inicial que será feita no *Data Mart* quando for acessada a página principal. Também é definido o arquivo XML do cubo, como mencionado por padrão será o FoodMart.xml
 - FoodMart.xml: Esse arquivo é um dos principais da configuração pois nele ficará definido quais são as relações de fato e dimensões, bem como os campos quais os campos serão disponíveis para o usuário.

Por fim vale lembrar que o mondrian possui conexões com diversos tipos de Banco de Dados, sendo: Mysql, PostgreSQL, Oracle entre outros. A limitação na verdade está na plataforma Java, se a mesma possui conectores para o Banco de Dados escolhido. No trabalho foi utilizado o Banco de Dados Mysql Server versão 5, disponível em: <http://www.mysql.com>.

7.3 ARQUIVO XML DE EXEMPLO

O Arquivo “graos.xml” apresentado na Figura 22 mostra como pode ser configurado as relações de dimensões e fatos obtidos do modelo dimensional para ser usado no mondrian.

```
<?xml version="1.0"?>
<Schema name="Graos">
  <Dimension name="Produto">
    <Hierarchy hasAll="true" primaryKey="IDProduto">
      <Table name="dim_produto"/>
      <Level name="Produto" column="nomeproduto" uniqueMembers="true"/>
    </Hierarchy>
  </Dimension>
  <Dimension name="Regiao">
    <Hierarchy hasAll="true" primaryKey="IDRegiao">
      <Table name="dim_regiao"/>
      <Level name="Estado" column="UF" uniqueMembers="false"/>
      <Level name="Regiao" column="nomepraca" uniqueMembers="false"/>
    </Hierarchy>
  </Dimension>
  <Dimension name="Time" type="TimeDimension">
    <Hierarchy hasAll="false" primaryKey="time_id">
      <Table name="Data_dia"/>
      <Level name="Ano" column="ano" type="Numeric" uniqueMembers="true" levelType="TimeYears"/>
      <Level name="Semestre" column="semestre" uniqueMembers="false" levelType="TimeSemesters"/>
      <Level name="trimestre" column="trimestre" uniqueMembers="false" levelType="TimeQuarters"/>
      <Level name="Mes" column="mes" uniqueMembers="false" type="Numeric" levelType="TimeMonths"/>
    </Hierarchy>
  </Dimension>
  <Cube name="Regional">
    <Table name="fato regional"/>
    <DimensionUsage name="Produto" source="Produto" foreignKey="IDProduto"/>
    <DimensionUsage name="Regiao" source="Regiao" foreignKey="IDRegiao"/>
    <DimensionUsage name="Time" source="Time" foreignKey="time_id"/>
    <Measure name="Valor_Vista" column="ValorVista" aggregator="avg"/>
    <Measure name="Valor_Prazo" column="ValorPrazo" aggregator="avg"/>
  </Cube>
</Schema>
```

Figura 22: Modelo de arquivo xml para o Mondrian.

Sendo as marcações “Dimension Name” o nome dado para a Dimensão, e as marcações “Cube Name” define o nome do Cubo ou Fato definido no modelo dimensional.

Outras marcações importantes são a “Hierarchy” que define qual descrição da dimensão é superior a outra, como por exemplo: Ano é superior a Semestre.

Já os itens que constam dentro do marcador “Cube” são referentes as dimensões e os dados numéricos que serão apresentados ao usuário. É possível através desses dados definir que tipos de operações serão executadas quando os mesmos forem agrupados.

A marcação “DimensionUsage” mostra quais da dimensão acima são usadas, e a marcação “Measure” apresenta quais campos da relação de fato será usada nesse cubo e qual a operação de agrupamento será feita, a usada nesse exemplo foi a “avg” que fará a média dos dados que são agrupados.

7.4 EXECUÇÃO DO MONDRIAN

A tela inicial do mondrian será conforme as alterações feitas no arquivo “mondrian.jsp”, então usando como exemplo uma configuração já feita teremos a Figura 23 como exemplo. Em que será destacada a barra superior de ferramentas.

Data Mart Pecuaria

	Measures		
Series	• ValorVista	• Min_Vista	• Max_Vista
-All Series	143,451	139,669	147,269
Bezerro	351,307	339,928	362,594
Boi Equivalente			
Boi Gordo	45,795	45,356	46,269
Boi Magro	563,678	548,366	579,225
Carcaca Casada Boi	3,141	3,112	3,172
Carcaca Casada Vaca	2,838	2,816	2,865
Dianteiro Boi	2,429	2,403	2,459
PA Boi	2,1	2,059	2,149
Traseiro Boi	4,008	3,982	4,036
Vaca	40,26	39,732	40,843

Slicer: [Year=2006]

[back to index](#)

Figura 23: Tela do mondrian com consulta inicial.

A barra de ferramenta do mondrian permite a, por exemplo, alterar as dimensões e campos da relação de fato que foram configurados inicialmente, a Figura 24 mostra a tela que é aberta ao clicar no ícone de cubo da barra de ferramenta e como é possível alterar as dimensões.

Data Mart Pecuaria

Figura 24: Cubo com as dimensões e medidas do *Data Mart*.

Outros ícones que se destacam nessa barra é a possibilidade de produção de gráficos disponíveis, como na Figura 25 é possível clicar no ícone para configurar as opções de gráficos ou no ícone do gráfico para incluir o gráfico na tela junto a relação.

Data Mart Pecuaria

Figura 25: Configuração e gráfico do Mondrian.

Por fim existem as opções de navegação do mondrian, que são os ícones de Drill UP e Drill Down que permitem subir os níveis de acesso aos dados no *Data Mart* ou diminuir. Outros ícones são responsáveis por impressão dos dados e de exportação, sendo possível exportar o conteúdo para Excel ou pdf.