

**Antonio César Centenaro**

**DESENVOLVIMENTO E IMPLANTAÇÃO DE UM DATA WAREHOUSE  
CORPORATIVO COM DATA MARTS DISTRIBUÍDOS EM UMA  
COOPERATIVA AGROINDUSTRIAL**

Dissertação apresentada ao Programa de  
Pós-Graduação em Engenharia de Produção  
da Universidade Federal de Santa Catarina  
como requisito parcial para obtenção do grau  
de Mestre em Engenharia de Produção.

Orientador: Prof. José Leomar Todesco, Dr.

**Florianópolis**

**2003**

**Antonio César Centenaro**

**DESENVOLVIMENTO E IMPLANTAÇÃO DE UM DATA WAREHOUSE  
CORPORATIVO COM DATA MARTS DISTRIBUÍDOS EM UMA  
COOPERATIVA AGROINDUSTRIAL**

Esta dissertação foi julgada e aprovada para obtenção de grau de **Mestre em Engenharia de Produção** no **Programa de Pós-graduação em Engenharia de Produção** da Universidade Federal de Santa Catarina

Florianópolis, 11 de julho de 2003.

Prof. Edson Pacheco Paladini, Dr.  
Coordenador do Programa

**Banca Examinadora**

---

Prof. José Leomar Todesco, Dr  
Universidade Federal de Santa  
Catarina  
**Orientador**

---

Prof. Aran Bey Tcholakian Morales, Dr.  
Universidade Federal de Santa Catarina

---

Prof. Oscar Ciro Lopez Vaca, Dr  
Universidade Federal de Santa Catarina

À minha esposa Regiane.

À minha filha Sarah.

## *Agradecimentos*

À Deus, por me permitir viver e por sempre iluminar o meu caminho.

À Universidade Federal de Santa Catarina.

Ao departamento de Engenharia de Produção e Sistemas.

Ao Prof. José Leomar Todesco, pelo valoroso trabalho de orientação e  
motivação durante o desenvolvimento desta pesquisa.

Aos professores do programa de pós-graduação em Engenharia de  
Produção e Sistemas.

A C.Vale pela oportunidade oferecida e em especial aos amigos da  
consultoria de informações pelo apoio no desenvolvimento deste projeto.

À minha esposa Regiane A. C. Centenaro, pela compreensão, carinho e  
apoio.

À minha filha Sarah C. Centenaro, por ter me dado tanta alegria e por  
desempenhar um papel único de motivação e incentivo, mesmo sem ter  
consciência ainda do que isto significa.

A todos que direta ou indiretamente contribuíram para a realização desta  
pesquisa.

## Resumo

CENTENARO, Antonio César. **Desenvolvimento e implantação de um data warehouse corporativo com data marts distribuídos em uma cooperativa agroindustrial**. 2002, 116f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-graduação em engenharia de produção, UFSC, Florianópolis.

A tecnologia de Data Warehouse vem despertando um grande interesse em corporações, que vêem nela a oportunidade de agilizar e qualificar seus processos decisórios. Contudo, algumas destas corporações possuem unidades de negócio distribuídas geograficamente e, devido a problemas de infra-estrutura de comunicação de dados, acabam por não implementar este tipo de solução ou, quando implementam, criam uma solução que não abrange toda a corporação. Este trabalho propõe o desenvolvimento e implantação de um data warehouse corporativo com data marts distribuídos à ser utilizado por empresas que tenham a necessidade de implementação de um sistema de apoio a decisão baseado em data warehouse distribuído e que possuam ou não deficiências tecnológicas de transmissão de dados. No trabalho são descritos conceitos de sistemas de informação e suas classificações quanto a finalidade, o modelo dimensional e o data warehouse. O resultado do trabalho é a implementação de um modelo de data warehouse corporativo com data marts distribuídos e construídos segundo a arquitetura incremental (BUS), em uma cooperativa agroindustrial – C.Vale.

Palavras-chave: Data Warehouse; Data Marts Distribuídos; Sistemas de Apoio a Decisão.

## Abstract

CENTENARO, Antonio César. **Desenvolvimento e implantação de um data warehouse corporativo com data marts distribuídos em uma cooperativa agroindustrial.** 2002, 116f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-graduação em engenharia de produção, UFSC, Florianópolis.

The technology of Data Warehouse has been arousing a great interest in some corporations, that find in it the opportunity of rushing and qualifying their deciding process. However, some of these corporations own business units apread geographically and, due to problems of super-structure of data communication, they end up by not implementing this kind of solution or, when they do it, they create a solution that doesn't range all the corporation. This paperwork proposes the development e implantation of a corporative data warehouse with distributed data marts to be used by companies which have the necessity of implementing of a decision support system based on the data warehouse distributed and which they whether own or not technological deficiencies of data broadcast. In the paperwork are described concepts of information systems and their classification as for the purpose, the dimensional model and the data warehouse. The result of the paperwork is the implement of a model of corporative data warehouse with data marts distributed and built according to the incremental architecture (BUS), in an agricultural and industrial cooperative - C.Vale.

Keywords: Data Warehouse; Data Marts Distributed; Decision Support Systems.

## Sumário

<b>1. INTRODUÇÃO .....</b>	<b>3</b>
<b>1.1 Apresentação.....</b>	<b>3</b>
<b>1.2 Objetivos .....</b>	<b>3</b>
1.2.1 Objetivo geral.....	3
1.2.2 Objetivos específicos .....	3
<b>1.3 Justificativa.....</b>	<b>3</b>
<b>1.4 Estrutura do Trabalho.....</b>	<b>3</b>
<b>2. TECNOLOGIA DE SISTEMAS DE INFORMAÇÃO.....</b>	<b>3</b>
<b>2.1 Introdução.....</b>	<b>3</b>
<b>2.2 Sistemas de Informação .....</b>	<b>3</b>
2.2.1 Sistemas de processamento de transações .....	3
2.2.2 Sistemas de informações gerenciais.....	3
2.2.3 Sistemas de apoio à decisão .....	3
2.2.4 Sistemas de informação executiva.....	3
<b>2.3 Data Warehousing .....</b>	<b>3</b>
2.3.1 Dados Operacionais Versus Dados Informacionais.....	3
2.3.2 Data Warehouse .....	3
2.3.3 Características do <i>Data Warehouse</i> .....	3
2.3.4 Data Mart .....	3
2.3.5 Arquitetura do <i>Data Warehouse</i> .....	3
2.3.6 Topologias de Implementação .....	3
2.3.7 Arquiteturas de Implementação .....	3
2.3.8 Abordagens de Implementação .....	3
<b>2.4 Modelagem Dimensional .....</b>	<b>3</b>
2.4.1 Navegação pelo modelo dimensional .....	3
2.4.2 Fatos.....	3
2.4.3 Transações e instantâneos .....	3
2.4.4 Agregados.....	3
2.4.5 Tabela de fatos sem fatos.....	3
2.4.6 Dimensões .....	3

2.4.7 Granularidade e Particionamento .....	3
2.4.8 Modelo Snowflake.....	3
<b>2.5 Data Warehouse Distribuído .....</b>	<b>3</b>
<b>3. A APLICAÇÃO .....</b>	<b>3</b>
<b>3.1 Introdução.....</b>	<b>3</b>
<b>3.2 A Cooperativa .....</b>	<b>3</b>
3.2.1 A cooperativa hoje .....	3
3.2.2 A Infra-estrutura Tecnológica.....	3
3.2.3 A Problemática da Centralização da Informação .....	3
<b>3.3 O Desenvolvimento da Aplicação.....</b>	<b>3</b>
3.3.1 Definição dos <i>Data Marts</i> .....	3
3.3.2 Extração, padronização e incorporação dos dados .....	3
3.3.3 Exportação e importação dos dados nos <i>data marts</i> distribuídos.....	3
<b>3.4 A Modelagem do <i>Data Warehouse</i>.....</b>	<b>3</b>
3.4.1 Dimensões .....	3
3.4.2 Fatos .....	3
3.4.3 Agregados.....	3
<b>3.5 Visualização dos dados .....</b>	<b>3</b>
<b>3.6 Resultados Obtidos .....</b>	<b>3</b>
<b>4. CONCLUSÃO E ESTUDOS FUTUROS .....</b>	<b>3</b>
<b>4.1 Conclusão .....</b>	<b>3</b>
<b>4.2 Estudos Futuros.....</b>	<b>3</b>
<b>5. REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>3</b>
<b>APÊNDICE A – Visualização dos dados.....</b>	<b>3</b>



## Lista de Figuras

Figura 1 - Processos básicos de uma estrutura de DW .....	3
Figura 2 - Uma arquitetura multicamadas para o <i>data warehouse</i> .....	3
Figura 3 - Abordagem top down .....	3
Figura 4 - Abordagem bottom up.....	3
Figura 5 - Abordagem incremental (BUS) .....	3
Figura 6 - O cubo de decisão .....	3
Figura 7 - Um modelo dimensional de banco de dados .....	3
Figura 8 - Drill down e Roll up .....	3
Figura 9 - Slice .....	3
Figura 10 - Dice.....	3
Figura 11 - Dimensões comuns em um modelo dimensional.....	3
Figura 12 - Snowflake do tipo pesquisa .....	3
Figura 13 - Snowflake em cadeia.....	3
Figura 14 - Snowflake do tipo atributo.....	3
Figura 15 - Fluxo dos dados nos sistemas operacionais existentes.....	3
Figura 16 - Fases do desenvolvimento de <i>data marts</i> incrementais .....	3
Figura 17 - Fluxo das consultas ao SAAG .....	3
Figura 18 - Fluxo dos dados extraídos para o SAAG.....	3
Figura 19 - Processo de exportação e importação dos dados no SAAG .....	3
Figura 20 - A Dimensão Tempo .....	3
Figura 21 - A Dimensão Filial .....	3
Figura 22 - A Dimensão Produto .....	3
Figura 23 - A Dimensão Grupo.....	3
Figura 24 - A Dimensão Conta .....	3
Figura 25 - A Dimensão Produtor.....	3
Figura 26 - A Dimensão Safra.....	3
Figura 27 - A Dimensão Auditoria .....	3
Figura 28 - Fato Consumo de Lenha.....	3
Figura 29 - Fato Recebimento.....	3
Figura 30 - Fato Orçamento .....	3
Figura 31 - Agregado Recebimento por Grupo de Estoque .....	3

Figura 32 - Tela inicial do SAAG .....	3
Figura 33 - Parametrização das consultas no SAAG .....	3
Figura 34 - Resultados de uma pesquisa no SAAG .....	3
Figura 35 - Exemplo de relatório obtido com o Microsoft Excel.....	3
Figura 36 - Exemplo de gráfico de recebimento gerado com Microsoft Excel.....	3
Figura 37 - Exemplo de gráfico de custos gerado com o Microsoft Excel .....	3

## Lista de Quadros

Quadro 1 - Diferenças entre dados de sistemas operacionais e <i>data warehouse</i> .....	3
Quadro 2 - Vantagens e desvantagens da abordagem <i>top down</i> .....	3
Quadro 3 - Vantagens e desvantagens da abordagem <i>bottom up</i> .....	3
Quadro 4 - Vantagens e desvantagens da abordagem incremental (BUS).....	3

## Lista de Abreviações

ASP	<i>Active Server Pages</i> . Servidor de Páginas Dinâmicas
BD	Banco de Dados
BPS	Bits por segundo
DBA	<i>Data Base Administrator</i> . Administrador de banco de dados
DM	<i>Data Mart</i>
DW	<i>Data Warehouse</i>
EIS	<i>Executive Information Systems</i> . Sistemas de Informações Executivas
JSP	<i>Java Server Pages</i> . Servidor de páginas JAVA
KBPS	Kilobits por segundo
LAN	<i>Local Área Network</i> . Rede local
MBPS	Megabits por segundo
ODBC	<i>Open Database Connectivity</i> . Conectividade Aberta à Banco de Dados
ODS	<i>Operational Data Staging</i> . Área de Estagiamento Operacional
OLEDB	Conjunto de componentes para acesso a banco de dados
OLTP	<i>On-Line Transaction Processing</i> . Processamento de transações on-line
SAD	Sistemas de Apoio a Decisão
SAAG	Sistema de Apoio a Atividade Gerencial
SGBD	Sistema Gerenciador de Banco de Dados
SI	Sistemas de Informações
SIG	Sistemas de Informações Gerenciais
SQL	<i>Structured Query Language</i> . Linguagem Estrutura de Consultas

TCP/IP	Transport Control Protocol / Internet Protocol. Protocolo de controle de transporte / Protocolo Internet. Protocolo padrão de comunicação
TI	Tecnologia de Informações
TPS	<i>Transaction Processing System</i> . Sistema de Processamento de Transação

## 1. INTRODUÇÃO

### 1.1 Apresentação

Nos últimos anos o relacionamento entre as empresas e seus clientes vem tomando uma nova direção. As empresas já não satisfazem seus clientes apenas conhecendo bem seus produtos ou serviços, existe a necessidade, tão importante quanto, de se conhecer também os próprios clientes, suas preferências, seus padrões, suas faixas de valores. Este aumento de interatividade entre empresa e cliente acabou gerando um relacionamento, relacionamento este que, quanto mais duradouro e interativo for, maiores benefícios ambos os envolvidos obterão, ou seja, a empresa terá subsídios para fornecer ao cliente produtos e serviços mais personalizados ou em maior conformidade com as necessidades dos mesmos visto que as mesmas conhecerão melhor sua clientela, seus gostos, valores, entre outros. O cliente por sua vez terá menos incentivo a procurar um novo fornecedor que supra suas necessidades visto que o mesmo se sentirá satisfeito com os serviços prestados pelo atual.

Além da necessidade de conquistar o cliente, as empresas de um modo geral necessitam de subsídios para gerir suas atividades. Subsídios estes que orientam as decisões da alta gerência de modo que as empresas possam se tornar ainda mais competitivas e conseqüentemente oferecerem um maior valor agregado em seus produtos e/ou serviços satisfazendo ainda mais os desejos de seus clientes que, referindo-se a cooperativas, são partes diretamente interessadas por fazerem parte da sociedade da empresa.

Esse novo enfoque comercial que está sendo dado ao relacionamento entre empresa e cliente aumentou a quantidade de informações geradas nas transações operacionais das organizações. As informações começaram a ser geradas e conseqüentemente armazenadas, formando grandes depósitos de dados conhecidos como *Data Warehouse* (DW), e que são considerados por muitos como sendo a memória da empresa.

O DW surgiu no mercado com o intuito de organizar os dados corporativos da melhor maneira de modo a dar subsídio de informações aos gerentes e diretores das empresas para as decisões de nível tático-estratégicas, através da utilização de um banco de dados paralelo aos sistemas operacionais da empresa. A questão chave para o sucesso destes sistemas está no correto entendimento dos negócios da organização e, principalmente, dos seus níveis decisórios.

Diretamente ligada à importância da organização dos dados corporativos, está a correta e eficiente disponibilização das informações geradas. Muitas organizações, principalmente às que possuem unidades remotas, possuem infra-estruturas de teleprocessamento um tanto quanto deficitárias e isto implica, muitas vezes, na não implantação de um sistema de DW ou na implantação de um modelo de DW distribuído de modo a levar a informação a estas unidades remotas.

A heterogeneidade de ambientes, multiplicidade de sistemas e tecnologias de transmissão de dados de baixa velocidade tem dificultado a utilização desta tecnologia nos níveis decisórios das empresas e em suas filiais e coligadas. O resultado disso são informações muitas vezes inconsistentes e imprecisas, geradas sobre planilhas e outros documentos não conexos e que geralmente possuem pequenas diferenças, irritantes e perigosas, principalmente quando utilizadas como subsídio para a tomada de decisões.

Neste trabalho, o objetivo é o desenvolvimento e a implantação de um *Data Warehouse* Corporativo com *Data Marts* Distribuídos de modo a fornecer o apoio necessário para que a empresa possa gerir melhor suas atividades frente aos desafios mercadológicos, melhorando cada vez mais o relacionamento com seus clientes e parceiros através de um contínuo aprimoramento de seus processos e métodos de produção e análise dos resultados, ou seja, uma gestão de qualidade mais eficiente e inovadora que acarrete em um crescimento competitivo através do aperfeiçoamento de seus produtos ou serviços oferecidos.

## 1.2 Objetivos

### 1.2.1 Objetivo geral

O objetivo deste trabalho é o desenvolvimento e a implantação de um *Data Warehouse* Corporativo com *Data Marts* Distribuídos em uma Cooperativa Agroindustrial – C.Vale.

### 1.2.2 Objetivos específicos

Os objetivos específicos do presente trabalho de pesquisa surgem do estudo das características das diferentes bases de dados dos Sistemas de Informação Legados e seus meios de transmissão, buscando:

- Estudar a contribuição das novas tecnologias e modelos de dados para suporte e evolução das bases de dados legadas para as novas plataformas tecnológicas;
- Estudar as características das diversas bases de dados dos sistemas legados existentes através de um estudo de caso;
- Identificar problemas e benefícios com relação à implementação da tecnologia de *data warehousing*;
- Definir um repositório de dados históricos para auxiliar os gerentes, contendo informações relevantes para a tomada de decisões;
- Extrair informações de diferentes origens, integrá-las e armazená-las em um repositório único central de *Data Warehouse*;
- Definir os níveis de granularidades necessários para a distribuição dos dados dentro deste modelo dimensional distribuído;



- Determinar os métodos e técnicas de transmissão e incorporação dos novos dados ao repositório já existente em cada unidade remota, de modo que não haja consultas a serem executadas no repositório central;
- Repassar à sociedade os conhecimentos e resultados obtidos nesta pesquisa.

### **1.3 Justificativa**

As empresas de médio e grande porte em sua maioria possuem uma base de sistemas legados que lhes oferece um certo nível, ou porque não dizer, um bom nível de informatização organizacional. Estes sistemas, quando de seu desenvolvimento, possuíam como objetivos a automação de processos e procedimentos que manipulassem grande quantidade de dados, visando à eliminação de inconsistências e re-trabalho. Também estes, muitas vezes, foram desenvolvidos sobre tecnologias que não permitem a retirada de informações de caráter tático-estratégicas nem a sua eficaz disponibilização.

Diante disso, muitas empresas possuem seus dados armazenados em sistemas extremamente robustos, porém não flexíveis, o que acaba tornando estas empresas lentas e despreparadas perante seus concorrentes. A expectativa desta integração de informações pautadas sobre os diversos sistemas legados existentes, bem como um meio de disponibilizar estes dados em toda a abrangência física da empresa de forma coerente mediante os recursos existentes, faz com que muitas empresas sonhem com uma possível “superioridade esmagadora” frente a sua concorrência.

### **1.4 Estrutura do Trabalho**

O primeiro capítulo traz uma visão geral do trabalho, seus objetivos, justificativas e estrutura.

No segundo capítulo, é apresentado um estudo sobre os diversos métodos e tecnologias existentes para elaboração de uma base de dados com caráter estratégico, apresentando-se as tecnologias de Sistemas de Informações (SI), Sistemas de Informações Gerenciais (SIG), Sistemas de Apoio a Decisão (SAD), Sistemas de Informações Executivas (EIS), métodos e tecnologias para a implementação e utilização de *Data Marts (DM)*, *Data Warehouse* e *Data Warehouse* distribuídos, sendo abordados conceitos, características e demais etapas para a implementação no modelo proposto.

No terceiro capítulo, tem-se um estudo sobre a cooperativa agroindustrial objeto de pesquisa do estudo de caso – C.Vale, a proposta de um modelo dimensional de *data warehouse* corporativo com *data marts* distribuídos e a apresentação e análise dos resultados após a implantação do modelo em duas unidades mais o repositório central.

O quarto capítulo trata das conclusões obtidas no desenvolvimento do trabalho, bem como faz referências e algumas recomendações para trabalhos futuros.

Finalmente é registrada a bibliografia utilizada neste trabalho bem como é apresentado um anexo contendo algumas figuras que representam as visualizações dos dados extraídas do sistema proposto.

## **2. TECNOLOGIA DE SISTEMAS DE INFORMAÇÃO**

### **2.1 Introdução**

Atualmente estamos inseridos num cenário de rápidas e constantes mudanças, onde a concorrência se torna cada vez mais acirrada. O consumidor torna-se cada vez mais exigente em relação ao custo e a qualidade do produto e dos serviços associados. O mercado agora é global e mais sensível, sendo as alterações no mesmo mais rápidas e representativas resultando, como imperativa, a importância da informação como instrumento de impacto decisivo nas perspectivas de rentabilidade e competitividade das empresas. Dentro deste cenário, a tecnologia da informação assume um papel de suma importância, ao permitir, de forma rápida e simples, a extração, organização, análise e circulação de informações necessárias a todos os níveis da empresa, em suporte aos objetivos estratégicos.

Mão-de-obra, matéria-prima, máquinas, etc., normalmente são considerados como recursos de uma empresa. Contudo, a informação também é um recurso essencial para os processos de planejamento, organização e controle de uma atividade de negócio.

Diariamente, dados sobre os mais variados aspectos dos negócios da empresa são gerados e armazenados, e passam a fazer parte dos recursos de informação das empresas, entretanto, essas informações, encontram-se em geral espalhadas em diversos sistemas e exigem um esforço considerável de integração para que possam dar suporte efetivo à tomada de decisão de gerentes e executivos.

Diante deste cenário encontram-se os sistemas de informações que possuem além da tarefa de automatizar processos e eliminar re-trabalhos, a função de capturar, organizar e analisar os dados relevantes para manter a vantagem competitiva das empresas.

## 2.2 Sistemas de Informação

Um sistema de informação é um conjunto interdependente de pessoas, estruturas organizacionais, software, hardware, processos e métodos interligados com o objetivo de facilitar o planejamento e o controle em empresas e outras organizações, organizando informações de forma que estas se tornem utilizáveis na coordenação do fluxo de trabalho de uma empresa (LAUDON e LAUDON, 1998).

Os sistemas de informação têm a função de coletar, manipular, armazenar e disseminar dados e informações pela empresa (STAIR, 1998), podendo estes, dados ou informações, serem pertinentes a assuntos internos à empresa ou relacionados ao ambiente que esta está inserida.

### 2.2.1 Sistemas de processamento de transações

Os sistemas de processamento de transações (TPS), representam a aplicação dos conceitos e tecnologia de informação em transações rotineiras, repetitivas e geralmente comuns de negócios (STAIR, 1998). Uma transação pode ser entendida como um evento que ocorre num negócio tal como compra, venda, pagamento, entre outros.

Os sistemas operacionais, sistemas de apoio às operações ou OLTP, como também são conhecidos os TPS, foram os primeiros aplicativos de computador a serem desenvolvidos na grande maioria das empresas. Eles têm a tarefa de monitorar e processar as funções básicas e rotineiras de uma organização, tais como processamento de folha de pagamento, faturamento, entre outros (STAIR, 1998, LAUDON e LAUDON, 1998).

Os sistemas de processamento de transações possuem foco no nível operacional da empresa, armazenando e processando fluxos de dados pertinentes a automação de processos de modo que estes sejam mais planejados e otimizados, visando garantir a integração e normalização, possuindo, em sua grande maioria, alguns relatórios para gerenciamento.

### 2.2.2 Sistemas de informações gerenciais

Sistema de Informação Gerencial é o processo de transformação de dados em informações que serão utilizadas na estrutura decisória da empresa, bem como proporcionarão a sustentação administrativa para otimizar os resultados esperados (OLIVEIRA, 1998).

Estes sistemas têm o propósito de fornecer informações necessárias à medição da eficiência operacional da organização, dando ênfase às necessidades gerenciais através de informações resumidas, obtidas a partir da filtragem e análise de dados altamente detalhados, extraídos das bases de dados dos sistemas de processamento transacionais e de fontes externas (STAIR, 1998, FALSARELLA e CHAVES, 2001).

O SIG, em contrapartida aos sistemas de informações operacionais, possui foco sobre a informação, direcionando a gerentes de nível médio (tático). Tais sistemas possuem fluxo estruturado de informações e geralmente possuem integração com as atividades dos TPS por função organizacional / comercial tais como SIG para produção, SIG para pessoal, entre outros, gerando e trabalhando com um grande número de relatórios e consultas gerenciais.

### 2.2.3 Sistemas de apoio à decisão

Os Sistemas de Apoio à Decisão, são sistemas que realizam o processamento analítico e provêm as informações necessárias ao usuário, permitindo a análise de situações e a tomada de decisões (INMON, 1997).

Os sistemas de apoio à decisão são uma ferramenta de produtividade pessoal e eficiência organizacional. Seu sucesso depende de uma série de fatores, entre eles, dois fundamentais são o treinamento / educação e a cultura organizacional (MEIRELLES, 1994).

Os processamentos analíticos dos SAD permitem ao usuário analisar uma grande quantidade de dados, normalmente históricos, verificando problemas e situações, de modo a identificar perfis, tendências e padrões, sendo a performance das consultas ou extrações de dados importantes, porém não crucial no desenvolvimento deste tipo de aplicação.

A arquitetura de um SAD engloba um planejamento de hardware, software e interface que atenda as necessidades dos usuários, as possibilidades da empresa, as expectativas da área de TI e se encaixem na cultura da empresa, portanto, os processos relacionados à análise, extração e armazenamento da base de dados, bem como o formato como essas informações serão disponibilizadas aos usuários de modo que o mesmo possa fazer o melhor uso da informação ali contida deve ser alvo de constantes análises para avaliar se os resultados obtidos com o sistema estão sendo satisfatórios.

Algumas das principais características de um SAD são o fato de serem mais utilizados na resolução de problemas de maior complexidade, problemas estes que são mais comuns no cotidiano da alta administração, deve possuir uma interface amigável para que o usuário (neste caso a alta gerência) não perca o interesse em utilizar a ferramenta devido à complexidade das consultas, devem ser flexíveis o bastante para se adaptar a mudanças que venham a ocorrer em um futuro próximo, devem fornecer subsídios para a implementação dos resultados obtidos e, não podem ficar limitados somente aos executivos, mas sim, devem ser aplicados a todos os níveis de gerenciamento da empresa.

Os bancos de dados originados por este tipo de sistemas, bancos de dados analíticos, possuem ainda a grande vantagem de poderem se tornar integradores das informações provenientes dos diversos sistemas operacionais, possibilitando uma visão global de toda a organização.

Um dos conceitos mais conhecidos dentro dos SAD é o conceito de *Data Warehouse*. Esse consiste em organizar os dados corporativos da melhor maneira, para dar subsídio de informações dos gerentes e diretores das empresas para as

decisões de nível tático-estratégicas. Tudo isso num banco de dados paralelo aos sistemas operacionais da empresa.

O DW é o ponto central da arquitetura de processamento de informações para sistemas de informática modernos suportando o processamento informacional de um SAD através de um alicerce sólido de integração de dados corporativos e históricos para a realização de análises gerenciais (INMON e HACKATHORN, 1997).

#### 2.2.4 Sistemas de informação executiva

Os Sistemas de Informação executiva, são um tipo especial de SAD destinado à tomada de decisões de alto nível da organização, oferecendo informações estruturadas, tanto internas quanto externas a respeito de aspectos da organização considerados fatores críticos de sucesso para a mesma (STAIR, 1998, FALSARELLA e CHAVES, 2001).

Enquanto um SAD oferece apoio para decisões focadas em uma área específica, geralmente oferecendo suporte à média e baixa gerência, um EIS oferece informações consolidadas de alto nível e análises multidimensionais a executivos de alto nível (GUPTA, 2001).

### 2.3 Data Warehousing

As organizações procuram ativamente tornarem-se mais competitivas e rentáveis. Para obter vantagem competitiva, essas companhias precisam acelerar o processo de tomada de decisão, devendo, para isso, reagir rapidamente às modificações do ambiente, normalmente através da análise, planejamento e execução de ações táticas ou estratégicas adequadas. Tais empresas se vêem em posição de ter que deixar de lado modelos fixos de gestão para se concentrar em conhecer melhor o seu negócio, ou seja, seus produtos e clientes.

Um dos pontos chave para acelerar a tomada de decisões é ter informações corretas, no momento oportuno e facilmente acessíveis. Muitas organizações após analisarem seus sistemas transacionais verificam que estes não estão prontos para atender essa necessidade de informação devido às suas características inerentes: não guardam históricos, são modelados e otimizados para fazer acesso registro a registro de forma que não estão capacitados a recuperar e consolidar grandes volumes de dados de forma eficiente, e, de uma maneira geral, não estão bem integrados entre si. Na grande maioria das vezes esses sistemas estão espalhados dentro da companhia em múltiplas e independentes plataformas e localizações, criando dificuldades adicionais de integração e análise dos dados, tornando o acesso às informações oportunas uma atividade difícil, por vezes impossível.

Devido a esta necessidade de gerar sistemas que pudessem dar suporte a decisões gerenciais, surgiu um conceito que vem sendo desenvolvido e popularizado: *Data Warehouse*. O DW é um conceito que se preocupa em extrair, integrar, limpar e dar consistência a dados provenientes tanto de sistemas operacionais da companhia quanto de dados externos tais como pesquisas de mercado e dados sobre a concorrência, ou seja, procura integrar e consolidar dados disponíveis em diferentes acervos ampliando o conteúdo informacional destes para atender às expectativas e necessidades de nível estratégico da empresa. Além disso, o DW dimensionaliza e consolida esses dados, organizado-os de forma a melhorar a performance das consultas.

A implementação de um *Data Warehouse* deve ser adaptada às características e às expectativas de cada empresa, mas o principal objetivo em todas elas é o de descobrir maneiras diferentes de atuar no mercado e quais as mudanças internas que devem ocorrer para atender as novas realidades.

### 2.3.1 Dados Operacionais Versus Dados Informacionais

É importante salientar as diferenças existentes entre os dados gerados, manipulados e armazenados em sistemas operacionais dos dados que serão



manipulados em *data warehouses*. O Quadro 1 descreve as principais diferenças entre os tipos de informações.

Quadro 1 - Diferenças entre dados de sistemas operacionais e *data warehouse*

Dados Operacionais	Dados em um <i>Data Warehouse</i>
Baseados em aplicações	Baseados em assuntos ou negócios
Detalhados	Resumidos ou refinados
Exatos em relação do momento de acesso	Representam valores de momentos já decorridos ou instantâneos
Acessados uma unidade por vez	Acessados um conjunto por vez
Voltados para transações	Voltados para análise
Alta disponibilidade	Disponibilidade atenuada
Não contemplam a redundância	A redundância não pode ser ignorada
Estrutura fixa: conteúdos variáveis	Estrutura flexível
Pequena quantidade de dados usada em um processo	Grande quantidade de dados usada em um processo
Atendem às necessidades cotidianas	Atendem às necessidades gerenciais
Alta probabilidade de acesso	Baixa, ou modesta probabilidade de acesso

Fonte: SELL, D. Uma arquitetura para distribuição de componentes tecnológicos para sistemas de informações baseados em *data warehouse*. Florianópolis, p. 26, 2001. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

### 2.3.2 Data Warehouse

Segundo Inmon (1997), *Data Warehouse* é uma coleção de dados orientados a assuntos, integrados, variáveis com o tempo e não voláteis para suporte ao processo gerencial de tomada de decisão.

Singh (2001), resume *Data Warehouse* como sendo uma tecnologia de gestão e análise de dados.

Para Cruz (2000), DW é um conjunto de *hardware* e *software* que formam o lugar onde as pessoas podem acessar dados estratificados e consolidados de forma consistente e rápida, a fim de evitar buscas redundantes e dispersivas pelos diversos repositórios genéricos existentes na organização.

De acordo com Inmon e Hackathorn (1997), o objetivo de um DW é fornecer uma imagem única da realidade do negócio. De uma forma geral, sistemas de DW compreendem um conjunto de programas que extraem dados do ambiente operacional da empresa, um banco de dados que os mantém, e sistemas que fornecem estes dados aos seus usuários.

Segundo Kimball (1998a), o DW fornece acesso a dados corporativos ou organizacionais, seus dados são consistentes podendo ser separados e combinados usando-se qualquer medição possível no negócio, um DW não consiste apenas em dados, mas em um conjunto de ferramentas para consultar, analisar e apresentar informações, é um local onde se publica dados confiáveis sendo a qualidade desses um impulso à reengenharia de negócios.

Segundo Machado (2000), construir um DW é construir armazéns de dados onde a história da empresa, seus clientes, fornecedores e operações se mantêm disponíveis e acessíveis para consultas e análises.

Logo, o DW caracteriza-se como um banco de dados contendo dados extraídos do ambiente operacional da empresa, que foram selecionados e depurados, tendo sido otimizados para processamento de consultas e não para processamento de transações. Em geral, um DW requer a consolidação de outros recursos de dados além dos armazenados em BDs relacionais, incluindo informações provenientes de planilhas eletrônicas, documentos textuais, entre outros.

Em decorrência disso tem-se dois ambientes: o operacional, onde as transações relacionadas ao dia-a-dia da empresa acontecem (busca-se otimizar ao máximo a performance de acesso às transações) e outro, onde após uma extração, uma transformação e uma carga desses dados oriundos dos sistemas operacionais,

formam um banco de dados histórico (priorizando-se a qualidade da informação em tempo hábil) que permite a análise dos dados, sendo este, um banco de dados somente para consulta e, em sua grande maioria, não é permitida alteração do seu conteúdo. O fato de existirem dois sistemas rodando em paralelo garante que operações efetuadas em um dos sistemas não afeta a performance do outro e vice-versa.

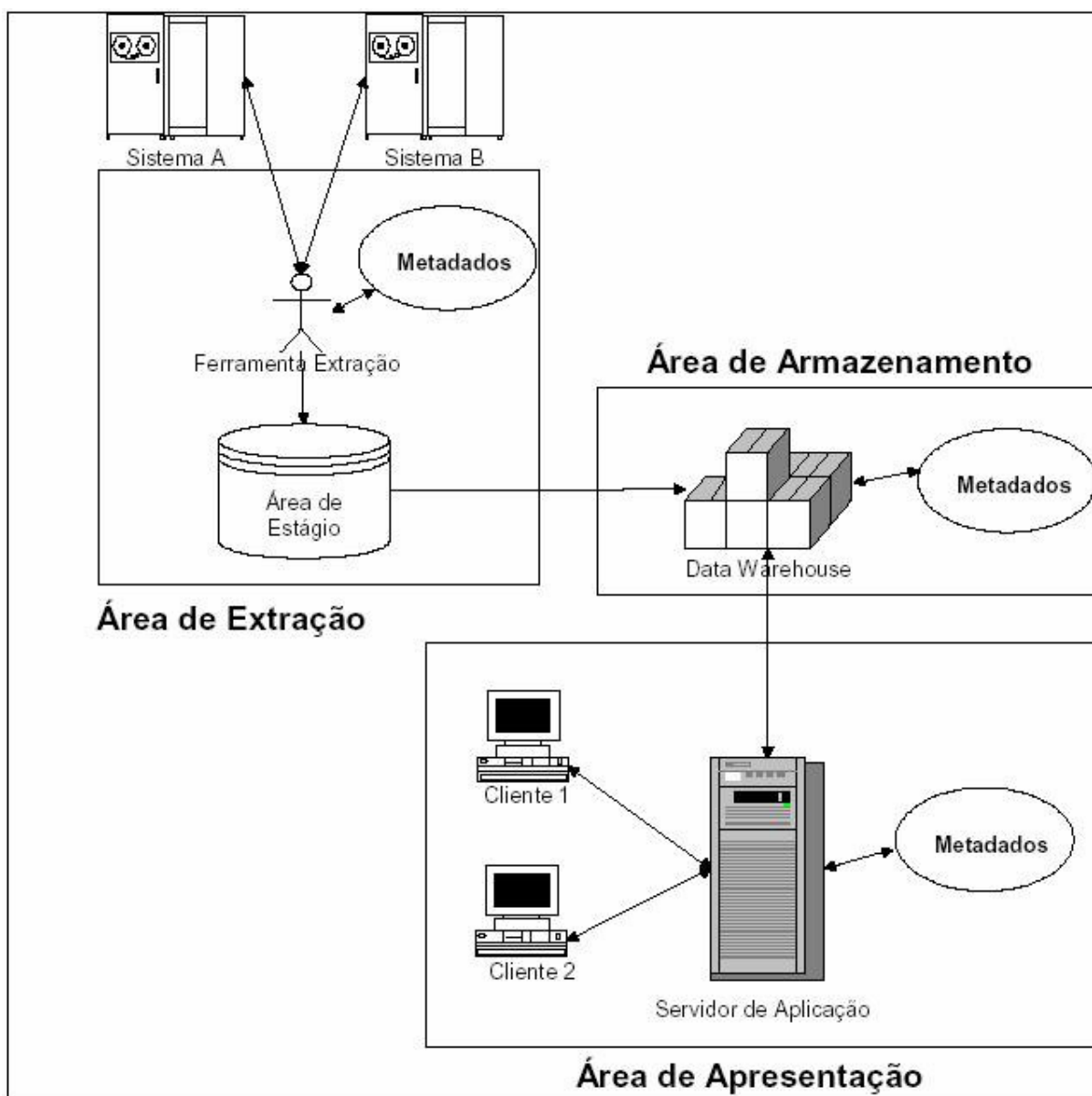
Observa-se através destas definições que o *Data Warehouse* é, portanto, um ambiente e não um produto como muitas empresas acreditam e outras vendem.

Basicamente, o DW pode ser dividido em três processos básicos: extração de dados dos sistemas operacionais, armazenamento dos dados e apresentação de informações. Estes três processos podem ser visualizados na Figura 1.

Sell (2001) afirma que:

- a extração de dados dos sistemas operacionais consiste da concepção ou aquisição e parametrização das ferramentas que realizarão as tarefas de coleta, limpeza, transformação e migração dos dados operacionais ao *data warehouse*. Estas tarefas, quando da sua realização, constituem um dos processos mais morosos e delicados no *data warehousing*.
- o armazenamento dos dados consiste da concepção do repositório das informações, sendo este o núcleo do ambiente de *data warehouse*. Neste estarão representados todos os dados extraídos dos sistemas operacionais, necessários para o processo de tomada de decisão.
- a apresentação de informações consiste da concepção ou aquisição e parametrização das aplicações clientes e do servidor que atenderá às requisições de dados junto ao *data warehouse* e disponibilizará as informações resultantes.

Figura 1 - Processos básicos de uma estrutura de DW



Fonte: SELL, D. Uma arquitetura para distribuição de componentes tecnológicos para sistemas de informações baseados em data warehouse. Florianópolis, p. 45, 2001. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

O DW possui uma estrutura distinta. Existem diferentes níveis de sintetização e detalhe que o demarcam e, adicionalmente, diferentes níveis de idade dos fatos (INMON e HACKATHORN, 1997). Além de possuir os dados, estando eles sintetizados ou não, o DW também se responsabiliza em armazenar informações sobre os dados que ele contém que são os Metadados, a principal função desses metadados é fornecer uma documentação imprescindível ao projeto de Data

Warehouse que juntamente com outros fatores pode representar o sucesso ou o fracasso deste projeto.

Vale a pena salientar que, como foi mencionado, um DW não contém apenas dados resumidos. É desejável prover ao usuário a capacidade de aprofundar-se num determinado tópico, investigando níveis de agregação menores ou mesmo o dado primitivo, permitindo também a geração de novas agregações ou correlações com outras variáveis de negócio. Limitar o conteúdo de um DW apenas a dados resumidos significa limitar os usuários apenas às consultas e análises que eles puderem antecipar frente a seus requisitos atuais, não deixando qualquer flexibilidade para novas necessidades.

### 2.3.3 Características do *Data Warehouse*

Segundo Machado (2000), o DW possui um conjunto de características que os distinguem dos sistemas convencionais. São elas:

- Extração de dados de fontes heterogêneas (existentes ou externas),
- Transformação e integração dos dados antes de sua carga,
- Normalmente requer máquina e suporte próprio,
- Visualização dos dados em diferentes níveis. Os dados do *Data Warehouse* podem ou não ser extraídos para um nível mais específico, os *Data Marts*, e a partir destes, para um banco de dados individual,
- Utilização de ferramentas voltadas para acesso com diferentes níveis de apresentação,
- Dados somente são inseridos, não existindo atualização, ou melhor, *updates*.

Segundo Inmon (1997), além das características acima, os DW também possuem orientação por assunto, integração, variação no tempo, não volatilidade e

localização. Vêm-se a seguir maiores detalhes sobre cada uma dessas características.

#### a) Orientação por assunto

Armazena as informações agrupadas por assuntos de interesse da empresa que são mais importantes, em contraste com os sistemas operacionais que são orientados a processos desenvolvidos para manter as transações realizadas diariamente (MACHADO, 2000).

O mundo dos sistemas operacionais é projetado em torno de aplicações e funções relacionadas com as atividades diárias da empresa como contas a pagar, vendas, controle de estoque, etc. O mundo do DW é organizado em torno dos principais assuntos da organização como faturamento, clientes, vendas, entre outros. O alinhamento do DW em torno destas áreas da corporação afeta o projeto e a implementação dos dados encontrados no mesmo. Mais importante ainda, os principais assuntos da corporação influenciam a estrutura-chave e a organização dos dados não-chave em torno desta (INMON e HACKATHORN, 1997).

#### b) Integração

Constitui uma das principais características do DW, segundo a qual se define a representação única para os dados oriundos dos diversos sistemas que irão compor a base de dados do *Data Warehouse*.

Segundo Kimball (1998b), dois elementos básicos do DW estão relacionados com a integração: à área de estagiamento de dados e o armazenamento de dados operacionais. O processo de limpeza, transformação e agregação ocorrem no estagiamento, enquanto compatibilização e integração nos próprios sistemas legados ocorrem no ODS.

Um clássico exemplo de integração que pode ser dado é a representação do estado civil de uma pessoa. Muitos sistemas tratam como: "S" solteiro, "C" casado e assim por diante. Outros sistemas por sua vez tratam como: 1 solteiro, 2

casado e consecutivamente. Para resolver este problema estabelece-se um padrão e dentro do DW utiliza-se somente uma das formas, ou seja, a forma padronizada.

### c) Variação no Tempo

Segundo Inmon (1997), todos os dados no DW são precisos em algum instante no tempo. Como eles podem estar corretos somente em um determinado momento, é dito que esses dados variam com o tempo.

Os dados de um DW são precisos em relação ao tempo e representam resultados operacionais em determinado momento de tempo, o momento em que foram capturados. Isto implica em que os dados de um DW não possam ser atualizados (MACHADO, 2000).

Deve-se considerar que não apenas os dados têm uma característica temporal, mas também os metadados, que incluem definições dos itens de dados, rotinas de validação, algoritmos de derivação, entre outros. Sem a manutenção do histórico dos metadados, as mudanças das regras de negócio que afetam os dados no DW são perdidas, invalidando os dados históricos.

Os dados armazenados podem ser classificados em duas categorias: dados detalhados atuais, que são os de maior interesse por refletir os acontecimentos mais recentes, são em grande volume por serem armazenados em menor nível de granularidade, permitem a utilização de técnicas como o *data mining* e descoberta de conhecimento, sendo que o horizonte de tempo para esse tipo de dados normalmente é de dois anos. A segunda categoria são os dados detalhados antigos, não possuem frequência de acesso e em decorrência disso muitas vezes são extraídos para outros meios de armazenamento, porém continuam fazendo parte do DW e sempre que necessário podem ser carregados.

Outro parâmetro importante a ser definido é o período de atualização dos dados oriundos dos sistemas operacionais para o DW. O problema é definir de quanto em quanto tempo esta atualização deve ocorrer. Segundo Inmon (1997), 24 horas deve-se passar entre o momento em que a alteração é observada pelo

ambiente operacional e sua repercussão no DW. O correto estabelecimento deste período evita problemas de informações incorretas no DW, pois os dados podem ainda não estar estáveis no ambiente operacional além de diminuir a complexidade da tecnologia envolvida na replicação.

#### d) Não volatilidade

Os dados originados nos sistemas operacionais que serão carregados para o DW geralmente passam por um processo de filtragem e transformação para se adequarem às necessidades de informações antes de serem propriamente carregados. Depois de efetuada esta carga (inicial e/ou incremental) os dados ficam disponíveis para consultas a serem realizadas pelos usuários, não sendo permitido a estes, alteração no conteúdo dos mesmos.

No ambiente operacional, ao contrário, os dados são em geral atualizados registro a registro, em múltiplas transações. Esta volatilidade requer um trabalho considerável para assegurar integridade e consistência através de atividades de *rollback*, recuperação de falhas, *commits* e bloqueios. Um DW não requer este grau de controle típico dos sistemas orientados a transações (CAMPUS, 2001).

Kimball e Ross (2002) sobre atualizações nos DW afirmam que os DW modernos também podem ser atualizados, mas, em geral, essas são atualizações de carga gerenciada e não atualizações transacionais.

#### e) Localização

Em um *data warehouse*, pode-se encontrar os dados armazenados fisicamente de três formas:

- armazenados em um único local, centralizando em um DW integrado onde se procura maximizar o poder de processamento e busca dos dados;



- distribuídos por áreas de interesse, também conhecida como estrutura federativa, onde se separa os dados contábeis em um servidor, dados financeiros em outro e assim por diante,
- armazenados por níveis de detalhes em que as unidades de dados são mantidas no DW. Os dados em um DW podem estar dispostos em diferentes níveis de detalhamento/resumo, ou seja, dados altamente resumidos podem ser armazenados em um servidor, dados atômicos podem ser armazenados em outros e níveis intermediários em outro, sempre lembrando que cada servidor poderá (ou porque não dizer deverá) ser parametrizado para otimizar ao máximo as consultas destinadas ao seu processamento.

#### 2.3.4 Data Mart

Para Inmon (1997), um *Data Mart* pode ser definido como um SGBD multidimensional que fornece uma estrutura bastante flexível de acesso a dados. Enquanto o DW extrai, transforma e limpa os dados dos sistemas transacionais, mantendo-os integrados em quantidades massivas e em seu nível mais baixo, o DM se serve destes dados, extraíndo dados para um departamento ou uma área de negócio, oferecendo flexibilidade e controle ao usuário final, pois com o DM é possível fatiar e agrupar dados de diversas maneiras.

Para Machado (2000) e Kimball (1998b), os dados do *Data Mart* são direcionados a um departamento ou a uma área específica do negócio e representam um subconjunto do DW corporativo.

Para Kimball (1998b), o conjunto de todos os *data marts* da organização, construídos de forma incremental, compartilhando dimensões e fatos comuns, segundo um planejamento prévio, formam o *data warehouse* lógico da organização.

O DM muitas vezes é visto como uma alternativa ao DW, pois custa menos e leva menos tempo para ser projetado e implementado. É criado para um grupo dirigido de usuários, normalmente um setor da empresa.

### 2.3.5 Arquitetura do *Data Warehouse*

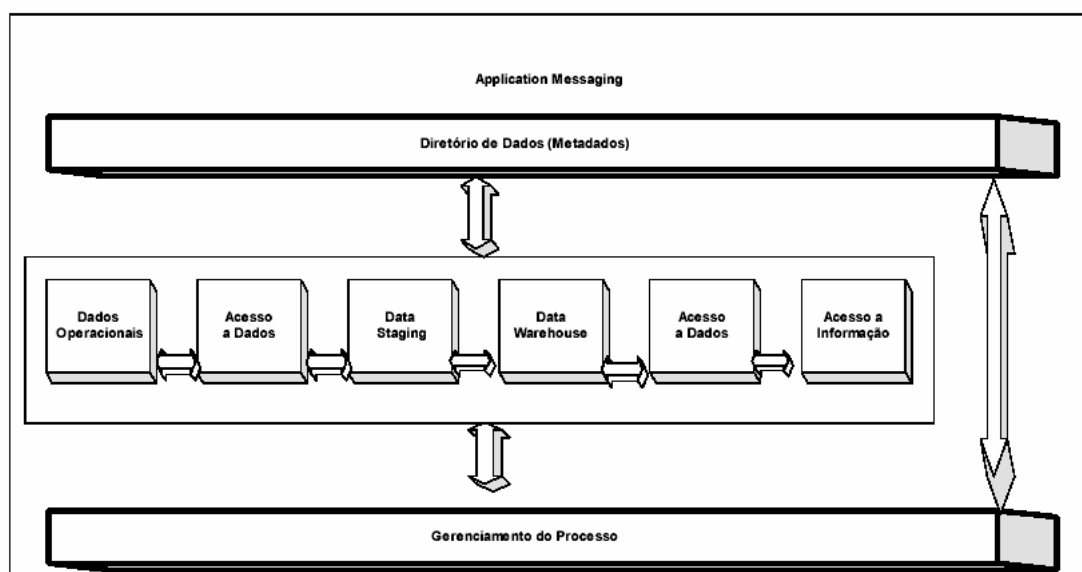
Uma arquitetura de *data warehouse* pode ser definida como a forma de representar toda a estrutura do ambiente de dados, comunicação, processamento e apresentação disponível para o usuário na empresa (SINGH, 2001).

Esta arquitetura é composta por algumas camadas interconectadas:

- Camada de acesso à informação
- Camada de acesso aos dados
- Camada de diretório de dados (Metadados)
- Camada de gerenciamento do processo
- Camada *Application Messaging*
- Camada de dados operacionais
- Camada DW (Física)
- Camada *Data Staging*

Estas camadas estão representadas na Figura 2.

Figura 2 - Uma arquitetura multicamadas para o *data warehouse*



Fonte: MELLO, J. A. B. Uma proposta de Modelo de Dados para Suporte ao Processamento Transacional e de Apoio Informacional Simultaneamente. Florianópolis, p. 31, 2002. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

a) Camada de acesso à informação

É a camada com a qual os usuários finais interagem. Representa as ferramentas que o usuário utiliza no dia a dia para acessar as informações contidas em um DW, envolve hardware e software que são utilizados na obtenção de relatórios, planilhas, gráficos entre outros. Nesta camada também estão inseridas as ferramentas de mineração de dados (*data mining*) e análises dimensionais, as quais fornecem ao usuário maior flexibilidade ao analisar um grande volume de dados.

b) Camada de acesso aos dados

A camada de acesso a dados é responsável por fazer a conexão entre a camada de dados operacionais e a camada de *data staging* sendo também responsável pela conexão entre a camada de DW e a camada de acesso à informação.

Esta camada faz uso extensivo da linguagem SQL e através de *drivers* apropriados, se comunica praticamente com qualquer SGBD e sistema de arquivos, permitindo assim o processo de extração de dados dos sistemas operacionais existentes.

Esta camada é composta por diferentes SGBDs, sistemas operacionais, hardware, protocolos de transmissão entre outros, com o objetivo de proporcionar acesso transparente, não importando a localização ou a plataforma em uso.

c) Camada de diretório de dados (Metadados)

Os metadados ou dicionário de dados como também são conhecidos, são os dados sobre os dados armazenados, ou seja, descrevem a estrutura do DW através do fornecimento de informações sobre certos dados do ambiente como fontes de dados, transformações sobre os dados, visões do usuário, autorização de acesso, estrutura dos dados entre outros.

Esta camada permite que o usuário do DW acesse seus dados de forma única, sem que seja necessário saber onde eles estão e em que formato estão

armazenados. Um exemplo de metadado pode ser uma sentença *CREATE* em SQL ou dados em um dicionário de dados.

O conceito de metadados não é novo, porém a sua importância no ambiente de DW é. Sem metadados, o DW e seus componentes associados no ambiente projetado são meramente componentes soltos, funcionando independentemente e com objetivos separados (INMON et. Al, 1999).

Um problema com os metadados é a padronização dos mesmos. Existem diversos fornecedores de soluções de *data warehouse*, cada um com sua arquitetura proprietária, o que obriga os profissionais do corpo técnico a manter uma documentação própria, a fim de gerenciar o *data warehouse* e fornecer os metadados de negócio de forma padronizada ao usuário final (KIMBALL, 1998a).

Segundo Brackett (1996) *apud* Sell (2001), existem basicamente dois tipos de metadados: os metadados técnicos, utilizados pelo corpo técnico (DBAs e desenvolvedores) para manutenção do *data warehouse*, e os metadados de negócio, utilizados pelos usuários finais, que fornecem uma descrição do negócio, regras e cálculos de formação dos dados e outros itens que auxiliem na interpretação dos dados e informações sobre frequência de atualização dos mesmos.

Segundo Mello (2002), em um ambiente de DW, o metadado é utilizado para fornecer informações sobre a localização do conteúdo do DW, mapeando os dados à medida que os mesmos são transformados do ambiente operacional para o ambiente de *data warehouse*, informações a respeito dos algoritmos utilizados na sumarização dos dados, informações sobre as estruturas dos dados, histórico sobre a extração e transformação dos dados e estatísticas sobre a utilização dos mesmos.

#### d) Camada de gerenciamento de processo

A camada organiza os diversos processos que atuam em um DW a fim de mantê-lo atualizado. Envolve todas as tarefas necessárias à construção e manutenção do *data warehouse*, de modo a garantir a consistência do mesmo.

e) Camada *application messaging*

Esta camada também é conhecida como camada de transporte, camada *middleware* ou camada de troca de mensagens entre aplicações. É responsável pelo transporte de informações do *data warehouse* pela rede da organização, podendo envolver outros mecanismos além de protocolos de rede.

A troca de mensagens pode ser utilizada, por exemplo, para isolar as aplicações, operacionais ou informativas, da complexidade dos formatos reais dos dados nas duas extremidades, podendo ser utilizada também na coleta de transações ou mensagens a serem processadas.

f) Camada de dados operacionais

Corresponde aos dados das bases de dados operacionais da organização junto com dados provenientes de outras fontes externas que serão tratados e integrados para compor o *data warehouse*.

g) Camada *data warehouse* (física)

Esta camada corresponde aos dados usados para fins "informativos", ou seja, o *data warehouse* propriamente dito. O *data warehouse* pode ser apenas uma visualização lógica ou uma cópia dos dados operacionais e dados externos livres de inconsistências e em um formato que proporcione um acesso rápido e flexível.

h) Camada *data staging*

Esta camada inclui todos os processos necessários à extração, transformação e limpeza dos dados antes que estes passem para o *data warehouse* físico. É uma camada onde os dados permanecem até serem limpos e padronizados antes do processo de carga no DW. Normalmente isto envolve programação complexa, mas cada vez mais estão sendo disponibilizadas ferramentas para facilitar

estes processos. Esta camada pode também envolver programas de análise da qualidade dos dados e filtros que identificam padrões nos dados operacionais.

### 2.3.6 Topologias de Implementação

Existem várias topologias que podem ser consideradas na implementação de um *Data Warehouse*. Estas incluem:

- Um nível (*One-tier*), consiste em uma arquitetura monolítica para o DW, onde o próprio DW, as ferramentas de análises e buscas e o próprio cliente de acesso residem sobre a mesma plataforma de hardware. Singh (2001) afirma que essa configuração é ideal para pequenas soluções que necessitem de recursos de suporte à decisão. É possível incluir todos os componentes da solução em um laptop e funcionar com eficiência mesmo sem um acesso a redes corporativas;
- Dois níveis (*Two-tier*), consiste em uma arquitetura cliente/servidor. Segundo Singh (2001), essa arquitetura simples foca o desempenho, integridade e administração do *data warehouse* de forma tradicional. Essa abordagem é atrativa porque utiliza sistemas existentes legados como servidores de base de dados e requer o mínimo de investimentos adicionais em hardware e software, porém, com um aumento significativo de usuários efetuando consultas ao DW, esta abordagem acaba por sofrer um sério problema de performance, visto que, como o processamento das consultas é feito no cliente, muitas vezes fica inviável trafegar pela rede enormes quantidades de registros para posterior processamento no cliente;
- Três níveis (*Three-tier*), esta abordagem difere da abordagem de dois níveis no momento em que aplica uma estrita separação entre a interface gráfica, a lógica de negócios e os dados. Esta abordagem não sobrecarrega a estação de trabalho cliente. Obtém-se uma melhora em termos de escalabilidade, porém, tem-se um aumento de custo e complexidade da solução. Cada uma das camadas desta abordagem pode rodar em um processador ou hardware distinto, podendo a camada de lógica de negócios muitas vezes rodar no

mesmo servidor de DW físico. Segundo Singh (2001), com essa topologia, os usuários não precisam saber onde e como os dados são armazenados, ou a complexidade dos bancos de dados para usar as aplicações. Ela também simplifica a geração de SQL e otimiza o desempenho;

- Quatro níveis (*Four-tier*), consiste em uma arquitetura que inclui o fornecimento de alguns dados ao *desktop* através de replicação de dados ou redirecionamentos para bancos de dados no *desktop*. Com isso muitas consultas podem ser feitas diretamente no cliente sem poluir a os dados nos DW local e central.

É importante deixar claro que não existe nenhuma topologia de implementação assumida como correta ou mais adequada, esta escolha deve ser pautada sobre as necessidades de cada organização.

### 2.3.7 Arquiteturas de Implementação

A escolha da arquitetura é uma decisão gerencial do projeto, e está normalmente baseada nos fatores relativos à infra-estrutura disponível, ao ambiente de negócios (porte da empresa), concomitantemente com o escopo de abrangência desejado, assim como a capacitação dos empregados da empresa e dos recursos disponibilizados ou projetados para investimento (MACHADO, 2000).

Para a correta escolha de qual arquitetura será utilizada, é necessário avaliar uma série de variáveis como tempo para a execução do projeto, o retorno do investimento a ser realizado, a velocidade dos benefícios da utilização das informações, a satisfação do usuário executivo e os recursos necessários à implementação de uma arquitetura.

As arquiteturas que serão apresentadas são: Global, Independente e Integrada.

a) Arquitetura global

O *Data Warehouse* é projetado e construído baseado nas necessidades da empresa como um todo. É considerado como um repositório comum de dados de suporte à decisão, disponível em toda a empresa (MACHADO, 2000).

A palavra “global” não está relacionada à localização física do DW e sim com o escopo de acesso e utilização das informações da empresa. Quanto à localização física do repositório, esta arquitetura pode ser fisicamente centralizada ou fisicamente distribuída nas instalações de uma organização.

Segundo Machado (2000), a arquitetura global habilita os usuários finais a utilizar visões corporativas de dados, que normalmente são requisitos de negócio; entretanto, este tipo de ambiente consome muito tempo de desenvolvimento e administração, assim como seu custo de implementação é muito alto.

#### b) Arquitetura independente

A arquitetura independente implica em *Data Marts* não conectados (*stand-alone*) controlados por um grupo específico de usuários e que atende somente às suas necessidades específicas e departamentais, sem foco corporativo algum (MACHADO, 2000).

A implementação desta arquitetura resulta sempre em uma implementação rápida, raramente impactando nos recursos de tecnologia de informação. Em contrapartida, sua integração corporativa é nula, não permitindo nenhuma visão global do negócio. Este tipo de *Data Mart* está disponível somente ao pessoal do departamento intitulado “proprietário” do *Data Mart*.

#### c) Arquitetura integrada

A arquitetura de *Data Marts* integrados é basicamente uma distribuição de implementação. Apesar de os DM serem implementados separadamente por grupos de trabalho ou departamentos, eles são integrados ou interconectados, provendo uma visão corporativa maior dos dados e informações (MACHADO, 2000).



Este tipo de implementação assemelha-se muito à arquitetura global no que diz respeito à integração.

Em decorrência da integração existente, usuários de um departamento podem acessar e utilizar dados de um DM de outro departamento. Toda esta integração e flexibilidade acabam por despendem um maior nível de complexidade de requisitos à solução. Um exemplo disto são dados comuns a múltiplos DM, onde se pode optar por compartilhamento ou replicação de dados entre os *data marts*.

### 2.3.8 Abordagens de Implementação

As abordagens de implementação se referem à forma como o *data warehouse* é projetado e implementado em relação às necessidades globais da empresa. As três abordagens de implementação descritas neste trabalho são a abordagem *top-down*, a abordagem *bottom-up* e a arquitetura BUS (incremental).

A opção por uma abordagem de implementação é influenciada por fatores como a infra-estrutura de tecnologia de informação, a arquitetura escolhida, o escopo da implementação, os recursos disponíveis e principalmente pela necessidade ou não de acesso corporativo dos dados, assim como pelo retorno de investimento desejado e velocidade de implementação (MACHADO, 2000).

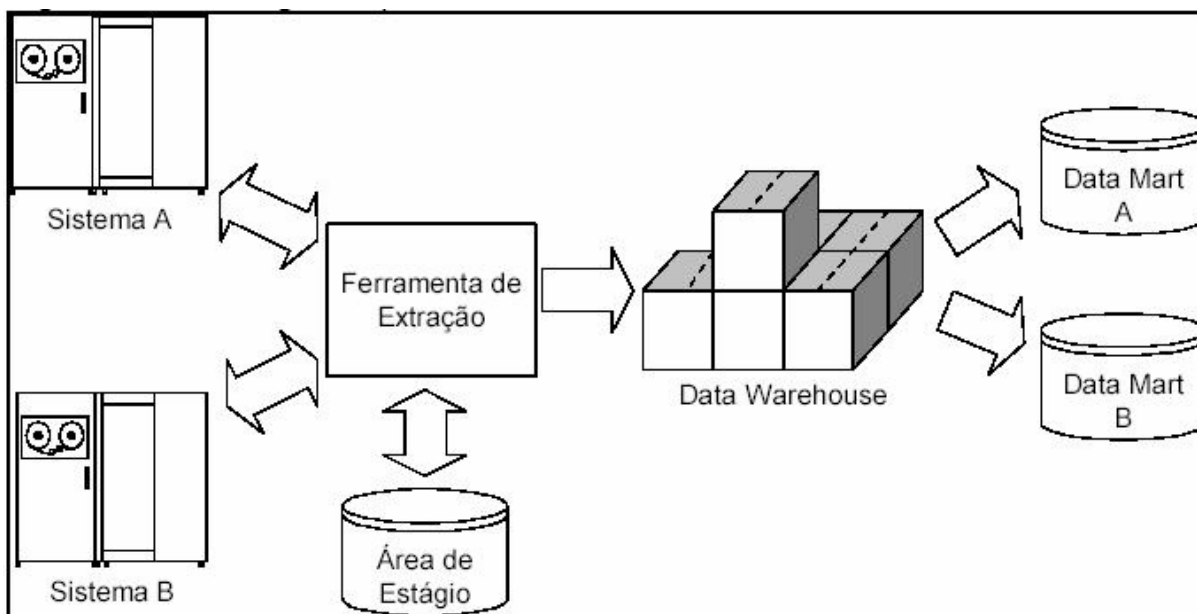
#### a) Abordagem *top-down*

Nesta abordagem, inicialmente serão levantados todos os requisitos, fontes de dados, padrões, e outros, dos departamentos da empresa participantes da iniciativa, para então dar seqüência na implementação de toda a estrutura (SELL, 2001).

Este tipo de abordagem foi proposta por Inmon (1997) e baseia-se em um *data warehouse* corporativo central, baseado no modelo relacional e totalmente normalizado. O processo de extração, transformação e carga e, conseqüentemente, a área de estágio de dados são implementados de forma única e integrada.

Este *data warehouse* corporativo contém todos os dados organizacionais e tem o propósito de servir de base de dados para os diversos *data marts* departamentais implementados com base no modelo dimensional (MELLO, 2002).

Figura 3 - Abordagem top-down



Fonte: SELL, D. Uma arquitetura para distribuição de componentes tecnológicos para sistemas de informações baseados em data warehouse. Florianópolis, p. 48, 2001. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

Segundo Machado (2000), nos *Data Marts* gerados, as informações geralmente estão em um maior nível de sumarização e, normalmente, não apresentam o nível histórico encontrado no *data warehouse* central.

Este tipo de abordagem, apesar de fornecer uma base de dados corporativa única, homogênea e totalmente integrada, traz consigo problemas relativos a altos custos de implementação e demora na apresentação de resultados (VASCONCELLOS, 1999 *apud* MELLO 2002, BALLARD et al. 2001). A Figura 3 destaca este tipo de abordagem. O Quadro 2 traz um comparativo entre as vantagens e desvantagens desta abordagem segundo Machado (2000).

Quadro 2 - Vantagens e desvantagens da abordagem *top-down*

Vantagens	Desvantagens
Herança de arquitetura – Todos os DM	Implementação muito longa – Os DW

<p>originados a partir de um DW utilizam a arquitetura e os dados desse DW, permitindo uma fácil manutenção</p>	<p>são, normalmente, desenvolvidos de modo iterativo por área de assunto, como, por exemplo, vendas, finanças e recursos humanos. Mesmo assim, são necessários, em média, 15 ou mais meses para que a primeira área de assunto entre em produção, dificultando a garantia de apoio político e orçamentário</p>
<p>Visão de empreendimento – O DW concentra todos os negócios da empresa, sendo possível, a partir dele, extrair níveis menores de informações</p>	<p>Alta taxa de risco – Não existem garantias para o investimento neste tipo de ambiente</p>
<p>Repositório de metadados centralizado e simples – O DW provê um repositório de metadados central para o sistema. Essa centralização permite manutenções mais simples do que aquelas realizadas em múltiplos repositórios</p>	<p>Heranças de cruzamentos funcionais – É necessária uma equipe de desenvolvedores e usuários finais altamente capacitados, para avaliar as informações e consultas que garantam à empresa habilidade para sobreviver e prosperar na arena de mudanças de competições políticas, geográficas e organizacionais</p>
<p>Controle e centralização de regras – A arquitetura <i>top-down</i> garante a existência de um único conjunto de aplicações para extração, limpeza e integração dos dados, além de processos centralizados de manutenção e monitoração</p>	<p>Expectativas relacionadas ao ambiente – A demora do projeto e a falta de retorno podem induzir expectativas nos usuários</p>

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

## b) Abordagem *bottom-up*

Nesta abordagem são implementados *data marts* individualmente, sem partir de uma análise de requisitos para toda a empresa, e com pouca perspectiva de integração entre eles (SELL, 2001).

Segundo Machado (2000), o propósito desta implementação é a construção de um *Data Warehouse* incremental a partir do desenvolvimento de *Data Marts* independentes.

Esta abordagem se baseia na construção de *data marts* dimensionais independentes, cada qual com sua respectiva área de estágio e processos de extração, transformação e limpeza e, metadados (MELLO, 2002).

O *data warehouse*, neste tipo de implementação é composto pelo conjunto de todos os *data marts* construídos, formando um *data warehouse* incremental lógico (VASCONCELLOS, 1999 *apud* MELLO (2002), BALLARD et Al. 2001). A Figura 4 destaca este tipo de abordagem. O Quadro 3 faz um comparativo entre as vantagens e desvantagens desta abordagem segundo Machado (2000).

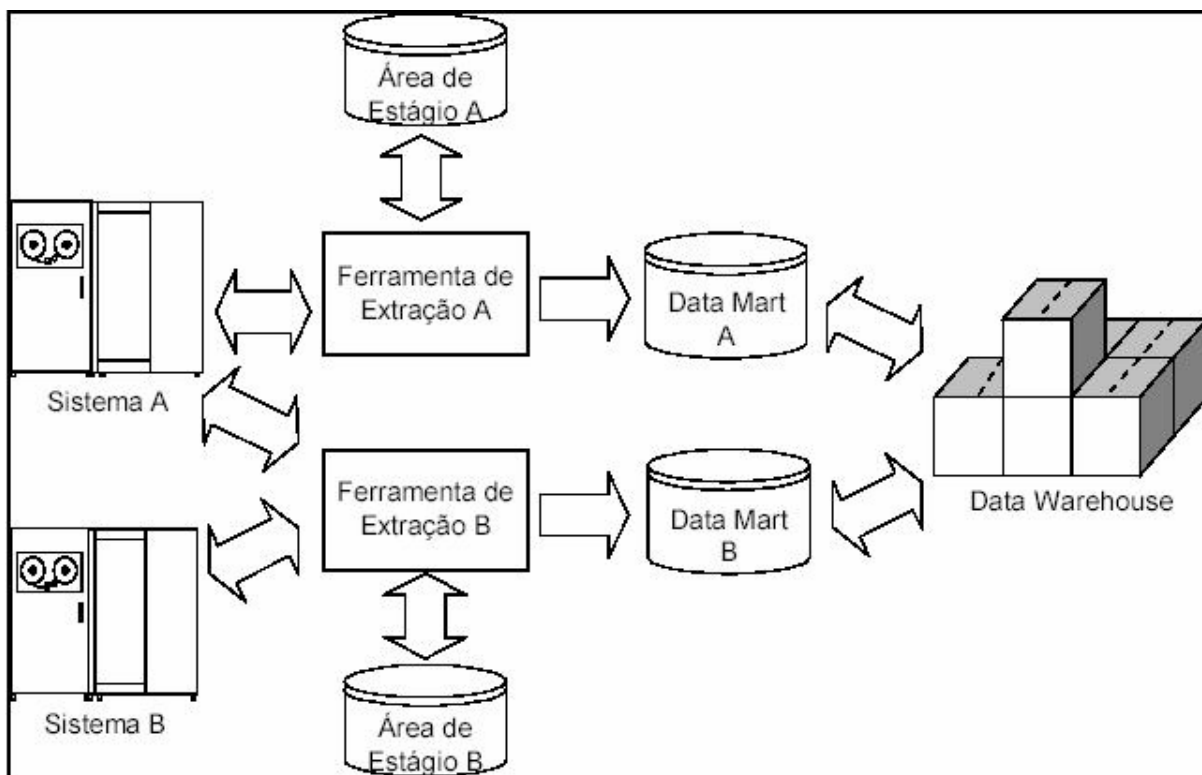
Quadro 3 - Vantagens e desvantagens da abordagem *bottom-up*

Vantagens	Desvantagens
<p>Implementação rápida – A construção dos DM é altamente direcionada, permitindo um rápido desenvolvimento. Normalmente, um DM pode ser colocado em produção em um período de seis a nove meses</p>	<p>Perigo de <i>legamarts</i> – Um dos maiores perigos no DW é a criação de DMs independentes. O advento de ferramentas de “<i>drag-and-drop</i>” facilitou o desenvolvimento de soluções individuais, de acordo com as necessidades da empresa. Estas soluções podem não considerar a arquitetura de forma global. Desta forma, os DMs independentes transformam-se em DMs legados, ou <i>legamarts</i>. Os <i>legamarts</i> dificultam,</p>

	quando não inviabilizam futuras integrações. Eles são parte do problema e não da solução
Retorno rápido – A arquitetura baseada em DM com incremento demonstra rapidamente seu valor, permitindo uma base para investimentos adicionais, com um nível mais elevado de confiança	Desafio de possuir a visão de empreendimento – Durante a construção dos DMs incrementais, é necessário que se mantenha um rígido controle do negócio como um todo. Este controle requer um maior trabalho ao extrair e combinar as fontes individuais do que utilizar um DW
Manutenção do enfoque da equipe – Um dos maiores desafios do desenvolvimento de um DW é a manutenção do mesmo enfoque por toda a equipe. A elaboração de DMs incrementais permite que os principais negócios sejam enfocados inicialmente, sem que haja gastos no desenvolvimento de áreas que não são essenciais ao problema	Administrar e coordenar múltiplas equipes e iniciativas – Normalmente, esse tipo de arquitetura emprega o desenvolvimento de DM em paralelo. Isto pode conduzir a uma rígida administração, tentando coordenar os esforços e recursos das múltiplas equipes, especialmente nas áreas de regas e semântica empresariais
Herança incremental – A estratégia de DMs incrementais obriga a entrega de recursos de informação, passo a passo. Isto permite à equipe crescer e aprender, reduzindo os riscos. A avaliação de ferramentas, tecnologias, consultores e vendedores só deve ser realizada uma vez, a não ser que existam restrições que impeçam o reaproveitamento	Maldição de sucesso – A arquitetura com DMs incrementais carrega a “maldição de sucesso”. Nestes casos, os usuários finais do DM encontram-se felizes, querendo mais informação para seus DMs. Ao mesmo tempo, outros usuários de outros DMs aguardam o incremento de seus DMs. Isto conduz a equipe de DMs a vencer desafios políticos, de recurso e de administração

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

Figura 4 - Abordagem bottom-up



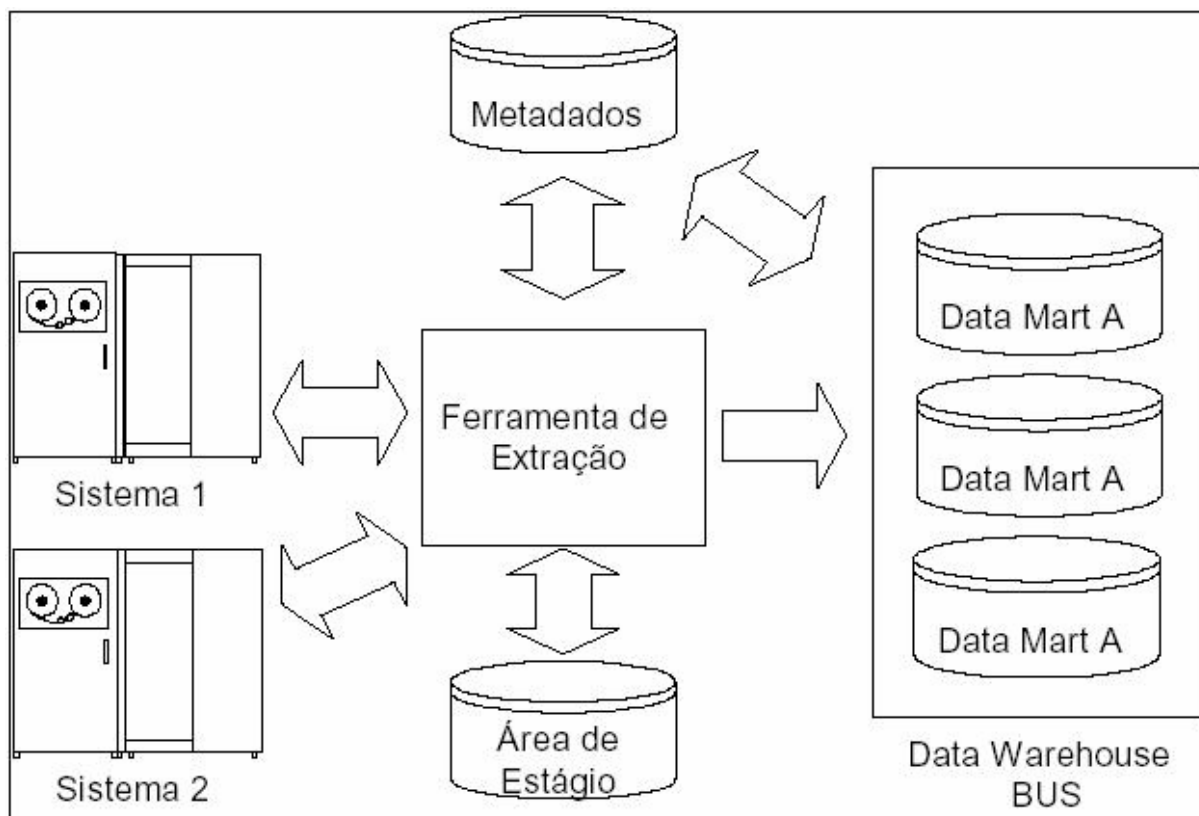
Fonte: SELL, D. Uma arquitetura para distribuição de componentes tecnológicos para sistemas de informações baseados em data warehouse. Florianópolis, p. 48, 2001. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

### c) Arquitetura incremental (BUS)

Esta abordagem parte da combinação do desenvolvimento *top-down* e *bottom-up* e baseia-se no princípio em que os requisitos devem ser bem definidos e que uma integração entre os *data marts* seja planejada antes de começar a construção dos mesmos, sendo esta construção um processo contínuo de desenvolvimento.

Nesta abordagem efetua-se a modelagem de dados do *data warehouse* de visão macro, sendo o passo seguinte à implementação de partes desse modelo, as quais são escolhidas por área de interesse e constituem os *Data Marts* (MACHADO, 2000).

Figura 5 - Abordagem incremental (BUS)



Fonte: SELL, D. Uma arquitetura para distribuição de componentes tecnológicos para sistemas de informações baseados em data warehouse. Florianópolis, p. 48, 2001. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

Segundo Sell (2001), inicialmente é realizada uma análise dos requisitos de forma global, de onde surge uma lista de *data marts* a serem implementados e como serão integrados. A partir daí, serão levantados os requisitos de um dos departamentos integrantes da iniciativa e implementado o *data mart* correspondente. O ciclo repete-se, incrementalmente, até que todos os *data marts* tenham sido implementados. O conjunto de todos os *data marts* desenvolvidos constitui o *data warehouse* da empresa. A Figura 5 destaca a abordagem de implementação incremental. O Quadro 4 faz um comparativo entre as vantagens e desvantagens desta abordagem segundo Sell (2001).

Quadro 4 - Vantagens e desvantagens da abordagem incremental (BUS)

Vantagens	Desvantagens
A apresentação dos primeiros	Complicações políticas por conta da

resultados é feita de modo mais rápido e barato do que a abordagem global	determinação da seqüência de implementação dos <i>data marts</i> e das prioridades de manutenção
A integração entre os <i>data marts</i> possibilita a unicidade de representação dos dados e informações mais confiáveis por não existirem redundâncias	Metadata mais complexo para gerenciar a distribuição e integração dos dados
Os mecanismos de extração são projetados uma única vez	Maior controle no nível de granularidade, padronização e nas manutenções das tabelas compartilhadas

Fonte: SELL, D. Uma arquitetura para distribuição de componentes tecnológicos para sistemas de informações baseados em data warehouse. Florianópolis, p. 41, 2001. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

## 2.4 Modelagem Dimensional

A modelagem dimensional permite ao usuário observar seu banco de dados no formato de um cubo contendo duas, três ou quantas dimensões forem possível e aplicáveis. Esta modelagem proporciona um ganho de tempo na consulta, uma melhor organização do sistema e principalmente a sua utilização de forma intuitiva para o usuário.

Modelagem dimensional é um nome novo para uma técnica antiga usada para criar banco de dados simples e compreensíveis (KIMBALL, 1998a).

Modelos multidimensionais proporcionam uma estrutura de sistemas de informação que permitem que uma empresa disponha de acesso muito flexível a dados, fatie e agrupe dados de qualquer número de maneiras explore dinamicamente o relacionamento entre dados resumidos e detalhados (INMON, 1997). Esses sistemas além de oferecerem flexibilidade em relação às necessidades dos usuários fornece um alto nível de controle.

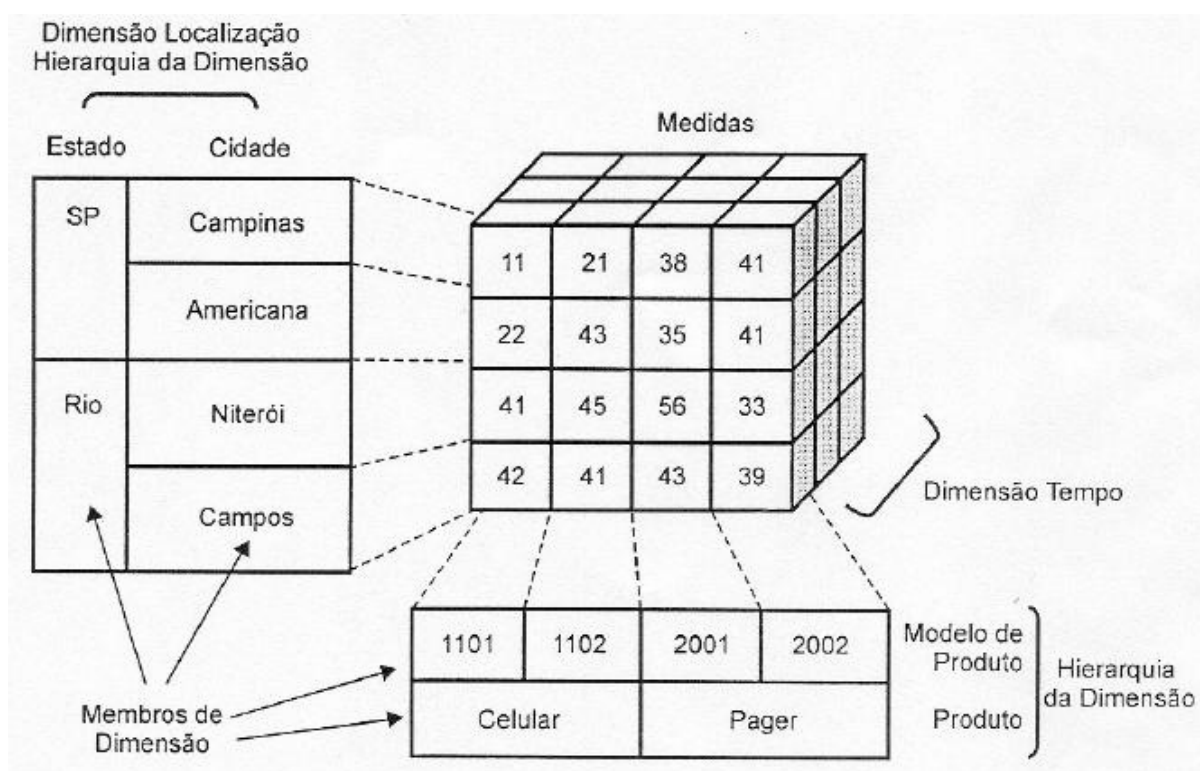


O modelo dimensional permite a visualização de dados na forma de um cubo, onde cada dimensão do cubo representa o contexto de um determinado fato, e a intersecção entre as dimensões representa as medidas do fato. A Figura 6 destaca a formação de um cubo de decisão.

Matematicamente, o cubo possui apenas três dimensões, entretanto, no modelo dimensional a metáfora do cubo pode possuir quantas dimensões forem necessárias para representar um determinado fato (MACHADO, 2000).

Os projetistas têm implementado o modelo dimensional usando um banco de dados multidimensional, ou relacional através do esquema estrela (*star join schema*) (KIMBALL, 1998a).

Figura 6 - O cubo de decisão



Fonte: MACHADO, F. N. Projeto de Data Warehouse: Uma visão multidimensional. Rio de Janeiro: Editora Érica, p.66, 2000.

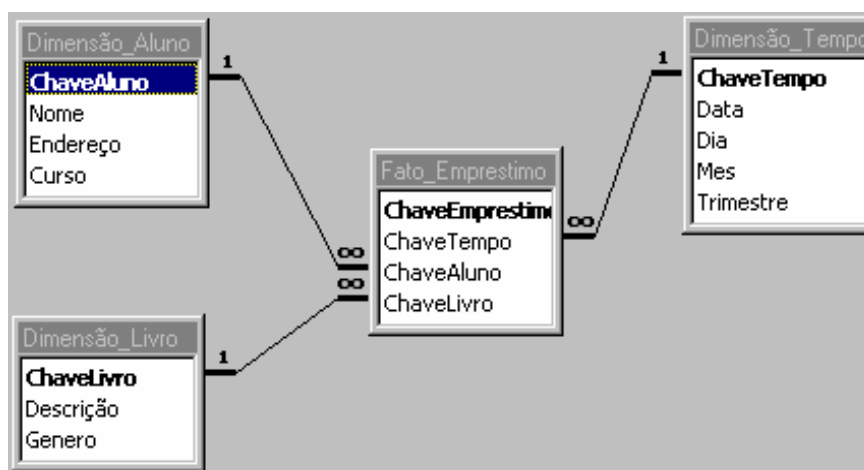
Harrison (1998), ressalta que o modelo dimensional é referenciado com frequência através do termo “esquema estrela” devido à organização de seu projeto lógico de banco de dados.

O Modelo dimensional combina tabelas de armazenamento de dados históricos em séries temporais, indexados em chaves dimensionais, descritas em tabelas dimensionais correspondentes (HARRISON, 1998). Ou seja, é organizado através de tabelas de fato, que são as ocorrências de transações de negócios da empresa, tabelas de dimensões que tem o papel descritivo nessas transações e medidas que são os atributos que quantificam um determinado fato.

Tabelas dimensionais irão guardar em sua maioria informações textuais, as quais ajudam a definir um componente da dimensão do negócio, elas possuem dados sobre produtos, mercados, períodos de tempo. As tabelas de fatos por sua vez contêm os dados temporais direcionados ao foco do negócio além de conter as informações decorrentes das tabelas de dimensões, isso garante a precisão do acesso aos dados através de uma estrutura de chave completa, eliminando assim pesquisas em tabelas e resulta em maior desempenho possível (HARRISON, 1998).

A Figura 7 expõe um modelo dimensional para um pequeno banco de dados de uma biblioteca composto por três tabelas de dimensões (Aluno, Livro e Tempo) e uma tabela de fatos (Empréstimo).

Figura 7 - Um modelo dimensional de banco de dados



### 2.4.1 Navegação pelo modelo dimensional

As operações básicas para navegação por um modelo dimensional são: *drill down*, *roll up*, *drill across* e *slice and dice*.

*Drill down* e *Roll up* são operações utilizadas para movimentar a visão verticalmente na hierarquia de uma dimensão. Através do *drill down* o usuário navega de um nível mais alto de detalhe até um nível mais baixo, enquanto que na operação *roll up* o usuário navega de um nível mais baixo de detalhe até o nível mais alto (BALLARD et al., 2001, MACHADO, 2000, SINGH, 2001). Para Kimball (1998a), efetuar *drill down* significa apenas navegar verticalmente na hierarquia de uma dimensão. Significa obter e remover rapidamente cabeçalhos de linha de qualquer uma das dimensões da tabela de fatos. A Figura 8 destaca as operações de *drill down* e *roll up*.

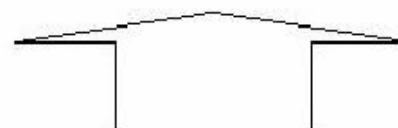
Figura 8 - Drill down e Roll up

Volume de Produção (em milhares)		1999			
		Trim. 1	Trim. 2	Trim. 3	Trim. 4
Região Sul	RS	78	67	22	56
	SC	90	67	88	99

Drill Down - Dimensão Tempo



Roll Up - Dimensão Tempo



Volume de Produção (em milhares)		Trimestre 1		
		Janeiro	Fevereiro	Março
Região Sul	RS	30	26	22
	SC	28	30	32

Fonte: MACHADO, F. N. Projeto de Data Warehouse: Uma visão multidimensional. Rio de Janeiro: Editora Érica, p.70, 2000.

A utilização de *Drill across* significa, para Kimball (1998a), a combinação de várias tabelas de fato que compartilham as mesmas dimensões em um único relatório.

*Slice and Dice* (fatiar e girar) são operações que “permitem acessar um *data warehouse* por meio de qualquer uma de suas dimensões de forma igual” (KIMBALL, 1998a), ou seja, através de estas operações é possível navegar através dos dados do cubo de decisão ao longo de qualquer dimensão (SINGH, 2001).

*Slice* corta o cubo mantendo a mesma perspectiva de visualização dos dados, permitindo que o usuário fixe a apresentação em um determinado detalhe (BALLARD et al., 2001, MACHADO, 2000). A Figura 9 destaca o processo de *slice*.

*Dice* significa mudar a perspectiva de visão, girando o cubo e invertendo uma ou mais dimensões (BALLARD et al., 2001, KIMBALL, 1998a, MACHADO, 2000). A Figura 10 destaca o processo de *dice*.

Figura 9 - Slice

Volumes		Celulares e Pagers		
		Janeiro	Fevereiro	Março
Região Sul	RS	30	26	22
	SC	28	30	32



Volumes		Celulares		
		Janeiro	Fevereiro	Março
Região Sul	RS	22	18	18
	SC	19	27	25

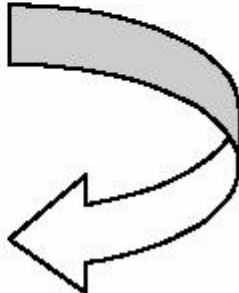
Fonte: MACHADO, F. N. Projeto de Data Warehouse: Uma visão multidimensional. Rio de Janeiro: Editora Érica, p.71-72, 2000.

Figura 10 - Dice

Volumes	1999			
	Pré-Pagos		Pós-Pagos	
	Masculino	Feminino	Masculino	Feminino
PR	30	40	80	50
SC	20	15	60	30

Dice

Volumes		1999	
		Pré-Pagos	Pós-Pagos
PR	Masculino	30	80
	Feminino	40	50
SC	Masculino	20	60
	Feminino	15	30



Fonte: MACHADO, F. N. Projeto de Data Warehouse: Uma visão multidimensional. Rio de Janeiro: Editora Érica, p.71-72, 2000.

#### 2.4.2 Fatos

Um fato usualmente é uma coisa sobre a qual não se conhece antecipadamente; é uma observação da realidade (KIMBALL, 1998a).

Segundo Kimball (1998a), os melhores fatos e os mais úteis são numéricos e caracterizam-se por serem continuamente valorados (diferentes a cada medida) e aditivos (modificam-se a cada combinação de atributos das tabelas de dimensão), pois praticamente todas as consultas podem ser feitas sobre estes fatos.

Fato é tudo aquilo que pode ser representado por meio de valores mensuráveis, geralmente numéricos (MACHADO, 2000, KIMBALL 1998a) e que são objeto de análise (BALLARD et al., 2001). Fatos, portanto, podem ser caracterizados por algo que ocorre de forma variável ao longo do tempo e podem ser expressos em valores mensuráveis (MELLO 2002).

Segundo Kimball (1998a e 1998b), os fatos podem ser aditivos, semi-aditivos e não aditivos:

- Fato aditivo: é aquele que pode ser adicionado a qualquer uma das dimensões. Uma tabela de fatos aditivos pode ser compactada facilmente através de processos de sumarização, produzindo consultas de apenas algumas linhas a partir de tabelas extremamente extensas.
- Fato semi-aditivo: é aquele que só pode ser adicionado ao longo de algumas dimensões, o que restringe o número de consultas apenas àquelas dimensões em que o fato pode ser adicionado.
- Fato não aditivo: é aquele que não pode ser adicionado a qualquer das dimensões. Para este tipo de fato, só se pode resumir registros através de contagens ou então consultá-los uma a uma.

#### 2.4.3 Transações e instantâneos

Freqüentemente, é necessário que existam duas versões de implementação de um determinado fato: uma representando cada transação individualmente e outra representando instantâneos retirados em períodos regulares de tempo (KIMBALL, 1998a).

A representação ao nível transacional é a mais fácil de se obter na implementação do modelo dimensional e a que permite análises mais ricas, que não podem ser obtidas a partir de fatos sumarizados (KIMBALL, 1998b). Ou seja, é o nível mais detalhado que se pode ter de um fato.

Algumas informações mais gerais, tais como a lucratividade de uma categoria de produtos em um intervalo de tempo, são difíceis, ou não podem, ser extraídas a partir da sumarização das transações. A solução para estes problemas é a criação de tabelas de fato contendo instantâneos (*snapshots*) retirados em determinados intervalos de tempo (MELLO, 2002). Estes instantâneos contêm não somente a sumarização das transações individuais, mas também informações indiretas, tais como a margem de lucro obtida no período (KIMBALL 1998a).

#### 2.4.4 Agregados

Segundo Kimball (1998a), um agregado é um registro de tabela de fatos que representa o resumo dos registros de nível básico da tabela de fatos.

Sell (2001) define agregados como sumários dos dados contidos no modelo original, organizados de forma a atender consultas rotineiras de forma mais ágil, constituindo-se como uma das formas mais eficientes de melhorar a performance das consultas em um DW.

Agregados são resumos construídos a partir de fatos individuais, inicialmente por questões de performance, ou quando o ambiente dos fatos é inexpressivo na menor granularidade (KIMBALL, 1998a e 1998b).

Segundo Singh (2001), os agregados permitem que as aplicações antecipem os resultados a serem pesquisados pelo usuário, eliminando a necessidade de se repetir cálculos comumente realizados.

Existem duas abordagens para armazenar agregados: definindo novas tabelas de fatos ou definindo campos nível. A utilização da primeira abordagem torna mais simples a manutenção, a carga e a utilização dos dados (KIMBALL, 1998a).

Múltiplos agregados podem ser construídos, representando os agrupamentos mais comuns dentro das dimensões do *data warehouse* a fim de aumentar a performance. A contrapartida ao aumento da performance é o aumento do consumo de espaço de armazenamento (KIMBALL, 1998b).

A definição de um agregado pode ser baseada na redução que este gerará no volume de dados, geralmente um bom agregado pode reduzir na razão de 10 ou mais vezes este volume.

Segundo Sell (2001), embora a definição de agregados ocorra mais freqüentemente durante a utilização por parte dos usuários, é válido projetá-los na

fase de definição do modelo de dados, para orientar dimensionamento de hardware e ajustes no próprio modelo, aproveitando também que os requisitos do usuário acabaram de ser definidos.

#### 2.4.5 Tabela de fatos sem fatos

Tabelas de fatos sem fatos são utilizadas para armazenar fatos que não podem ser associados a uma medida numérica. Tais tabelas podem ser utilizadas no rastreamento de eventos, como a frequência dos alunos nas salas de um escola, ou todos os elementos envolvidos em um acidente coberto por uma seguradora (MELLO 2002).

Tabelas de fatos sem fatos podem ser utilizadas para cobrir eventos que não ocorreram, como relacionar os produtos que não foram vendidos em uma determinada promoção (KIMBALL, 1998a e 1998b).

#### 2.4.6 Dimensões

Dimensões determinam o contexto em que ocorreram os fatos. No modelo dimensional, cada dimensão está associada a um ou mais fatos, sendo estas usualmente mapeadas em entidades não numéricas e informativas (BALLARD et al., 2001).

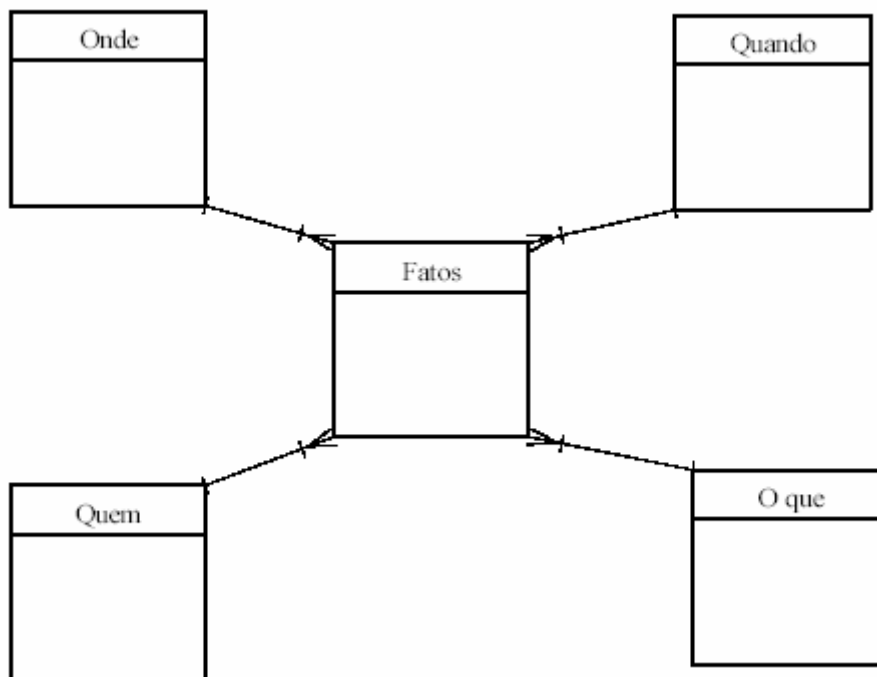
Os atributos que compõem a tabela de dimensões são as principais fontes de restrições em consultas SQL e normalmente estão presentes em cabeçalhos de linhas no conjunto resposta de consultas realizadas pelos usuários (KIMBALL, 1998a).

A maioria dos fatos envolve pelo menos quatro dimensões básicas: onde, quando, quem e o que. A dimensão Onde determina o local onde o fato ocorreu (local geográfico, filial). A dimensão Quando é a própria dimensão tempo. A dimensão Quem determina que entidades participaram do fato (cliente, fornecedor, funcionário). A dimensão O Quê determina qual é o objeto do fato (produto, serviço)



(MACHADO, 2000). A Figura 11 mostra um modelo estrela com as quatro dimensões comuns.

Figura 11 - Dimensões comuns em um modelo dimensional



Fonte: MACHADO, F. N. Projeto de Data Warehouse: Uma visão multidimensional. Rio de Janeiro: Editora Érica, p.95, 2000.

A maioria das dimensões se modifica lentamente ao longo do tempo. Para que se possam rastrear as modificações que uma dimensão pode sofrer, utiliza-se um dos três métodos a seguir (KIMBALL, 1998b e 1998a):

- Substituir os valores antigos do registro modificado por valores novos e, assim, perder a capacidade de rastrear o histórico da dimensão;
- Adicionar um novo registro à dimensão com os valores novos do atributo, segmentando o histórico entre a antiga e a nova dimensão;
- Criar novos campos para os valores atuais e os valores antigos, mantendo um histórico da dimensão em um único registro, mas reduzindo a capacidade de manter o histórico da dimensão.

Existem casos em que algumas dimensões podem conter milhões de entradas, por exemplo, a dimensão cliente em uma companhia telefônica, neste

caso a navegação por esta dimensão pode tornar-se demorada. Pode-se, neste caso, utilizar índices nos atributos que sejam objetos de navegação (KIMBALL, 1998a).

Freqüentemente, os campos mais utilizados em uma dimensão grande possuem um domínio pequeno, ou seja, assumem uma pequena quantidade de valores. Em uma dimensão cliente, estes atributos podem ser atributos demográficos, como sexo, faixa etária e classe social. Neste caso, pode-se optar pela criação de uma minidimensão separada da dimensão cliente para aumentar a eficiência da navegação (KIMBALL, 1998a).

Atributos como o número da nota fiscal de venda, aparentemente, deveriam fazer parte da tabela de fatos. Em um banco de dados relacional, ele seria o atributo determinante do cabeçalho da nota fiscal. Em um banco de dados dimensional, normalmente, todos os atributos determinados pelo número da nota foram armazenados em dimensões próprias e faria parte da chave primária dos itens da nota. Ainda assim, pode-se utilizar este atributo para agrupar os fatos pelo documento original. Atributos deste tipo são representados como dimensões descaracterizadas, isto é, chaves de dimensão sem uma dimensão correspondente (KIMBALL, 1998a).

Em um *data warehouse* bancário, a dimensão conta, contendo atributos do correntista, é uma dimensão suja, pois em caso de conta conjunta, a conta aparecerá várias vezes, uma para cada correntista (KIMBALL, 1998a). Logo tem-se que dimensões sujas ocorrem sempre que uma entidade aparece diversas vezes, podendo haver ligeiras diferenças em alguns de seus atributos.

A cada chave primária da tabela dimensão deve haver exatamente uma chave estrangeira em cada uma das tabelas de fatos que utilizam esta dimensão para descrever um acontecimento e, diferentemente dos fatos, as dimensões podem possuir campos numéricos que não variam continuamente a cada nova amostra.

Utilizar uma chave candidata como chave primária de uma dimensão pode causar problemas, caso esta chave não seja absolutamente estável ao longo

do tempo. Uma modificação em uma chave de uma dimensão pode ocasionar um grande volume de mudanças nas tabelas de fato relacionadas a esta dimensão. A solução para este problema é utilizar chaves artificiais absolutamente estáveis ao longo do tempo. Uma chave artificial é um campo inteiro e auto-incremental, que aumenta a cada novo registro incluído na tabela de dimensão (KIMBALL, 1998b).

Kimball (1998a) afirma que modelos dimensionais reais no mundo dos negócios contêm entre 4 e 15 dimensões, sendo raros os casos com 2 ou 3 e, segundo o autor, modelos com 20 ou mais dimensões devem ser estudados para se verificar as dimensões supérfluas e/ou combiná-las.

#### 2.4.7 Granularidade e Particionamento

A granularidade refere-se ao nível de detalhe ou resumo com o qual serão armazenados os dados no *data warehouse* (INMON, 1997), quanto maior o detalhamento, mais baixo será o nível de granularidade e vice-versa. A definição da granularidade afeta diretamente o volume de dados do *data warehouse*, bem como a qualidade e performance das consultas a serem feitas. Como exemplo podemos citar que uma granularidade alta garante maior rapidez nas consultas feitas, porém diminui a riqueza de informações que se pode extrair, enquanto uma menor granularidade possibilita a extração de qualquer informação, mas acarreta maior volume de dados, conseqüentemente maior tempo de resposta à consulta e maior investimento em hardware.

Um *data warehouse* pode ser implementado em níveis duais de granularidade ao longo do tempo. É possível manter as informações mais recentes em um baixo nível de granularidade, aumentando assim as possibilidades de extração de informações. À medida que os dados vão ficando obsoletos, é possível resumir-los em um alto nível de granularidade de forma a manter a performance (INMON, 1997).

Segundo Sell (2001), o nível adequado de granularidade deve ser definido de tal forma que atenda as necessidades do usuário, tendo como limitação os recursos disponíveis, ou seja, é necessário encontrar um ponto de equilíbrio.

Opcionalmente, pode-se optar por múltiplos níveis de granularidade em um *data warehouse*, partindo-se inicialmente de um baixo nível de granularidade, passando por níveis intermediários até um alto nível de granularidade, dependendo das necessidades de informação e do custo de mantê-las no *data warehouse* (BALLARD et al., 2001).

As tabelas de fato de um *data mart* na arquitetura incremental, assim como as dimensões padrão, devem estar no menor nível de granularidade possível, proporcionando facilidade na alimentação da base de dados e continuidade na extração de informações ao longo do tempo (KIMBALL, 1998b).

Já o particionamento consiste em dividir os dados em unidades físicas menores, de acordo com sua utilização ou lógica de negócio, de forma a obter uma maior flexibilidade no gerenciamento do banco de dados.

Segundo Inomn (1997), o particionamento proporciona:

- Facilidade de reestruturação, indexação e reorganização;
- Facilidade de recuperação;
- Facilidade de monitoramento;
- Escalabilidade no *data warehouse*;
- Portabilidade dos elementos do *data warehouse*.

Em geral, os critérios de particionamento dependem dos requisitos do negócio e restrições físicas da base de dados e podem ser os seguintes (INMON, 1997):

- Períodos;
- Geografia (local);
- Área de negócio ou linha de produtos;
- Unidade organizacional;
- Qualquer combinação dos critérios acima.

### 2.4.8 Modelo Snowflake

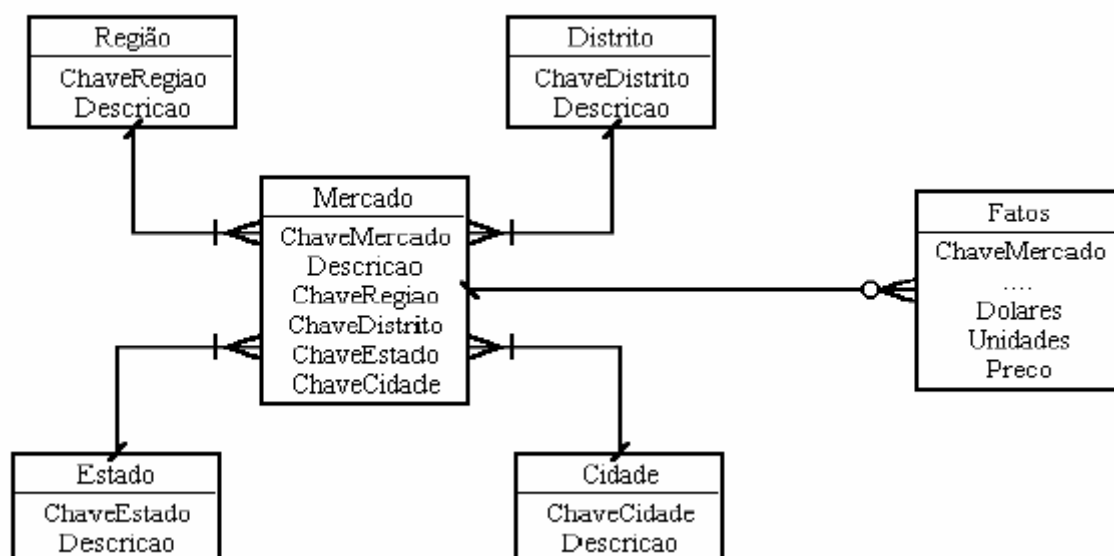
Este modelo incorpora tabelas dimensionais principais conectadas às tabelas de fato e tabelas dimensionais de extensão, onde são armazenadas as descrições das dimensões. Estas tabelas de extensão são obtidas normalizando-se as dimensões (HARRISON, 1998, KIMBALL 1998a, SINGH, 2001). As tabelas dimensionais principais contêm chaves ligando-as às tabelas de extensão. Segundo Harrison (1998), a desvantagem *snowflake* é a necessidade de múltiplos *joins* em consultas SQL, caso as consultas necessitem combinar diversas dimensões, o que aumenta a complexidade da instrução e reduz o desempenho.

Ainda segundo Harrison (1998), existem três tipos básicos de projetos *snowflake*: pesquisa, em cadeia e atributo.

#### a) Pesquisa

Este modelo utiliza uma tabela de dimensão principal conectada diretamente à tabela de fatos e tabelas de extensões conectadas a esta tabela de dimensão principal. A Figura 12 destaca um *snowflake* do tipo pesquisa.

Figura 12 - Snowflake do tipo pesquisa



Fonte: MELLO, J. A. B. Uma proposta de Modelo de Dados para Suporte ao Processamento Transacional e de Apoio Informacional Simultaneamente. Florianópolis, p. 52, 2002.

Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

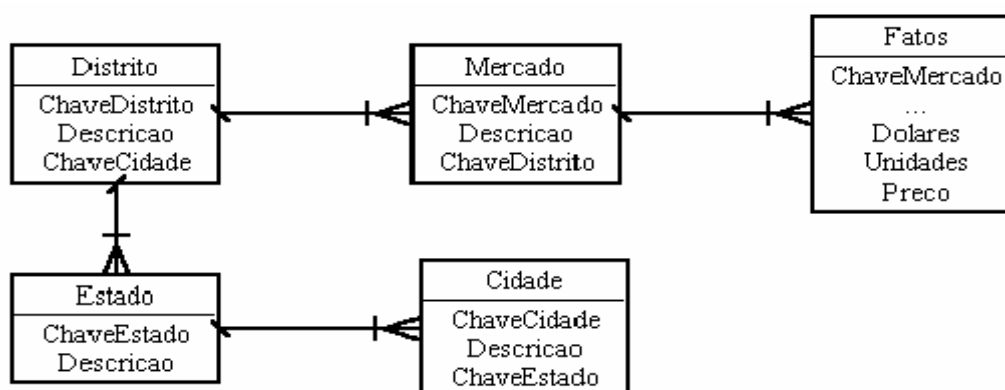
### b) Em cadeia

Este modelo coloca as tabelas de extensão encadeadas, começando com uma tabela dimensional principal conectada à tabela de fatos. A próxima dimensão é conectada a tabela dimensional principal e assim sucessivamente até a última dimensão. Este modelo mantém um alto nível de integridade de dados por que as dimensões são completamente normalizadas. A Figura 13 destaca um tipo *snowflake* em cadeia.

### c) Atributo

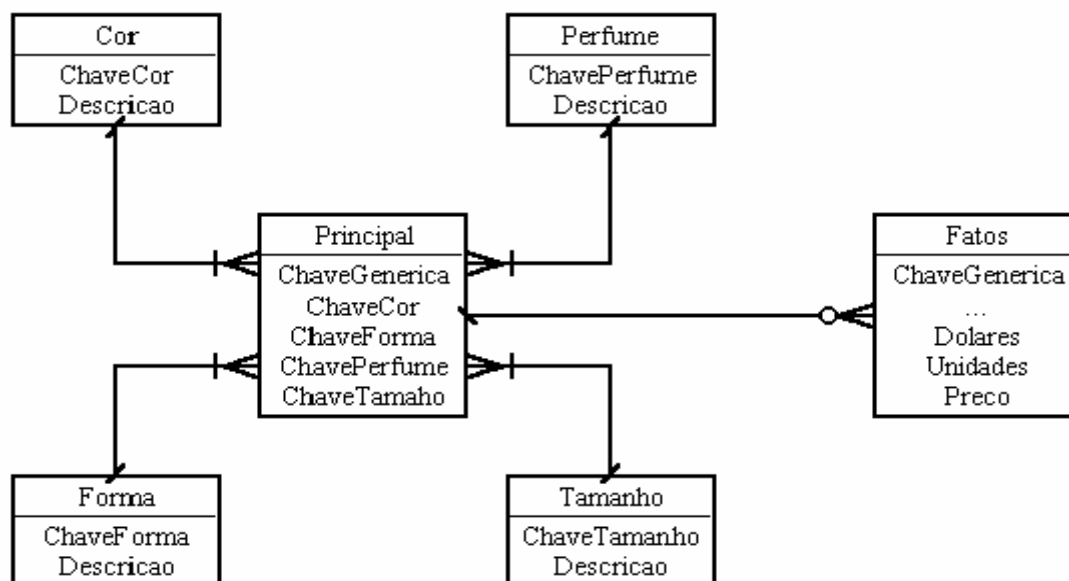
Este modelo permite construir uma dimensão principal artificial a partir de vários grupos de atributos dimensionais não relacionados. A Figura 14 destaca um *snowflake* do tipo atributo.

Figura 13 - Snowflake em cadeia



Fonte: MELLO, J. A. B. Uma proposta de Modelo de Dados para Suporte ao Processamento Transacional e de Apoio Informacional Simultaneamente. Florianópolis, p. 52, 2002. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

Figura 14 - Snowflake do tipo atributo



Fonte: MELLO, J. A. B. Uma proposta de Modelo de Dados para Suporte ao Processamento Transacional e de Apoio Informacional Simultaneamente. Florianópolis, p. 53, 2002. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

## 2.5 Data Warehouse Distribuído

As topologias de processamento existentes nas empresas muitas vezes influenciam a arquitetura do *data warehouse*.

Em organizações onde todo, ou grande parte do processamento é feito na matriz, dificilmente haverá necessidade de existir um *data warehouse* distribuído. Por outro lado, quando uma quantidade significativa de processamento ocorre no nível local (unidade remota), justifica-se a existência de uma estrutura de DW distribuído.

O principal motivo para implementar *data warehouses* distribuídos é que a integração dos dados da empresa inteira não faz sentido (SINGH, 2001).

De acordo com Inmon (1997), um *data warehouse* local contém dados que são de interesse exclusivo para o nível local. Ele exerce a mesma função que qualquer outro *data warehouse*, exceto pelo fato de que o escopo do *data*

*warehouse* é local. Não há coordenação de dados ou estrutura de dados de um *data warehouse* local para outro. As fontes de dados são os próprios sistemas locais.

Este tipo de implementação de *data warehouse* sugere que seus componentes sejam distribuídos em diferentes bancos de dados físicos, ocasionando com isso, um maior grau de redundância e gerando procedimentos mais complexos de carga e manutenção.

Para Singh (2001), os DW distribuídos possuem as seguintes características:

- A atividade ocorre em nível local;
- A maior parte do processamento é feita localmente;
- O site local é autônomo;
- Cada *data warehouse* local possui estrutura e conteúdos exclusivos;
- Os dados são exclusivos e de suma importância apenas para esta localidade;
- A maior parte dos dados é local e não replicada;
- Qualquer interseção de dados entre *data warehouses* locais é coincidência;
- O *data warehouse* local serve a diferentes regiões;
- O *data warehouse* local serve a diferentes comunidades técnicas;
- O escopo do *data warehouse* é limitado ao local;
- O *data warehouse* local também contém dados históricos e está integrado apenas no local.

O *data warehouse* global tem como escopo a corporação ou empresa enquanto que os *data warehouses* locais inseridos nele possuem como escopo a filial a que estão servindo.

Segundo Inmon (1997), existe mais de um tipo de *data warehouse* distribuído e, constitui um erro imaginar que o *data warehouse* distribuído é uma proposta binária, em vez disso, há graus de *data warehouse* distribuído.



Uma típica implementação de DW distribuído é a construção de vários *warehouses* locais, que fornecem uma visão local dos dados de cada site, e um *data warehouse* global que fornece uma visão global da empresa como um todo. Neste tipo de implementação é importante que qualquer dado contido no *data warehouse* local não seja encontrado no DW global e vice-versa, sendo por este motivo, desnecessário a alimentação do DW global a partir dos diversos DW locais. A existência de dados redundantes, ou seja, a mesma informação está presente no DW global e no DW local, mesmo que em uma área de estagiamento local temporária – tipo de implementação que sugere a armazenagem dos dados localmente em uma base temporária mesmo depois dos dados terem sido transferidos para o DW global - é uma indicação de que o escopo dos diferentes *warehouses* não foi apropriadamente definido sendo apenas uma questão de tempo o aparecimento de sistemas em teia de aranha.

Segundo Inmon (1997), ao construir um *data warehouse* global é preciso reconhecer que existem algumas anomalias. No que diz respeito aos níveis de dados, o *data warehouse* global não se encaixa na estrutura clássica de um *data warehouse*. Uma delas é que os dados detalhados residem no nível local, ao passo que os dados levemente resumidos residem no nível global centralizado.

Logo, a decisão pela implementação de um DW distribuído deve ser pautada sobre a topologia de processamento existente na organização e que a correta definição dos escopos dos DW locais e global será o ponto chave para o sucesso ou fracasso da solução.

### 3. A APLICAÇÃO

#### 3.1 Introdução

Diante do atual cenário econômico, o agro-negócio solidifica-se cada vez mais como uma atividade promissora e rentável. As exportações de produtos agrícolas, agropecuários e seus derivados aumentam a cada ano e à medida que o tempo passa, esta atividade ganha mais respeito e respaldo das equipes de governos visto a importância da mesma na balança comercial do país.

Apesar de movimentar uma respeitável quantia monetária, geralmente as principais atividades envolvidas com o agro-negócio não estão situadas em grandes centros urbanos / industriais, é nas zonas rurais que o agro-negócio possui seu alicerce. Tem-se então uma grande atividade econômica inserida em um contexto pequeno e muitas vezes desprivilegiado em relação a tecnologias de comunicação.

Esta deficiência de comunicação acarreta o uso de tecnologias de transmissão de dados muitas vezes obsoletas ou de baixa velocidade. Os custos envolvidos para se criar um ambiente de transmissão que possibilite um maior tráfego de informações acaba por inibir a construção de alguns tipos de sistemas que necessitam de uma maior banda passante.

A C.Vale em decorrência de suas atividades, acaba por estar inserida em um ambiente com todas as características acima citadas e, conseqüentemente, até então não dispunha de um sistema de apoio à decisão que fornecesse subsídios de informações para as gerências remotas pautarem suas decisões e gerirem suas atividades.

O Sistema de Apoio a Atividade Gerencial (SAAG), nome dado ao ambiente modelado, trata-se da implementação de um *data warehouse* centralizado com *data marts* distribuídos que visa suprir esta falta de informações de caráter gerencial existente na matriz e nas filiais da empresa.

O presente capítulo apresenta a cooperativa, destaca sua infra-estrutura tecnológica, relata a problemática da centralização da informação, apresenta o SAAG e descreve suas principais características de modelagem e implementação, bem como os resultados obtidos.

### **3.2 A Cooperativa**

Em meados da década de 50, agricultores oriundos do Rio Grande do Sul e de Santa Catarina começaram a se instalar em Palotina. Aos poucos, novas levas de colonizadores foram chegando e com isso aumentaram-se as áreas de lavoura. Este aumento de produção permitiu aos agricultores iniciar a comercialização da produção. Como existiam poucas estradas, que ficavam intransitáveis em períodos de chuva, e não havendo armazéns para estocar a produção, muitas vezes a produção era perdida antes mesmo de conseguir compradores.

Tais problemas fizeram com que estes colonizadores encontrassem alternativas viáveis para contornar estas dificuldades. Isto foi o passo inicial para a criação de uma cooperativa de produção.

Em 07 de novembro de 1963, as 14:00 horas, numa das dependências da Prefeitura Municipal de Palotina, Estado do Paraná, reuniram-se agricultores, para a constituição da Cooperativa que, em apenas 07 dias após o encontro, estava constituída legalmente e pronta para o início de suas atividades.

Assim nasceu a Cooperativa Agrícola Mista de Palotina Ltda. - CAMPAL, com sede em Palotina - PR.

Em 31 de março de 1974, os associados decidiram alterar a razão social da Cooperativa Agrícola Mista de Palotina Ltda. - CAMPAL, para Cooperativa Agrícola Mista Vale do Piquiri Ltda. - COOPERVALE, ampliando para os municípios de Assis Chateaubriand, Terra Roxa e Francisco Alves a área de ação.

A unidade da sede do município de Assis Chateaubriand foi inaugurada em 01 de fevereiro de 1975 e em 13 de dezembro do mesmo ano aconteceu a filiação à Cooperativa Central Regional Iguaçu Ltda. - COTRIGUAÇU. A ativação da unidade de Santa Rita do Oeste em Terra Roxa aconteceu em 06 de janeiro de 1976.

Em 27 de fevereiro de 1981 assumiu administrativamente a Cooperativa Agrícola Mista de Diamantino no Estado do Mato Grosso. As unidades de Diamantino, Nova Mutum e Novo Horizonte, ambas daquele estado, foram oficialmente ativadas em 24 de julho de 1981, marcando o início das atividades da COOPERVALE no centro-oeste do país.

A ativação da unidade de Abelardo Luz em Santa Catarina aconteceu em 17 de setembro de 1984 e o entreposto de Encantado do Oeste no município de Assis Chateaubriand, foi inaugurado em 1º de fevereiro de 1985. Também em Assis Chateaubriand, foi ativada em 06 de fevereiro de 1985 a unidade de Vila Nice. O entreposto da localidade de Vila Maripá, no município de Palotina -PR (hoje pertencente ao município de Maripá), foi ativado em 27 de abril de 1985. Em 26 de julho do mesmo ano foi ativada a unidade de Terra Nova em Assis Chateaubriand e em 1º de outubro, também de 1985, o entreposto de Faxinal dos Guedes em Santa Catarina.

Em outubro de 1997 iniciaram-se as atividades do Abatedouro de Aves, bem como a construção do matrizeiro e alojamento de matrizes, construção do incubatório, ampliação e modernização da fábrica de ração visando atender o projeto avícola com o início de suas atividades.

A partir de 1997 intitulou-se que a cooperativa se chamaria Coopervale/C.vale por um período transitório, o que faria com que a marca de seus produtos (C.Vale), fosse consolidada junto ao mercado. Depois de consolidada a marca, somente seria intitulada C.Vale.

Em 1998, a C.Vale abriu entrepostos em Alto Piquiri, Brasilândia, e Paulistânia. Em 2001 iniciaram-se as atividades nas unidades de Sorriso e Sinop no Mato Grosso, consolidando ainda mais a atuação da cooperativa nesta região.

Ainda em 2001 foram inauguradas as unidades de Rio Brillhante e Fátima do Sul no Mato Grosso do Sul, iniciando a atuação neste estado e também se iniciaram as atividades nas unidades de Corpus Cristi, Catuetê e La Paloma no país vizinho Paraguai. Também em 2001 iniciaram-se as exportações de carne de frango para países da Europa e outros, exportações estas que, atualmente, consomem 30% de toda produção realizada.

Em 2002 foram iniciadas as atividades na unidade de Amambaí, Dourados e Caarapó, ambas no Mato Grosso do Sul e também a unidade de Santa Carmem no Mato Grosso. Também em 2002 tiveram início às atividades nas unidades de São Francisco, Silverópolis, Bairro Catarinense, Paulistânia e Navegantes (todas no PR), sendo esta, mais uma unidade beneficiadora de fécula de mandioca.

Iniciaram-se ainda em 2002 as atividades do projeto de suinocultura, com a produção de leitões, matrizeiro e avozeiro, contanto inicialmente com 4.000 matrizes para a produção de leitões.

A história da C.Vale pode ser dividida em quatro fases:

A primeira etapa caracterizou-se pelo trabalho de criação da cooperativa, a chamada organização documental, que exigiu o esforço dos pioneiros para atrair agricultores para dentro da organização e conseguir os registros para o seu funcionamento.

A segunda fase foi a da organização da produção. A preocupação era prestar e manter o atendimento básico ao associado. Aproveitando as facilidades do crédito fácil e barato, a C.Vale ampliou sua estrutura, construindo novas unidades e aumentando sua área de ação. Essa política prolongou-se até meados da década de 80.

A partir desta fase iniciou-se um período de transição. As modificações na economia começavam a reduzir a rentabilidade do setor primário e traziam novas exigências aos produtores e às empresas. Era preciso modernizar-se e buscar alternativas de renda aos associados e a C.Vale.

Em 1995 foram tomadas medidas de reformulação com base em um plano de modernização, dando início à quarta fase da C.Vale caracterizada principalmente pela agregação de valores aos produtos. A empresa definiu que seu objetivo seria a industrialização e optou por dar a arrancada através da transformação da soja e do milho – seus dois principais produtos – em carne de frango. Com o mesmo propósito, a C.Vale implantou uma indústria de amido modificado de Mandioca.

### 3.2.1 A cooperativa hoje

A C.Vale cresceu nos últimos oito anos aproximadamente cinco vezes, consolidando-se como uma das maiores empresas do estado do Paraná, estando enquadrada dentro das 500 maiores empresas do Brasil (segundo a revista exame) e como a segunda maior Cooperativa singular do Brasil.

A C.Vale hoje está classificada como a 52<sup>a</sup> maior empresa da região sul e a 294<sup>a</sup> maior do Brasil.

A cooperativa ganhou posições no levantamento “Melhores e Maiores”, publicado na edição de julho/2002 da revista Exame, que analisou os balanços financeiros das 500 maiores empresas do país. A cooperativa subiu 61 posições em volume de vendas, aparecendo em 294<sup>o</sup> lugar, com faturamento de 258,7 milhões de dólares. No estudo anterior, a C.Vale estava na 355<sup>a</sup> colocação. Os dados apontam, ainda, que a cooperativa teve o 9<sup>o</sup> maior aumento de vendas do Brasil em 2001, no segmento atacado e comércio exterior, com uma variação de 31,6% comparativamente ao ano anterior, já descontada a inflação.

O levantamento também mostra que a C.Vale subiu da 57<sup>a</sup> para a 52<sup>a</sup> colocação entre as maiores empresas da região Sul em volume de vendas. Entre as

empresas do segmento comercial do Sul do Brasil, a cooperativa teve o 10º melhor desempenho em 2001, crescendo cinco posições em relação ao ano 2000.

De acordo com os dados extraídos do balanço do ano de 2002, a empresa conta com 55 unidades, cresceu em número de colaboradores de 564 em 1996 para 2.547 em 2002, em número de associados de 5.191 em 1996 para 6.886 em 2002, em número de toneladas de produção agrícola recebida de 697.172 em 1996 para 1.120.398 em 2002 e principalmente em faturamento realizado de R\$ 211.154.503,00 em 1996 para R\$ 844.594.432,00 em 2002, tendo-se uma previsão deste faturamento para 2003 na ordem de R\$ 1.050.000.000,00.

A revista Exame detalhou também, o desempenho das principais empresas paranaenses. A C.Vale classificou-se em terceiro lugar em crescimento de vendas (31,6%) e terceiro lugar em retorno sobre os investimentos (10,8%). Os dados revelam, ainda, que a cooperativa obteve o terceiro melhor desempenho em liquidez corrente – a empresa tem R\$ 1,44 a receber para cada real investido – e o quinto lugar em investimento no imobilizado, com um índice de 22,3%.

A C.Vale conquistou em 2002 o terceiro melhor desempenho entre as cooperativas que atuam em avicultura e suinocultura no Brasil, conforme levantamento da Fundação Getúlio Vargas para a revista Avicultura Industrial, o troféu A Granja como destaque nacional na produção de trigo pela décima segunda vez, além da certificação ISO2001:2000 para os processos de recebimento, beneficiamento, armazenamento e comercialização de sementes de soja.

No ano de 2003 continuam as reestruturações, modernizações, adequações tecnológicas, padronizações e melhorias na sede e unidades, além da continuidade de abertura de novas unidades de recebimento de produção, tendo-se como meta transformar o crescimento da empresa em um crescimento em escala.

Está prevista a construção e instalação da nova fábrica de rações, da desativadora de soja e a ampliação das atividades do abatedouro de aves que hoje possui capacidade de abate de 150 mil aves/dia para 300 mil aves/dia o que permitirá um aumento do percentual de carne de frango exportado para Europa e

outros. Este crescimento acarretará em desenvolvimento e maior geração de empregos para a região, maior rentabilidade para seus associados, além de contribuir para um melhor posicionamento da empresa frente à concorrência.

Está previsto o início das atividades da central de armazenagem de cereais, onde irá ser centralizada toda a produção agrícola da C.Vale para posterior destinação da mesma (seja para produção de ração e outros derivados, sejam para exportação e/ou revenda). Estima-se que nos próximos dois anos esteja em funcionamento o abatedouro de Suínos que irá abater a produção de suínos fomentada pelo projeto de suinocultura, além de indústria de esmagamento de grãos com a fabricação de óleos, margarina e produtos derivados em geral.

Os números do levantamento “Melhores e Maiores” comprovam o acerto da C.Vale em investir na industrialização e no aumento do recebimento de produtos, esperando com este perfil no mercado se tornar a melhor empresa no segmento de alimentos.

Estes fatores acabam instigando e gerando cada vez mais a necessidade de se desenvolver sistemas que apóiem às gerências nos processos decisórios, passando a ter condições de oferecer novos benefícios aos associados, além de possibilitar o crescimento e agilidade necessários para manter a competitividade.

### 3.2.2 A Infra-estrutura Tecnológica

A C.Vale possui a grande maioria de seus sistemas legados em um grande porte Unisys. O NX6832 é um equipamento de alta performance e escalabilidade que busca a integração entre os ambientes mainframe e plataformas abertas.

Este equipamento procura utilizar o que há de melhor nos dois ambientes, ou seja, a confiabilidade do grande porte com as facilidades do mundo de plataformas abertas, atuando fortemente em parceria com a Microsoft.



O banco de dados utilizado é o DMSII, um banco de dados extremamente seguro devido a suas estruturas de controle, *logs* de recuperação e gerenciadores de transações, sendo que toda esta confiabilidade pode ser acessada através do mundo aberto através de *drivers* ODBC ou OLEDB.

A linguagem de programação utilizada no grande porte é LINC, que ao ser compilado gera código Cobol versão 85.

Por se tratar de um ambiente de grande porte, fornecer vários níveis de segurança, fazer uso de ferramentas de alta produtividade e requerer uma baixa largura de banda para transmissão de dados, este ambiente consolidou-se como a base dos principais, e por que não dizer vitais, sistemas operacionais da empresa. Porém todos estes benefícios acarretaram alguns adiamentos de investimentos, gerando assim uma obsolescência nos equipamentos de transmissão existentes, que em muitos casos, não ultrapassa a marca de 9600 BPS. Este mesmo ambiente dificulta, porém não impossibilita, o desenvolvimento de sistemas com interface gráfica, além de não prover recursos para implementação de sistemas mais aprimorados de análise de informações.

Essa estrutura de comunicação até meados de 1994 era quase que totalmente baseada em terminais caracteres, conectados ao grande porte através de linhas de baixa velocidade (19200 BPS no máximo) e por um protocolo proprietário chamado *Pool-Select*. Já as unidades remotas estavam (e muitas hoje continuam) conectadas através de rádios de baixa frequência que trafegam no máximo a 9600 BPS. As unidades de Santa Catarina e Mato Grosso eram conectadas através de canais de dados de 2400 BPS, sendo hoje conectadas através de canais dedicados que trafegam protocolo TCP/IP encapsulado sobre o protocolo *Frame-Relay*, chegando a velocidade de 128 KBPS.

A partir de meados de 1994 começaram a serem instalados microcomputadores que por suas vezes começaram a ser interligados, tanto ao grande porte através de emuladores de terminais, quanto uns aos outros através de LANs (*Local Área Network*). Começa a surgir uma nova tecnologia de transmissão de dados dentro da empresa.

A partir deste momento surge a utilização do protocolo de comunicação TCP/IP que proporcionou um ganho de produtividade excepcional tanto para a equipe de colaboradores da empresa quanto para a equipe de desenvolvimento de sistemas que passou a utilizar uma ferramenta de desenvolvimento gráfica que armazena todo o repositório de programas em um servidor Intel e posteriormente efetua a transferência dos códigos fontes para o grande porte, onde serão compilados.

Hoje a C.Vale possui uma rede de comunicação mista. Algumas unidades ainda estão conectadas através de linhas de baixa velocidade (rádios ou *links* dedicados), outras estão conectadas em linhas de média velocidade (*links frame-relay* de 64/128 KBPS) e outros estão conectados através de *links* de alta velocidade (10 / 100 MBPS).

O mundo aberto propicia ao mainframe a capacidade de interagir com as principais tecnologias de mercado e principalmente da Internet. É possível, por exemplo, portar rapidamente todo um sistema baseado em cobol no grande porte para um sistema em ASP/JSP através da geração de componentes java que irão compor a aplicação e efetuar a interação com o mesmo. Esta interação só é válida no sentido de padronizar o software cliente, pois como os sistemas continuam a serem executados no grande porte, não é possível aprimorar o nível das consultas executadas onde as limitações continuam prevalecendo.

Vale salientar que o grande porte, que até então só era acessado através de terminais burros e baseados em caracteres, hoje é disponibilizado na Internet através das principais tecnologias de plataformas abertas, além disso, a troca de arquivo entre os dois ambientes, que era feita através de softwares específicos de troca de arquivos com conversão de dados e tudo mais, passou a ser efetuado naturalmente através de ferramentas TCP/IP ou um simples arrastar e soltar do mouse.

Todos estes benefícios estão disponíveis para as unidades conectadas através de tecnologias de transmissão de média e alta velocidade, sendo, nas

demais, utilizado os métodos tradicionais de trabalho, onde isto significa que, não existe comunicação a não ser através do grande porte, ou seja, somente através de emuladores de terminais ou trocas de arquivos através de softwares específicos é possível implementar soluções que necessitem de interação entre os dois mundos.

Porém tudo isso não impediu o crescente aumento da utilização de microcomputadores na empresa. A mesma passou de 25 equipamentos em 1994 para aproximadamente 500 até o presente momento, sendo implementados sistemas que fazem uso dos dois ambientes respeitando-se as limitações existentes.

Muitos dos serviços do grande porte foram portados para este novo mundo da comunicação TCP/IP, porém ainda se têm limitações quanto à performance e recursos para o desenvolvimento de sistemas de gestão.

Este rápido crescimento e as constantes mudanças que ocorrem nas políticas e cenários econômicos, obrigam os colaboradores a buscarem dados mais elaborados para servirem de base à tomada de decisões. Esta busca muitas vezes é saciada através de simples consultas ou relatórios desenvolvidos no próprio grande porte, porém, qualquer consulta mais elaborada não é possível de ser implementada desta forma, visto o consumo de recurso necessário. Isto instiga os funcionários da empresa a iniciar a busca por alternativas para a solução deste problema, sendo a mais utilizada a construção de planilhas eletrônicas através do re-trabalho de digitação ou através de troca de arquivos entre os dois ambientes (sendo esta, a troca de arquivos, amplamente utilizada através de consultas ODBC).

Esta solução além de não atender totalmente aos desejos dos usuários, acarreta sérios problemas de inconsistências de dados, não integração e fragilidade, levando em alguns casos a tomada de decisões pautadas sobre dados que muitas vezes não expressam a realidade da empresa ou setor específico dentro da mesma.

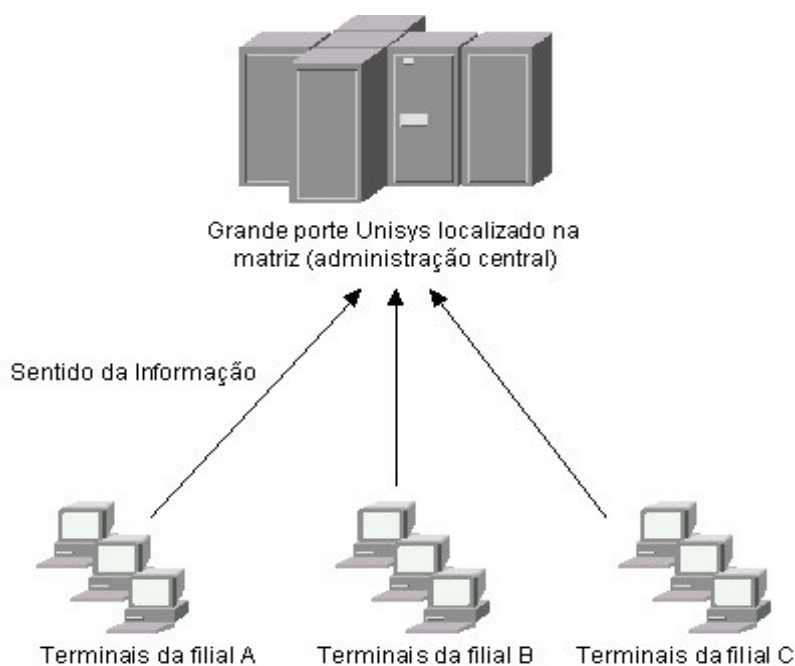
Deve-se mencionar que alguns sistemas foram optados por não se desenvolver na empresa como é o caso do Recursos Humanos, Controle de Matriz e Incubatório de Aves, estes sistemas estão baseados sobre um banco de dados Microsoft SQL Server.

Também alguns sistemas de apoio e controle estão sendo desenvolvidos em Visual Basic com bancos de dados Microsoft Access, Microsoft SQL Server e Interbase / Firebird.

Na Figura 15 está caracterizada a fluxo atual dos dados dos sistemas operacionais da empresa.

Diante deste cenário diversificado de tecnologias de transmissão, processamento e armazenamento de dados, não é possível o desenvolvimento de um sistema de apoio à decisão totalmente integrado/interligado o que obriga a empresa a buscar alternativas para a implementação de sistemas que possibilitem o trabalho distribuído, porém de forma integrada e consistente e que forneça subsídios para a correta tomada de decisão tanto na administração central quanto nas unidades.

Figura 15 - Fluxo dos dados nos sistemas operacionais existentes



### 3.2.3 A Problemática da Centralização da Informação

Os sistemas operacionais responsáveis pela automação de processos possuem como uma de suas principais características manipular um número reduzido de registros (poucas centenas em casos extremos). Em se tratando de um

sistema geograficamente distribuído, este número tende a reduzir-se em decorrência a todas as complicações inerentes as tecnologias de transmissão de dados envolvidas.

Levando-se em consideração os empecilhos de comunicação, atrelados às tecnologias de banco de dados não relacionais, torna-se praticamente impossível disponibilizar através do ambiente de grande porte consultas mais elaboradas e/ou sofisticadas que tendem a manipular muitos registros chegando na casa de milhares ou milhões destes.

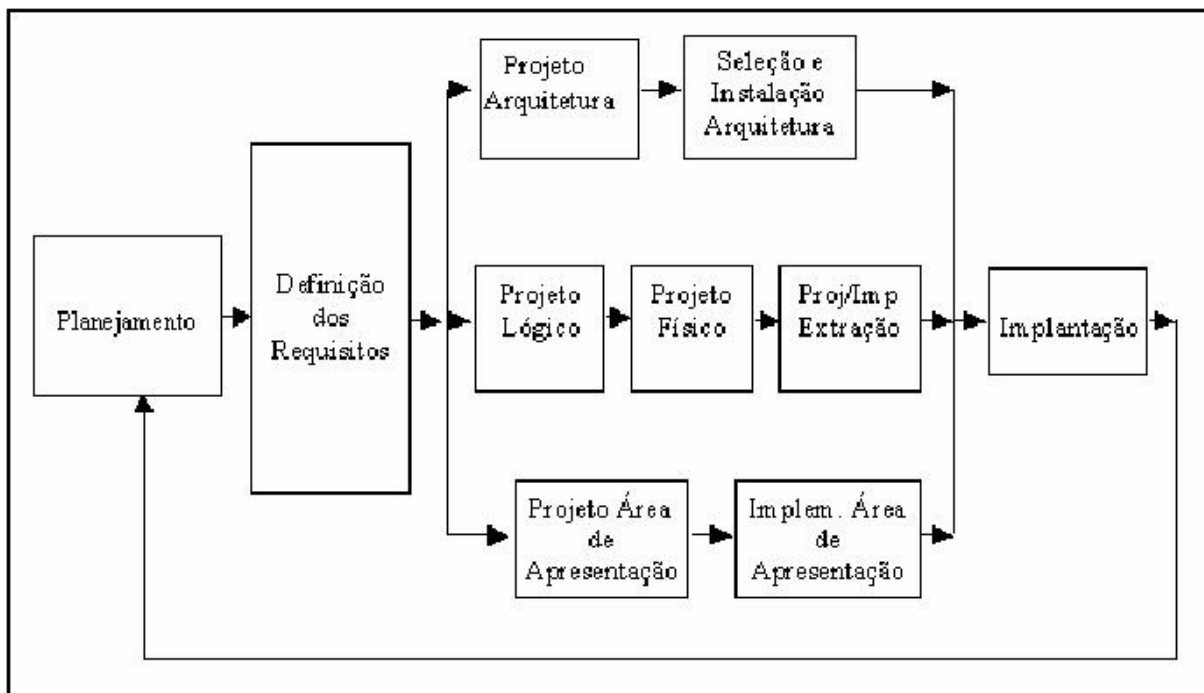
Diante de tudo isso, como os dados somente são destinados à automação de processos, cabe as unidades manipular os sistemas e alimentá-los com os dados operacionais pertinentes a cada filial. As gerências remotas por sua vez limitam-se a simples consultas aos sistemas legados, consultas estas que muitas vezes acabam por não subsidiar corretamente as decisões a serem tomadas, devido ao fato de não ser possível à utilização de recursos mais aprimorados.

Como visto na Figura 15, temos somente um caminho da informação. A mesma parte das fontes alimentadoras dos sistemas, que são os terminais remotos, não retornando em forma de relatórios analíticos para serem utilizados como lastro para a correta tomada de decisão.

### **3.3 O Desenvolvimento da Aplicação**

O SAAG foi desenvolvido utilizando-se a abordagem de implementação incremental (BUS) proposta por Kimball (1998a), onde os requisitos do negócio foram levantados e as dimensões criadas em conformidade. A Figura 16 descreve as fases do desenvolvimento de *data marts* incrementais.

Figura 16 - Fases do desenvolvimento de *data marts* incrementais

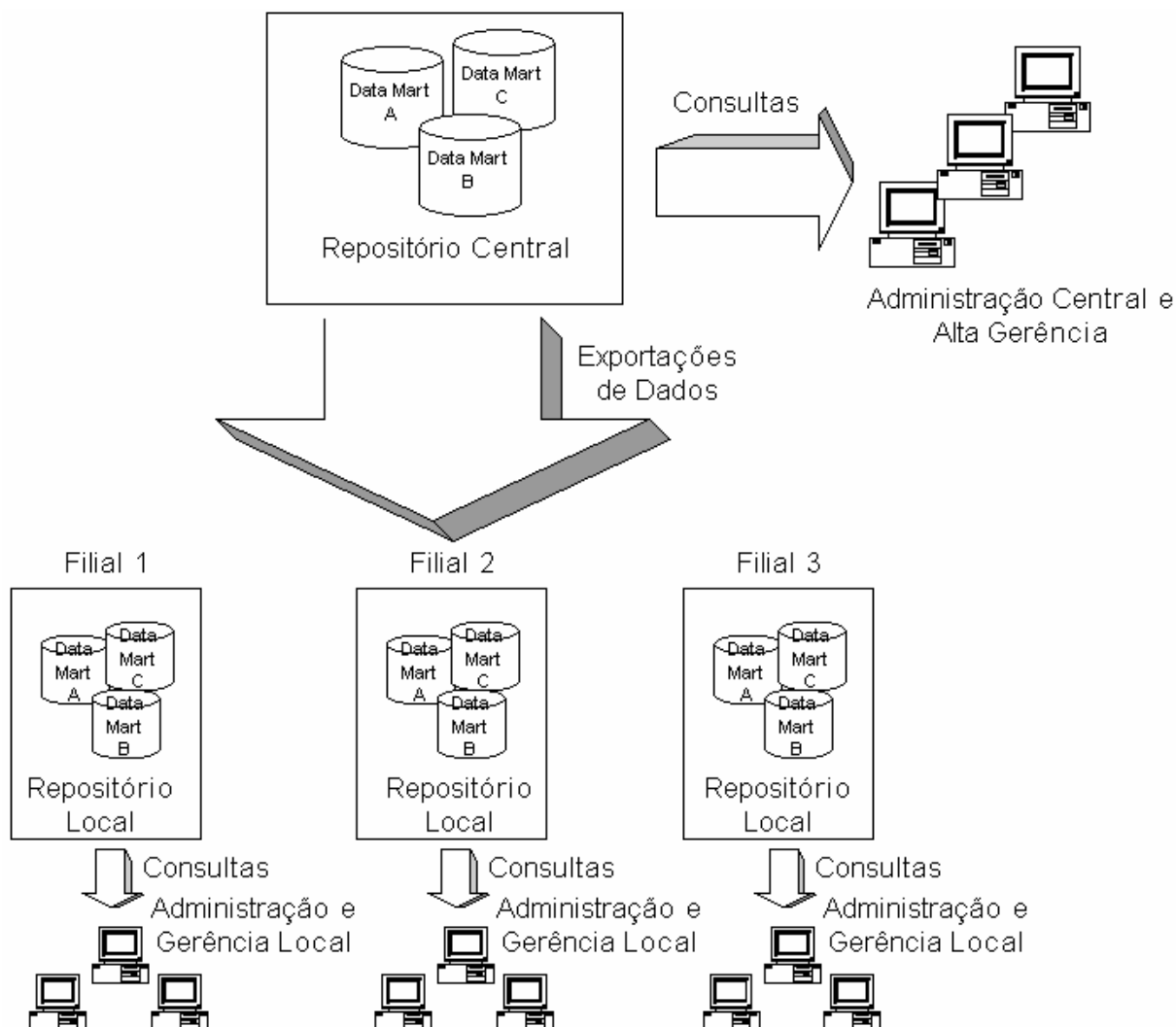


Fonte: SELL, D. Uma arquitetura para distribuição de componentes tecnológicos para sistemas de informações baseados em data warehouse. Florianópolis, p. 41, 2001. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

Inicialmente os *data marts* criados situam-se em um repositório central localizado na matriz da empresa. Este repositório é acessado somente pela administração central e pela alta gerência da empresa. Os dados contidos nos diversos *data marts* centrais criados são então replicados para os diversos *data marts* distribuídos situados nas filiais, sendo o acesso a estes, restrito a gerência local da unidade, não existindo consultas a serem executadas fora deste. A Figura 17 destaca o fluxo das consultas a serem efetuadas tanto no repositório central quanto nos *data marts* distribuídos.

Tanto a área de estágio quanto o repositório central são implementados utilizando-se o Microsoft SQL Server 2000, já os repositórios locais distribuídos, são implementados utilizando-se bancos de dados Interbase / Firebird.

Figura 17 - Fluxo das consultas ao SAAG



### 3.3.1 Definição dos *Data Marts*

A C.Vale nos últimos três anos além de aumentar significativamente seu faturamento, implantou um grande número de novas unidades o que desencadeou a necessidade de se aprimorar os controles sobre as atividades exercidas tanto nas filiais quanto na matriz.

Diante disto iniciou-se o desenvolvimento de alguns sistemas para atender esta necessidade de informações gerenciais. Tais sistemas foram desenvolvidos sobre a mesma plataforma dos atuais softwares de automação de processos, uma plataforma proprietária e pouco flexível no que se refere à extração de informações de caráter gerencial, o que acabou por inviabilizar tal solução.

Baseando-se nas definições do sistema que havia sido desenvolvido e na expressiva movimentação financeira, optou-se dentre as diversas áreas de atividade da empresa, pelo desenvolvimento de *data marts* relacionados à atividade de recebimento e armazenagem de cereais. Tais *data marts* deveriam atender não só o que já havia sido desenvolvido, mas também os requisitos adiados devido à inflexibilidade da plataforma anterior.

### 3.3.2 Extração, padronização e incorporação dos dados

A C.Vale possui seus os principais sistemas corporativos operando em um *Mainframe* Unisys. Estes sistemas são codificados em uma linguagem proprietária chamada LINCII e armazenam os dados operacionais em um banco de dados também proprietário chamado DMSII.

Os dados são extraídos deste ambiente através rotinas que utilizam *drivers* ODBC / OLEDB, troca de arquivos e camadas intermediárias de software que permitem o acesso aos dados armazenados diretamente como se fosse um terminal simulando o acesso de um usuário.

A modelagem é composta por uma única área de estágio, que corresponde ao local onde são armazenados os dados extraídos dos sistemas operacionais. Esta área é composta por uma coleção de tabelas organizadas de modo a possibilitar o reflexo dos dados no ambiente operacional e nos *data marts*.

Todos os processos de extração, transformação, padronização e sumarização dos dados oriundos dos sistemas operacionais existentes são executados sobre esta área de estágio única, e estão centralizados na matriz, facilitando o gerenciamento destes processos e evitando possíveis conflitos decorrentes de padronizações incorretas.

A freqüência de extração bem como a quantidade de registros movimentada varia de acordo com os dados a serem extraídos e o período do ano. No que se refere a dados de orçamento, a extração é efetuada mensalmente e



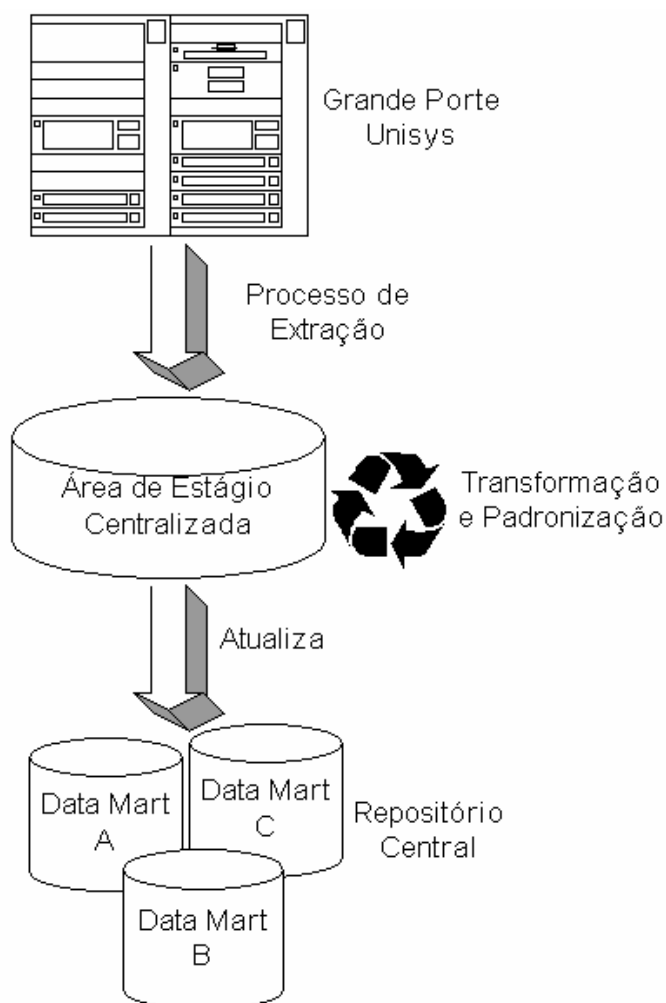
movimenta em média 25.000 registros. Os dados das dimensões são extraídos diariamente e movimentam em média 2.000 registros. Os dados de consumo de lenha são extraídos diariamente e movimentam em média 500 registros nos períodos em que não há colheita de cereais e 1.500 em épocas de colheita. Os dados de recebimento são extraídos diariamente e movimentam em média 500 registros em épocas que não há colheita e 8.000 em épocas de colheita de cereais.

O passo seguinte é a incorporação ao repositório central a partir da área de estágio. O repositório central contém todos os dados na menor granularidade possível, sendo utilizado pela administração central da empresa para efetuar consultas que darão uma visão global das atividades da mesma. Esta característica, dos dados estarem na menor granularidade, permite uma melhor flexibilidade nas consultas a serem efetuadas, bem como fornece um meio simplificado de extração e disponibilização dos dados a serem exportados. No momento da incorporação, é gerada uma nova instância na dimensão auditoria e associada aos fatos que estão sendo carregados, tal associação serve para controlar as exportações e importações bem como possíveis retransmissões dos dados. A Figura 18 destaca o fluxo dos dados partindo dos sistemas operacionais passando pela área de estágio e sendo incorporadas ao repositório central.

### 3.3.3 Exportação e importação dos dados nos *data marts* distribuídos

Uma vez incorporados ao repositório central, os dados são exportados, através de troca de arquivos, para posterior carga nos *data marts* distribuídos. Cada unidade receberá somente os dados pertinentes ao desenvolvimento de suas atividades bem como os dados das dimensões que são replicados na íntegra igualmente a todas as filiais. Desta forma, todas as consultas necessárias são efetuadas a nível local, não sendo permitido, nem necessário, consultas ao repositório central.

Figura 18 - Fluxo dos dados extraídos para o SAAG



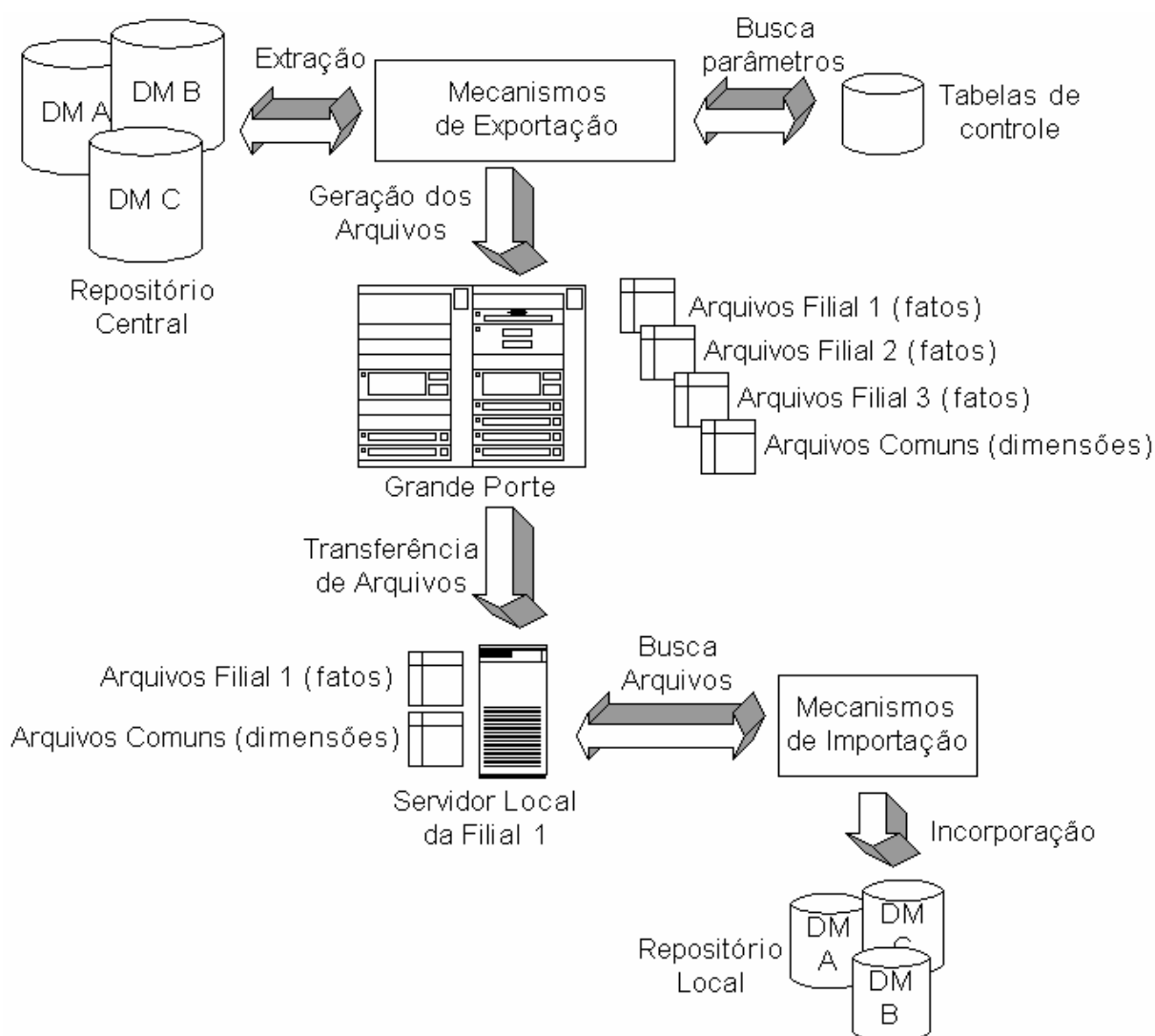
Para controlar as exportações, criaram-se algumas tabelas de controle responsáveis por indicar que tipo de canal de comunicação cada unidade utiliza bem como quais os agrupamentos existentes entre as mesmas. Um agrupamento consiste em subordinar a gerência de uma ou mais unidades a outra, desta forma, a filial que gerencia recebe além dos seus dados, dados pertinentes as unidade gerenciadas enquanto estas, as filiais gerenciadas, recebem somente seus dados.

Todas as trocas de arquivos envolvidas são realizadas através de uma área em disco no grande porte, desta forma tanto as unidades que possuem canais de comunicação de baixa velocidade quanto as que possuem canais maiores, conseguem ter acesso aos arquivos de atualização sendo, a forma de acesso a esses arquivos, a única diferença existente entre os processos de importação. As unidades que possuem canais de baixa velocidade acessam seus arquivos através

de ferramentas proprietárias enquanto as demais acessam através de compartilhamento de disco.

Estando cada filial de posse dos arquivos de atualização, os dados são importados aos DM distribuídos e disponibilizados aos usuários. Vale a pena mencionar que os fatos agregados não são exportados visto que os dados necessários para compô-los estão disponíveis localmente. Esta importação é controlada através de uma dimensão auditoria conforme veremos adiante neste capítulo. A Figura 19 destaca o processo de exportação e importação dos dados.

Figura 19 - Processo de exportação e importação dos dados no SAAG



Caso haja a necessidade de ser efetuada alguma retransmissão, deverão ser selecionadas as unidades à que os dados serão retransmitidos bem como as instâncias da dimensão auditoria que nortearão extração dos fatos.

### **3.4 A Modelagem do *Data Warehouse***

#### 3.4.1 Dimensões

##### a) Tempo

Esta dimensão fornece uma identidade temporal aos fatos que a mesma está relacionada sendo utilizada em todos os fatos desta solução. A Figura 20 destaca os atributos desta dimensão.

##### b) Filial

Esta dimensão caracteriza em que filial ocorreu o fato a que a mesma está relacionada sendo utilizada em todos os fatos desta solução. A Figura 21 destaca os atributos desta dimensão.

##### c) Produto

Esta dimensão caracteriza o produto que está sendo recebido sendo utilizada somente nos fatos de recebimento. A Figura 22 destaca os atributos desta dimensão.

##### d) Grupo

Esta dimensão caracteriza o grupo de estoque a que o fato está relacionado sendo utilizada em todos fatos desta solução. A Figura 23 destaca os atributos desta dimensão.

## e) Conta

Esta dimensão caracteriza a conta orçamentária a que o fato está relacionado sendo utilizada somente nos fatos de orçamento. A Figura 24 destaca os atributos desta dimensão.

## f) Produtor

Esta dimensão caracteriza o produtor que está efetuando a entrega de cereais sendo utilizada somente nos fatos de recebimento. A Figura 25 destaca os atributos desta dimensão.

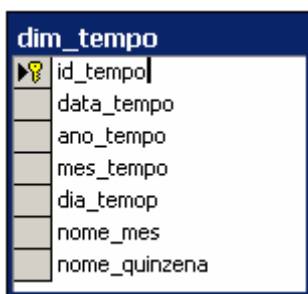
## g) Safra

Esta dimensão caracteriza a estação do ano em que está sendo efetuado o recebimento sendo utilizada somente nos fatos de recebimento. A Figura 26 destaca os atributos desta dimensão.

## h) Auditoria

Esta dimensão é responsável pelo controle das atualizações (exportações, importações e retransmissões) dos fatos. É utilizada em todos os fatos da solução exceto nos agregados, pois os mesmos não são exportados e sim gerados localmente. A Figura 27 destaca os atributos desta dimensão.

Figura 20 - A Dimensão Tempo



dim_tempo	
id_tempo	
data_tempo	
ano_tempo	
mes_tempo	
dia_tempo	
nome_mes	
nome_quinzena	

Figura 21 - A Dimensão Filial

dim_filial	
▶ 🔑	id_filial
▢	codigo_empresa
▢	nome_empresa
▢	codigo_unidade
▢	nome_unidade
▢	codigo_regiao
▢	codigo_uf
▢	nome_uf

Figura 22 - A Dimensão Produto

dim_produto	
▶ 🔑	id_produto
▢	codigo_empresa
▢	nome_empresa
▢	codigo_grupo
▢	nome_grupo
▢	codigo_produto
▢	nome_produto

Figura 23 - A Dimensão Grupo

dim_grupo	
▶ 🔑	id_grupo
▢	codigo_empresa
▢	nome_empresa
▢	codigo_grupo
▢	nome_grupo

Figura 24 - A Dimensão Conta

dim_conta	
▶ 🔑	id_conta
▢	codigo_empresa
▢	nome_empresa
▢	codigo_conta
▢	grau_conta
▢	nome_conta

Figura 25 - A Dimensão Produtor

dim_produtor	
▶ 🔑	id_produtor
▢	codigo_empresa
▢	nome_empresa
▢	matricula_produtor
▢	nome_produtor

Figura 26 - A Dimensão Safra

dim_safra	
▶ 🔑	id_safra
▢	codigo_safra
▢	nome_safra

Figura 27 - A Dimensão Auditoria

dim_auditoria	
▶ 🔑	id_auditoria
▢	data_auditoria
▢	hora_auditoria

### 3.4.2 Fatos

#### a) Consumo de Lenha

Este fato representa o consumo de lenha utilizado para efetuar a secagem da produção agrícola recebida em cada unidade. Através desta tabela de fatos é possível identificar qual unidade está sendo mais eficiente na utilização dos recursos térmicos disponíveis para suas atividades, bem como, auxiliar na reestruturação de processos que por ventura estejam sendo executados de forma incorreta ou ineficiente. Esta tabela de fatos é composta por quatro dimensões: tempo, filial, grupo e auditoria e contém 4.356 registros no repositório central. A Figura 28 destaca os atributos deste fato bem como relaciona as dimensões associadas ao mesmo.

## b) Recebimento de Cereais

Este fato representa o recebimento de cereais provenientes das lavouras dos produtores. Através destes fatos é possível identificar o aumento ou o decréscimo da entrega de cereais em cada unidade, direcionando para os motivos que estão levando a estes resultados. Também é possível identificar os níveis de padronização dos cereais recebidos, pois quanto mais padrão forem, menores serão os custos para padronizá-los e conseqüentemente mais lucrativa será a unidade e a empresa como um todo. Esta tabela de fatos utiliza sete dimensões: tempo, filial, grupo, produto, safra, produtor, auditoria e contém 430.600 registros no repositório central. A Figura 29 destaca os atributos deste fato bem como relaciona as dimensões associadas ao mesmo.

## c) Orçamento

Este fato representa a relação entre o orçamento previsto e o realizado, identificando aumentos ou decréscimos de lucros, custos, entre outros. Também é utilizada para acompanhar as metas da empresa no que tange valores previstos e realizados e quais motivos estão interferindo no bom andamento destas. Esta tabela de fatos utiliza cinco dimensões: tempo, filial, conta, grupo, auditoria e contém 750.489 registros no repositório central. A Figura 30 destaca os atributos deste fato bem como relaciona as dimensões associadas ao mesmo.

Figura 28 - Fato Consumo de Lenha





Figura 29 - Fato Recebimento

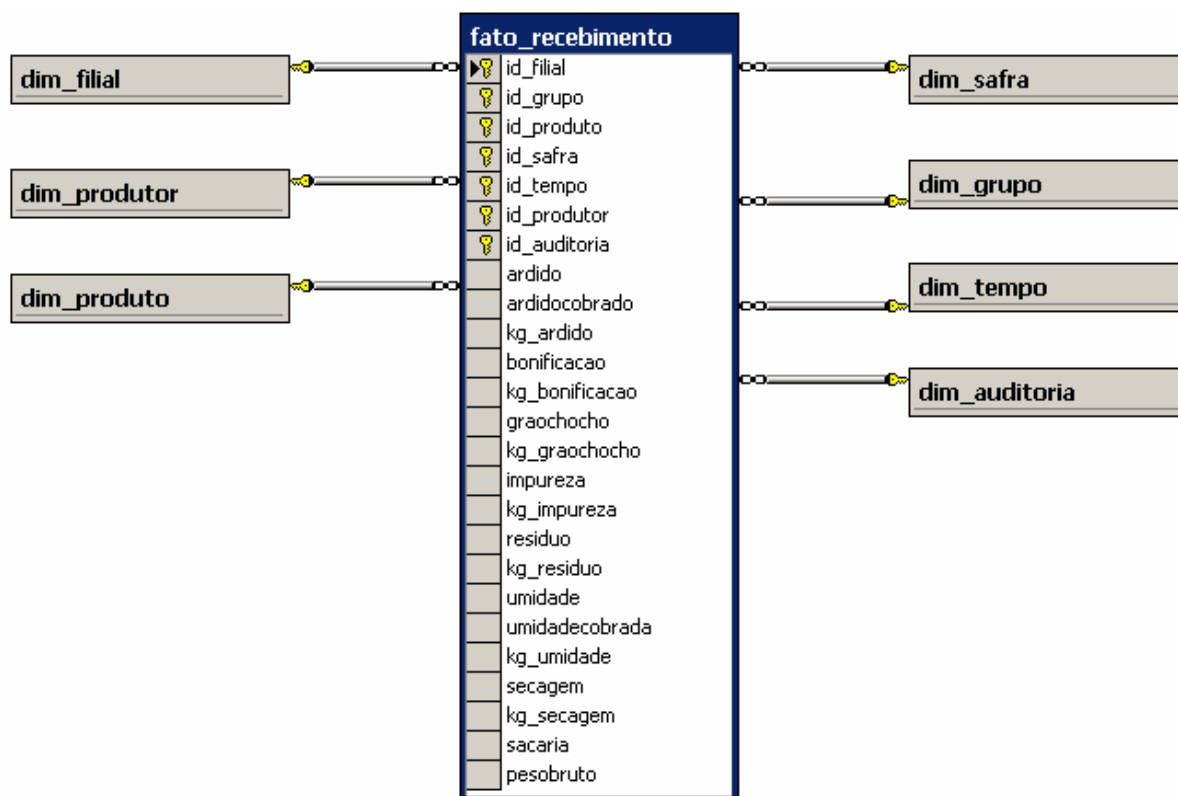


Figura 30 - Fato Orçamento



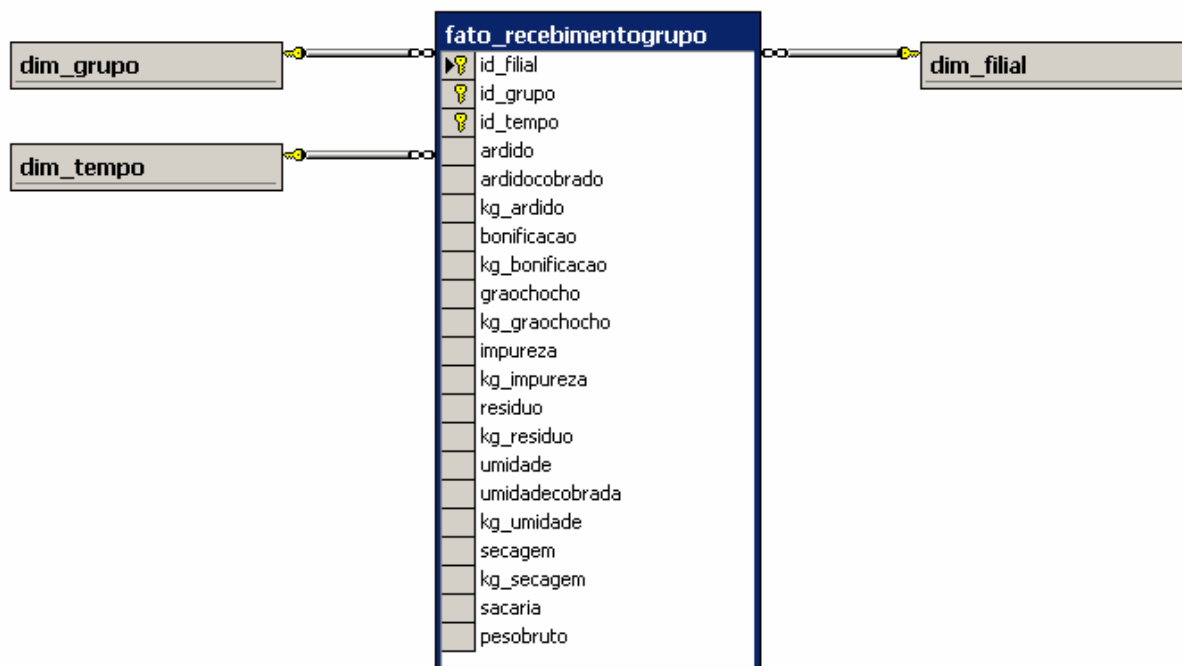
### 3.4.3 Agregados

#### a) Recebimento por Grupo de Estoque

Este agregado da tabela de fatos de recebimento sumariza os dados utilizando as dimensões filial, grupo e tempo. Este agrupamento foi feito em

decorrência do alto número de consultas que são efetuadas com base nessas dimensões. Este agregado possui 30.251 registros. A Figura 31 destaca os atributos deste agregado bem como relaciona as dimensões associadas ao mesmo.

Figura 31 - Agregado Recebimento por Grupo de Estoque



### 3.5 Visualização dos dados

A visualização dos dados tanto no *data warehouse* central quanto nos *data marts* distribuídos é efetuada de modo local, ou seja, os usuários da administração central somente consultam dados existentes no repositório central enquanto que as gerências das filiais somente consultam seu repositório local.

A visualização dos dados oriundos do DW é basicamente efetuada através do uso de uma ferramenta desenvolvida especificamente para a aplicação proposta. Este software faz parte da solução nomeada SAAG e foi desenvolvido na linguagem de programação Microsoft Visual Basic - Versão 5.0.

Esta ferramenta possibilita a elaboração de algumas consultas pré-estabelecidas sendo bastante flexível na parametrização das mesmas (esta

parametrização permite a inserção de novos agrupamentos e colunas bem como a remoção dos mesmos). Além de consultas as tabelas de fatos individualmente, também foram implementados um cruzamento entre as informações contidas nas tabelas de fatos consumo de lenha e recebimento bem como uma composição de dados entre as tabelas de fatos de orçamento e recebimento que origina uma matriz de indicadores necessária para avaliar o desempenho das unidades e da empresa como um todo. Tal composição dos dados faz uso além das tabelas de dimensão e fatos, de algumas tabelas adicionais responsáveis por indicar como deverá ser efetuada a mesclagem dos dados.

Também são utilizados, em menor escala, o Microsoft Excel e o Microsoft Query para extração e visualização dos dados contidos no DW. Tais ferramentas possibilitam a geração de gráficos e relatórios impressos, características estas que a ferramenta desenvolvida citada anteriormente não contempla. Algumas consultas mais rotineiras encontram-se pré-programadas para facilitar o uso enquanto outras ficam completamente a revelia do usuário, podendo este criar novas consultas no momento em que houver necessidade e de acordo com as necessidades emergentes.

### **3.6 Resultados Obtidos**

Após a construção e implantação do SAAG, alguns processos puderam ser analisados. Tais análises resultaram em novos processos ou readequação dos processos antigos que, em muitos casos, ocasionaram a substituição de equipamentos (fornalhas) e/ou troca dos materiais utilizados nos processos de secagem (lenha para cavaco e outros). Os novos investimentos desencadeados foram viabilizados através de análises de dados extraídos da ferramenta.

Segundo alguns gerentes de unidades, em muitos casos constatou-se que diversos custos poderiam ser diminuídos se os percentuais de umidade constatados nos produtos que estavam sendo entregues pelos produtores fossem diminuídos ou aumentados. No processo de armazenagem e negociação de grãos, estipula-se um grau de umidade ideal, desta forma, produtos que estejam acima devem ser secos enquanto produtos que estiverem abaixo devem ser umedecidos.

Tanto para o produtor quanto para a cooperativa, quanto maior for o padrão de unidade dos produtos, maiores serão os lucros visto que, os custos envolvidos na padronização são absorvidos por ambas as partes.

Os responsáveis pela unidade de Palotina afirmam que o processo de recebimento de cereais pode ser avaliado de uma forma mais abrangente. A fidelidade dos produtores pode ser observada e conseqüentemente medidas puderam ser tomadas. Procurou-se conscientizar os produtores quanto aos benefícios inerentes a fidelidade junto à cooperativa. Tais benefícios incluem um melhor preço no momento da comercialização, o fortalecimento da empresa frente à concorrência e um maior retorno ao próprio produtor que também faz parte do quadro social da empresa e que por sua vez recebe a sua parte nos lucros da mesma.

A administração central pode avaliar melhor o desempenho de cada unidade bem como agir em prol de melhorias, re-alinhando as atividades das mesmas e garantindo os lucros da empresa como um todo. Os índices operacionais que muitas vezes levavam dias para serem obtidos, hoje em questão de minutos estão disponíveis. Com o SAAG, este processo além de agilidade gerou maior confiança e segurança nas informações geradas, afirma o responsável pela análise dos índices operacionais.

Os resultados puderam ser avaliados (lucros, despesas, entre outros). Acréscimos de despesas puderam ser controlados e investimentos em atividades que alavancam maior faturamento puderam ser priorizadas.

## 4. CONCLUSÃO E ESTUDOS FUTUROS

### 4.1 Conclusão

A falta de informações para gerir qualquer atividade empresarial está se tornando um dos fatores mais críticos relacionados à sobrevivência das empresas. O cenário econômico globalizado não tolera mais decisões pautadas somente sobre a experiência profissional dos executivos envolvidos muito menos perdoa decisões equivocadas tomadas com base em informações errôneas geradas a partir de fontes duvidosas e imprecisas.

Diante deste contexto, este trabalho propôs um modelo de aplicação de *Data Warehouse* Centralizado com *Data Marts* Distribuídos visando suprir tanto a necessidade de informações quanto a qualidade e confiabilidade das mesmas.

Com a implantação do SAAG e a conseqüente disponibilização de informações de caráter gerencial às unidades eliminou-se o desencontro de dados e, a problemática da centralização da informação, que era um dos fatores de maior relevância e motivação para o desenvolvimento deste modelo, foi solucionada.

Segundo depoimentos, antes da utilização da solução proposta, as informações estratégicas eram compostas através do uso de planilhas eletrônicas. Tais planilhas ajudavam nos processos decisórios, porém necessitavam de muito re-trabalho e levavam muito tempo para serem elaboradas devido à quantidade de unidades e fatores além de, perigosamente em muitos casos, possuírem diferenças que poderiam levar a decisões erradas. Essas informações estratégicas agora podem ser extraídas com muito mais agilidade e segurança, fornecendo um base sólida para as decisões serem embasadas, pois o tempo, que anteriormente era utilizado para a elaboração das planilhas, hoje é utilizado para análises mais aprimoradas que desencadeiam controles e atitudes mais eficientes.

O setor agrícola possui suas margens de lucros bastante reduzidas e, em especial a C.Vale, prioriza seus investimentos nas áreas de negócios da empresa, sendo, os investimentos destinados às áreas de apoio, e dentro destas encontra-se

a área de TI, bastante limitados. Investir em um projeto inédito e desconhecido, com certeza não faz parte do cenário da C.Vale.

Desenvolver um sistema de apoio à decisão que englobe toda a empresa por si só caracteriza-se como um desafio a qualquer equipe de tecnologia da informação, desenvolver um sistema desta natureza considerando-se o não envolvimento de custos adicionais, o uso dos meios de transmissão de dados existentes com suas deficiências tecnológicas e, a abrangência global da empresa, tornaram-se um grande desafio durante a construção da solução apresentada.

A resistência e o descrédito por parte de alguns profissionais de TI frente a uma solução até o momento inexistente na empresa, bem como a quebra do paradigma da desnormalização do modelo de dados foram fatores que aumentaram a complexidade e as dificuldades no decorrer do desenvolvimento da aplicação.

Muitos afirmam que o analfabeto do terceiro milênio é o analfabeto digital, ou seja, aquele que não atualiza seus conhecimentos quanto ao uso da tecnologia da informação no desenvolvimento de suas atividades pessoais e profissionais. A falta de conhecimento e o medo do uso da informática, juntamente com o contexto da centralização da informação que perdurou por anos, foram fatores agravantes no que se refere ao apoio durante o desenvolvimento da pesquisa, pois, muitos profissionais, não sentiam-se à vontade e impuseram resistências quanto a utilizar o sistema proposto no auxílio de suas atividades.

Os atuais sistemas OLTP da C.Vale são baseados em um grande porte UNISYS. Tais sistemas por serem desenvolvidos com ferramentas de produtividade voltadas à automatização de processos não permitem flexibilidade para a construção de relatórios mais elaborados o que acabava por não permitir a extração de algumas informações bem como inibe a criatividade dos usuários.

Com a implantação deste modelo, além do problema citado no parágrafo anterior, também foi reduzida a dependência dos usuários quanto aos profissionais de TI, visto que, para cada visualização e/ou combinação diferente dos dados, era necessário o desenvolvimento de novos programas, muitas vezes utilizados somente

em situações momentâneas. Tais consultas passaram a ser efetuadas diretamente pelo usuário não dependendo de profissional algum e proporcionando grande flexibilidade na combinação, visualização e composição dos dados, além de permitir um acompanhamento mais eficiente visto que os mesmos, os dados, passaram a estar disponíveis diariamente para a elaboração das consultas necessárias.

A solução proposta, segundo usuários da administração central, favoreceu um controle mais aprimorado sobre os custos de recebimento de grãos através de análises no que tange o recebimento, armazenamento e processamento de cereais. Modelos de gerência, que resultaram em maior lucratividade, puderam ser detalhados e conseqüentemente replicados entre unidades com mesmo perfil. Esta solução permitiu também a identificação de deficiências e procedimentos errôneos existentes nas unidades que levavam às mesmas a ter um desempenho abaixo do esperado.

## 4.2 Estudos Futuros

Na seqüência deste trabalho pretende-se estudar a viabilidade de implementação de uma estrutura de agentes responsáveis pela replicação dos dados entre o repositório central e as unidades distribuídas visto que, num período de dois anos estima-se que toda a infra-estrutura de comunicação de dados terá sido modernizada e as velocidades atingirão pelo menos 64 KBPS de banda passante rodando protocolo TCP/IP.

Também se pretende avaliar a viabilidade de implantação de uma estrutura de agentes responsáveis por identificar as consultas mais utilizadas em cada unidade bem como o aumento ou decréscimo de desempenho da filial e, com base nestes dados, eliminar, criar novas ou replicar consultas existentes entre os *data marts* distribuídos de acordo com os desempenhos obtidos.

Por se tratar de um projeto inovador dentro da empresa, praticamente cada área de negócio da empresa possui pelo menos uma atividade a ser modelada em um *data mart*. Tais áreas já solicitaram tal desenvolvimento e estima-se que nos

próximos dois anos esta tecnologia esteja totalmente consolidada dentro da empresa.



## 5. REFERÊNCIAS BIBLIOGRÁFICAS

BALLARD, C. et Al. **Data Modeling Techniques for Data Warehousing**. IBM Corporation. Disponível em: <http://www.redbooks.ibm.com>. Acesso em 30/11/2001.

BRACKETT, M. H. **The Data Warehouse Challenge. Taming Data Chaos**. New York: John Wiley & Sons Inc., 1996.

CRUZ, T. **Sistemas de Informações Gerenciais**. 2ª ed. São Paulo: Atlas, 2000.

DOMENICO, J. A. **Definição de um Ambiente Data Warehouse em uma Instituição de Ensino Superior**. Florianópolis, 2001. 137 f. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

DWBRASIL. **Data Mart**. Disponível em <http://www.dwbrasil.com.br/html/dmart.html>. Acesso em 30/11/2001.

FALSARELLA, O. M. e CHAVES, E. O. C. **Sistemas de Informação e Sistemas de Apoio a Decisão**. Disponível em: <http://chaves.com.br/TEXTSELF/COMPUT/sad.htm>. Acesso em 30/07/2001.

FURLAN, J.D.; IVO, I. M. e AMARAL, F. P. **Sistemas de Informação Executiva – EIS**. São Paulo: Makron Books, 1994.

GUPTA, V. R. **Na Introduction to Data Warehousing**. Disponível em: <http://system-services.com/dwintro.asp>. Acesso em 30/01/2002.

HARRISON, T. **Intranet Data Warehouse**. São Paulo: Berkeley, 1998.

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

INMON, W. H. e HACKARTHORN, R. D. **Como usar o Data Warehouse**. Rio de Janeiro: IBPI Press, 1997.

INMON, W. H.; WELCH, J. D. e GLASSEY, K. L. **Gerenciando Data Warehouse: Técnicas Práticas para Monitorar Operações de Performance, Administrar Dados e Ferramentas, Gerenciar Alterações e Crescimento**. São Paulo: Makron Books, 1999.

KIMBALL, R. **Data Warehouse Toolkit**. São Paulo: Makron Books, 1998a.

KIMBALL, R. **The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing Data Warehouses**. New York: Wiley Computer Publishing, 1998b.

KIMBALL, R. e MERZ R. **Data Webhouse**. Construindo o data warehouse para a Web. Rio de Janeiro: Campus, 2000.

KIMBALL, R. e ROSS M. **The Data Warehouse Toolkit**. Guia completo para modelagem dimensional. Rio de Janeiro: Campus, 2002.

LAUDON, K. C. e LAUDON, J. P. **Sistemas de Informação**. Rio de Janeiro: Livros Técnicos e Científicos, 1998.

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

MEIRELLES, F. S. **Informática - Novas Aplicações com Microcomputadores**. São Paulo: Makron Books, 1994.

MELLO, J. A. B. **Uma Proposta de Modelo de Dados para Suporte ao Processamento Transacional e de Apoio Informacional Simultaneamente**. Florianópolis, 2002. 101 f. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

OLIVEIRA, D. P. R. **Sistemas de Informações Gerenciais**. São Paulo: Atlas, 1998.

PRESSMAN, R. S. **Engenharia de Software**. São Paulo: Makron Books, 1995.

SELL, D. **Uma Arquitetura para Distribuição de Componentes Tecnológicos de Sistemas de Informações Baseados em Data Warehouse**. Florianópolis, 2001. 90 f. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

SING, H. S. **Data Warehouse**. Conceitos, Tecnologias, Implementação e Gerenciamento. São Paulo: Makron Books, 2001.

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

THOMSEN, E. **OLAP**. Construindo Sistemas de Informações Multidimensionais. Rio de Janeiro: Campus, 2002.

VASCONCELLOS, J. M. Implementando um Data Warehouse Incremental. **DEVELOPERS CIO MAGAZINE**, Rio de Janeiro, n.38, p. 18-20, abr. 1999.

## APÊNDICE A – VISUALIZAÇÃO DOS DADOS

Figura 32 - Tela inicial do SAAG

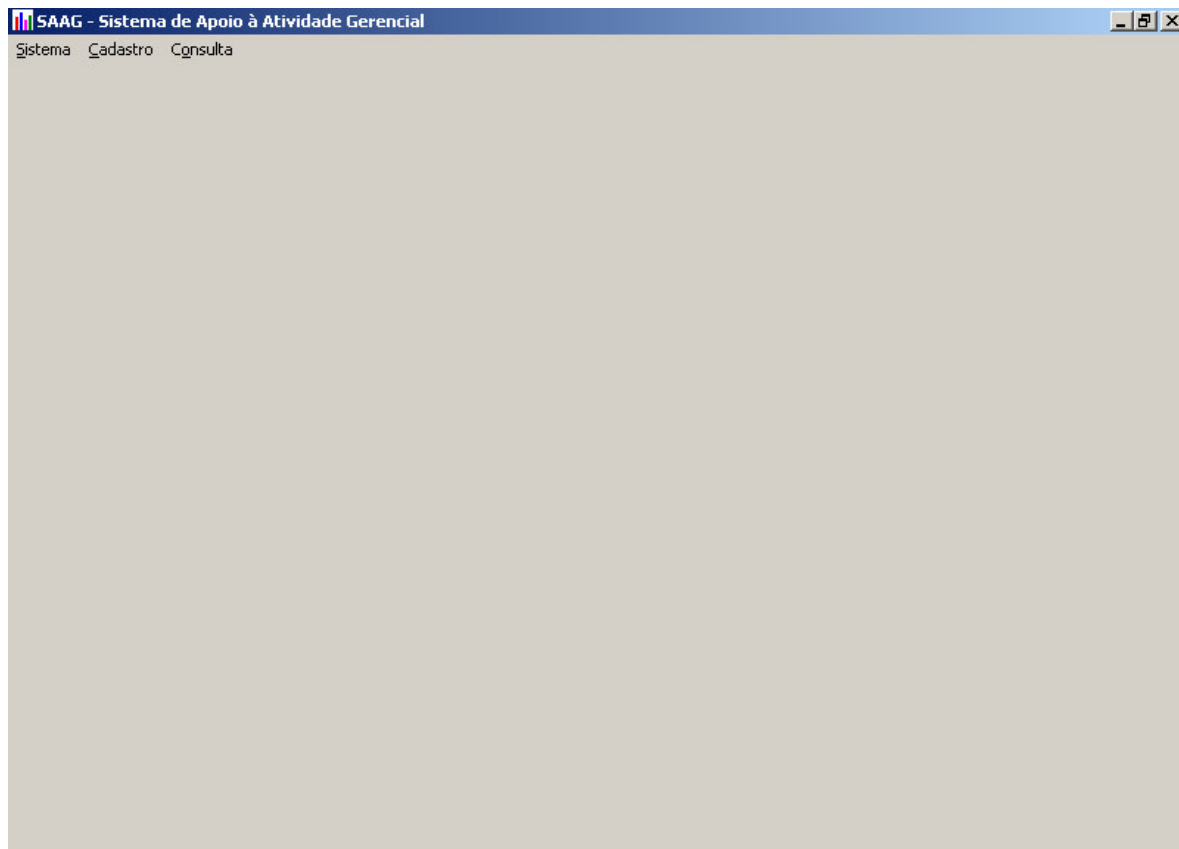


Figura 33 - Parametrização das consultas no SAAG

SAAG - Sistema de Apoio à Atividade Gerencial

Sistema Cadastro Consulta

### Recebimento

Agrupado por	Critérios
<input type="checkbox"/> Empresa	Empresa
<input type="checkbox"/> UF	Uf
<input checked="" type="checkbox"/> Filial	Filial 01,03
<input type="checkbox"/> Grupo	Grupo
<input checked="" type="checkbox"/> Produto	Produto
<input type="checkbox"/> Produtor	Produtor
<input type="checkbox"/> Safra	Safra
<input type="checkbox"/> Ano	Ano
<input checked="" type="checkbox"/> Mes	Mês
<input type="checkbox"/> Quinzena	Quinzena

**Período**

Data Inicial: 01/01/2002      Data Final: 31/05/2003

Pesquisa      Sair

Start    Usuário    Saag - Microsoft...    SAAG - Sistema ...    Documento1 - M...    14:34

Figura 34 - Resultados de uma pesquisa no SAAG

Consulta Recebimento					
Unidade	Produto	Mês	KG Brutos	KG Líquidos	
001 - PALOTINA	001 - SOJA COMERCIAL	01 - JANEIRO	3.775.300,00	3.611.684,00	
001 - PALOTINA	001 - SOJA COMERCIAL	02 - FEVEREIRO	67.406.250,00	64.525.717,00	
001 - PALOTINA	001 - SOJA COMERCIAL	03 - MARÇO	66.845.015,00	64.840.335,00	
001 - PALOTINA	001 - SOJA COMERCIAL	04 - ABRIL	1.710.300,00	1.662.317,00	
001 - PALOTINA	001 - SOJA COMERCIAL	05 - MAIO	538.120,00	530.305,00	
001 - PALOTINA	001 - SOJA COMERCIAL	06 - JUNHO	50.080,00	44.061,00	
001 - PALOTINA	001 - SOJA COMERCIAL	07 - JULHO	50.040,00	49.727,00	
001 - PALOTINA	001 - SOJA COMERCIAL	08 - AGOSTO	580,00	566,00	
001 - PALOTINA	001 - SOJA COMERCIAL	09 - SETEMBRO	14.730,00	14.730,00	
001 - PALOTINA	001 - SOJA COMERCIAL	10 - OUTUBRO	6.940,00	6.940,00	
001 - PALOTINA	001 - SOJA COMERCIAL	11 - NOVEMBRO	1.480,00	1.472,00	
001 - PALOTINA	001 - SOJA COMERCIAL	12 - DEZEMBRO	2.920,00	2.908,00	
001 - PALOTINA	003 - MILHO COMERCIAL	01 - JANEIRO	6.230.259,00	5.554.009,00	
001 - PALOTINA	003 - MILHO COMERCIAL	02 - FEVEREIRO	16.942.968,00	15.323.039,00	
001 - PALOTINA	003 - MILHO COMERCIAL	03 - MARÇO	1.398.520,00	1.343.180,00	
001 - PALOTINA	003 - MILHO COMERCIAL	04 - ABRIL	518.885,00	497.183,00	
001 - PALOTINA	003 - MILHO COMERCIAL	05 - MAIO	66.250,00	63.260,00	
001 - PALOTINA	003 - MILHO COMERCIAL	06 - JUNHO	1.070.370,00	925.444,00	
001 - PALOTINA	003 - MILHO COMERCIAL	07 - JULHO	9.323.730,00	7.923.575,00	
001 - PALOTINA	003 - MILHO COMERCIAL	08 - AGOSTO	15.564.470,00	13.569.050,00	
001 - PALOTINA	003 - MILHO COMERCIAL	09 - SETEMBRO	350.170,00	318.341,00	
001 - PALOTINA	003 - MILHO COMERCIAL	10 - OUTUBRO	115.730,00	105.073,00	
001 - PALOTINA	003 - MILHO COMERCIAL	11 - NOVEMBRO	516.770,00	513.184,00	
001 - PALOTINA	003 - MILHO COMERCIAL	12 - DEZEMBRO	1.546.880,00	1.546.829,00	
001 - PALOTINA	005 - ARROZ COMERCIAL	02 - FEVEREIRO	7.000,00	6.111,00	
001 - PALOTINA	007 - ALGODÃO COODET	04 - ABRIL	530,00	517,00	
001 - PALOTINA	010 - SORGO COMERCIAL	05 - MAIO	21.090,00	17.357,00	
001 - PALOTINA	011 - SOJA SEMENTE A	06 - JUNHO	151.020,00	144.956,00	
001 - PALOTINA	011 - SOJA SEMENTE A	07 - JULHO	68.130,00	65.982,00	
001 - PALOTINA	012 - TRIGO SEMENTE	08 - AGOSTO	95.460,00	86.482,00	

Sair

Figura 35 - Exemplo de relatório obtido com o Microsoft Excel

The image shows a screenshot of the Microsoft Excel 2003 interface. The title bar reads 'Microsoft Excel - saag.xls'. The menu bar includes 'Arquivo', 'Editar', 'Exibir', 'Inserir', 'Formatar', 'Ferramentas', 'Dados', 'Janela', and 'Ajuda'. The toolbar contains various icons for file operations, editing, and formatting. The status bar at the bottom shows 'A1' and 'fx Filial'. The main data area is a table with the following content:

	A	B	C	D	E
1	Filial	Grupo	Ano	Quantidade	
2	ABATEDOURO DE AVES CVALE	ABATEDOURO	2002	11.423.840,00	
3	ABATEDOURO DE AVES CVALE	ABATEDOURO	2003	731.750,00	
4	ABELARDO LUZ	MILHO EM GRAOS	2002	150.000,00	
5	ABELARDO LUZ	SOJA EM GRAOS	2002	60.000,00	
6	ALTO SANTA FE	MILHO EM GRAOS	2002	273.380,00	
7	ALTO SANTA FE	SOJA EM GRAOS	2002	53.760,00	
8	AMIDONARIA	FECULAVAMIDO DE MANDIOCA	2002	5.887.880,00	
9	AMIDONARIA	FECULAVAMIDO DE MANDIOCA	2003	238.430,00	
10	AMIDONARIA NAVEGANTES	FECULAVAMIDO DE MANDIOCA	2002	1.484.680,00	
11	AMIDONARIA NAVEGANTES	FECULAVAMIDO DE MANDIOCA	2003	192.470,00	
12	ASSIS CHATEAUBRIAND	MILHO EM GRAOS	2002	684.640,00	
13	ASSIS CHATEAUBRIAND	PROD.SEMENTE DE TRIGO	2002	42.630,00	
14	ASSIS CHATEAUBRIAND	SOJA EM GRAOS	2002	235.130,00	
15	ASSIS CHATEAUBRIAND	TRIGO EM GRAOS	2002	350.410,00	
16	CANDEIA	MILHO EM GRAOS	2002	642.572,00	
17	CANDEIA	PROD.SEMENTE DE NABO	2002	17.400,00	
18	CANDEIA	TRIGO EM GRAOS	2002	93.040,00	
19	DIAMANTINO II	MILHO EM GRAOS	2002	105.720,00	
20	ENCANTADO D OESTE	MILHO EM GRAOS	2002	465.100,00	

Figura 36 - Exemplo de gráfico de recebimento gerado com Microsoft Excel

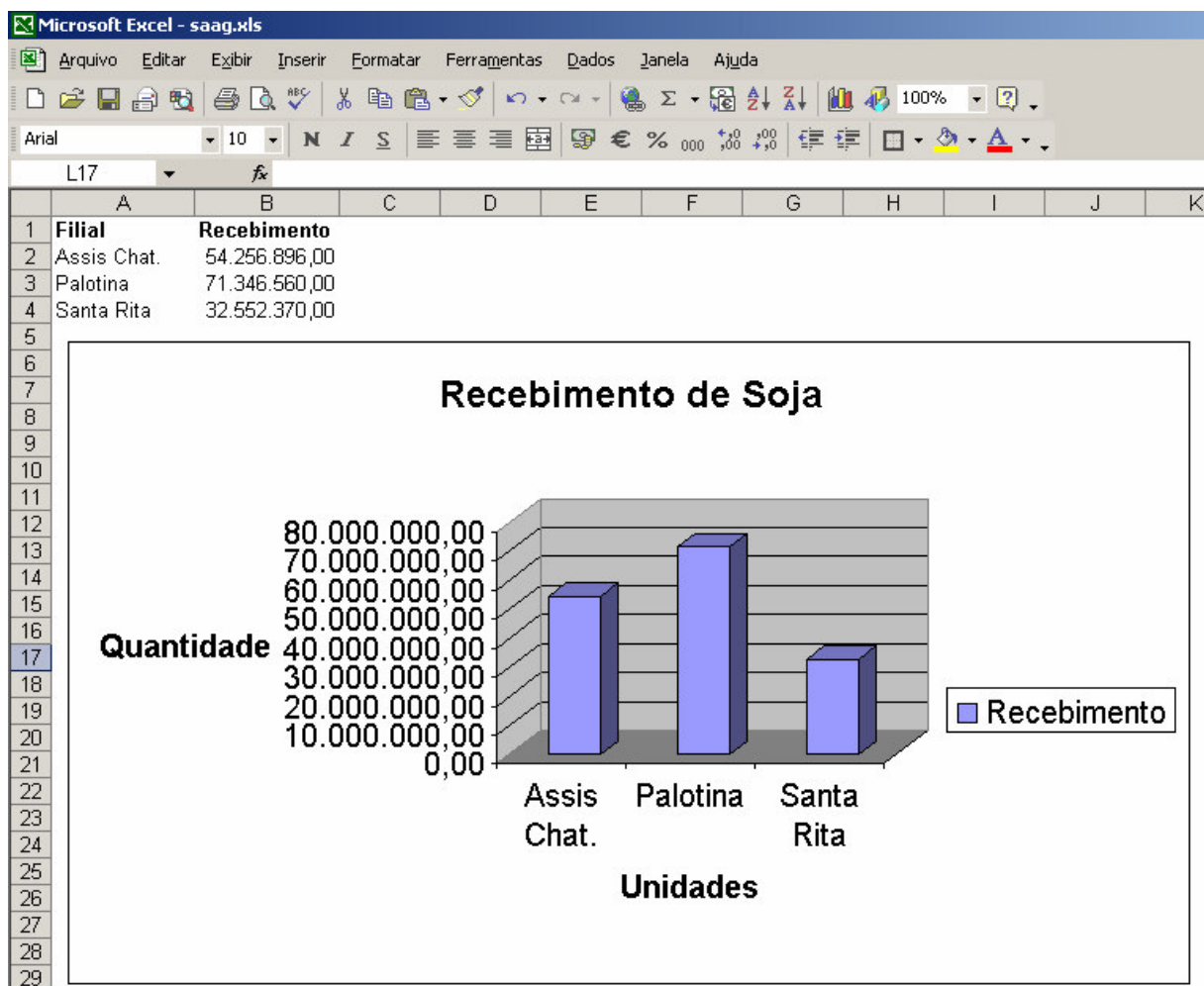


Figura 37 - Exemplo de gráfico de custos gerado com o Microsoft Excel

