

Universidade Federal de Santa Catarina
Programa de Pós-Graduação em
Engenharia de Produção

DEFINIÇÃO DE UM AMBIENTE DATA WAREHOUSE
EM UMA INSTITUIÇÃO DE ENSINO SUPERIOR

Jorge Antonio Di Domenico

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

Florianópolis
2001

Jorge Antonio Di Domenico

**DEFINIÇÃO DE UM AMBIENTE DATA WAREHOUSE
EM UMA INSTITUIÇÃO DE ENSINO SUPERIOR**

Esta dissertação foi julgada e aprovada para a
obtenção do título de **Mestre em Engenharia de
Produção** no Programa de Pós Graduação em
Engenharia de Produção da
Universidade Federal de Santa Catarina

Florianópolis, 28 de fevereiro de 2001.

Prof. Ricardo Miranda Barcia, Ph.D.
Coordenador do Curso

BANCA EXAMINADORA

Prof. Roberto Carlos dos Santos Pacheco, Dr.
Orientador

Prof. Alexandre L. Gonçalves, M. Eng.
Tutor de Orientação

Prof. João Bosco da Mota Alves, Dr.

Prof. José Leomar Todesco, Dr.

Agradecimentos

Ao Prof. Roberto Carlos dos Santos Pacheco, Dr. por ter oportunizado minha entrada no Programa de Pós-Graduação em Engenharia de Produção, aceitando-me, inicialmente como seu orientando, o qual me conduziu e orientou com excelência durante todos os momentos necessários.

Ao Prof. Ricardo Miranda Barcia, Ph.D. coordenador geral do PPGEP, por ter viabilizado esse curso através do ensino a distância (na modalidade presencial virtual).

À UNOESC - Campus de Chapecó, em especial aos amigos Prof. Odilon Luiz Poli, Prof. Gilberto Luiz Agnolin, Profa. Silvana Marta Tumelero e Profa. Arlene Renck, pelo apoio no desenvolvimento dessa pesquisa, bem como pela iniciativa de viabilizarem esse convênio via FUNCITEC.

Ao amigo Alexandre Gonçalves do laboratório Stela, pelos seus apontamentos, os quais vieram enriquecer em muito esse trabalho.

Ao amigo e colega José Alexandre De Toni (Zéca), pelo companheirismo e troca de idéias a cerca da presente pesquisa.

A minha esposa e filhos pelo grande apoio e carinho, fundamentais para conclusão deste trabalho.

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

SUMÁRIO

Agradecimentos.....	iii
SUMÁRIO.....	iv
LISTA DE FIGURAS.....	viii
LISTA DE TABELAS.....	ix
LISTA DE ABREVIATURAS.....	x
RESUMO.....	xi
ABSTRACT.....	xii
1 INTRODUÇÃO.....	13
1.1 Objetivos.....	14
1.1.1 Objetivo Geral.....	14
1.1.2 Objetivos Específicos.....	14
1.2 Justificativa.....	15
1.3 Metodologia.....	16
1.4 Estrutura do trabalho.....	17
2 ESTRUTURA ORGANIZACIONAL UNIVERSITÁRIA.....	18
2.1 Introdução.....	18
2.2 A Organização Universitária.....	19
2.3 Modelos de Organizações Universitárias.....	21
2.3.1 Modelo Burocrático	21
2.3.2 Modelo Colegiado	22
2.3.3 Modelo Político	22
2.3.4 Anarquia Organizada	23
2.4 Estrutura Universitária.....	24
2.5 Tipos de Organização Universitária.....	25
2.6 Processo Decisório nas Organizações Universitárias.....	26
2.6.1 Modelo de Faculdades	28
2.6.2 Modelo de 4 Níveis	29

2.6.3 Modelo baseado em Áreas Afins (Centros)	30
2.6.4 Modelo Departamental	31
2.7 Estrutura de decisão na Administração Superior.....	32
2.8 O papel da informação no Sistema Universitário Brasileiro.....	34
2.9 Considerações Finais	36
3 TECNOLOGIA DE DATA WAREHOUSE	38
3.1 Conceitos Básicos	39
3.1.1 Data Warehouse	39
3.1.2 Dados Operacionais X Dados no Dw	40
3.1.3 Elementos Básicos de DW	41
3.2 Características do Data Warehouse.....	44
3.2.1 Orientação por assunto	44
3.2.2 Integração	44
3.2.3 Variação no tempo	46
3.2.4 Não Volatilidade	47
3.2.5 Localização	48
3.3 Arquitetura do Data Warehouse	48
3.3.1 Arquitetura Top-Down	49
3.3.2 Arquitetura Bottom-up	51
3.3.3 Data Warehouse Distribuídos	52
3.4 Granularidade.....	53
3.4.1 Níveis duais de granularidade	55
3.5 Modelo de dados	57
3.5.1 Modelo estrela	58
3.5.2 Variação do modelo estrela	61
3.6 Metadados.....	61
3.6.1 Definição	62
3.6.2 Classificação	62
3.6.3 Fonte de metadados	64
3.6.4 Problemas na criação dos Metadados	66
3.7 Data Mart.....	67
3.8 O Desenvolvimento do Data Warehouse.....	70

3.8.1 As Funções em um Ambiente de Data Warehouse	70
3.8.2 Estratégias no desenvolvimento de um Data Warehouse	71
3.8.3 Desenvolvimento de um Data Warehouse utilizando o modelo Estrela	73
3.8.4 Os dados operacionais	74
3.9 - Povoando o Data Warehouse	76
3.9.1 Processo de Extração dos dados do ambiente operacional	76
3.9.2 Processo de Filtrar, Transformar e Integrar os dados extraídos	77
3.10 Recuperação de Informações do Data Warehouse	78
3.11 Passos para a implantação do Data Warehouse.....	79
3.12 Considerações Finais	80
4 MODELO DE IMPLANTAÇÃO DE DW EM UMA INSTITUIÇÃO DE ENSINO SUPERIOR: CASO DA UNOESC - CHAPECÓ	82
4.1 Metodologia de Implantação.....	83
4.1.1 Análise da Estrutura Organizacional da Universidade	83
4.1.2 Levantamento dos Sistemas Legados	83
4.1.3 Proposição do Modelo	84
4.2 Estrutura Organizacional da UNOESC	84
4.2.1 Histórico e Abrangência	84
4.2.2 Estrutura Administrativa	85
4.2.3 Estrutura Organizacional da Administração Central	86
4.2.4 Estrutura Organizacional dos Campi	86
4.3 Estrutura Organizacional da UNOESC - Campus Chapecó	88
4.3.1 Estudo da Estrutura Organizacional da UNOESC - Campus de Chapecó	89
4.4 Fluxo de informações na UNOESC - Campus Chapecó para a tomada de decisão na Gestão Acadêmica	95
4.5 Sistemas Legados da UNOESC	96
4.6 Análise do Sistema Acadêmico Existente.....	99
4.7 Modelo Proposto.....	100
4.8 Considerações Finais	104
5 PROTÓTIPO DE UM DATA MART PARA A UNOESC CHAPECÓ	105

5.1 Bottom-up X Top-down.....	105
5.2 O Projeto	106
5.2.1 Identificação da área de negócio	106
5.2.2 Processo de Negócio a ser modelado	106
5.2.3 Levantamento de requisitos	107
5.2.4 Medidas Envolvidas	109
5.2.5 Granularidade	109
5.2.6 Modelagem Dimensional	110
5.3 Implementação	113
5.3.1 Ferramentas de Implementação	114
5.3.2 Passos da Implementação	115
5.3.3 Extração e Análise dos Resultados	121
5.3.4 Considerações Finais	127
6 CONCLUSÕES E RECOMENDAÇÕES	129
7 FONTES BIBLIOGRÁFICAS	131
8 ANEXOS.....	134
8.1 Anexo A: Interface do Oracle Discoverer for Users®	134
8.2 Anexo B: Definição dos Atributos das tabelas do DM.....	134
8.3 Anexo C: Interface do Oracle Discoverer Administration®	135
8.4 Anexo D: Scripts do Esquema do BD gerado para o DM	135
8.5 Anexo E: Scripts de Criação Primary Key, Foreign Key e Constraint	136

LISTA DE FIGURAS

Figura 1 – Metodologia utilizada.....	16
Figura 2 - A IES e o ambiente externo.....	19
Figura 3 - Estrutura do Modelo de Faculdade	29
Figura 4 - Estrutura do Modelo de 4 Níveis	30
Figura 5 - Estrutura do Modelo baseado em Áreas Afins	31
Figura 6 - Estrutura do Modelo Departamental.....	32
Figura 7 - Arquitetura Top-Down segundo Firestone (2000)	50
Figura 8 - Arquitetura Bottom-up segundo Firestone (2000)	52
Figura 9 - Componentes do DW. Adaptado de Inmon (1997).....	54
Figura 10 - Nível Dual de Granularidade. Adaptado de Inmon (1997).....	55
Figura 11 - Modelo Dimensional do tipo Estrela. Adaptado de Kimball (1996). 59	
Figura 12 - Repositório de Metadados segundo David (1999)	63
Figura 13 - Arquitetura de Data Marts Departamentais (arq. <i>Top-down</i>).....	68
Figura 14 - Arquitetura de Data Marts Departamentais (arq. <i>Bottom-up</i>)	69
Figura 15 - A falta de Integração dos Dados de Diferentes Aplicações.....	74
Figura 16 - Sistemas de Informação Existentes	97
Figura 17 - Modelo Proposto para a UNOESC.....	103
Figura 18 – Esquema Estrela do Protótipo.....	112
Figura 19 – Diagrama Entidade Relacionamento	116
Figura 20 – Esquema Físico obtido a partir do modelo E-R	117
Figura 21 - Processo de Carga implementado no protótipo	119
Figura 22 - Índices de trancamentos por área do conhecimento.....	122
Figura 23 - Representação gráfica dos índices de trancamentos.....	122
Figura 24 - Índice de Reprovação por Grande Área do Conhecimento.....	123
Figura 25 - Gráfico com o Índice de Reprovação	124
Figura 26 - Índice de Trancamentos por Sexo.....	125
Figura 27 - Conceito Médio por Grande Área do Conhecimento.....	126
Figura 28 - Gráfico de Conceito Médio por Grande Área do Conhecimento ..	127

LISTA DE TABELAS

Tabela 1 - Características dos Processos Universitários.....	24
Tabela 2 – Definição dos Elementos Básicos de um Data Warehouse (kimball et. al. 1998).....	41
Tabela 3 - Integração de Dados	45
Tabela 4 - Exemplo de Metadados Formais.....	64
Tabela 5 - Funções em um Data Warehouse.	71
Tabela 6 - Diagnóstico dos Sistemas Existentes.....	98
Tabela 7 - Variáveis Internas.....	102
Tabela 8 - Variáveis Externas.....	102
Tabela 9 – Indicadores solicitados pela alta administração da área acadêmica.	109
Tabela 10 – Elementos do Modelo Multidimensional Desenvolvido.....	111

LISTA DE ABREVIATURAS

COBOL	Common Business Oriented Language.
DM	Data Mart
DOS	Sistema Operacional
DW	Data Warehouse
EIS	Sistemas de Informação executivos
E-R	Entidade/Relacionamento
IES	Instituição de Ensino Superior
MEC	Ministério da Educação e Cultura
OLAP	Processamento Analítico on-line
OLTP	Processamento de transações on-line
PASCAL	Linguagem de programação com enfoque científico
SADs	Sistemas de apoio à decisão
SGBD	Sistema de gerenciamento de banco de dados
SQL	Structure Query Language

RESUMO

O objetivo deste trabalho é propor uma metodologia para o desenvolvimento de um ambiente Data Warehouse, voltado a uma Instituição de Ensino Superior (IES). A proposta tem bases no estudo das diferentes formas organizacionais de IES e na identificação dos respectivos níveis decisórios. O resultado do trabalho é a proposta de uma metodologia de implantação do Data Warehouse da IES, segundo a arquitetura *bottom-up* de construção de Data Marts, centrada nos diferentes níveis organizacionais e decisórios da empresa. A metodologia foi aplicada na Universidade do Oeste de Santa Catarina - UNOESC Chapecó, resultando em um modelo de DW para a instituição, testado em um sistema OLAP sobre os processos de trancamento e conceitos por cursos da universidade.

ABSTRACT

The present study is aimed at proposing a methodology to develop a Data Warehouse environment for high education institutions. The proposal is based on a study of different organizational forms of such institutions and on the identification of their respective decision levels. As a result of the study, a methodology for the implementation of a Data Warehouse is proposed, according to the bottom-up architecture of the Data Marts construction, centered on different organizational and decisory level in the company. The methodology was applied in the Universidade do Oeste de Santa Catarina – UNOESC Chapecó, resulting in a data warehouse model for institution, tested in a OLAP system regarding the cancelling and grading processes in the university's courses.

1 INTRODUÇÃO

As bases de dados legadas são encontradas em um número expressivo de organizações públicas e privadas. Geralmente trata-se de repositórios que estão sendo manipulados por sistemas de informação em operação há uma longa data. Cumprem um papel decisivo nas organizações, contemplam milhões de linhas de código fonte e, em geral e freqüentemente, escritas em COBOL.

Os sistemas legados constituem-se em causadores de sérios problemas às IES, pois são essenciais às atividades das mesmas e foram concebidos com o enfoque orientado à aplicação e não orientados a assuntos, o que não permite aos dirigentes obterem informações relevantes sobre dados históricos para a tomada de decisões, devido às características legadas por eles. Sendo assim, existe uma carência muito grande com relação a dados históricos, estáveis e sumarizados que suportem o processo decisório da Instituição.

A partir da década de 90, a Tecnologia da Informação incluía a área de Data Warehouse (Inmon, 1997), criada com o objetivo de construir repositórios de informações agregadas, visando atender os níveis decisórios mais altos das organizações, utilizando como matéria-prima as bases de dados legadas. A questão chave para o sucesso destes sistemas está no correto entendimento dos negócios da organização e, principalmente, nos seus níveis decisórios.

Neste trabalho, o objetivo é propor um modelo de implantação do Data Warehouse em Instituições de Ensino Superior. A heterogeneidade de ambientes, multiplicidade de sistemas e diferentes modelos organizacionais têm dificultado a utilização desta tecnologia nos níveis decisórios das instituições. O resultado é a falta de sistemas de apoio à decisão e conseqüentemente planejamento e gestão ad-hoc (Anjos, 1999).

Para testar a metodologia em um caso real, o presente estudo será realizado na Universidade do Oeste de Santa Catarina - UNOESC, instituição multicampi, fundada em 1992. A área de abrangência é o grande oeste catarinense e se estende desde a parte inferior do Alto Vale do Rio do Peixe, até o extremo-oeste, cujo maior centro é Chapecó. Abrange uma extensão geográfica equivalente a 28,5% de todo o território catarinense. Devido à estrutura multicampi da UNOESC e à natureza exploratória deste trabalho, o escopo de atuação dar-se-á junto ao Campus de Chapecó através de um estudo nos sistemas de informação legados e da definição de um ambiente de Data Warehouse com enfoque na gestão acadêmica, com objetivo de possibilitar aos gestores uma ferramenta de auxílio no processo de apoio à decisão.

Sendo a UNOESC uma das maiores universidades do estado, verifica-se a necessidade de deixar esse estudo como referência para outras IES, bem como difundir o uso da tecnologia de DW em universidades.

1.1 Objetivos

1.1.1 Objetivo Geral

O objetivo deste trabalho é propor uma metodologia para o desenvolvimento de um ambiente Data Warehouse, voltado a uma Instituição de Ensino Superior.

1.1.2 Objetivos Específicos

Os objetivos específicos do presente trabalho de pesquisa surgem do estudo das características das diversas bases de dados dos Sistemas de Informação Legados, 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 através de um estudo de caso de dimensão real;
- Identificar problemas com relação à implementação da tecnologia data warehousing;
- Definir um repositório de dados históricos para auxiliar os dirigentes de IES, contendo informações relevantes para a tomada de decisões;
- Extrair informações de diferentes origens, integrá-las e armazená-las em um repositório único central Data Warehousing;
- Repassar à sociedade o conhecimento e resultados obtidos na pesquisa.

1.2 Justificativa

Todas as organizações de médio e grande porte, estabelecidas há mais de dez anos, possuem suas atividades apoiadas por um certo nível de informatização operacional. Este é constituído, em sua grande maioria, por sistemas de informações proprietários, que tinham como objetivo, na época de sua concepção, automatizar tarefas repetitivas sobre grandes volumes de dados, visando a minimizar os erros e agilizar os processos operacionais de manipulação da informação.

Hoje essas organizações enfrentam problemas da manipulação destas informações as quais se encontram armazenadas local ou remotamente em uma grande variedade de bases de dados legadas. A radical evolução que ocorreu na informática gerou uma demanda por quantidade e qualidade da informação para a tomada de decisões. Existe um grande desconhecimento sobre o que consta nessas bases de dados, isto porque o ambiente em que estão implementadas é tipicamente composto por uma linguagem de programação de 3º geração como o COBOL, Pascal e, eventualmente, um SGBD hierárquico, que não suportam restrições de integridade sobre os dados.

1.3 Metodologia

Para alcançar os objetivos propostos foram superadas várias etapas que juntas integram o trabalho desenvolvido. A primeira etapa do trabalho consistiu no levantamento de subsídios teóricos (livros, artigos e internet) sobre a estrutura organizacional universitária, bem como o processo decisório nessas organizações. Uma vez conhecidos os modelos de organizações universitárias e a estrutura de decisão, foi necessário um embasamento teórico sobre a tecnologia de Data Warehouse e Data Mart através do levantamento da teoria e prática em Data Warehouse. Conhecendo as técnicas mais utilizadas, o próximo passo foi a aplicação da metodologia estudada através de levantamento dos sistemas existentes e da concepção de uma proposta de um modelo de DW para a UNOESC. Tendo desenvolvido o modelo, o próximo passo foi a sua validação, mediante a construção de um protótipo de um data mart departamental. Ao final, apresentam-se as conclusões obtidas da pesquisa, bem como sinalizam-se algumas recomendações para trabalhos futuros. A Figura 1 ilustra a metodologia utilizada na pesquisa.

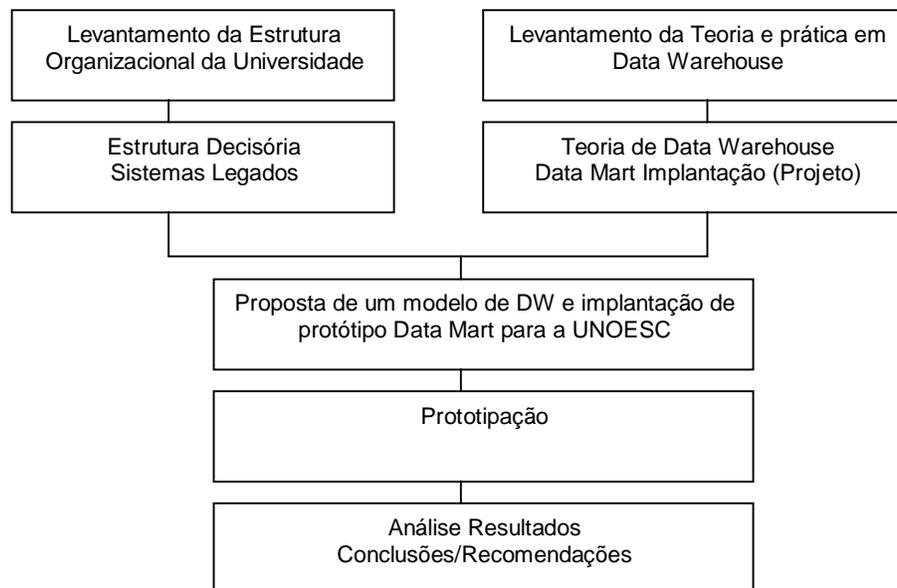


Figura 1 – Metodologia utilizada

1.4 Estrutura do trabalho

O primeiro capítulo traz uma visão geral do trabalho, seus objetivos, justificativa, metodologia e a estrutura.

No segundo capítulo, tem-se um estudo da estrutura organizacional universitária, onde são analisadas as diferentes estruturas e organizações de Instituições de Ensino Superior, bem como o processo decisório nessas organizações.

No terceiro capítulo, é apresentado um estudo sobre a tecnologia de Data Warehouse - Data Mart, onde são abordados conceitos, características e demais etapas necessárias para o projeto de um Data Warehouse.

O quarto capítulo traz a proposta para implantação de um Data Mart na UNOESC, através de um estudo de caso.

No quinto capítulo, é validada parte do modelo proposto, através do desenvolvimento de um protótipo.

O sexto capítulo trata das conclusões obtidas no desenvolvimento do trabalho, bem como sinaliza algumas recomendações para trabalhos futuros.

Finalmente é registrada a bibliografia utilizada nesse trabalho.

2 ESTRUTURA ORGANIZACIONAL UNIVERSITÁRIA

2.1 Introdução

O complexo mundo das organizações tem alterado os modelos clássicos de administração e mudado as formas antes não flexíveis e exatas, por formas imprecisas e flexíveis. Estruturas organizacionais, funções administrativas e hierarquia adquirem características de flexibilidade até há pouco desconhecidas ou não praticadas (Finger, 1997). A evolução do meio social, econômico e político, no qual as organizações encontram-se inseridas, são os fatores que provocam essas mudanças.

As mudanças que ocorrem no ambiente afetam também as organizações universitárias. Sendo assim, estas devem estar sensíveis a tais transformações. Anual e semestralmente os estudantes, que ingressam em uma Instituição de Ensino Superior, trazem consigo novas exigências, novos critérios e projetos, aos quais a universidade deve estar atenta.

A universidade tem um papel importante com a sociedade. Cabendo a ela definir o produto mais adequado, bem como entender, em termos de quantidade e qualidade, as necessidades dessa sociedade, na qual se encontra inserida.

Nesse capítulo, tem-se por objetivo realizar o levantamento das diferentes estruturas e organizações universitárias, visando identificar a hierarquia de decisão neste tipo de instituição, bem como identificar o papel e a natureza da informação utilizada nas IES.

Não há a pretensão de apresentar, neste capítulo soluções para as tomadas de decisões nas instituições de ensino superior, e sim situar a problemática do processo decisório nas organizações universitárias, já que o processo de tomada de decisão exige um conhecimento prévio da instituição e do ambiente interno e externo. Sendo assim, o trabalho iniciasse-a com um

estudo específico das organizações universitárias e conclui com uma análise do processo decisório nessas organizações.

2.2 A Organização Universitária

As organizações universitárias, apesar de terem finalidades específicas e objetivos diferentes, possuem semelhanças, pois têm estruturas similares, podendo ser administradas segundo seus princípios, conforme os modelos propostos pelas teorias da Administração (Finger, 1997). Analisadas pelo ângulo da administração, uma indústria, uma loja, Museu, a Universidade são semelhantes. Do ponto de vista administrativo, não há algo que individualize ou singularize essas organizações, pois são todas passíveis de serem administradas de forma semelhante, isto é, organizáveis, planejáveis sob influências do ambiente no qual estão inseridas. Entende-se por ambiente o conjunto de elementos que não pertencem à organização, conforme exemplifica a Figura 2.

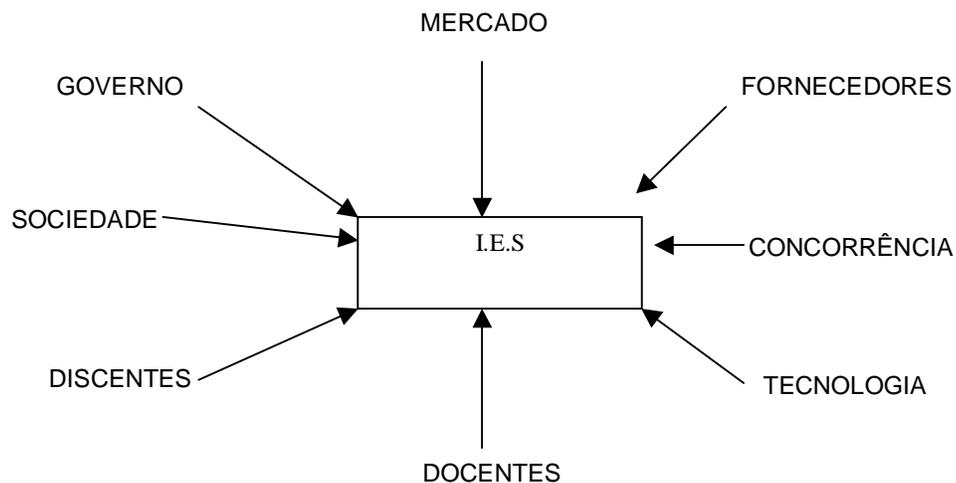


Figura 2 - A IES e o ambiente externo

Sob muitos aspectos a universidade não é diferente de muitas empresas em qualquer área, com exceção da missão específica. Não difere no que diz respeito ao trabalho e encargos do dirigente, ao planejamento e à estrutura da organização. Entretanto, a universidade é, na sua essência, diferente das empresas nos seus "negócios". É diferente na sua atividade fim, possui valores diferentes e contribui de forma diferente à sociedade. O desempenho é a área em que a universidade tem um diferencial significativo da empresa. Finger (*apud* Drucker, Peter F. Produtividade - A chave do futuro. Diálogo nº 4, vol 26, 1993) considera que:

"As atividades de conhecimento e de serviços não devem ser tratadas apenas como trabalho. Na fabricação e transporte de coisas o foco quanto aumentar a produtividade está no trabalho. Na atividade de conhecimentos e de serviços, ela tem de estar no desempenho. Para ser específico, em algumas tarefas de conhecimento e serviços, o desempenho significa qualidade".

A administração Universitária segue as premissas da administração de empresas, busca um modelo eficiente e procura ampliar sua validade, através da elaboração de proposições sobre estruturas organizacionais e os critérios de avaliação do seu funcionamento, considerados elementos que definem o desempenho da maioria das organizações.

Segundo Finger (*apud* Cohen e March In Baldrig 1980), a liderança e as decisões na organização universitária permitem rotulá-la como uma "*anarquia organizada*", onde existe pouca coordenação e controle, onde cada indivíduo é um autônomo tomador de decisões. Destacam-se como características especiais:

- Ambigüidade de objetivos - os objetivos são vagos e difusos;
- Clientela - alunos com necessidades específicas e diversificadas;
- Tecnologia problemática - utilização de uma variedade de métodos e técnicas para atender a uma clientela diferenciada.;
- Profissionalismo - seus profissionais desenvolvem funções não rotinizáveis, gozando de autonomia no trabalho e manifestando dupla lealdade:
 - à profissão a que pertencem,
 - à organização para qual trabalham.

- Vulnerabilidade ao ambiente - sensibilidade aos fatores ambientais externos que poderá afetar a sistemática e padrões de administração universitária;
- A natureza política que predomina nas decisões;
- A existência de uma estrutura fragmentada e descentralizada onde as decisões estão diluídas em órgãos colegiados;
- A dificuldade de se mensurar os produtos resultantes da ação organizacional;
- A ausência de padrões de performance e compromisso com o resultado.

2.3 Modelos de Organizações Universitárias

Levando em conta as características das organizações universitárias, pesquisas demonstram que, desde os anos 70, as instituições têm-se baseado em quatro modelos organizacionais ou processo decisório, a saber (Hardy, 1996): Modelo Burocrático, Modelo Colegiado, Modelo Político e Anarquia Organizada.

A seguir, cada um destes modelos é revisto, segundo as características de Finger (Finger, 1997).

2.3.1 Modelo Burocrático

Entende-se por "organização burocrática" uma determinada organização formal, cuja ênfase está centrada no aspecto racional de sua organização, fazendo uso tanto de estruturas formais quanto das informais, com o objetivo de assegurar o máximo de eficiência técnica em relação aos fins pré-estabelecidos, e que adquire uma certa prática rotinizada e ritualizada na execução dos papéis e das funções.

Na universidade, as normas e regras e outras características burocráticas servem a várias funções. Com isso ocorre um tratamento igual a todos os membros da organização, e evita favoritismo.

O aparato administrativo que corresponde à dominação legal é o burocrático. Tem seus princípios baseados nas leis e na ordem legal. A burocracia é a forma de organização típica da sociedade democrática e das grandes empresas.

O dirigente, segundo as características de seu modelo burocrático, é visto como um herói no topo do poder. Seu papel é resolver problemas, propor soluções e fazer escolhas racionais. Os dirigentes são analistas racionais que utilizam os meios mais eficientes para alcançar os objetivos desejados (Finger, 1997).

2.3.2 Modelo Colegiado

O termo colegiado nas universidades é oriundo de duas fontes principais: em primeiro lugar, há a idéia de uma comunidade composta por letrados na qual o processo decisório dá-se por obtenção de consenso entre os seus integrantes; em segundo lugar, observa-se a idéia de autoridade profissional, fundamentada na competência e não na posição.

No modelo de colegiado, a organização não é caracterizada por uma estrutura hierárquica e por procedimentos racionais que enfatizam a precisão e a eficiência na tomada de decisões. A instituição se organiza de uma forma democrática, com responsabilidade compartilhada. Nas decisões predominam o consenso e a satisfação da maioria dos participantes. Uma condição essencial para que a organização mantenha esse modelo é que seja pequena, permitindo uma constante interação entre seus membros (Finger, 1997). A colegialidade é entendida como descentralização dentro da subunidade, isto é, nesse modelo ocorre um alto grau de influência dos membros do corpo docente no processo decisório.

2.3.3 Modelo Político

No modelo político, o poder é difuso e mutável. Existe dualidade de controle, e os conflitos entre as autoridades administrativas são constantes,

pois os valores são múltiplos, não consensuais, baseados em interesses próprios.

O dirigente desempenha um papel mais político do que administrativo: é um negociador, um mediador que procura meios para estabelecer ações viáveis para a organização (Finger, 1997). Uma boa parte de seu tempo é destinada às atividades políticas que visam o uso e aquisição do poder. Com relação ao processo de tomada de decisões, reflete mais um jogo político, no qual o resultado depende da posição privilegiada de seus participantes.

2.3.4 Anarquia Organizada

O chamado modelo "lata de lixo", pelo qual as universidades foram visualizadas como anarquias organizadas, citado por Hardy (*apud* Cohen e March, 1974; March e Olsen, 1976), é diferente, tanto da abordagem burocrática como política.

Anarquia Organizada ocorre em organizações ou situações de decisões com características difusas múltiplas, objetivos indefinidos, tecnologia desconhecida e falta de conexão entre problemas e soluções (Finger, 1997).

Finger (*apud* Cohen e March 1983) considera que, nas organizações universitárias, cada indivíduo é visto como um tomador de decisões. As decisões não são pretendidas ou controladas por ninguém, elas são conseqüências do Sistema. Nesse modelo, teoricamente, cada professor é capaz, pode selecionar seu material de trabalho, sem ser submetido a regras ou regulamentos. Sendo assim, a universidade seria uma anarquia, do ponto de vista ortodoxo da administração, pois cada um dos professores terá liberdade para melhor desenvolver seu trabalho, sem nenhum tipo de imposição administrativa.

Não existem receitas para uma estrutura organizacional universitária. A existência de modelos não significa o estabelecimento de uma proposta teórica. Não há como afirmar que um modelo seja mais eficiente que outro. Cada modelo apresentado explica melhor a realidade de uma

universidade, ou parte dela, mas universidade alguma reflete, de forma pura, qualquer modelo. São os fatores ambientais em cada universidade que podem determinar aquele que mais se adapte às suas realidades (Finger, 1997). A Tabela 1 sintetiza o processo decisório em cada modelo.

Tabela 1 - Características dos Processos Universitários.

Modelo	Tomada de decisão por	Concepção
Burocrático	Padrões de interações estruturadas	Estruturalista
Colegiado	Consenso	Recursos Humanos
Político	Resolução de conflitos	Política
Anarquia Organizada	Acidente	Simbólica

2.4 Estrutura Universitária

O estudo das formas atuais de estruturas universitárias é bastante abrangente. O escopo desse estudo concentra-se no processo decisório dentro da estrutura universitária.

O processo de tomada de decisão em uma universidade é diferente de outras empresas cujos objetivos é essencialmente a geração de lucros. Ocorre uma sobreposição de atividades que poderiam ser chamadas de legislativas e executivas dentro de vários órgãos da estrutura. Num certo sentido, na independência do nível hierárquico que integra a estrutura burocrática universitária, o que já não ocorre numa organização empresarial.

Segundo Finger (1997), as organizações universitárias, do ponto de vista da atividade administrativa, possui uma área de atuação específica que pode ser o ensino, a pesquisa, a extensão e a administração. O autor classifica a administração universitária em três níveis, a saber:

1 - Administração Superior - onde se enquadram os Conselhos Superiores (Universitário, Curadores, Ensino, Pesquisa e Extensão) responsáveis pela formação das diretrizes, organização e supervisão de sua implementação e decisões. Enquadram-se também na Administração Superior o Reitor, Vice-Reitor, Pró-Reitores e suas respectivas estruturas administrativas de apoio.

2 - Administração Acadêmica - abrange a Direção dos Centros, Faculdades, Escolas, Institutos, Chefias de Departamento e Coordenadores de Cursos.

3 - Administração da Atividade-meio - órgão que desenvolve atividades não instrucionais e que serve de apoio à área-fim da organização (Departamento de Recursos Humanos, Departamento Financeiro, entre outros).

2.5 Tipos de Organização Universitária

Segundo Finger (1997), as universidades brasileiras podem ser classificadas, quanto à propriedade dos meios de produção, em públicas (federais ou estaduais) ou privadas, e em autarquias ou fundações de direito público; e quando particulares, sob a forma de fundações ou associações.

O Conselho de Reitores das Universidades Brasileiras denominado de CRUB, reunido em 1998 (Crub, 1998), classifica as Instituições de Ensino Superior (IES) em cinco tipos: Federal, Estadual, Municipal, Particular e Comunitária. A Lei 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação (LDB), nos seus artigos 16, 17, 18, 19 e 20 define cada uma das classificações, a saber:

- *Federal* - Instituição de ensino superior mantida pela União.
- *Estadual* - Instituições de ensino superior mantidas pelo poder público estadual e pelo Distrito Federal.
- *Municipal* - São consideradas Públicas quando as Instituições de ensino superior são criadas ou incorporadas, mantidas e administradas pelo poder público; e privadas, quando as instituições de ensino superior são mantidas e administradas por pessoas físicas ou jurídicas de direito privado.

- *Particular* - Instituições de ensino superior mantidas e administradas por pessoas físicas ou jurídicas de direito privado.
- *Comunitária* - Instituições de ensino superior instituídas por grupos de pessoas físicas ou por uma ou mais pessoas jurídicas, inclusive cooperativas de professores e alunos que incluam na sua entidade mantenedora representantes da comunidade.

No aspecto que se refere à sobrevivência, ou seja, ao grau de dependência da universidade em relação aos recursos, Finger (1997) identifica três tipos de organização universitária:

1 - Universidade cuja sobrevivência *depende parcialmente* dos recursos de fonte própria. Nesse grupo enquadram-se as Universidades Confessionais e as Universidades Municipais, chamadas Universidades Comunitárias. Os recursos são oriundos de cobrança de anuidade e de outras fontes como o poder público, por exemplo. Essas universidades possuem geralmente uma mantenedora que lhes garante a sobrevivência.

2 - Universidade cuja sobrevivência *depende totalmente* dos recursos de fonte própria. Encontra-se nesse grupo as universidades chamadas empresariais, ou seja, aquelas criadas e dirigidas por sócios-proprietários, as quais dependem das anuidades de seus alunos e/ou de outros rendimentos próprios.

3 - Universidade cuja sobrevivência *independe* dos recursos de fonte própria, ou seja, são todas as universidades públicas cuja sobrevivência é garantida pelo poder público.

2.6 Processo Decisório nas Organizações Universitárias

O processo de tomada de decisão em organizações universitárias é altamente complexo e parece não se apoiar em um modelo específico. Segundo Finger (1997), dependendo das características da organização, este processo parece obedecer a uma determinada lógica de ação que, com base em determinados valores, orienta a atuação dos indivíduos no alcance dos objetivos: organizacionais, grupais ou pessoais.

Essa lógica de ação parece ser orientada por tipos de racionalidade que integram premissas de valor, envolvidas nos processos antecedentes à decisão e à ação, e estão vinculadas aos elementos éticos e valorativos que exprimem os objetivos a atingir.

O processo decisório da universidade contempla diferentes tipos de racionalidade, como a política, a social e a econômica. Em virtude da natureza das organizações universitárias, ocorre uma tendência ao predomínio de decisões baseadas na racionalidade política. Em uma IES é notório a participação fluida de vários grupos de interesse (docentes, discentes, funcionários) na tomada de decisões.

Nas universidades que dependem, na sua totalidade de recursos de fonte própria, as leis de mercado interferem nas decisões. Em consequência, os seus dirigentes, em suas tomadas de decisões, consideram com muito cuidado as consequências econômicas e políticas de tais atos; já nas universidades comunitárias, que dependem parcialmente de recursos de fonte própria, a sobrevivência parece não ser essencial para a tomada de decisão, uma vez que são mantidas por uma associação (leiga ou religiosa). O processo decisório pode ser identificado como sendo burocrático quanto à racionalidade, e em algumas universidades particulares, observam-se características econômicas e políticas.

No caso da universidade pública, que independe de recursos de fonte própria, a sobrevivência está assegurada pelo Estado e seus integrantes fazem uso da racionalidade política, principalmente em organizações burocráticas, onde é fundamental a obediência às normas, às regras e à hierarquia. Este tipo de universidade utiliza-se de decisões políticas, negociações, barganhas, beneficiando o tomador de decisões, em termos individuais ou grupais.

A teoria dos "centros de decisão" tem aplicação na administração universitária, talvez como em nenhum outro setor de atividade. A universidade é em si um conjunto de centros de decisão, representados ora pelos colegiados, ora pelos departamentos, coordenadores nos pontos estratégicos do processo decisório (Finger, 1997). O fluxo de decisões desenvolve-se entre

esses centros numa interação constante, em sentido ascendente ou descendente, sempre em busca da otimização do processo decisório

"A teoria da Administração Universitária, talvez, mais do que qualquer outro tipo de organização, tem que apoiar-se em centros de decisões em todos os níveis do sistema universitário" (Ribeiro, 1977)

O departamento é visto como a célula da organização universitária, onde se oriundam as iniciativas e os projetos didático-científicos. Deve ser a base inicial das decisões acadêmicas.

As atividades dos departamentos, em função do princípio de interação orgânica, devem ser compartilhadas e coordenadas por diferentes órgãos em função das teorias e dos propósitos da universidade como um todo (Finger, 1997).

Segundo Finger (*apud* Newton Sucupira, I encontro de Reitores das Universidades Públicas, "A Condição Atual da Universidade e a Reforma Universitária Brasileira"), o autor aborda, em seu trabalho, alguns modelos possíveis dentro da estrutura formal universitária:

2.6.1 Modelo de Faculdades

Esse modelo mantém o sistema tradicional de Faculdades, Escolas ou Institutos e Departamentos, conforme exemplificado na Figura 3.

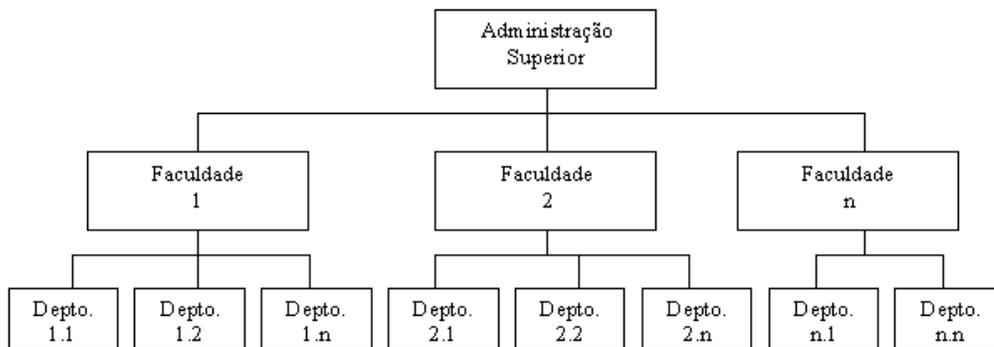


Figura 3 - Estrutura do Modelo de Faculdade

O modelo contempla *três níveis decisórios*:

- a) *Departamento*, onde se elaboram programas e projetos didáticos e científicos.
- b) a *Faculdade*, que compatibiliza as decisões departamentais no plano administrativo.
- c) a *Administração Superior da Universidade*, instância última do comando e integração, em função da política universitária.

Fazendo uma análise do ponto de vista teórico, esse modelo é possível de ser trabalhado e poderá garantir iniciativa e flexibilidade de ação ao Departamento, desde que lhe seja atribuída relativa autonomia. Mas haja vista a tradição brasileira de independência das faculdades, o modelo tende a reforçar a faculdade em detrimento do departamento. Daí por que, ainda hoje, na prática, em muitas faculdades o departamento tem apenas uma existência nominal.

2.6.2 Modelo de 4 Níveis

O modelo na Figura 4 é formado por *quatro níveis de decisão*:

- a) *Departamento*;

- b) Faculdade;
- c) Centros coordenando Faculdades afins;
- d) Administração Superior

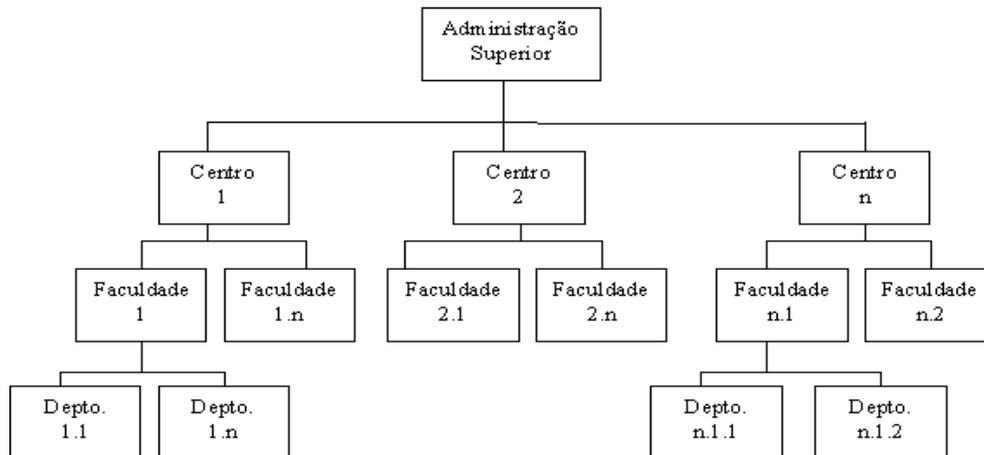


Figura 4 - Estrutura do Modelo de 4 Níveis

Como se observa, entre o departamento e a administração superior existem dois níveis intermediários de integração. Segundo Finger (*apud* Newton Sucupira), o modelo apresenta sérios inconvenientes com relação à flexibilidade e à mobilidade das operações para adquirir eficácia. Uma decisão gerada no âmbito departamental terá, muitas vezes, de atravessar os planos da Faculdade e do Centro, até que seja aprovada pelos órgãos superiores, o que muitas vezes implica em perda de oportunidade.

2.6.3 Modelo baseado em Áreas Afins (Centros)

No terceiro modelo, os departamentos correspondentes a áreas afins do conhecimento são integrados por grandes unidades que são os *centros*. Eles se limitam a efetuar uma primeira integração entre o departamento e o escalão superior da universidade a Figura 5 exemplifica essa estrutura.

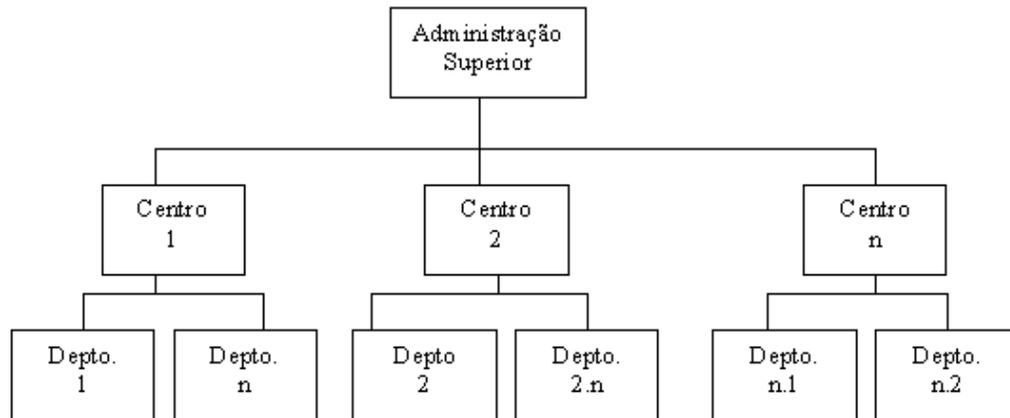


Figura 5 - Estrutura do Modelo baseado em Áreas Afins

2.6.4 Modelo Departamental

No quarto modelo, como mostra a Figura 6, os departamentos são diretamente integrados pela administração superior, eliminando-se toda coordenação por escalão intermediário. Esse modelo é operável numa universidade pequena e, é inadequado, para universidades maiores e complexas, onde a necessidade de integração é maior.

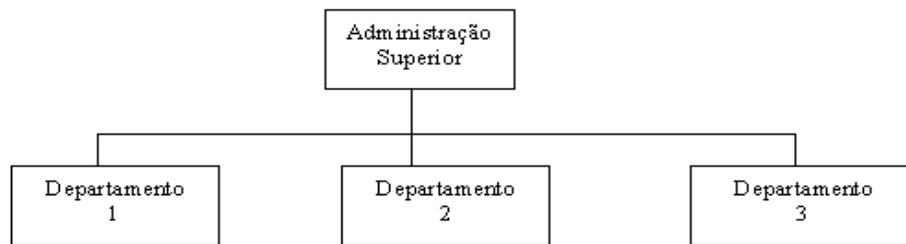


Figura 6 - Estrutura do Modelo Departamental

As decisões tomadas pelos dirigentes, nos diversos níveis organizacionais, diferem em tipo e amplitude e de universidade para universidade. Nos níveis de cúpula, o estabelecimento de objetivos e estratégias globais está incluído entre as mais importantes decisões a serem tomadas. Os dirigentes intermediários geralmente estão mais envolvidos com as decisões de planos operacionais e políticos que dirigirão esforços para a realização de objetivos globais.

As normas e os procedimentos também variam de uma universidade para outra, desde as mais exigentes e rígidas, até as mais flexíveis. Elas são prescritas em regimentos, estatutos, resoluções, manuais, etc.

2.7 Estrutura de decisão na Administração Superior

Segundo Finger (*apud* Newton Sucupira) a partir dos níveis estruturais acima é possível descrever características dos órgãos das universidades brasileiras no que se refere à constituição e atribuições.

- *Conselho Administrativo* - exerce funções consultivas e deliberativas na área administrativa. Tem como funções principais aprovar o quadro pessoal,

propor e fixar o orçamento geral da universidade, apreciar o planejamento anual das atividades e orientar o pessoal administrativo.

- *Conselho Universitário* - é um dos órgãos de maior poder na estrutura universitária, supervisionando, inclusive, as decisões tomadas em outras instâncias. Tem como atribuições aprovar o regimento, participar na eleição de reitores, indicar os representantes para outros colegiados e reconhecer os órgãos estudantis.
- *Conselho Curador* - Por lei, é obrigada a existência desse conselho, cuja atribuição básica é a execução econômico-financeira da universidade.
- *Conselho de Ensino, Pesquisa e Extensão* - é órgão superior normativo, deliberativo, consultivo da universidade em matéria didático-científica e acadêmica no ensino, pesquisa e extensão. Possui, também, como atribuições aprovar seus próprios regimentos, participar na eleição do reitor e vice-reitor, entre outras.
- *Reitoria* - é órgão executivo superior da universidade que coordena, supervisiona e superintende de todas as atividades da instituição. O Reitor tem legitimamente o direito de tomar decisões "*ad referendum*" dos demais colegiados, inclusive exercendo o poder disciplinador em toda a universidade. As universidades, em geral, mantêm órgão de assessoria e comissões permanentes ou provisórias na linha de suporte técnico administrativo como: Assessoria Jurídica, de Informática, entre outras; e órgãos complementares, como comissão de Vestibular, por exemplo.
- *Pró-Reitorias* - As Pró-Reitorias são nomeadas pelo reitor nas áreas de administração, planejamento, ensino, e pesquisa e extensão. Dependendo do porte da universidade, algumas destas são aglutinadas em áreas afins com o objetivo de enxugar a estrutura. É importante lembrar que em alguns casos os Pró-Reitores são legitimados pelo voto direto, cabendo ao reitor acatar ou não a legitimidade do processo de escolha.
- *Conselho Departamental* - é órgão deliberativo e consultivo da administração setorial da universidade em matéria administrativa, didático-pedagógica e científica.

- *Diretoria* - órgão executivo das unidades universitárias, cuja função é supervisionar e coordenar todas as atividades das respectivas unidades.
- *Chefia de Departamento* - é órgão executivo em nível da administração básica da universidade que promove, coordena e supervisiona as atividades de competência do departamento.

Finger (1997) cita que as decisões mais significativas são tomadas em nível Reitoria, Conselho Universitário e Conselho de Ensino e Pesquisa. Percebe-se, também que as decisões de caráter administrativo ficam a cargo da reitoria. A contratação de professores, decisões de caráter acadêmico, como mudança de grade curricular, criação de novos cursos, ficam a cargo do Conselho de Ensino e Pesquisa. Assim, a participação, envolvendo os funcionários de nível hierárquico mais baixo: professores e corpo discente, no processo decisório, é quase nula, comprovando que a direção e o poder nas universidades centralizam-se na cúpula da administração universitária.

2.8 O papel da informação no Sistema Universitário Brasileiro

Com o advento da globalização e a velocidade das inovações tecnológicas, a procura por informação se tornou um ponto comum de toda a sociedade. Surge, então o conceito de uma *sociedade da informação*, ou seja, possuir informação ou ao menos ter o privilegio garantido de acesso a ela passa a ser um diferencial nesse início de novo século.

Os esforços que estão sendo feitos para organizar e analisar dados, de forma a que sejam disponibilizados como informação, com valor agregado, para subsidiar processos de tomada de decisão, formam a estrutura orgânica da nova sociedade do conhecimento (Alvim, 1998).

As universidades hoje precisam ter a preocupação em obter qualidade de informação, e não quantidade. O importante é obter a informação correta, ou melhor, adequada à determinada necessidade, em um tempo correto e com um custo compatível. Os gestores universitários estão vivendo

em um ambiente inteiramente novo, no qual vêm-se obrigados a tomar decisões, não só em maior número, mas de forma cada vez mais rápida.

Alvin (1998) cita palavras do Papa João Paulo II, "a riqueza das nações funda-se muito mais sobre a propriedade do conhecimento, da técnica e do saber, do que sobre a propriedade dos recursos naturais". Hoje a tecnologia nas IES é um fator alavancador para competir e também constitui-se um elemento diferenciador que garante a sobrevivência dessas instituições. Diante disso cada vez mais faz-se necessário canalizar esforços para o processo de capacitação tecnológica, que precisa ser entendido como algo dinâmico e fazer parte das estratégias dos gestores, que envolvem especificamente a decisão de adaptar ou adquirir tecnologia, seja através do próprio desenvolvimento ou através da compra para atender as necessidades e expectativas de seus clientes e colaboradores.

A adoção de sistemas de informação é uma maneira de se minimizar o grau de incerteza e risco em uma IES e de propiciar decisões melhores. O processo decisório nos últimos anos tem sido estudado por vários teóricos e pesquisadores, resultando na chamada "teoria da decisão", a qual foi estruturada para possibilitar aos decisores analisar melhor uma situação complexa. A condição necessária para a decisão é geralmente decorrente de um problema que surge de alguma falha no processo de atingir os objetivos e metas da instituição. Sendo assim, é preciso desenvolver alternativas de solução para o problema. Ainda Borges (apud Flippo, Edwin B., Musinger, Gay M. Management 5. Ed Boston, 1970), afirma que "é difícil avaliar que informação é necessária ao decisor, bem como atribuir o seu valor em termos de contribuição para decisões mais acertadas". Os dirigentes tendem a ser inundados com memorandos e relatórios, cujo conteúdo pode ser trivial, não sendo relevante às decisões que devem ser tomadas. Para a informação ser útil, deve ser compreendida e absorvida pelo decisor.

Hoje, milhares de informações são manipuladas por sistemas de informação, que a cada dia têm um papel mais importante na vida dos gestores universitários. Porém, tanto a literatura quanto a prática têm mostrado que os sistemas de informações tem atendido em parte as expectativas dos usuários.

Borges (*apud* Sapiro, Arão. Inteligência empresarial informacional RJ, v.33, 1993, p. 106-124) cita que:

"As grandes empresas estão gastando mais dinheiro do que nunca na obtenção de informação, mas muito dinheiro é perdido na construção de ineficientes centros informacionais, baseados em bancos de dados não apropriados ou carregados de informações não relevantes"

Diante da afirmativa, deduz-se que não basta somente definir a informação necessária, mas também especificar quando e em que formato a informação deverá ser disponibilizada. Os usuários devem participar do processo de definição de necessidades e avaliar o valor que a informação terá como ferramenta para a tomada de decisão. Portanto, antes de se planejar e implementar sistemas de informação, é fundamental que se conscientize o usuário, de que, para que tal sistema atenda às suas necessidades informacionais, cabe a ele explicitá-las. Para isso, torna-se imperativo saber definir com clareza as suas necessidades.

Quando à tarefa, é construir sistemas de informação executivas, ou seja, quando o usuário necessita de informações agregadas e transformadas para apoiar decisões gerenciais, o usuário tem papel diferenciado. Nestes casos, o ciclo de vida é alterado pelo fato de o usuário desconhecer todas suas necessidades, no início do projeto. É o conhecimento gradual que faz com que ele informe ao projetista do DW o que é importante. Portanto, os dirigentes universitários devem acompanhar o desenvolvimento do DW e constantemente ajudar o projetista a trabalhá-lo para buscar melhores sistemas informacionais.

2.9 Considerações Finais

Os estudantes e a sociedade estão exigindo da universidade uma revisão de seus programas, dos serviços e das suas operações, a fim de assegurar um desenvolvimento que responda às suas expectativas.

Não existem receitas para uma estrutura organizacional. Não há como afirmar que um modelo seja mais eficiente que outro. Cada um dos modelos que foram citados explica melhor a realidade de uma universidade, ou

parte dela, mas universidade alguma reflete, de forma pura, qualquer modelo (Finger, 1997).

A estrutura formal é composta por níveis hierárquicos sobrepostos que formam uma pirâmide, tendo a reitoria no topo e os executores na base.

Os objetivos estratégicos e operacionais não são associados de forma clara e/ou não formam um consenso, o que muitas vezes torna difícil escolher as atividades que serão realizadas e as que serão abandonadas ou reorganizadas devido à falta de pontos de referência, a fim de estabelecer prioridades. Além disso, observa-se que as universidades não dispõem de informações necessárias e qualificadas para determinar quais as atividades que deverão ser prioritárias no processo administrativo, ou seja, há uma carência com relação a informações pertinentes, completas, históricas, confiáveis, atualizadas e de qualidade.

Como o processo decisório implica numa escolha entre as diversas opções possíveis, de fato, os dirigentes precisam ter em mãos os elementos que permitam uma análise da situação e avaliação dos benefícios e inconvenientes de cada caminho considerado. Esta participação e definição (incluindo revisão) dos subsídios às decisões devem continuar ao longo do ciclo de vida do desenvolvimento do DW.

3 TECNOLOGIA DE DATA WAREHOUSE

Com a rápida evolução da tecnologia de informação e a disseminação do uso de computadores ligados entre si, praticamente todas as organizações de médio e grande porte fazem algum uso de sistemas informatizados para realizar seus processos mais importantes. Com o passar do tempo, acaba-se gerando uma quantidade enorme de dados relacionados aos negócios, mas desintegrados entre si. Estes dados armazenados em uma ou mais plataformas são um recurso, mas de modo geral, raramente servem como subsídio estratégico no seu estado original para o processo decisório.

Os sistemas tradicionais de informática não são projetados para gerar e armazenar as informações estratégicas. Suas bases são formadas de dados cruciais a operação da organização. Em termos de decisão, os dados de certa forma são vazios e sem valor transparente para o processo gerencial das organizações. Estas decisões normalmente são tomadas baseadas na experiência dos administradores, quando pode, também, ser fundamentadas em fatos históricos que foram armazenados pelos diversos sistemas de informação utilizados pelas organizações.

Um Data Warehouse é projetado para que os dados possam ser armazenados e acessados de maneira que não fiquem restritos a tabelas e linhas relacionais. Como o DW está separado dos bancos de dados operacionais, as consultas dos usuários não impactam nestes sistemas, que ficam resguardados de alterações indevidas ou de perdas de dados. O DW contempla a base e os recursos necessários para um Sistema de Apoio à Decisão (SAD) e, principalmente, sistemas de Informações Executivas eficientes, fornecendo dados integrados e históricos que atendem desde a alta direção, que necessita de informações mais resumidas, até as gerências de baixo nível, em que os dados detalhados ajudam a observar aspectos mais táticos da empresa. Nele, os executivos podem obter de modo imediato respostas para perguntas que normalmente não as possuem em seus sistemas

operacionais e, com isso, tomar decisões com base em fatos, e não apenas em intuições ou especulações.

Desta forma, um DW provê um banco de dados especializado, que gerencia o fluxo de informações a partir de banco de dados corporativos e fontes de dados externas à organização.

O presente capítulo tem por objetivo:

- Apresentar um estudo sobre a tecnologia de Data Warehouse, especificando conceitos, características, ferramentas e metodologia de implantação.
- Identificar, especificamente para as estruturas universitárias, os passos que devem seguir para implantação de DW.

3.1 Conceitos Básicos

3.1.1 Data Warehouse

Na bibliografia encontramos muitos conceitos sobre DW, como os apresentados a seguir:

- Segundo Inmon (1997), Data Warehouse é uma coleção de dados orientados por assuntos, integrados, variáveis com o tempo e não voláteis, para dar suporte ao processo gerencial de tomada de decisão.
- Segundo Harjinder & Rao (1996), Data Warehouse é um processo em andamento que aglutina dados de fontes heterogêneas, incluindo dados históricos e dados externos para atender à necessidade de consultas estruturadas e ad-hoc, relatórios analíticos e de suporte à decisão.
- Segundo Barquini (1996), Data Warehouse é uma coleção de técnicas e tecnologias que juntas disponibilizam um enfoque pragmático e sistemático para tratar com o problema do usuário final de acessar informações que estão distribuídas em vários sistemas da organização.

- Segundo Kimball et. al. (1998), Data Warehouse é uma fonte de dados consultáveis da organização, formado pela união de todos os data marts correspondentes.

3.1.2 Dados Operacionais X Dados no Dw

Com o objetivo de tornar mais claro o entendimento sobre data warehouse faz-se necessário um estudo comparativo entre o conceito tradicional de banco de dados (BD) e data warehouse (DW). "Um banco de dados é uma coleção de dados operacionais armazenados e utilizados pelo sistema de aplicações de uma empresa específica", (Batini & Lenzerini, 1986) os dados mantidos por uma empresa são chamados de "operacionais" ou "primitivos". Batini & Lenzerini (1986) refere-se aos dados no banco de dados como "dados operacionais", distinguindo-os de dados de entrada, dados de saída e outros tipos de dados. Baseado na definição de Batini & Lenzerini sobre dados operacionais pode-se dizer que um DW é, na verdade, uma coleção de dados derivados dos dados operacionais para sistemas de suporte à decisão. Estes dados derivados são, muitas vezes, referidos como dados "gerenciais", "informacionais" ou "analíticos" (Inmon, 1997).

Os bancos de dados operacionais mantêm armazenadas as informações necessárias para viabilizar as operações diárias da empresa. São utilizados por funcionários para registrar e executar operações pré-definidas, por isso seus dados podem sofrer constantes mudanças, conforme as necessidades atuais da empresa. Por não ocorrer redundância nos dados, e as informações históricas não ficarem armazenadas por muito tempo, este tipo de BD não exige grande capacidade de armazenamento.

Já um Data Warehouse armazena dados analíticos, destinados às necessidades da gerência no processo de tomada de decisões. Isto pode envolver consultas complexas que necessitam acessar um grande número de registros. Por isso, é importante a existência de muitos índices criados para acessar às informações da maneira mais rápida possível. Um DW armazena informações históricas de muitos anos, o que implica em uma grande

capacidade de processamento e armazenamento dos dados que se encontram de duas maneiras, detalhados e resumidos.

Com base no exposto acima, conclui-se que o Data Warehouse é um recurso que as organizações possuem para analisar informações históricas, podendo fazer uso dessas para melhoria dos processos atuais e futuros. DW é formado por resumo de dados extraídos de um ou vários sistemas de computação, normalmente utilizados há vários anos e que continuam em operação. São construídos para que os dados possam ser armazenados e acessados de forma que não sejam limitados por tabelas e linhas estritamente relacionais.

3.1.3 Elementos Básicos de DW

Desde seu surgimento no início dos anos 90, a área de data warehouse tem perdido a precisão na definição de seus termos (kimball, et. al. 1998). A Tabela 2 apresenta os conceitos dos principais elementos componentes de um data warehouse, segundo a visão integrada de Kimball et. al (1998).

Tabela 2 – Definição dos Elementos Básicos de um Data Warehouse (kimball et. al. 1998)

Elemento Básico	Definição
Sistemas de Origem	<i>“Sistema operacional de registros cuja função é capturar as transações do negócio”. pág. 14</i>
Área de Estagiamento de dados	<i>“Área de armazenamento e conjunto de processos que limpam, transformam, combinam, retiram duplicações, retêm, arquivam e preparam os dados fonte para uso no data warehouse” pág. 16</i>
Servidor de Apresentações	<i>“Máquina física de destino onde estão armazenados e organizados os dados do data warehouse para consultas diretas dos usuários finais, dos geradores de relatórios e de outras aplicações” pág.16</i>
Modelo Dimensional	<i>“Disciplina específica para modelagem de dados que é uma alternativa ao modelo de entidades-relacionamento (model E/R)” pág.17</i>

Processos do Negócio	<i>“Conjunto coerente das atividades do negócio da organização, que fazem sentido aos usuários de negócio do data warehouse”</i> pág. 18
Data Mart	<i>“Um subconjunto lógico do data warehouse completo”</i> pág.18
Armazenamento de Dados Operacionais (ODS)	<i>Ponto de integração com os sistemas operacionais da organização. Criados para integrar em nível operacional os diferentes sistemas da organização, sem, contudo, incluir consultas gerenciais, que ficam ao nível do DW. págs. 19-20.</i>
OLAP	<i>“Atividade genérica de consultar e apresentar dados textuais ou numéricos de data warehouses, bem como uma forma dimensional específica de consultar e apresentar que é exemplificado por um número de ‘vendedores OLAP’. Trata-se de uma tecnologia não-relacional e geralmente baseada em cubos multidimensionais de dados.”</i> pág.21
ROLAP (OLAP Relacional)	<i>“Conjunto de interfaces ao usuário e de aplicações que dão características multidimensional a bancos de dados relacionais”.</i> pág. 21
MOLAP (OLAP Multidimensional)	<i>“Conjunto de interfaces ao usuário, aplicações com base de dados proprietária que são fortemente multidimensionais”</i> pág. 21
Aplicação para Usuário Final	<i>“Coleção de ferramentas que consultam, analisam e apresentam informações desejadas com vistas às necessidades do negócio da organização”</i> pág. 21
Ferramenta de Controle de Acesso aos Dados para Usuário Final	<i>“Cliente do Data Warehouse (...). Uma ferramenta de controle de acesso aos dados para o usuário final. Pode ser simples como sistemas de consultas ad-hoc ou complexas e sofisticadas com mineração de dados ou aplicações de modelagem”.</i> pág. 21
Ferramentas de Consultas Ad-Hoc	<i>“Tipo específico de ferramenta de acesso dos dados que induz o usuário final a formar suas próprias consultas, manipulando diretamente tabelas relacionais e suas funções”.</i> pág. 22
Aplicações de Modelagem	<i>“Tipo sofisticado de ferramenta cliente do data warehouse com capacidades analíticas de transformar ou compreender as saídas do data warehouse” (e.g. data mining, modelos de previsão, modelos de comportamento, etc)</i> pág. 22
Metadados	<i>“Toda informação no ambiente do data warehouse que não é dado real em si mesmo”</i> pág. 22

O trabalho de Kimball et. al. permite apresentar uma visão geral dos elementos básicos de um data warehouse. A construção de um data

warehouse, em organizações universitárias deve, também, refletir abordagens e soluções para cada um destes elementos.

Os sistemas de origem são formados pelos sistemas legados da universidade, geralmente voltados para suas áreas meio (pessoal, financeiro, estoques, etc) e fim (vestibular, inscrições, cursos, egressos, etc.). Devido à sua natureza operacional, tais sistemas são inadequados à formação de relatórios temporais e de difícil utilização pelos dirigentes da instituição.

Na construção de elementos como *área de estagiamento de dados, servidor de apresentações e armazenamento de dados operacionais (ODS)* o projetista deve estar atento para a infraestrutura disponível e, principalmente, para a arquitetura de informações da organização. Em instituições de ensino superior, a multiplicidade de plataformas, distanciamento físico das bases e heterogeneidade das aplicações acentua a necessidade de cuidado redobrado na definição destes elementos.

Um *processo de negócio* da organização envolve um conjunto de recursos de informações úteis e com um tema coerente com as atividades da instituição. Cada processo de negócio pode requerer um ou mais *data mart*. Em organizações universitárias, os processos de negócio podem ser identificados nas atividades meio e fim da instituição. De fato, dependendo da disponibilidade de seus sistemas legados, uma IES pode construir data marts para suas áreas de compra, recursos humanos e gestão financeira ou para avaliação acadêmica, fluxo de egressos ou perfil pedagógico de seu quadro docente. Estas decisões também terão impacto na escolha dos demais elementos do data warehouse, em especial a opção pelo modelo OLAP, e das aplicações voltadas ao usuário final.

Nas próximas seções, aprofunda-se o levantamento sobre a tecnologia de data warehouse, com relação às suas características e arquitetura de implementação. Alguns dos elementos básicos são revistos em maior detalhe.

3.2 Características do Data Warehouse

O conjunto de dados que compõe o Data Warehouse deve ser classificado por assunto e também faz-se necessário a sua integração de representação, a fim de agilizar as consultas. Também é necessário definir a granularidade e a forma como os dados ficarão armazenados. Outro aspecto importante é ter consciência de que dados em um DW não são modificados, pois representam as informações em um determinado instante de tempo e podem estar fisicamente armazenados de diferentes formas. Segundo Inmon (1997), essas são as principais características do ambiente DW. A seguir, uma breve abordagem de cada uma das características.

3.2.1 Orientação por assunto

Um Data Warehouse armazena dados importantes da organização de acordo com o interesse das pessoas que irão fazer uso dessas informações. Uma instituição de ensino, por exemplo, tem entre seus objetivos, prestar serviços educacionais à comunidade. E um de seus interesses é conhecer o perfil de seus tomadores de serviço, isto faz com que o DW contemple dados dos alunos que fazem uso de seus serviços.

Ao iniciar a construção de um data warehouse, deve-se discutir com os usuários finais quais os seus objetivos, definir quais são as informações importantes para o processo de análise.

3.2.2 Integração

A integração constitui-se em 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. O processo de análise dos sistemas operacionais e dos dados que eles contêm, exige um tempo relativamente grande devido à falta de padrões de

codificação. Cada analista de sistemas define a mesma estrutura de dados de várias formas, ocasionando que os dados que representam a mesma informação, sejam representados de diversas maneiras dentro dos sistemas utilizados pela organização ao longo do tempo.

Um exemplo desse problema é a representação do estado civil. Em um sistema pode-se definir um campo de uma posição alfanumérica, onde signifique C = “Casado”, S = “Solteiro” D = “Desquitado”, V = “Viúvo” e O = “outros”. E em outro sistema a mesma informação pode ser representada por 1,2,3,4 e 5, respectivamente. Com a integração dos dados, este problema desaparece, conforme ilustra a Tabela 3, pois deve ser utilizada uma única representação para esta informação.

Tabela 3 - Integração de Dados

Sistema	Ambiente Operacional	Data Warehouse
Aplicação X	Representa estado civil como: C,S,D,V,O	1,2,3,4,5
Aplicação Y	Representa estado civil como: 1,2,3,4,5	1,2,3,4,5

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

Na organização universitária, a formação do ODS vai depender do estágio de seus sistemas legados. Dependendo do nível de integração de suas bases operacionais, o DW pode utilizar diretamente a área de estagiamento. Quando há incompatibilidade de sistemas e necessidade de integração por assuntos, o projetista deve considerar a construção do ODS.

3.2.3 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.

A variação no tempo pode apresentar-se de três maneiras:

- Em um Data Warehouse é comum que as informações sejam representadas em horizontes de tempo superiores a cinco anos. Nas aplicações operacionais, o período de tempo é mais curto. Varia entre 2 e 3 meses, pois é necessário uma resposta rápida às exigências das tarefas diárias, o que só pode ser conseguido com o processamento de poucos dados;
- Assim como os dados, os metadados, que incluem definições dos itens de dados, rotinas de validação, algoritmos de derivação, etc., também possuem elementos temporais para que, com eventuais mudanças nas regras do negócio, a empresa não perca dados históricos;
- Os dados armazenados corretamente no DW não serão mais atualizados, tendo-se assim uma imagem fiel da época em que foram gerados.

Os dados ainda podem ser separados em duas categorias:

- Os *dados detalhados atuais* são os dados de maior interesse por refletir os acontecimentos mais recentes. São em grande volume, porque são armazenados no nível mais baixo de granularidade e devem ficar armazenados em disco, um meio de acesso rápido, mas de difícil gerenciamento. Os dados detalhados atuais fornecem uma visão do comportamento recente e podem permitir a utilização de técnicas como mineração de dados e descoberta de conhecimento. O horizonte de tempo, para esses dados, normalmente é de dois anos.

- *Os dados detalhados antigos* são aqueles que não são acessados freqüentemente e por isso normalmente ficam armazenados em meios de armazenamento de baixo custo, pois possuem um grande volume por ficarem em um nível de detalhe consistente com os dados detalhados atuais. Mesmo não ficando armazenados em outros meios, como fitas, por exemplo, eles continuam fazendo parte do DW e podem ser carregados sempre que surgir necessidade de extrair informações.

Uma definição imprescindível refere-se ao período de atualização dos dados que diz respeito ao tempo necessário para que uma alteração sobre dados do ambiente operacional se reflita no DW. Uma vez que os dados tenham sido refletidos no ambiente operacional, as alterações precisam ser passadas para o DW. O problema é definir de quanto em quanto tempo isto deve ocorrer. Inmon (1997) sugere que pelo menos 24 horas deve se passar entre o momento em que a alteração é observada pelo ambiente operacional e sua repercussão no DW. Existem algumas razões para a existência deste intervalo. A primeira delas consiste no fato de que quanto mais rigidamente o ambiente operacional for emparelhado com o DW, mais dispendiosa e complexa será a tecnologia necessária. A segunda é que este intervalo de tempo possibilita a estabilização dos dados, diminuindo a chance de o DW receber informações incorretas.

3.2.4 Não Volatilidade

Uma das características do DW é a não volatilidade, ou seja, não existem alterações de dados, ocorrem a carga inicial e as consultas posteriores.

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 data warehouse não requer este grau de controle típico dos sistemas orientados a transações (Campos, 2001).

3.2.5 Localização

Em um Data Warehouse, pode-se encontrar os dados armazenados fisicamente de três formas:

- Armazenados em um *único local*, centralizando o banco de dados em um DW integrado, procurando maximizar o poder de processamento e agilizando a busca dos dados;
- *Distribuídos* por áreas de interesse, o que pode ser chamado de arquitetura federativa, com dados financeiros em um servidor, dados de marketing em outro e dados de manufatura em um terceiro lugar;
- Armazenados por *níveis de detalhes* em que as unidades de dados são mantidas no DW. Pode-se armazenar dados altamente resumidos em um servidor, dados resumidos em um nível de detalhe intermediário em um segundo servidor, e os dados mais detalhados (atômicos) em um terceiro servidor. Os servidores da primeira camada podem ser otimizados para suportar um grande número de acessos e um baixo volume de dados, enquanto servidores nas outras camadas podem ser adequados para processar grandes volumes de dados, mas baixo número de acessos.

3.3 Arquitetura do Data Warehouse

Um dos requisitos do DW é o de ele ser capaz de responder a consultas avançadas de maneira rápida, sem deixar de mostrar detalhes importantes à resposta. Para isso, ele deve possuir uma arquitetura que lhe permita coletar, manipular e apresentar os dados de forma eficiente e rápida. Mas construir um DW eficiente, que servirá de suporte a decisões para a empresa, exige mais do que simplesmente descarregar ou copiar os dados dos sistemas atuais para um banco de dados maior. Deve-se considerar que os

dados provenientes de vários sistemas podem conter redundâncias e diferenças. Então, antes de passá-los para o DW é necessário aplicar filtros sobre eles.

A constituição de um DW pode ser feita basicamente com dois modelos de arquitetura: (a) arquitetura top-down e (b) arquitetura bottom-up. Em termos práticos, o projetista pode construir um DW integrado, conjugando a base de informações para todos os negócios da organização (i.e., modelo top-down), ou pode partir da construção de data marts incrementais que juntos formarão o DW da organização. A seguir, comenta-se cada uma destas abordagens.

3.3.1 Arquitetura Top-Down

Introduzida por Inmon (1997), esta é a primeira arquitetura de data warehouse. A Figura 7 exemplifica tal arquitetura. O primeiro passo corresponde à extração, transformação, migração e carregamento de dados oriundos dos sistemas legados e/ou de fontes externas. No processo de extração, transformação e migração, os dados são retirados de suas origens e armazenados na área de *estagiamento de dados*. Em seguida, esses dados e os metadados necessários são carregados para dentro do Data Warehouse. Uma vez constituído o DW, os Data Marts são oriundos a partir de resumos do DW e dos metadados. O data warehouse é constituído por dados atômicos e também contém dados históricos detalhados. Em contraste, os data marts contêm dados altamente resumidos e levemente resumidos.

O modelo *top-down* pode fazer uso do modelo de dados E-R (Entidade-Relacionamento) normalizado (Firestone, 2000). Em contrapartida, o Data Mart usa o modelo de dados *esquema estrela* que é abordado no próximo tópico.

Na arquitetura *top-down*, a integração entre o data warehouse e os data marts é automática, desde que se mantenha uma disciplina na construção, partindo da premissa de que os data marts são subconjuntos do data warehouse. Existem algumas ferramentas para gerar data marts a partir do data warehouse, um exemplo é a Microstrategy's DSS Server.

As principais críticas ao modelo *top-down* referem-se aos custos de implantação, demora na obtenção de resultados parciais. Além disso, o modelo *top-down* apresenta alto grau de dificuldade no planejamento e implementação de um modelo único, não diferenciável e mestre à toda organização (Kimball et. al. 1998).

Nas organizações universitárias, em particular, pode ser extremamente difícil obter tal visão integrada, mesmo que seus sistemas legados sejam ampliados.

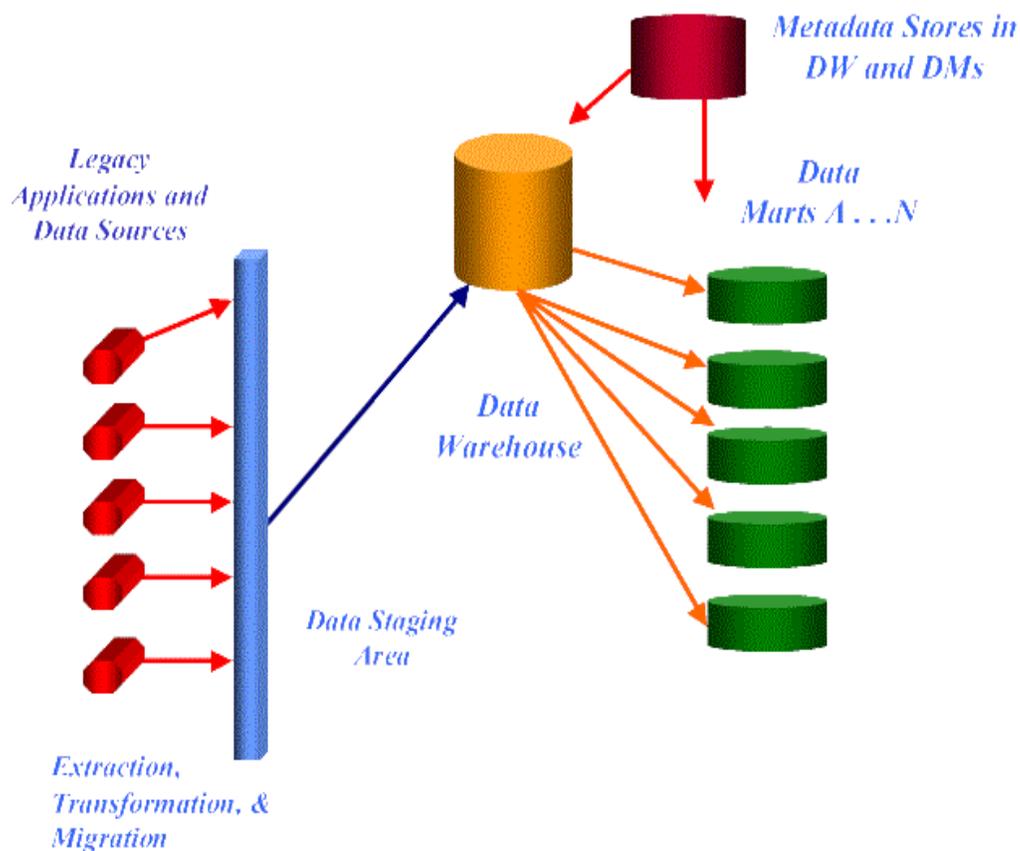


Figura 7 - Arquitetura Top-Down segundo Firestone (2000)

3.3.2 Arquitetura *Bottom-up*

A segunda arquitetura de DW é a *bottom-up*. A arquitetura *top-down* levou muito tempo para ser implementada. Politicamente era inaceitável e o processo de implantação muito caro e demorado. A arquitetura *bottom-up* oferece uma alternativa de construção incremental a Figura 8 apresenta a referida arquitetura (Firestone, 2000).

A idéia central desta arquitetura é construir o data warehouse de forma **incremental**. Com o passar do tempo, independente do desenvolvimento dos data marts. Na literatura, essa arquitetura foi introduzida por Ralph kimball. O processo tem início com a construção de um ou mais data marts. Também existe a área de estagiamento de dados, que fica separada para cada Data Mart.

Os data marts geralmente não usam modelos de dados E-R (Entidade-Relacionamento) de forma normalizada (Firestone, 2000). Geralmente é reconhecido que quando os Data Marts utilizam bases de dados relacionais, fazem uso do modelo *esquema estrela*, ou seja, a modelagem multidimensional (kimball et. at. 1998).

Na arquitetura *top-down*, os data marts usam dados levemente resumidos e altamente resumidos, mas na arquitetura *bottom-up* usa dados atômicos e detalhados, incluindo dados históricos. Considerando que os data marts são blocos de suporte do DW, devem conter no conjunto todos os dados que apareceram no projeto do data warehouse.

O modelo *bottom-up* difere do modelo *top-down* também no sentido que não provê nenhum componente de metadado comum para os data marts. Esta é a diferença mais importante entre as duas arquiteturas do ponto de vista de integração (Firestone, 2000).

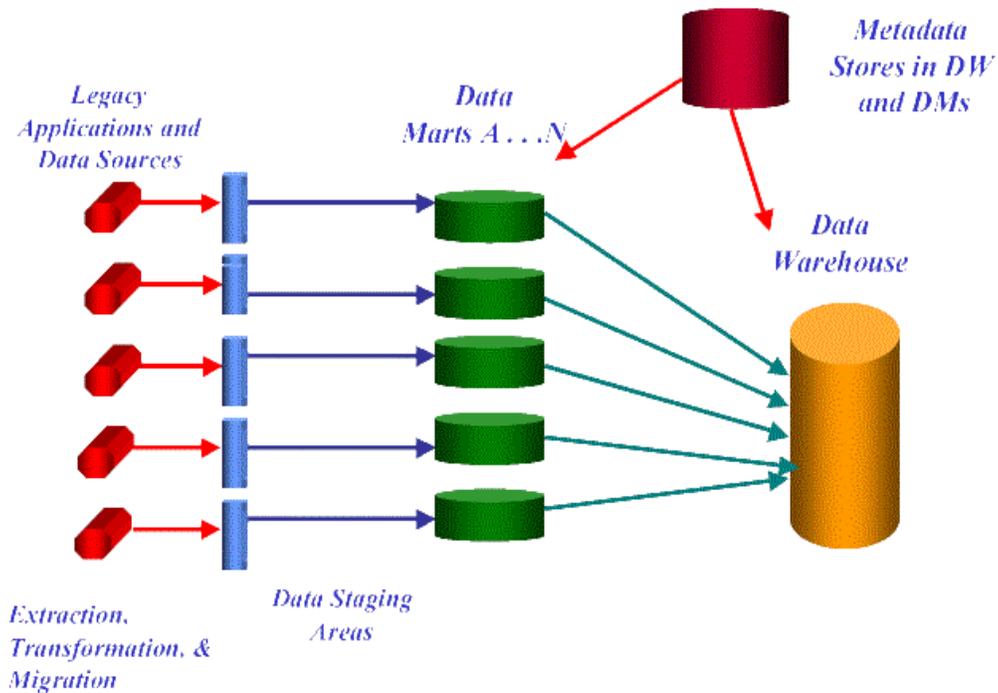


Figura 8 - Arquitetura Bottom-up segundo Firestone (2000)

3.3.3 Data Warehouse Distribuídos

Outra área de tendências em DW é a possibilidade de integração de diferentes DW, consistindo dezenas ou centenas de servidores separados, com sistemas operacionais diferentes e bancos de dados distintos, incluindo uma variedade de ferramentas OLAP (Kimball et. at. 1998). Mesmo na visão tradicional de construção de DW, uma empresa pode optar pelo modelo distribuído quando os processamentos operacionais de informações ocorrem tanto no nível local como global. Essa situação é comum em empresas com sede matriz e sedes filiais (Inmon, 1997).

Os data warehouses distribuídos podem ser muito adequados às instituições de ensino superior, tanto quando se concentra na visão multi-campi, como quando se vislumbra a arquitetura de informação da universidade distribuída em DW constituídos por nível organizacional (ex. DW por centros).

3.4 Granularidade

Granularidade diz respeito ao nível de detalhe ou de resumo contido nas unidades de dados existentes no DW. Quanto maior o nível de detalhes, menor o nível de granularidade.

Existem diferentes níveis de detalhe no Data Warehouse. Uma das abordagens para o projeto de estrutura de dados de um DW foi proposta por Inmon (1997), que contempla três níveis de detalhe para armazenamento de dados:

- Dados de detalhe corrente
- Dados levemente resumidos (o departamental ou datamart)
- Dados altamente resumidos (o individual)
- Para cada nível de detalhe, um horizonte de tempo é estabelecido. A granularidade constitui-se em uma das principais questões do projeto de um DW. O nível de granularidade afeta profundamente o volume de dados que residem no data warehouse, e, ao mesmo tempo, interfere no tipo de consulta que pode ser atendida. A granularidade é relevante para diferentes componentes do DW (Figura 9).

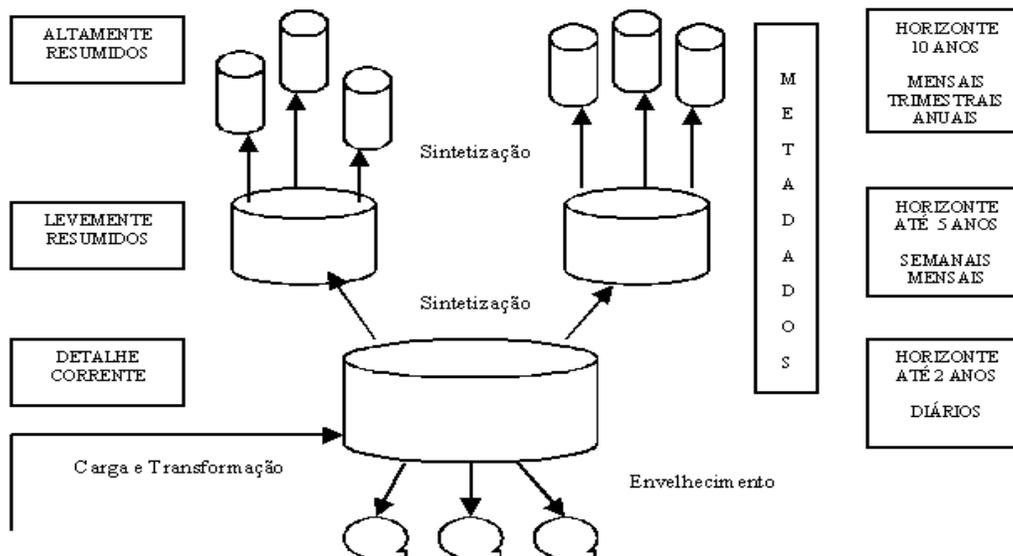


Figura 9 - Componentes do DW. Adaptado de Inmon (1997)

A preocupação mais importante em relação aos dados encontra-se no nível de detalhe corrente, visto que estes dados refletem os acontecimentos mais recentes e que são de maior interesse para a instituição. Os dados atuais, além de permitirem uma visão do comportamento recente, podem ser utilizados para aplicação de técnicas como descoberta de conhecimento e mineração de dados. Neste nível de detalhamento, os dados são armazenados em séries diárias com horizonte de tempo de até 2 anos.

Os dados levemente resumidos compreendem um nível intermediário e são separados do detalhe de baixo nível encontrado nos dados detalhados atuais. O conteúdo dos dados sofre modificações em relação aos dados de detalhe corrente. Neste nível, trabalha-se com horizonte de tempo de até 5 anos e os dados são armazenados em séries semanais e mensais.

O seguinte nível de dados proposto no projeto de estrutura do DW é o dos dados altamente resumidos, onde são compactos e de fácil acesso. Normalmente, neste nível residem as informações destinadas aos *Sistemas de Informação Executivas*, ao passo que os dois níveis anteriores

compreendem informações destinadas aos *Sistemas de Apoio à Decisão* (Inmon, 1997).

3.4.1 Níveis duais de granularidade

Na maioria das vezes, o que ocorre é uma grande demanda por eficiência no armazenamento e recuperação de dados, bem como a possibilidade de analisar dados com um nível de detalhe maior. Segundo Inmon (1997), quando uma organização possui grande quantidade de dados no DW, faz sentido pensar em dois ou mais níveis de granularidade na parte detalhada do data warehouse. Na realidade, a necessidade de existência de mais de um nível de granularidade é tamanha que a opção de projeto que consiste em níveis duais de granularidade deveria ser padrão para quase todas as empresas (Inmon, 1997).

Para tornar mais claro, a Figura 10 exemplifica a aplicação do nível dual de granularidade.

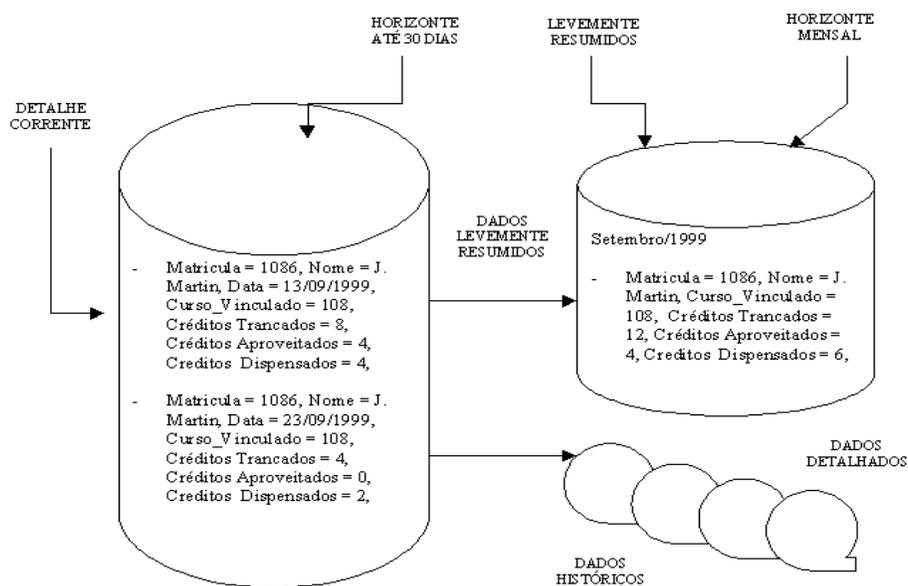


Figura 10 - Nível Dual de Granularidade. Adaptado de Inmon (1997)

Chamado nível "dual" de granularidade, o desenho apresentado na Figura 10, - uma Instituição de Ensino - se enquadra no requisito da maioria das empresas. No nível operacional está localizada uma grande quantidade de detalhes. A maioria desses detalhes é necessária para os sistemas de matrícula e financeiro. À medida que os dados fluem do armazenamento operacional de 30 dias de dados, eles são resumidos, por acadêmico, na forma de campos apropriados para utilização pela análise dos SADs. O registro referente a J. Martin apresenta o número de créditos trancados, aproveitados e dispensados.

Há um volume de dados significativamente menor no banco de dados *levemente resumido*. Obviamente, existe um limite para o nível de detalhe que pode ser acessado no banco de dados *levemente resumido*.

A segunda camada de dados existentes no data warehouse é o nível mais baixo de granularidade, onde são armazenados dados históricos, ou seja, todos os detalhes oriundos do ambiente operacional são armazenados. Há uma verdadeira montanha de dados neste nível. Por essa razão, faz sentido armazenar os dados em um meio como fita magnética.

Ao criar dois níveis de granularidade, o projetista do DW está contemplando o atendimento a processamento de dados *levemente resumidos* que são compactos e de fácil acesso, bem como para ocasiões em que um maior nível de detalhe deve ser investigado que geralmente gira em torno de 5 por cento do tempo ou menor, enquanto que 95 por cento do processamento do SAD é feito no nível *levemente resumido*.

Em virtude dos custos, da eficiência, da facilidade de acesso e da possibilidade de atender a qualquer consulta que possa ser respondida, para a maior parte das instituições, o nível dual de dados torna-se a melhor opção arquitetônica para o nível de detalhe do data warehouse.

Quando a granularidade de um DW é estabelecida adequadamente, os demais aspectos de projeto e implementação fluem tranqüilamente; quando essa granularidade não for estabelecida, todos os demais aspectos se complicam.

O ponto de equilíbrio na escolha dos níveis adequados de granularidade, como citado anteriormente, se dá entre o gerenciamento do volume de dados e a armazenagem dos dados, segundo um nível de granularidade tão alto que a utilização detalhada dos dados não possa ser feita.

3.5 Modelo de dados

Como visto no item 3.1.3, a modelagem de dados do DW consiste na metodologia pela qual as informações do DW são representadas e armazenadas.

Na literatura encontra-se diferentes enfoques sobre modelagem de dados. Embora haja mais de um modelo para construir um Data Warehouse com sucesso a modelagem dimensional tem-se firmado como meio mais efetivo para o projeto do DW. Nesse trabalho será apresentado, de forma resumida, o modelo dimensional de dados de Ralph Kimball, que pode ser pesquisado mais detalhadamente em (Kimball, 1996).

Perguntas complexas que envolvam questões de análise dos negócios de uma empresa, geralmente requerem uma visão dos dados de várias perspectivas. As respostas a esse tipo de pergunta podem levar a tomada de decisões acertadas ou não. As ferramentas com base em SQL ajudam na pesquisa de dados relacionados a este tipo de consulta, o que geralmente ocorre é que os dados não são conseguidos em um tempo curto, principalmente devido à falta de flexibilidade da ferramenta.

Um bom exemplo seria uma rede de lojas que esteja querendo melhorar o desempenho de suas vendas ou saber se suas promoções estão trazendo resultado. Para responder a essas perguntas, faz-se necessário examinar os dados sobre as vendas disponíveis nas lojas. Uma análise deste tipo requer uma visão histórica baseada no volume de vendas, sob múltiplas perspectivas, como, por exemplo: a quantidade de vendas por produto,

quantidade por marca, quantidade de vendas por filial, total de vendas por período de tempo.

Chamam-se *dimensões* as diferentes perspectivas envolvidas, no caso produto, marca, filial e mês. Considera-se também um conjunto de *medidas*, tal como vendas ou despesas com promoção. A seguir, avaliam-se as agregações destas medidas, segundo diversas dimensões. Essas agregações ficam armazenadas para acesso futuro. Por exemplo, calcula-se a média de todas as vendas por todos os meses por filial. A maneira de como estas agregações são armazenadas pode ser vista em termos de dimensões e coordenada, dando origem ao termo *multidimensional*.

A modelagem multidimensional é o nome de uma técnica de projeto lógico, freqüentemente usada para data warehouse, cujo principal objetivo é apresentar o dado em uma arquitetura padrão, que possibilite acesso de alta-performance (Kimball, 1996). Cada eixo no espaço multidimensional corresponde a um campo ou coluna de uma tabela relacional e cada ponto um valor correspondente à interseção desses campos ou colunas.

Dados multidimensionais podem ser armazenados e representados em estruturas relacionais. Para isso é necessário utilizar formas específicas de modelagem, como o modelo "Estrela" descrito a seguir.

3.5.1 Modelo estrela

Conforme Valente (1996), tradicionalmente, modelos de bases de dados relacionais apresentam tabelas com relacionamentos complexos e com múltiplas uniões circulares entre dois pontos do modelo. Para a maioria dos usuários que utilizam ferramentas para compor suas consultas, é necessário que o acesso à base de dados seja simples o suficiente para facilitar o acesso direto à base de dados. Para acomodar as necessidades de todos os usuários e facilitar a atualização do DW, o projetista deve criar um modelo que o usuário final possa facilmente entender em termos do negócio.

O principal tipo de modelo dimensional é o chamado modelo Star (Estrela), onde existe uma tabela dominante no centro chamada de *tabela de*

fatos, com múltiplas junções, conectando-a a outras tabelas, sendo estas chamadas tabelas de dimensão. Cada uma das tabelas secundárias possui apenas uma junção com a tabela central. O modelo Estrela na Figura 11 tem a vantagem de ser simples e intuitivo, mas também utiliza enfoques de indexação e união de tabelas. Para Kimball (1996), o modelo entidade-relacionamento não é o mais adequado para se analisar os dados no ambiente gerencial. O modelo dimensional é o mais apropriado para este ambiente.

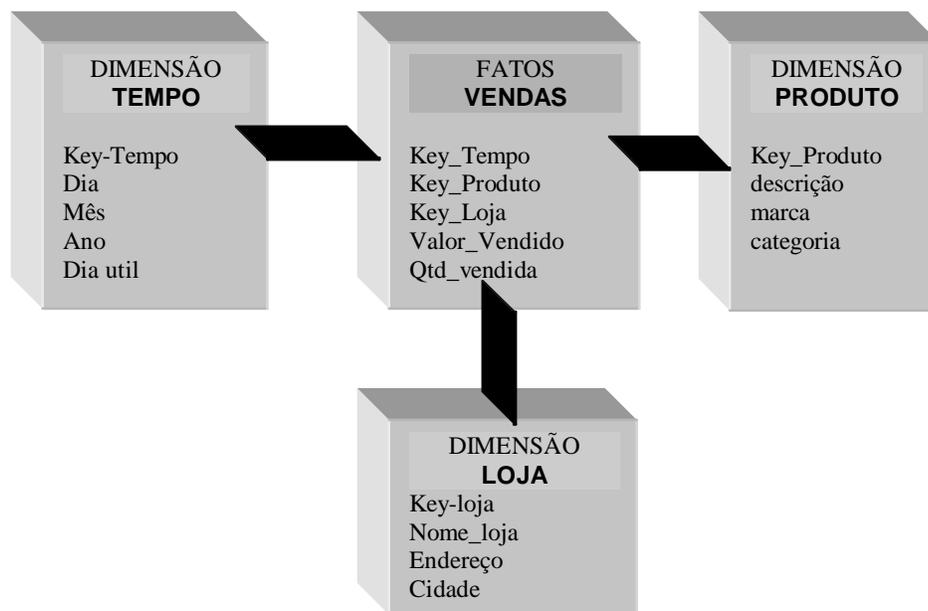


Figura 11 - Modelo Dimensional do tipo Estrela. Adaptado de Kimball (1996)

A tabela de fatos contém milhares (ou milhões) de valores e medidas do negócio da empresa, como transações de vendas ou compras. Cada uma destas medidas é tomada segundo a interseção de todas as dimensões. No exemplo da loja, as medidas numéricas são valor vendido e a quantidade de unidades vendidas. Uma característica importante da tabela de fatos é a esparsidade, ou seja, se não existe um cruzamento para alguns valores das dimensões, a tabela de fatos não armazena zeros.

As tabelas de dimensão armazenam as descrições textuais das dimensões do negócio. Cada uma dessas descrições textuais permite definir um componente da respectiva dimensão. Uma das principais funções dos atributos de tabelas de dimensão é servir como fonte para restrições em uma consulta ou como cabeçalhos de linha no conjunto de respostas do usuário. Tabelas de dimensões tendem a utilizar tipo de caracteres ao invés de numéricos, de forma que suas linhas são muito mais longas mas em pouca quantidade, ocupando pequena percentagem de espaço em disco. As tabelas de fatos podem utilizar até 95% da área destinada ao data warehouse (Barquini, 1996).

Um fator importante relacionado à tabela de fatos é que, como representa os relacionamentos muitos-para-muitos entre as tabelas de dimensão, esta tem como chave primária uma chave composta de todas as chaves estrangeiras das tabelas de dimensão (Kimball, 1996).

Para um bom desempenho do modelo Estrela, é necessário que os projetistas saibam antecipar, na modelagem do DW, as consultas mais freqüentes a serem realizadas pelos usuários.

Segundo Kimball (1996), o modelo dimensional apresenta várias vantagens no que diz respeito à sua utilização para o DW. Dentre estas estão:

- Arquitetura padrão e previsível;
- Dimensões do modelo são equivalentes, ou seja podem ser vistas como pontos simétricos para a tabela de fatos;
- Modelo dimensional é flexível, pois permite à inclusão de novos elementos de dados, bem como mudanças que ocorram no projeto;
- Facilidade na alteração das tabelas de fatos e dimensão;
- Todas as aplicações que existiam antes da mudança permanecem rodando sem problemas.

3.5.2 Variação do modelo estrela

Outro tipo de estrutura bastante comum, conforme abordado em Campos (1997), é o modelo de dados floco de neve ou "*Snowflake*", que consiste em uma extensão do modelo estrela apresentado no tópico anterior, onde cada uma das "pontas da estrela" passa a ser o centro de outras estrelas. O modelo surge da des-normalização e redução de cardinalidade do modelo estrela "quebrando-se" a tabela original ao longo de hierarquias existentes em seus atributos. Este modelo pode ser ilustrado imaginando-se a classificação de um automóvel, onde a dimensão do produto possui uma hierarquia definida: categoria se divide em marca, e marca se divide em produtos. Kimball (1996) aconselha os projetistas a resistirem à tentação de transformar modelo Estrela em modelos Floco de Neve, devido ao impacto da complexidade deste tipo de estrutura sobre o usuário final, enquanto que o ganho em termos de espaço de armazenamento seria pouco relevante.

3.6 Metadados

O último componente a ser considerado na estrutura do DW são os metadados. Os Metadados compõem o centro nervoso do Data Warehouse. Sem os metadados, o Data Warehouse e seus componentes associados no ambiente projetado são meramente componentes soltos, funcionando independentemente e com objetivos separados (Inmon, 1999). O conceito de metadados não é novo, porém, o papel dos metadados e sua importância no ambiente de apoio à tomada de decisão certamente é novo (David, 1999). No ambiente operacional, os usuários interagem com as informações através de telas e formulários, permitindo a estes desconhecer como as informações são armazenadas nos Bancos de Dados. Os metadados são tratados quase como algo a ser deixado para mais tarde, sendo relegados normalmente, em nível de importância, à documentação de sistemas. Já no ambiente DW, ao contrário dos ambientes operacionais, os usuários são analistas e executivos que buscam por informações relacionadas ao negócio. A procura por dados neste

caso deve ser a mais clara e lógica possível. Neste caso, os metadados assumem um papel de suma importância, a qual eles nunca tiveram, em um ambiente operacional.

3.6.1 Definição

Metadado é tipicamente definido na literatura como sendo "dados sobre dados", segundo Inmon (1997), metadados podem ser caracterizados "como um diretório que auxilia os analistas de SAD a localizarem os componentes do DW". Mantém informações sobre a estrutura dos dados, segundo a visão do programador e segundo a visão do analista de SAD. Mantém ainda informações sobre o modelo de dados, especificação dos arquivos (chaves e atributos), histórico de extrações, controle de acesso, etc. Uma melhor definição seria a de que metadados é o significado dos dados, ou seja, de que metadado é uma abstração dos dados, ou ainda, dados de mais alto nível que descrevem dados e um nível inferior. Sem metadados, os dados não têm significado. São exemplos de metadados as descrições de registros em um programa de aplicação ou o esquema de um banco de dados descrito em seu catálogo ou ainda as informações contidas em um dicionário de dados.

3.6.2 Classificação

Os metadados podem ser classificados em dois tipos (David, 1999), os metadados técnicos (*technical metadata*), usados pelo pessoal de desenvolvimento e manutenção, e os metadados de negócio (*business metadata*), usados pelos analistas de negócio. Existem ainda os metadados centrais (*core metadata*), usados pelos administradores de dados e do banco de dados, que alguns autores classificam como metadados técnicos.

Os metadados técnicos provêm aos desenvolvedores e administradores de banco de dados todas as descrições técnicas dos dados e suas operações. Especificam nomes de atributos, tipos de dados, fontes a partir dos quais os dados são extraídos, regras de transformação, etc.

Os metadados de negócios são comumente usados por usuários finais e são derivados e inferidos por especificações existentes. É uma ligação entre o Data Warehouse e usuários de negócios, que são executivos ou analistas de negócio e tendem a ser menos técnicos, mas necessitam ter uma visão clara das regras do negócio. Desta forma, os metadados de negócios provêm uma descrição de negócio dos objetos da informação. Podem especificar o cálculo de um valor, o cálculo de custo de um produto, etc. A Figura 12 exemplifica os tipos de metadados.

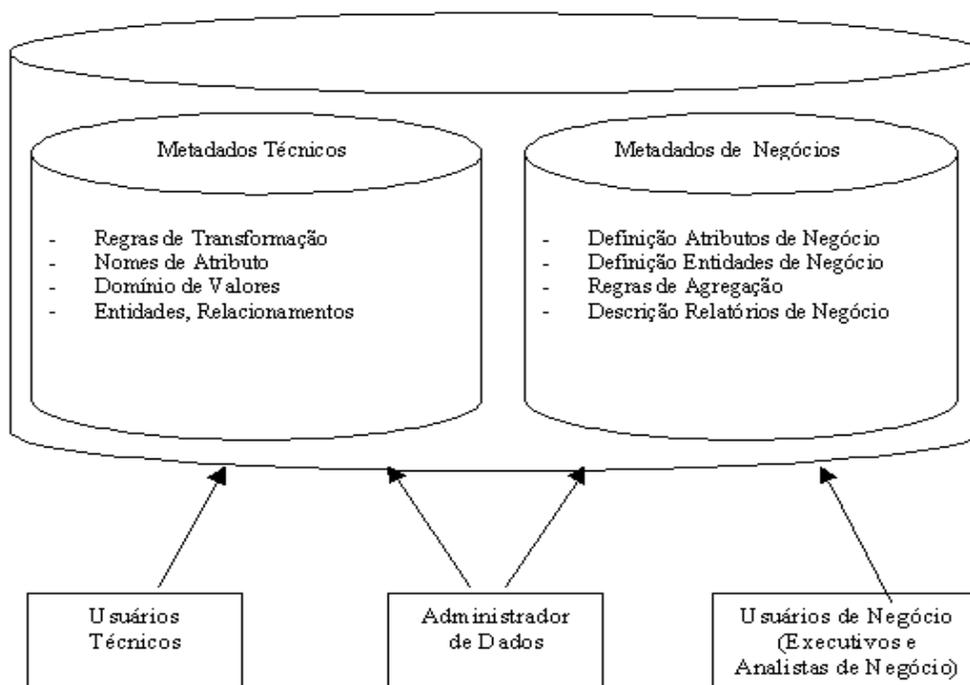


Figura 12 - Repositório de Metadados segundo David (1999)

Os metadados centrais representam a forma pela qual os dados são tratados no sistema. São relacionados ao desempenho do Data Warehouse. São úteis para a geração de consultas, de modo que elas tenham um bom desempenho. Exemplos são listas das tabelas acessadas, colunas sobre as quais o *JOIN* é realizado, estatísticas de *IO* para uma consulta, etc.

Os metadados técnicos e os de negócios fazem parte dos processos existentes na empresa, e, portanto, são caracterizados como metadados formais uma vez que se encontram formalizados e documentados. Já os metadados que não estão documentados, que estão somente na cabeça das pessoas, são caracterizados como metadados informais.

É necessário que a maior quantidade de metadados informais, a partir das entrevistas com administradores de dados, analistas de negócios, sejam documentados e transformados em metadados formais.

3.6.3 Fonte de metadados

Existem dois tipos de fontes de metadados – formal e informal. Estas fontes são usadas, tanto para os metadados técnicos como para os metadados de negócios (David, 1999).

a) Fontes formais

São fontes em que os metadados foram documentados, discutidos e acordados pelos tomadores de decisões da empresa, ou seja, estes metadados fazem parte dos processos formais da empresa. Comumente são armazenados em ferramentas ou documentos que são mantidos e distribuídos na organização. A Tabela 4 apresenta exemplos de fonte de metadados formais.

Tabela 4 - Exemplo de Metadados Formais

Tipos Formais de Metadados	Localizações Comum
Ferramentas de Extração e Transformação	Prism Warehouse Executive, Vmark DataStage, D2K Tapestry, ETI Extract
Modelos de Dados Lógico e Físico	Ferramentas CASE (Cayenne, Power Designer, Erwin) e Aplicações (SAP, Peoplesoft)
Políticas de Negócios	MS Word, MS Excel
Dicionário de Dados	Ferramentas Case, MS Access, MS Excel

Os dados oriundos de ferramentas CASE, provenientes do desenvolvimento de sistemas, tanto transacionais, como de suporte à decisão, constituem-se uma fonte de metadados de excelente qualidade, pois são estruturados, tendendo com isso a possibilidade de uma integração automática entre a origem dos metadados e o repositório do ambiente em construção.

Os metadados provenientes de outras aplicações, dependendo do nível de estruturação, podem também ser de boa qualidade e, via de regra, podem ser carregados automaticamente do ambiente, onde ela foi desenvolvida.

Quando não contamos com uma documentação eficiente de um sistema, somos obrigados a extrair as informações de um nível mais baixo, ou seja, do fonte dos programas que implementam seus processos e os objetos que armazenam seus dados.

b) Fontes Informais

Metadados informais são os que fazem parte do conhecimento corporativo, regras e procedimentos que não estão dentro dos procedimentos padrões da empresa (David, 1999). Este tipo de informação encontra-se somente na cabeça dos empregados, ou está em anotações informais nas escrivatinhas dos mesmos. Eles não estão formalmente documentados. Normalmente metadados informais provêm informações tão valiosas que eles devem ser documentados para serem considerados meta dados formais. Uma vez que todas as organizações diferem, fica difícil dizer onde se podem encontrar os metadados informais. Abaixo uma lista mais comum dos metadados informais.

- Administrador de dados
- Regras de negócio
- Definições do negócio
- Lista de produto dos competidores
- Transformações e Sumarizações

Entrevistas com profissionais da empresa que entendam do negócio, bem como com o administrador de dados, são fontes de vital importância, muito embora não tenhamos os dados de forma estruturada. Deles pode-se obter informações sobre todos os metadados informais, bem como as regras de transformação para gerar o dado no ambiente de suporte à decisão, requisitos para testes dos dados e indicadores da qualidade dos dados. Estes dados serão inseridos manualmente no repositório.

Definições do projeto deste sistema são informações que ficarão armazenadas no repositório de metadados, mas que precisam ser definidas. É o caso da definição dos responsáveis por cada dado, pelos metadados, e os perfis de acesso de cada usuário às informações do ambiente de suporte à decisão. O próprio ambiente de suporte à decisão, ao longo de sua utilização, gerará informações a respeito de atualizações, início e final de processamento, tempo de uma consulta, datas de carga e atualização.

3.6.4 Problemas na criação dos Metadados

Diferentes pessoas usam diferentes métodos para obter e criar novos dados. Esses, freqüentemente, não são documentados corretamente, ou seja, muitas vezes são documentados com uma variedade de ferramentas resultando na criação de dados discrepantes (Brackett, 1996).

A tarefa de obter e manter eficientes metadados não deve ser considerada de fácil solução. Existe uma carência de bons metadados, o que muitas vezes limita a utilização plena dos recursos dos dados, e leva à criação de novos dados discrepantes, os quais não são adequadamente documentados. Assim, cria-se a tendência de um ciclo de criação de dados discrepantes na organização.

A qualidade de dados obtidos no DW representa um dos maiores riscos que devem ser gerenciados quando são projetados os metadados. Risco é definido como prejuízos, danos ou perdas por usar os dados inapropriadamente. Segundo Brackett (1996), para melhorar a qualidade dos dados dever-se-á incluir três elementos chaves:

- Definir as expectativas dos usuários com relação aos dados. Essas expectativas são definidas, usando-se os metadados e qualidade de dados métricas para mensurar as características dos dados (exatidão, legibilidade, consistência), apropriadas para cada uso.
- Definir os riscos em termos do que pode causar falhas na qualidade dos dados, e iniciar ações ou projetos para minimizar esses riscos.
- Incorporar validações, verificações e certificações como uma prática integral para analisar e detectar problemas potenciais de qualidade de dados (Brackett, 1996).

3.7 Data Mart

Na construção de um Data Warehouse Corporativo Integral, há uma gama de fatores que afetam a complexidade como, por exemplo, a construção do projeto que é lenta e cara, com o objetivo de equilibrar os gastos e oferecer resultados em prazos mais curtos, é possível construir Data Marts, que na verdade são pequenos DW departamentais.

Uma das principais vantagens na construção de Data Marts encontra-se na redução do tempo de implementação (em média de 120 dias cada), e o fator preço.

Na literatura encontramos vários conceitos. Segundo Berry (1997) "Data mart é um sistema especializado que fornece juntos os dados necessários para departamento ou uma aplicação relacionada". Segundo Inmon (1997), os Data Marts são subconjuntos de dados da empresa armazenados fisicamente em mais de um local, geralmente divididos por departamento (Data Marts "departamentais")

Organizações cujas exigências são mais modestas, ou seja, aquelas que necessitam construir um DW para departamentos específicos,

podem optar por construir pequenos Data Marts, os quais utilizam uma arquitetura baseada em rede.

Existem diferentes alternativas para se implementar os Data Marts (Oneil, 1997), originalmente eram desenvolvidos a partir de um DW central. A Figura 13 exemplifica tal situação, segundo a arquitetura *top-down*.

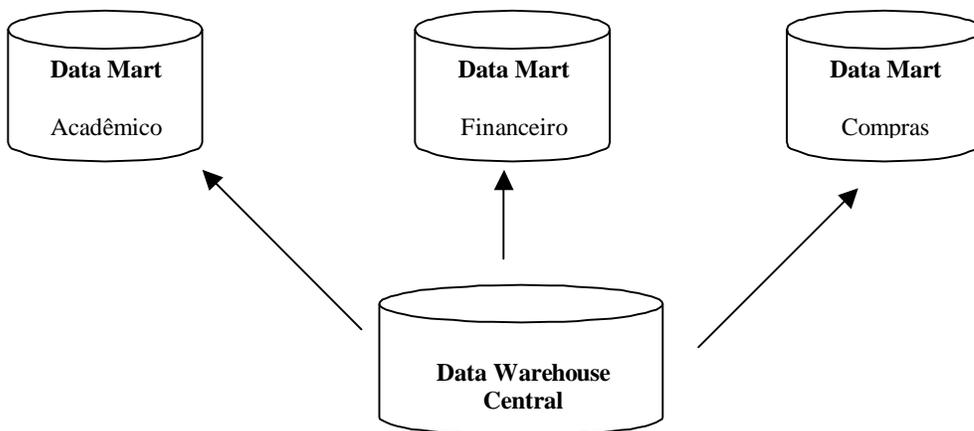


Figura 13 - Arquitetura de Data Marts Departamentais (arq. *Top-down*).

Nesta arquitetura, grupos de usuários acessam diretamente os Data Marts de seus respectivos departamentos. Apenas aquelas análises que necessitam de uma visão global da organização são realizadas sobre o DW. Segundo Inmon (1997), os Data Marts se diferenciam do DW pelos seguintes fatores:

- São personalizados: Atendem às necessidades de um departamento específico ou grupos de usuários;
- Menor volume de dados: Por atenderem a um único departamento, armazenam um menor volume de dados;
- Histórico limitado: Os Data Marts raramente mantêm o mesmo período histórico que o DW. Geralmente, o DW mantêm um histórico que abrange de 5 a 10 anos, enquanto que os Data Marts devem optar em manter o mesmo período, porém com os dados em um nível maior de granularidade,

ou um menor período, com os dados armazenados no mesmo nível de granularidade do DW;

- Dados sumarizados: Os Data Marts geralmente não mantêm os dados no mesmo nível de granularidade do DW, ou seja, os dados são, quase sempre, sumarizados quando passam do DW para os Data Marts.

Alguns problemas podem ocorrer com Data Marts, tais como:

- Dificuldade de trabalhar com dados muito abrangentes;
- Data Mart possibilita um suporte à decisão limitado a universos restritos;
- Corre-se o risco de desvio do modelo original, pois pode acontecer um crescimento desestruturado;
- Em virtude de estar em constante aperfeiçoamento pode vir a ocorrer replicações de informações em vários locais, o que dificulta uma futura integração entre os Data Marts em um único Data Warehouse.

Como se viu na sessão 3.3, a forma mais comum de construção de data marts está baseada na arquitetura bottom-up, como ilustra a figura 14.

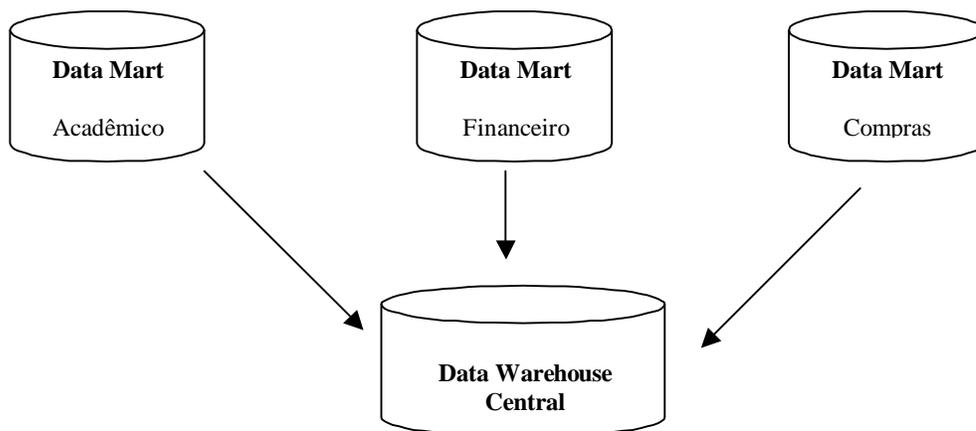


Figura 14 - Arquitetura de Data Marts Departamentais (arq. Bottom-up)

3.8 O Desenvolvimento do Data Warehouse

Para o desenvolvimento de um DW não existe uma receita pronta, o que existe são ferramentas que contemplam etapas desde a extração e análise dos dados até a construção propriamente dita e, também, a fase de gerenciamento do Data Warehouse.

Ao iniciar a construção, é importante observar o valor do investimento, que geralmente situa-se na casa dos milhões de dólares, em virtude do projeto estar ligado aos negócios da empresa. Além do pessoal da parte técnica é imprescindível a participação constante dos profissionais da área executiva, pois qualquer passo em falso pode causar graves prejuízos e levar a empresa a consultar informações não confiáveis, e, conseqüentemente, vir a tomar decisões erradas.

A seguir, serão abordadas as funções que as pessoas desempenham dentro do ambiente do DW, bem como algumas fases e processos necessários para a obtenção do modelo de dados do DW. Um fato importante é que, dependendo do tamanho do projeto, é possível desenvolver um *Data Mart* que poderá ser definitivo, dependendo do escopo do projeto, ou poderá mais tarde ser agrupado a outros data marts com o objetivo de formar um único DW.

3.8.1 As Funções em um Ambiente de Data Warehouse

Para a construção de uma Data Warehouse é necessário desenvolver uma série de funções. Muitas vezes, devido à dimensão do projeto e ao tipo de tecnologia utilizada, podem ser necessárias várias pessoas para realizar as funções previstas. Barquini (1996) descreve as responsabilidades nos processos do DW na (tabela 5).

Vale lembrar que nem todas as funções são necessárias durante todo o tempo de vida de um DW. Estas funções podem variar conforme o

estágio em que se encontra o DW, bem como, podem ser agrupadas para que uma só pessoa realize várias delas ao mesmo tempo.

Tabela 5 - Funções em um Data Warehouse.

Função	Responsabilidades
Gerente do Data Warehouse	<ul style="list-style-type: none"> ▪ Define as estratégias pertinentes ao DW ▪ Planeja e gerencia o DW ▪ Comunica os objetivos do DW para a equipe de desenvolvimento
Arquiteto de Dados	<ul style="list-style-type: none"> ▪ Desenvolve o modelo de dados ▪ Analisa as exigências de dados ▪ Desenha as estruturas dos dados ▪ Define as visões gerenciais para os dados
Administrador de Metadados	<ul style="list-style-type: none"> ▪ Define os padrões de metadados ▪ Gerencia o repositório dos metadados
Administrador do Banco de Dados	<ul style="list-style-type: none"> ▪ Cria as estruturas físicas no BD ▪ Monitora o carregamento dos dados e a performance das consultas
Usuário de nível gerencial	<ul style="list-style-type: none"> ▪ Descreve os dados necessários ▪ Especifica as regras de negócio ▪ Testa os resultados das transformações dos dados
Analista de processos e aplicações	<ul style="list-style-type: none"> ▪ Desenvolve as aplicações de suporte à decisão
Especialista em Aplicações Operacionais	<ul style="list-style-type: none"> ▪ Indica onde estão os dados nos sistemas operacionais
Analista e programador de conversões	<ul style="list-style-type: none"> ▪ Indica as fontes de dados para o DW ▪ Desenvolve os programas para selecionar e carregar os dados
Especialista em suporte técnico	<ul style="list-style-type: none"> ▪ Desenvolve as atividades técnicas como instalar e configurar máquinas
Instrutor	<ul style="list-style-type: none"> ▪ Treina os usuários para acessar o DW

3.8.2 Estratégias no desenvolvimento de um Data Warehouse

O processo inicial de construção de um DW, basicamente, consiste na extração dos dados de fontes internas, bem como a transformação e limpeza dos mesmos. Para que o projeto do DW apresente sucesso, é

importante escolher uma estratégia correta, a qual deve ser adequada às características e necessidades específicas do ambiente onde o DW será implementado. É importante fazer uma escolha fundamentada em pelo menos três dimensões:

- *Escopo do Data Warehouse* (departamental, empresarial): É importante estabelecermos o escopo, quanto maior for o escopo mais valor o DW tem para a empresa, conseqüentemente mais cara e trabalhosa será sua criação e manutenção. Em virtude disso, muitas empresas optam pela metodologia *bottom-up*, começando com a construção de Data Marts e só após obterem um *feedback* de seus usuários passam a expandir o escopo com a finalidade de integrar os Data Marts em um único Data Warehouse.
- *Grau de redundância de dados*: Com relação à redundância de dados encontramos três níveis:
 - *Data Warehouse Virtual*: Tem como objetivo prover os usuários finais com facilidades no que se refere à extração das informações diretamente dos bancos de produção, com isso não ocorrendo redundância, mas poderá sobrecarregar o ambiente operacional;
 - *Data Warehouse Centralizado*: é um único banco de dados físico contendo todos os dados para uma determinada área funcional, um departamento ou uma empresa, sendo usado onde existe uma necessidade comum de informações;
 - *Data Warehouse Distribuído*: é aquele em que seus componentes encontram-se distribuídos por diferentes bancos de dados físicos. Normalmente possui um grau de redundância elevado e, conseqüentemente, apresenta procedimentos mais complexos de carga e manutenção.
- *Tipo de usuário alvo*: É importante selecionar um grupo de usuários, prover ferramentas adequadas, construir um protótipo e deixar que os usuários analisem de forma minuciosa as amostras dos dados. Somente após o aval do grupo no que se refere aos requisitos e funcionamento é que o DW será de fato carregado com dados dos sistemas operacionais da empresa e dados externos.

É necessário ter consciência que ainda não existe uma metodologia consolidada para o desenvolvimento de um DW. Campos (1997) afirma que o que se vê na literatura e nas histórias de sucesso de implementações em empresas, são propostas no sentido de construir um modelo dimensional a partir do modelo de dados corporativo ou departamental (base dos bancos de dados operacionais da empresa), de forma incremental.

3.8.3 Desenvolvimento de um Data Warehouse utilizando o modelo Estrela

O modelo estrela armazena e representa dados multidimensionais em estruturas relacionais. A seguir estão relacionados os principais tópicos sugeridos para um desenvolvimento, utilizando o modelo Estrela.

Segundo Kimball (1996), desenvolver um DW é uma questão de juntar as necessidades dos seus usuários com a realidade dos dados disponíveis. Kimball elenca um conjunto de nove pontos fundamentais no projeto da estrutura de um Data Warehouse, que constituem definições a serem feitas, as quais correspondem as etapas do projeto. Portanto, em um projeto que utiliza o modelo do tipo estrela, deve-se fazer as seguintes definições.

- Os processos e, por conseqüência, a identidade das tabelas de fatos;
- A granularidade de cada tabela de fatos;
- As dimensões de cada tabela de fatos;
- Aos fatos, incluindo fatos pré-calculados;
- Os atributos das dimensões;
- Como acompanhar mudanças graduais em dimensões;
- As agregações, dimensões heterogêneas, minidimensões e outras decisões de projeto físico;
- Duração histórica do banco de dados;
- A urgência com que se dá a extração e carga para o DW.

3.8.4 Os dados operacionais

Inmon (1997) afirma "É tentador pensar que a criação do DW consiste em apenas extrair dados operacionais e inseri-los no Data Warehouse. Nada poderia estar mais longe da verdade.". Esta afirmação faz sentido, pois um DW é construído com informações oriundas de várias aplicações operacionais com dados não integrados entre si. Quando estas aplicações já existentes foram concebidas, não foi prevista a possibilidade de uma integração futura. Cada aplicação possui seu conjunto único e particular de requisitos e, durante o processo de desenvolvimento, as demais aplicações não eram levadas em conta. A Figura 15 exemplifica a falta de integração dos dados de diferentes aplicações.

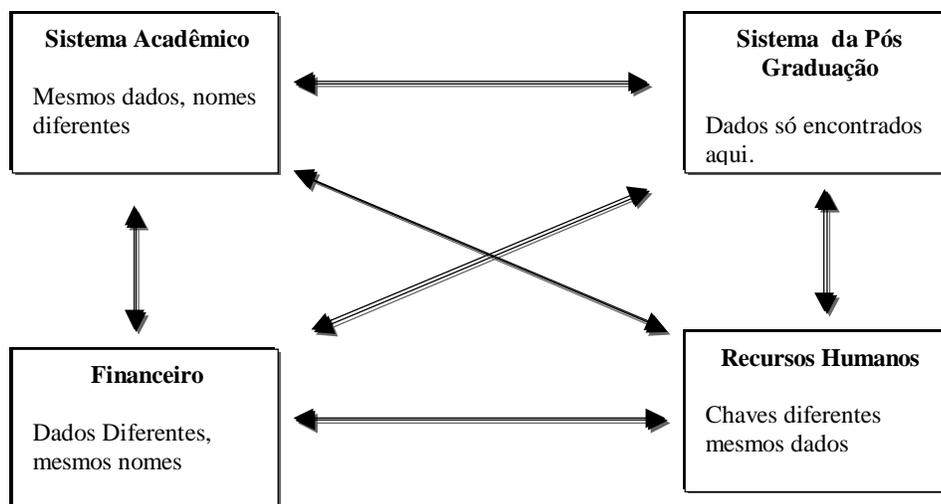


Figura 15 - A falta de Integração dos Dados de Diferentes Aplicações

Outro importante problema da passagem de dados do ambiente operacional para o ambiente do DW diz respeito ao acesso eficiente aos dados dos sistemas existentes. Não é simples definir um método que indique, para os programas que estão varrendo as bases de dados operacionais, se um arquivo já foi ou não varrido anteriormente. Há uma enorme quantidade de dados no ambiente de sistemas existentes, e a tentativa de efetuar varreduras

completas, toda vez que é feita uma varredura para o DW, não é nada econômica e pouco realista.

Para carregar e transformar os dados oriundos de sistemas operacionais é necessário um esforço de processamento muito grande. Por isso estes processos são realizados normalmente à noite, deixando os dados disponíveis para o dia seguinte. Muitos dados do ambiente operacional não passam para o DW, pois o processo de carga efetua uma filtragem neles, impedindo que dados redundantes ou sem valor para análise façam parte do sistema.

Existem três tipos de cargas que podem ser feitas do ambiente operacional para o DW: a carga dos dados históricos, a carga dos valores correntes e o carregamento das alterações (Inmon, 1997).

- *Dados históricos:* O carregamento dos dados históricos do sistema operacional para posterior descarregamento no DW não é uma boa maneira de acessar os dados dos sistemas existentes, pois, além de sobrecarregar os sistemas on-line, também pode acessar dados que já foram carregados para o DW;
- *Dados de valor corrente:* O carregamento de dados de valor corrente no ambiente operacional é feito através da criação de um ou mais arquivos seqüenciais com os valores atuais do sistema operacional. Este arquivo posteriormente será processado e descarregado no DW, o que não ocasionará danos ao ambiente on-line;
- *Alterações:* O carregamento de alterações do DW, com base nas atualizações que tenham ocorrido no ambiente operacional desde a última atualização do DW, consiste no maior desafio ao arquiteto de dados. Não é fácil realizar o rastreamento eficiente e o tratamento das alterações que estão sendo efetuadas sobre o ambiente operacional sem sobrecarregá-lo.

3.9 - Povoando o Data Warehouse

A etapa de povoamento é uma das partes mais críticas na construção de um DW. Esta etapa contempla os processos de *extração, filtragem, transformação e migração dos dados*. Em virtude de não existir um produto capaz de oferecer suporte adequado a todas as fases desta etapa, ela constitui-se uma tarefa crítica. As ferramentas, também conhecidas como "componentes back end", normalmente são especializadas em um dos processos necessários, o que indica que sempre será imprescindível a utilização de dois ou mais componentes "back end" para concluir esta etapa. Existem no mercado algumas ferramentas de extração, filtragem, transformação e migração, as quais têm capacidade de acessar bases dos sistemas operacionais, e algumas são limitadas na sua execução.

3.9.1 Processo de Extração dos dados do ambiente operacional

A primeira tarefa a ser executada pelo componente back end é a extração dos dados operacionais da empresa, geralmente envolvendo fontes de dados autônomas, distribuídas e/ou heterogêneas. Para esta etapa, deve ser considerada uma área de trabalho temporária em que os dados extraídos serão colocados para receberem os tratamentos necessários antes de serem atualizados no DW (área de estagiamento de dados). Quando se trabalha com uma área temporária, a lógica utilizada para "limpar", transformar e integrar os dados é comum a todas as origens, independente do método de acesso utilizada por cada uma delas (Bohn, 1997). Esta área de trabalho temporária poderá estar no mesmo servidor do DW ou não.

Não existe uma regra que defina qual intervalo de tempo deve ser adotado para realizar as extrações. Na maioria das vezes, os dados são extraídos periodicamente como, por exemplo, no final do dia, final do mês ou final de um período fiscal.

As rotinas de extração devem ser capazes de isolar somente aqueles dados que foram inseridos e atualizados desde a última extração. Este processo é conhecido como *refresh*. A melhor política de *refresh* deve ser avaliada pelo administrador do DW, que levará em conta características, como as necessidades dos usuários finais, tráfego na rede e períodos de menor sobrecarga, tanto das origens dos dados quanto do DW. Deve-se considerar que os períodos de sobrecarga podem variar para cada origem de dados.

Existem várias opções de extração dos dados do ambiente operacional para o DW, dependendo das características de cada origem dos dados (Chaudhuri & Dayal, 1997).

- *Origens cooperativas*: origens que suportam gatilhos (triggers), fazendo que notificações de mudanças na base de dados possam ser programadas para ocorrer automaticamente;
- *Origens com log*: origens que mantêm um log o qual pode ser consultado, de forma que mudanças possam ser extraídas deste log (ex. Sybase Replication Server e Oracle Replication Server);
- *Origens consultáveis (time stamps)*: origens que são modeladas para indicar dados novos e modificados através de time stamps. Desta forma, apurações periódicas podem ser realizadas diretamente sobre estas origens, a fim de isolar as alterações que se tem interesse;
- *Origens de instantâneos (snapshots)*: origens que não suportam triggers, logs ou consultas. Neste caso, a solução é realizar periódicos snapshots off-line, onde as mudanças são detectadas, comparando os sucessivos snapshots.

3.9.2 Processo de Filtrar, Transformar e Integrar os dados extraídos

Uma vez que os dados são extraídos e colocados na área de trabalho temporária, estes devem passar por uma série de tratamentos.

O primeiro destes tratamentos refere-se à limpeza ou filtragem dos dados. O objetivo é garantir a integridade dos dados através de programas ou rotinas especiais que tentam identificar anomalias e resolvê-las, deixando os

dados em um estado consistente antes de serem instalados no DW. A correção de erros de digitação, a descoberta de violações de integridade, a substituição de caracteres desconhecidos, a padronização de abreviações, podem ser exemplos de limpeza de dados. Algumas ferramentas no mercado conseguem identificar inconsistências e resolvê-las de forma automática.

O segundo passo é colocar os dados em uma forma homogênea, aplicando uma metodologia de comparação de representações, que inclua os critérios a serem utilizados na identificação de semelhanças e conflitos de modelagem (Batini & Lenzerini, 1986). Conflitos de modelagem podem ser divididos em (Ribeiro, 1995): semânticos e estruturais. Conflitos semânticos são todos aqueles envolvendo o nome ou palavra associado às estruturas de modelagem. Por exemplo, mesmo nome para diferentes entidades ou diferentes nomes para a mesma entidade. Conflitos estruturais englobam os conflitos relativos às estruturas de modelagem escolhidas, tanto no nível de estrutura propriamente dito como no nível de domínios. O principal tipo de conflito estrutural são os conflitos de domínio de atributo que se caracterizam pelo uso de diferentes tipos de dados para os mesmos campos.

Após identificados os conflitos de modelagem, deve-se criar as regras de mapeamento de representações equivalentes e de conversão para os padrões estabelecidos pelo DW. Para criar estas rotinas, existem atualmente no mercado uma variedade de ferramentas especializadas, conhecidas como (*data migration*). Estas ferramentas permitem que sejam especificadas regras de transformações simples como, por exemplo, converter o código F e M para "masculino" e "feminino" antes que o registro seja atualizado no DW.

3.10 Recuperação de Informações do Data Warehouse

No processo de recuperação de informações do DW, é importante que a *aplicação para o usuário final* kimball et. al. (1998) abordado no item 3.1.3, contemple ferramentas eficientes que permitam realizar consultas e

análises no DW desde as mais simples às mais complexas. Depois de definido e projetado o escopo do projeto e depois de construído o repositório de dados, é que se deve chegar às ferramentas de front-end responsáveis pelo meio de campo entre as bases de dados e os usuários finais da área executiva. Elas não podem ser muito complexas porque não serão utilizadas por profissionais da área técnica, mas precisam ser robustas o suficiente para dar agilidade no acesso às informações estratégicas.

Existem várias maneiras de recuperar informações de um DW. As formas de recuperação mais comuns no mercado hoje são os relatórios, as consultas, os EIS, ferramentas que utilizam características OLAP e as ferramentas de Data Mining. A nova tendência dessas soluções é a integração com o ambiente Web, permitindo maior agilidade em consultas estáticas e dinâmicas (Moraes, 1999). Um exemplo é a ferramenta WEBDB da ORACLE, ideal para aplicações desta natureza.

3.11 Passos para a implantação do Data Warehouse

Um DW não é um produto que se compra, mas sim um projeto que envolve análise e implementação, com a participação de várias tecnologias. Sendo assim, a observância de alguns passos é de suma importância.

Os passos a seguir foram retirados do artigo "How to Build an Effective Data Warehouse", do suplemento especial "New Insights In Data Warehousing Solutions", da HP-Information Week HP, out/96, onde o autor sugere sete etapas para a criação de um DW, que pode ser inicialmente um data mart (DW para um assunto específico, departamental) até mesmo um DW no nível corporativo.

- Passo 1 - Os primeiros resultados devem estar disponíveis a curto prazo. É importante traduzir rapidamente as necessidades do negócio em uma especificação que possa ser construída em etapas. Essa abordagem minimiza riscos e o tempo de apresentação dos resultados iniciais.

- Passo 2 - Mesmo para especialistas, construir um DW pode ser um desafio de integração de sistemas. Os dados de produção e de fontes externas precisam ser mapeados para o modelo de dados do DW. Precisa haver um sincronismo entre os dados operacionais e os dados de tomada de decisão. Para análises multidimensionais, os dados precisam ser transferidos e sincronizados em um banco de dados multidimensional.
- Passo 3 - A escolha do banco de dados de suporte ao DW precisa ser criteriosa. Alguns critérios que devem ser analisados são: desempenho na carga e indexação dos dados, tempo de resposta, capacidade de armazenamento, paralelismo e escalabilidade.
- Passo 4 - Deve-se considerar as ferramentas disponíveis no mercado que normalmente ajudam a compor um ambiente DW com a finalidade de prover: interfaces amigáveis, geração de relatórios, análises multidimensionais, acesso via Web e data mining.
- Passo 5 – Deve-se construir um DW que possa ser expandido, mantendo níveis aceitáveis de desempenho até centenas de gigabytes.
- Passo 6 - O ambiente DW deve ser aberto para permitir que os componentes ou ferramentas identificadas no passo 4 possam ser substituídas por outras mais atuais e eficientes.
- Passo 7 – Deve-se considerar o sistema de armazenamento que fisicamente gerencia o tráfego, alocação, backup e restauração dos dados.

3.12 Considerações Finais

Este capítulo apresentou uma visão geral da tecnologia de Data Warehouse. Surgida no começo do anos 90, a tecnologia de DW tem por objetivo organizar os dados operacionais, oriundos de sistemas legados de uma organização, visando atender as necessidades de suporte à decisão da alta administração.

Os principais elementos do DW incluem desde os sistemas de origem (sistemas legados ou operacionais) até as aplicações sobre o ambiente construído. A arquitetura de construção do ambiente DW surge ou da

abordagem *top-down*, onde o DW é construído inteiramente e Data Marts são extraídos do mesmo, ou da abordagem *bottom-up*, em que o ambiente DW é resultado da integração de Data Marts orientados por assunto.

O estudo mostrou a maior tendência de aplicação da abordagem *bottom-up*. No caso das instituições de ensino superior, em especial, a arquitetura que parte dos Data Marts é mais adequada tanto do ponto de vista econômico como político-cultural na organização. Isto deve-se à natureza multidisciplinar e distribuída dos sistemas legados.

O capítulo também abordou a modelagem multidimensional com que são formalizadas as informações no DW e a organização dos metadados, salientando sua importância no entendimento do ambiente DW. Finalmente, abordam-se as formas de construção do DW e os profissionais envolvidos neste processo.

A questão central neste trabalho é como levar a tecnologia de DW para o ambiente de decisões gerenciais da universidade. No próximo capítulo, o levantamento apresentado é a base para o processo de construção de um DW em uma organização universitária.

4 MODELO DE IMPLANTAÇÃO DE DW EM UMA INSTITUIÇÃO DE ENSINO SUPERIOR: CASO DA UNOESC - CHAPECÓ

O tema Data Mart para a Gestão Universitária é tão recente que muitas IES o desconhecem no âmbito administrativo. Todavia, a revolução que a informática vem operando nas organizações em geral abre novas perspectivas também à gestão universitária, que pode contar com esse poderoso instrumento para o aumento da sua produtividade, por meio de agilidade e eficiência no processo decisório.

Como visto no capítulo 3, *Data mart* é uma abordagem mais "focada" na tecnologia de DW em que os dados estão mais próximos aos usuários. Ao invés de se projetar um armazém de dados centralizado, capaz de atender todas as áreas de decisão da instituição, utiliza-se uma abordagem departamental, preservando-se a possibilidade de se integrar posteriormente diversos *data marts*, formando um DW completo.

O objetivo deste capítulo é apresentar um modelo para a implantação de um Data Mart que possa contribuir para o melhor desempenho dos gestores universitários na Universidade do Oeste de Santa Catarina – UNOESC – Chapecó, em suas atividades de planejamento e tomada de decisões.

O foco do presente trabalho, para efeito de aplicabilidade, centrar-se-á no aspecto de *gestão acadêmica*, tendo como premissa que esta área é a responsável por conduzir as ações para resultados positivos ante os compromissos fundamentais da UNOESC, principalmente, o de ensino e aprendizado.

Teve-se em mente aproveitar ao máximo a estrutura e recursos já disponíveis, não apenas por questão de economicidade, mas também para melhorar estrategicamente as possibilidades de sua efetiva implantação.

4.1 Metodologia de Implantação

O desenvolvimento do ambiente DW na UNOESC - Chapecó tem por objetivo rever o processo de construção de aplicativos de apoio à tomada de decisão em instituições universitárias. Para tal, o DW desenvolvido seguiu as seguintes etapas:

4.1.1 Análise da Estrutura Organizacional da Universidade

Aborda-se o histórico, a abrangência, a estrutura administrativa e o organograma da IES, com o intuito de identificar os diferentes níveis decisórios da organização e o estágio atual de informações e subsídios à gestão acadêmica.

4.1.2 Levantamento dos Sistemas Legados

Após conhecer e documentar os níveis organizacionais da universidade, foi realizado levantamento dos principais sistemas legados que atendem a instituição. Para cada área da universidade, registra-se a área de competência, sistemas utilizados, módulos disponíveis, nível de decisão e principais restrições atuais destes sistemas. Nesta etapa, também se registram o fluxo de informações entre os níveis organizacionais e o mapeamento dos relacionamentos entre os vários sistemas existentes. Após a análise global dos sistemas legados, aprofunda-se o estudo do(s) sistema(s) operacional(is) que gera(m) os dados que servirão de fontes para o Data Mart a ser desenvolvido. Outra análise importante nesta etapa é o nível de utilização atual por parte da alta administração, o que subsidiará as definições do ambiente DW.

4.1.3 Proposição do Modelo

Com o levantamento da estruturada universidade, dos sistemas legados e análise da área em que se vai desenvolver o Data Mart, propõe-se o modelo de construção do ambiente DW da universidade. Seguindo a orientação dos estudos do capítulo 3, propõe-se a construção segundo a arquitetura *bottom-up*. Com isto, o ambiente DW será formado a partir de Data Marts incrementais.

O modelo do primeiro Data Mart deve prever a seleção de variáveis estratégicas às decisões da organização, tanto no nível interno como externo.

4.2 Estrutura Organizacional da UNOESC

4.2.1 Histórico e Abrangência

A UNOESC é uma Universidade multicampi, fundada em 1992, a partir da fusão de três Fundações Educacionais que atuavam no ensino superior do meio oeste ao extremo oeste catarinense, desde o início da década de 70.

A Fundação Educacional e Empresarial do Alto Vale do Rio do Peixe - FEMARP, com sede na cidade de Videira; a Fundação Universitária do Oeste Catarinense - FUOC, sediada na cidade de Joaçaba; a Fundação de Ensino do Desenvolvimento do Oeste - FUNDESTE, com sede na cidade de Chapecó e extensões nas cidades de São Miguel d'Oeste e Xanxerê. A unificação dos três Centros de Ensino Superior, mantidos respectivamente por essas três instituições isoladas, deu origem, em 1991, ao patrimônio científico-cultural e à estrutura organizacional e administrativa da nova universidade catarinense, autorizada no mesmo ano pelo Conselho Federal de Educação, reconhecida em 1995 pelo Conselho Estadual de Educação de Santa Catarina

e credenciada pelo decreto presidencial de 14 de agosto de 1996 (Proens, 1999).

Possui sede jurídica na cidade de Joaçaba, onde se encontra a Fundação UNOESC, mantenedora da Universidade e sede administrativa (Reitoria) em Chapecó.

A Fundação UNOESC, estatutariamente, é a entidade mantenedora da Universidade do Oeste de Santa Catarina, constituída pela FUOC, de Joaçaba, FEMARP, de Videira, e FUNDESTE, de Chapecó.

A Fundação UNOESC é administrada pela Assembléia Geral, que é o órgão superior, soberano e de deliberação, composta por representante de todos os segmentos das Fundações que a instituíram; pelo Conselho Curador, a quem compete o acompanhamento e a fiscalização econômica - financeira; e pela Presidência. Compete ao Presidente zelar para que a Instituição alcance sua finalidade, bem como presidir a Assembléia Geral.

Dos Campi da Universidade nascem políticas e propostas de ação que, debatidas no Conselho Universitário - CONSUN, são traduzidas em diretrizes e normas que dão a direção do caminhar da Universidade.

A UNOESC é uma *instituição comunitária* com abrangência em todo o Oeste de Santa Catarina, numa dimensão geográfica de 27.230 Km², na qual reside atualmente uma população de cerca de um milhão de habitantes. Alcança um território correspondente a 1/3 do Estado de Santa Catarina. Atende um número de aproximadamente 14.000 alunos de graduação, 1.500 alunos de pós-graduação e inúmeras entidades da região, através de seus programas de extensão.

4.2.2 Estrutura Administrativa

Criada por Lei Municipal, reconhecida pelo CEE e credenciada pelo MEC, a UNOESC é uma universidade de natureza jurídica privada, mas de caráter comunitário e filantrópico, ou seja, segundo citado no capítulo 2 (Finger, 1997), classifica-se como uma universidade cuja sobrevivência depende em

grande parte de recursos de fonte própria e possui uma mantenedora que é a Fundação UNOESC, com sede jurídica na cidade de Joaçaba.

4.2.3 Estrutura Organizacional da Administração Central

Conforme abordado no capítulo 2, a UNOESC assemelha-se ao modelo baseado em áreas afins, ou seja, utiliza a estrutura de centros com seus respectivos cursos. Esse modelo foi implantando recentemente e pode-se dizer que se encontra em processo de amadurecimento gradativo.

- *Reitoria*: Reitor e Chefia de Gabinete
- *Conselho Universitário - CONSUN*: é o órgão de instância superior em matéria administrativa, financeira, acadêmica e disciplinar, de natureza normativa, consultiva e deliberativa da Universidade.

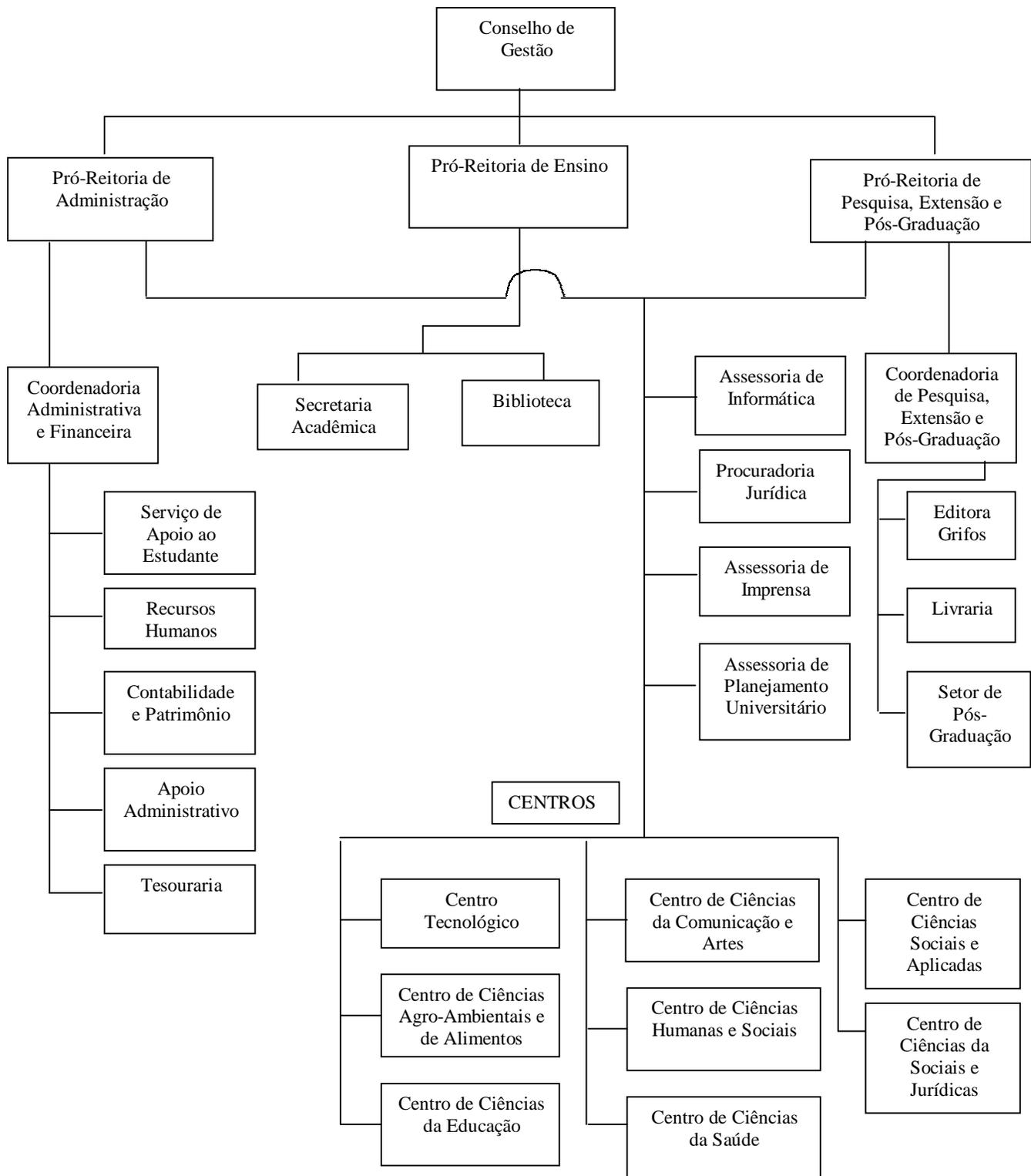
4.2.4 Estrutura Organizacional dos Campi

- *Conselho de gestão*: é o órgão colegiado de natureza consultiva, normativa e deliberativa do campus, em matéria acadêmica, administrativa, financeira e disciplinar.
- *Pró-Reitorias*: as Pró-Reitorias são órgãos executivos incumbidos de administrar, planejar, coordenar e avaliar as atividades dos campi.
- *A Pró-Reitoria de Administração do campus*: é órgão executivo, incumbido de administrar, planejar, coordenar, descentralizar, delegar competências e controlar as atividades administrativas, financeiras, patrimoniais, comerciais e de suporte do campus, promovendo sua compatibilização com a Administração Central da Universidade.
- *A Pró-Reitoria de Ensino*: é órgão executivo incumbido da gestão das atividades fim de ensino do campus, em consonância com as políticas e diretrizes da Universidade.

- *A Pró-Reitoria de Pesquisa Extensão e Pós-Graduação*: é órgão executivo, incumbido da gestão das atividades fins de pesquisa, de pós-graduação e de extensão, em consonância com as políticas, objetivos e diretrizes da Universidade.

- *Conselho de centro*: é instância deliberativa que gera e acompanha as políticas de ensino, de pesquisa, de extensão, de pós-graduação e de administração afetas à sua área de atuação, de acordo com os objetivos da Universidade.

4.3 Estrutura Organizacional da UNOESC - Campus Chapecó



4.3.1 Estudo da Estrutura Organizacional da UNOESC - Campus de Chapecó

a) Coordenadoria Administrativa e Financeira

É função executiva, de confiança da Pró-reitoria de Administração, tem como atribuições:

- a) coordenar as atividades desenvolvidas pelos setores subordinados à Pró-Reitoria de Administração (Serviço de Apoio ao Estudante - SAE, Recursos Humanos - RH, Contabilidade/Patrimônio, Apoio Administrativo e Tesouraria) articulando as ações dos mesmos);
- b) ser o preposto do Pró-Reitor de Administração junto à justiça do Trabalho, justiça comum, nas ações trabalhistas e de cobrança movidas pela UNOESC - Campus de Chapecó contra alunos inadimplentes e tem competência para movimentar contas bancárias em conjunto com o Pró-Reitor de Administração.

b) Serviço de Apoio ao Estudante (SAE):

O serviço de Apoio ao Estudante foi criado com o objetivo de atender aos alunos economicamente carentes de recursos financeiros, oportunizando e possibilitando o acesso à continuidade dos estudos, sobretudo nos cursos de graduação da UNOESC. Para tal, conta com os programas:

- a) Programa de bolsas de estudo federal - Crédito Educativo (CREDUC)
- b) Programa de bolsa-trabalho (CIEE)
- c) Implantação do banco de recursos humanos
- d) Implantação de programa de aulas particulares
- e) Programa de atendimento psicológico
 - Sistemas de Informação utilizados: Controle Acadêmico e Tesouraria
 - Módulos: Cadastro e Consultas
 - Dados de Identificação do acadêmico - Nome, endereço, curso vinculado, disciplinas matriculadas, disciplinas cursadas, disciplinas reprovadas.

- Situação financeira - Valores em aberto, valores pagos, valores de bolsas de auxílio recebido.
- Restrições atuais: Apenas são manipulados os dados relativos a consultas sobre cadastros. Observou-se que o sistema não manipula dados históricos.

c) Recursos Humanos

O setor de Recursos Humanos realiza o processo de admissão, folha de pagamento, controle de ponto, rescisão, recolhimento de impostos e atualização de carteiras de trabalho.

- Sistema de Informação utilizado: Vetorh
- Módulos: Cadastro, Tabelas, Cálculos, Impostos, Consultas, Relatórios, Importação e Exportação.

O sistema mantém informações cadastrais relacionadas ao colaborador como, por exemplo, nome, endereço, data de admissão, último salário, carga horária contratual, regime de trabalho, faixa de IRPF e demais informações usuais em um sistema de recursos humanos.

- Nível de Decisão: Operacional
- Restrições atuais: A obtenção de informações gerenciais requer conhecimentos adicionais em nível de programação para obter resultados, sendo que em muitas vezes não é possível obter devido às limitações do software como o não armazenamento de informações ao longo do tempo, ou seja, não armazena séries históricas.

d) Contabilidade e Patrimônio

A contabilidade realiza a contabilização das informações financeiras, e o patrimônio controla as entradas e saídas e localização dos bens patrimoniais.

- Sistema de Informação utilizado:
- *Hércules - Contábil:*

- Módulos: Cadastros, Lançamentos, Baixas, Tabelas, Consultas, Relatórios, Importação e Exportação.

Através desses módulos realiza o registro contábil, lançamentos e baixas. O possui incorporada uma série de tabelas bases, e possibilita a emissão de vários relatórios, bem como realiza a importação/exportação de dados com outros sistemas.

- Nível de Decisão: Operacional e Gerencial
- Restrições atuais: Com relação ao Hércules, este apresenta relatórios complexos que dificulta o entendimento do usuário, conseqüentemente a tomada de decisão. Utilização de planilhas de cálculos para compilar os dados oriundos dos relatórios com objetivo de tornar mais claros os dados em nível gerencial.

- *ADTC - Patrimonial*

- Módulos: Cadastros, Lançamentos, Baixas, Tabelas, Consultas, Relatórios, Importação e Exportação

ADTC – Patrimonial contempla todos os módulos pertinentes a um sistema de controle patrimonial, e possibilita a emissão de vários relatórios, bem como realiza a importação/exportação de dados com outros sistemas.

- Nível de Decisão: Operacional
- Restrições atuais: Com o ADTC, não foi possível levantar as suas restrições, uma vez que ele encontra-se em fase de implantação.

e) Apoio Administrativo

O serviço de Apoio Administrativo tem por finalidade dar suporte às atividades meio e fim da Universidade. Realiza o processo de orçamentos e compras, bem como coordena os serviços de manutenção do campus.

- Sistema de Informação utilizado: Contas a Pagar e Compras
- Módulos: Cadastros, Lançamentos, Baixas, Consultas, Relatórios, Importação e Exportação.

- Nível de Decisão: Operacional
- Restrições atuais: Os referidos sistemas apresentam o módulo de consultas comprometido, em face da pouca flexibilidade na obtenção de informações, devido às limitações da linguagem em que foi implementado, necessitando de constante interferência da área de informática.

f) Tesouraria

A tesouraria é responsável pelo controle financeiro das operações econômicas da instituição, pagamentos, recebimentos de contas a pagar, a receber e da conta corrente bancária.

- Sistema de Informação utilizado: Tesouraria e Contas a Pagar
- Módulos: Cadastros, Lançamentos, Baixas, Consultas, Relatórios, Importação e Exportação.
- Nível de Decisão: Operacional
- Restrições atuais: Os sistemas foram desenvolvidos há uma longa data e não são práticos na obtenção de informações, devido a linguagem não ser flexível. Observou-se que os dirigentes têm uma carência muito grande de informações no processo decisório.

g) Secretaria Acadêmica (SERCA)

Integrante da estrutura do Campus. A SERCA é o órgão de apoio administrativo, cuja função é centralizar a administração da vida acadêmica nos processos de admissão, matrícula, integralização curricular e avaliação da aprendizagem, efetuando o controle, registro e a certificação de todos os atos acadêmicos pertinentes à graduação, pós-graduação e extensão.

- Sistema de Informação utilizado: Controle Acadêmico.
- Módulos: Cadastros, Consultas, Relatórios, Horários e Matrícula.
- Nível de Decisão: Operacional.
- Restrições atuais:

- ✓ Sistema Acadêmico: Carência muito grande com relação a informações históricas, o que dificulta a tomada de decisões. Necessidade constante de novos programas para atender a demanda de dados estatísticos. Há pouca flexibilidade na adaptação de novos procedimentos devido à arquitetura do mesmo. Falta de informações gerenciais, como cursos com maior evasão, cursos deficitários, cursos ociosos, análise da viabilidade de implantação de novos cursos.
- ✓ Pós-Graduação: A falta de informações históricas e sumarizadas, muitas vezes, impossibilita os processos acadêmicos de gestão em nível de pós-graduação.

h) Biblioteca

A biblioteca é responsável pelo controle do acervo bibliográfico. Realiza o controle de empréstimos e disponibiliza consultas por autor, título e assunto em terminais de auto-atendimento.

- Sistema de Informação utilizado: Pergamum
- Módulos: Catalogação, Consulta, Empréstimo e Compras.
- Nível de Decisão: Operacional
- Restrições Atuais:
 - Pergamum: O módulo de consulta é lento devido ao volume de dados na base. Informações históricas são de difícil apuração, dependência forte do proprietário do software. Esse software encontra-se em fase de implantação. Até o momento não foi possível detectar, na íntegra, suas restrições.
 - Microlsis: Consultas limitadas, interface voltada a ambiente caractere, falta módulo de empréstimo e compras.

i) Assessoria de Informática

Compete à Assessoria de Informática:

- a) prestar assessoria às Pró-Reitorias, Centros, Coordenações e Setores em atividades de informática do Campus;
- b) efetuar a manutenção dos sistemas existentes e o desenvolvimento de novos programas;
- c) prestar apoio aos usuários em "software" e "hardware"
- d) fazer planejamento de redes e redimensionamento dos equipamentos.

A Assessoria de informática presta suporte em questão de software e hardware às Pró-Reitorias, aos Centros, setores e coordenações integrantes da estrutura organizacional da UNOESC.

j) Assessoria de Planejamento Universitário

Tem por finalidade assessorar as atividades de planejamento acadêmico do Campus na área de ensino, auxiliando as Pró-Reitorias, os Departamentos, Coordenações de Curso, Secretaria Acadêmica, Centro Educacional do Ensino Fundamental e Médio, nas matérias referentes a projeto de criação de cursos, processos de reconhecimento de cursos, alteração de grades curriculares, normatização de estágios curriculares, orientações sobre procedimentos acadêmicos de acordo com a legislação, acompanhamento de tramitação dos processos e a manutenção e atualização de arquivos sobre a legislação da área educacional

- Sistema de Informação utilizado: Credenciamento
- Módulos: Cadastros Consultas e Relatórios
- Nível de Decisão: Operacional
- Restrições Atuais: As Consultas são limitadas, devido a linguagem não oferecer muita flexibilidade.

e) Setor de Pós-Graduação

Sua principal função é auxiliar aos coordenadores dos cursos de pós-graduação e extensão, desde a digitação dos projetos, sua execução e emissão dos certificados, bem como o atendimento da demanda que necessita de informações relativas aos cursos e eventos do Campus.

- Sistema de Informação utilizado: Pós-Graduação e SEP (Controle Extensão e Planejamento).
- Módulos: Cadastros, Consultas, Relatórios, Horários, Matrícula e Certificados.
- Nível de Decisão: Operacional.
- Restrições atuais:
 - Pós-Graduação: A falta de informações históricas e sumarizadas, muitas vezes, impossibilita a tomada de decisões. Falta desenvolver os módulos de:
 - aproveitamento de disciplinas;
 - controle e acompanhamento de monografias;
 - módulo financeiro.
 - SEP (Controle da Extensão e Pós-Graduação): Sistema desenvolvido há vários anos. Requer uma nova modelagem; não apresenta flexibilidade na obtenção de informações.

4.4 Fluxo de informações na UNOESC - Campus Chapecó para a tomada de decisão na Gestão Acadêmica

Funcionalmente, a Universidade do Oeste de Santa Catarina - Campus de Chapecó, pode ser dividida em dois grandes setores: acadêmico e administrativo.

Para efeito do presente trabalho, que está voltado à tomada de decisão, é objeto de consideração apenas o **setor acadêmico**, responsável

primeiro pelas atividades-fim da instituição, a saber: ensino, pesquisa, extensão e pós-graduação.

A instância máxima deliberativa, normativa e consultiva do setor acadêmico é o conselho de gestão colegiado superior, presidido pelo Pró-Reitor de Administração, integrado pelo Pró-Reitor de Ensino, Pró-Reitor de Pesquisa, Extensão e pós-graduação, diretores de centros, um representante dos coordenadores de curso de cada centro, um representante docente de cada centro, dois representantes técnico-administrativos e três representantes da comunidade externa.

As Pró-Reitorias de Administração, Ensino, Pesquisa e Extensão e Pós-Graduação são órgãos superiores do Campus, compostas por três pró-reitores. Cabe a eles a direção superior do Campus.

O Conselho de Gestão e o Conselho de Centro integram a Administração Superior do Campus e são disciplinados por estatuto.

A Pró-Reitoria de Ensino, instância máxima executiva do setor acadêmico, é ocupada por um membro do corpo docente, com titulação mínima de especialista e tem por atribuição liderar o processo de gestão acadêmica.

Cotidianamente, o pró-reitor reúne-se com os coordenadores de curso e diretores de centro, individual e coletivamente, e com eles planeja as atividades futuras da instituição, estabelece metas, avalia o desempenho e decide questões específicas em nível superior.

Em suma, pode-se considerar o pró-reitor acadêmico, pró-reitor de administração e os coordenadores de curso como os principais gestores acadêmicos da UNOESC e para efeito do presente trabalho, como os principais usuários potenciais do Data Mart proposto.

4.5 Sistemas Legados da UNOESC

No levantamento das bases de dados e sistemas de informação existentes, ficou claro que a UNOESC, hoje, possui várias "*ilhas de informação*". Conforme exemplifica a Figura 16, há incompatibilidade de linguagens entre os sistemas e especialmente, o isolamento de dados entre os

setores, causando, com isso, redundância de dados. Também constatou-se que esse quadro está mudando com a implantação de uma nova base de dados única. Na Tabela 6, é apresentado um breve diagnóstico dos sistemas existentes.

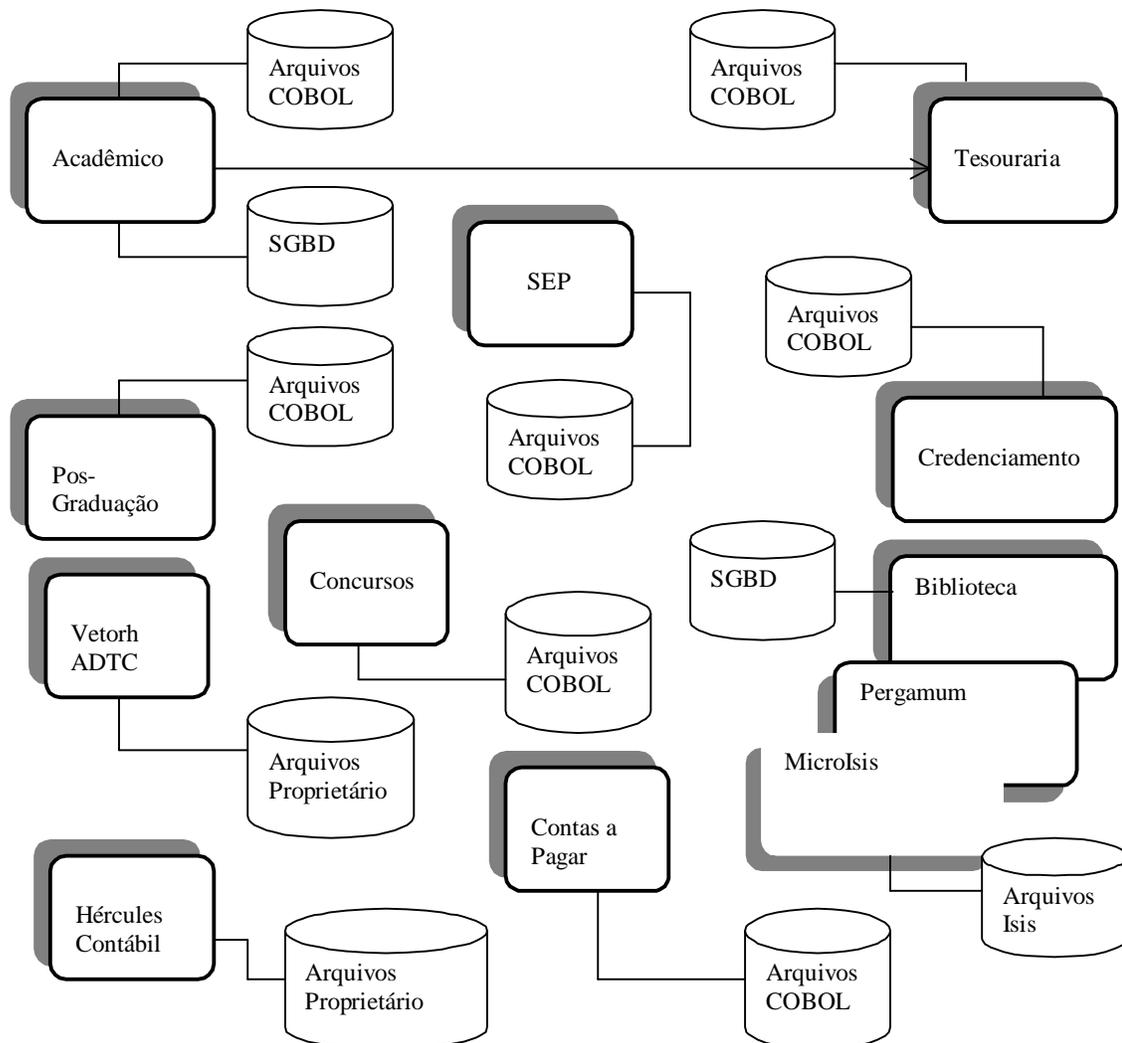


Figura 16 - Sistemas de Informação Existentes

Tabela 6 - Diagnóstico dos Sistemas Existentes

Nome Do Sistema	Plataforma	Base de Dados	Linguagem	Descrição	Localização na Universidade
Acadêmico	Netware 5.0	Arquivos Comuns	COBOL	Realiza o Controle dos Acadêmicos, desde o ingresso na instituição até a sua colação de grau, controle de horário, matrículas e históricos	Assessoria de Informática
	NT Server 4.0	SGBD Oracle	Forms, Reports e WebDb da Oracle		
Pós-Graduação	Netware 5.0	Arquivos Comuns	COBOL	Realiza o Controle dos Alunos da Pós-Graduação	Assessoria de Informática
Tesouraria	Netware 5.0	Arquivos Comuns	COBOL	Realiza o controle dos recebimentos de mensalidades e taxas. Possui uma forte integração com o Sistema de Controle Acadêmico	Assessoria de Informática
Credencia-mento	Netware 5.0	Arquivos Comuns	COBOL	Tem por finalidade registrar os professores que estão credenciados e/ou autorizados a ministrar aulas, possui uma integração direta com o Sistema de Controle Acadêmico	Assessoria de Informática
Pergamum	Windows NT Server 4.0	SGBD Sybase	Delphi	Mantém registro do acervo bibliográfico, controle de empréstimo e consultas. Este sistema realiza o controle do acervo e disponibiliza consultas à comunidade acadêmica, roda em paralelo com o sistema Pergamum	Assessoria de Informática
Microisis	Netware 5.0	Arquivos Proprietários	ISIS		
SEP Setor de Extensão e Planejamento	Netware 5.0	Arquivos Comuns	COBOL	Realiza o Registro de Controle dos Cursos de Extensão e Pós-Graduação, bem como a emissão de certificados	Assessoria de Informática
Controle de Concursos	Netware 5.0	Arquivos Comuns	COBOL	Tem por objetivo controlar a realização de concursos desde a inscrição até a seleção dos aprovados	Assessoria de Informática
ADTC	Netware 5.0	Arquivos Comuns	C++	Realiza o controle do patrimônio	Assessoria de Informática
Vetorh	Netware 5.0	Arquivos Comuns	C++	Registro de Empregado, Folha de Pagamento, Cartão Ponto e Rescisão.	Assessoria de Informática
Contas a Pagar	Netware 5.0	Arquivos Comuns	COBOL	Controle de compras e obrigações a pagar	Assessoria de Informática
Hercules - Contábil	Netware 5.0	Arquivos Comuns	C++	Controle Contábil e Gerencial da Universidade	Assessoria de Informática

Com relação à localização na Universidade, todos os sistemas trabalham com a arquitetura de dados centralizada no servidor, onde os dados são migrados através da rede por demanda do sistema cliente e não ficam residentes na máquina do usuário.

4.6 Análise do Sistema Acadêmico Existente

O sistema informatizado de gerenciamento dos registros acadêmicos, atualmente em uso na UNOESC, denomina-se *Sistema Acadêmico* e foi desenvolvido pela área interna de informática.

Trata-se de um sistema concebido sobre arquivos convencionais, utilizando linguagem COBOL, apresentado em um ambiente do tipo DOS, ou seja, com mínimos recursos gráficos e bastante frágil no que se refere à integridade de dados. Atualmente encontra-se rodando em rede sobre a plataforma NOVELL. Paralelamente, encontra-se em processo de desenvolvimento um novo sistema acadêmico com enfoque OLTP. Parte do sistema já está rodando em paralelo com o sistema atual, sendo que este sistema foi concebido através do uso de ferramentas CASE e utiliza um SGBD relacional da ORACLE.

Desde a sua implantação na antiga FUNDESTE, esse sistema vem recebendo ampliações e adaptações para se adequar à realidade da Instituição.

O sistema é composto de um módulo denominado acadêmico, que subdivide-se em: controle operacional, matrícula e consultas. Este módulo é composto pelos seguintes cadastros, aqui denominados variáveis: instituição, cursos, professores, disciplinas, turmas e alunos.

Todas as informações dessas bases de dados ficam armazenadas em uma única máquina denominada de servidor, cujo uso é compartilhado de acordo com os níveis de prioridade e habilitações de acesso entre seus usuários, que podem realizar consultas e/ou alterações, dependendo de como foram cadastrados.

A instituição possui em torno de 260 microcomputadores com diversas configurações, todos estão ligados em rede com acesso à Internet, acessando servidores internos específicos. Recentemente, foi implantado um novo servidor de uso exclusivo da biblioteca para o gerenciamento da movimentação do acervo, independentemente do servidor principal, o que possibilitou mais segurança ao setor e menor congestionamento na rede em geral.

O tráfego de informações se dá através de uma rede local que utiliza protocolos de comunicação TCP/IP, NETBEEU e IP/SPX, com acesso a servidores Novell, NT, Unix e Linux. Sua distribuição física é feita através de cabos de Fibra Óptica para interligação dos blocos e cabos par trançado categoria 5 para ligação internas dos micros com os *hubs* e *switchs*.

O principal setor usuário do *Sistema Acadêmico* é a secretaria acadêmica, que o utiliza para as rotinas de secretaria, tais como: digitação de notas, frequência, emissão diário de classe, cálculos de médias e outros.

Os gestores acadêmicos propriamente ditos, a saber, os coordenadores de curso, pró-reitor de administração e a pró-reitoria de ensino, não costumam acessar essas bases de dados, por duas razões básicas: porque as informações não têm uma apresentação amigável (ambiente não gráfico) e, principalmente, porque as informações disponíveis atendem apenas a parte de suas necessidades e vêm acompanhadas de muitas outras informações. Para eles, a princípio, sem utilidade, ou seja, as informações encontram-se muito detalhadas e não são apresentadas de forma sumarizadas, a fim de auxiliar no processo de tomada de decisão.

4.7 Modelo Proposto

Como descrito no capítulo 3, o ambiente DW, para uma instituição universitária, deve ser construído segundo a arquitetura *bottom-up*. Assim,

propõe-se, a seguir, o modelo do Data Mart inicial da universidade, tomando-se a área acadêmica como ponto de partida, dado o caráter estratégico de seus resultados para a instituição.

O modelo proposto de um Data Mart é parte de um visão estratégica mais ampla, que tem por objetivo a melhoria no processo de ensino-aprendizado e aumento de sua competitividade perante o mercado.

Como foi apresentado no capítulo 3 o projeto de um Data Mart vem de encontro à proposta, já que o mesmo se propõe a trabalhar com áreas específicas da UNOESC. Partiu-se do princípio que, ao invés de se projetar um armazém de dados centralizado, capaz de atender todas as áreas da instituição, será adotado uma abordagem departamental, do tipo **Bottom-Up**, tendo como premissa a possibilidade de integração posterior dos diversos DM em um DW completo.

A utilização do modelo Bottom-up possibilita a apresentação de resultados aos gestores acadêmicos em um curto espaço de tempo, iniciando-se com a construção gradativa de data marts por áreas elencadas como prioritárias.

Para tanto, são consideradas as variáveis mais significativas para a leitura dos ambientes interno e externo à instituição, tendo como foco a gestão acadêmica, a saber, as necessidades e expectativas de sua clientela, as reais condições de sua capacidade produtiva, bem como as restrições e oportunidades de mercado.

As informações que alimentarão o Data Mart, em parte, já estão presentes no novo banco de dados da instituição. As demais deverão ser obtidas a partir de uma ampliação dos cadastros já existentes.

A Tabela 7 indica as variáveis internas prováveis que podem vir a servir como base ao modelo proposto, bem como o setor encarregado da sua coleta, organização e alimentação para o Data Mart. A Tabela 8 sinaliza para as variáveis externas que, ao longo do tempo virão a enriquecer o Data Mart.

Tabela 7 - Variáveis Internas

VARIÁVEIS	ITENS	FONTE
ALUNOS	Dados pessoais	Acadêmico
	Disciplinas Cursadas	
	Frequências	
	Escola de origem	
	Situação Financeira na Instituição	Tesouraria
PROFESSORES	Dados Pessoais	Recursos Humanos
	Disciplinas Credenciadas	Assessoria de Planejamento
	Titulação	Recursos Humanos
	Regime de Trabalho (horista, dedicação exclusiva, dedicação parcial)	
	Controle de Ponto (faltas, justificativas, reposições)	
	Ocorrências especiais (processos em trâmite, sanções disciplinares, etc)	Pós Graduação
	Qualificação (Cursos que frequenta, previsão conclusão)	
	Produção científica (trabalhos publicados, participação em eventos científicos)	Coordenação de Curso
Instituições em que ministra aula		
DISCIPLINAS	Objetivo Geral	Secretaria Acadêmica
	Objetivos específicos	
	Ementa	
	Conteúdo	
	Carga Horária	
	Programa/cronograma	
	Metodologia de ensino	
	Metodologia de avaliação	
CURSO	Grades curriculares	Secretaria Acadêmica
	Candidatos/vagas (séries históricas)	
	Turmas/alunos	
	Estatísticas de desistências; trancamentos de matrícula	Biblioteca
	Biblioteca (acervo em geral, acervo específico, estatística de utilização).	Apoio Administrativo
	Instalações laboratoriais (equipamentos, instalações e materiais; estatísticas de utilização)	

Tabela 8 - Variáveis Externas

VARIÁVEIS	ITENS	FONTE
MERCADO EDUCACIONAL - Oferta (séries históricas relativas aos principais concorrentes)	Dados da Instituição concorrente, Localização, Vagas ofertadas, Número de candidatos, valor de mensalidades praticadas, condições de trabalho dos docentes, nível salarial dos docentes	Pesquisa Institucional
MERCADO EDUCACIONAL - Demanda (séries históricas relativas aos candidatos)	Número de candidatos, perfil econômico social e cultural, origem escolar (ensino médio), Expectativas Quanto ao curso, Expectativas quanto a profissão	Pesquisa Institucional
EX-ALUNOS	Dados Pessoais, Colocação no mercado, nível salarial, interesse em novos cursos, grau de aceitação da UNOESC	Pesquisa Institucional

A inclusão de uma nova informação a partir de determinado setor indicado nas Tabelas 7 e 8 implica em procedimentos administrativos adicionais, nem sempre simples tais como estabelecimento de novas rotinas, controle, etc. Além do que, se faz necessário um trabalho de esclarecimento e sensibilização dos colaboradores envolvidos para que o resultado seja satisfatório.

Com relação aos dados externos, está prevista a criação de um setor voltado para a *Pesquisa Institucional*. Na realidade, trata-se de um setor a ser estruturado para a realização de pesquisas (na acepção acadêmica do termo) e que estaria incumbido de obter e organizar informações mais complexas provenientes do ambiente externo. A Figura 17 exemplifica o modelo proposto.

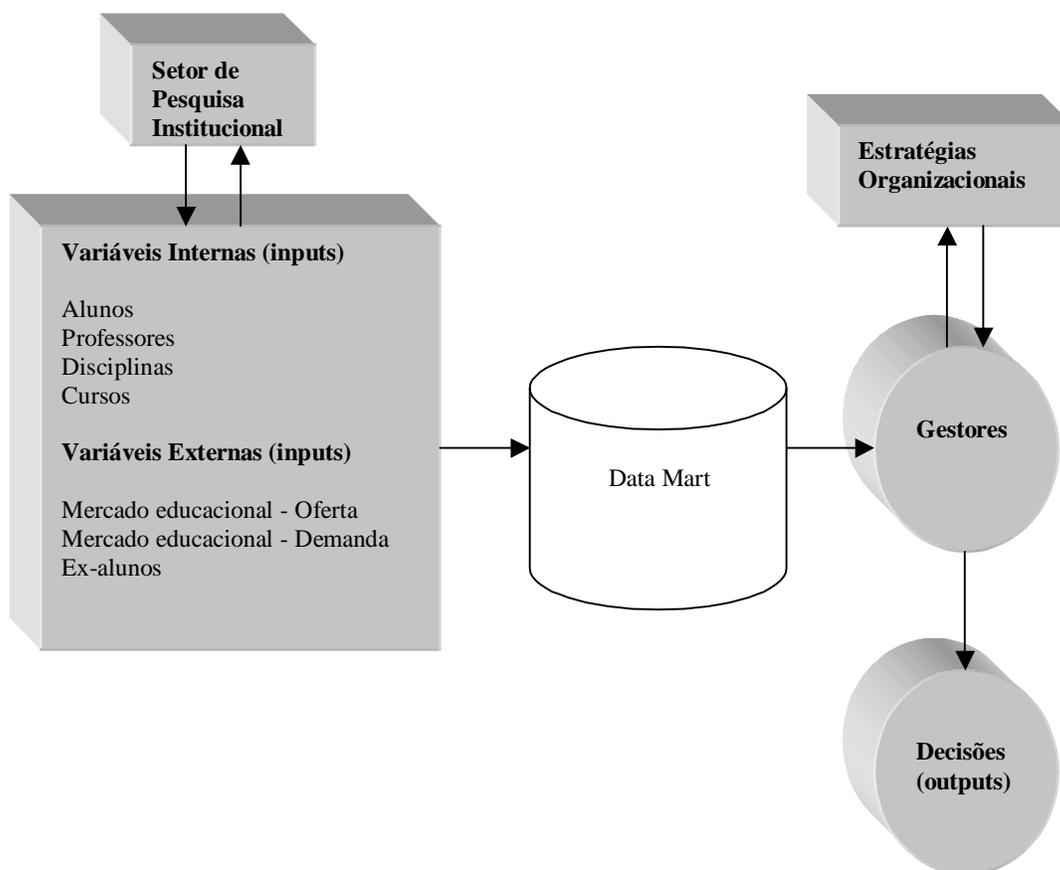


Figura 17 - Modelo Proposto para a UNOESC

Esses dois conjuntos de variáveis compõem os *inputs* do Data Mart. Os *outputs* do DM são decisões do gestor acadêmico, materializadas em propostas de projetos pedagógicos, de planos estratégicos e outros produtos em nível gerencial.

A essência da eficácia do sistema está na correta escolha das variáveis e dos itens de informação que deverão estar contidos no banco de dados voltado ao processo decisório.

4.8 Considerações Finais

Este capítulo apresentou o estudo da Universidade do Oeste de Santa Catarina, campus de Chapecó, com vistas à construção de um ambiente DW para esta instituição de ensino superior.

O estudo da estrutura organizacional e administrativa da universidade aponta para uma organização de três níveis nas atividades-fim (conselho – pró-reitorias – centros). O modelo decisório apresenta tanto aspectos do modelo colegiado e burocrático.

O levantamento dos sistemas legados aponta heterogeneidade de sistemas e plataformas, caracterizando diversas “ilhas de informação”. Esta situação faz que a direção da universidade não seja atendida com subsídios à tomada de decisão.

Considerando as arquiteturas estudadas no capítulo 3 e a natureza das instituições de ensino superior, propõe-se para a UNOESC a arquitetura *bottom-up* e sugere-se que o Data Mart inicial contemple o sistema acadêmico, incluindo as variáveis internas e externas à universidade.

5 PROTÓTIPO DE UM DATA MART PARA A UNOESC CHAPECÓ

No capítulo 4, apresentou-se um modelo de construção do ambiente DW para a Universidade do Oeste de Santa Catarina (UNOESC – Chapecó). O modelo tem base no levantamento da estrutura administrativa da universidade e na opção pela arquitetura bottom-up de construção.

Neste capítulo, apresenta-se o processo de desenvolvimento do primeiro Data Mart da UNOESC, segundo o modelo proposto no capítulo 4. O objetivo é implementar o processo completo da construção do ambiente DW, considerando a Tecnologia de Data Warehouse (capítulo 3), e o modelo proposto para a UNOESC (capítulo 4), dadas suas características institucionais e estrutura atual de suas informações.

Inicialmente, retoma-se a discussão sobre a arquitetura a adotar no DW. A seguir, detalha-se o projeto do ambiente DW proposto, seguindo-se a implementação do Data Mart Acadêmico da universidade. São implementados quatro exemplos de análises possibilitadas pelo Data Mart desenvolvido.

5.1 Bottom-up X Top-down

Como se viu no capítulo 3, os primeiros projetos sobre Data Warehouse referiam-se a uma arquitetura do tipo centralizada. Embora fosse interessante fornecendo uniformidade, controle e maior segurança, a implementação desta abordagem não é uma tarefa fácil. Requer uma metodologia rigorosa e uma completa compreensão dos negócios da empresa. Esta abordagem pode ser longa e dispendiosa e por isto sua implementação exige um planejamento bem detalhado. Com o aparecimento de data mart ou warehouse departamental, a abordagem descentralizada passou a ser uma das opções de arquitetura data warehouse. Os data marts podem ser concebidos de duas maneiras. A primeira é *top-down*, e a outra é a *bottom-up*.

A tecnologia usada, tanto no DW como no Data Mart, é a mesma. As variações que ocorrem são mínimas, sendo em volume de dados e na complexidade de carga. A principal diferença é a de que os Data Marts são voltados somente para uma determinada área. Já o DW é voltado para os assuntos da instituição toda. Portanto, cabe a cada instituição avaliar a sua demanda e optar pela melhor solução. No caso específico deste projeto optou-se pela criação de um protótipo de Data Mart do tipo **Bottom-up**, o que se mostra mais adequado às IES.

5.2 O Projeto

Conforme citado no Capítulo 3, não existe uma receita pronta para a construção de Data Mart. O que existe são ferramentas que contemplam as várias etapas desde o processo de extração e análise dos dados até a construção propriamente dita. Abaixo, são elencadas as principais fases que foram percorridas para a concepção desse protótipo.

5.2.1 Identificação da área de negócio

É importante ter bem clara qual a área de negócio a ser modelada. Através de um levantamento minucioso em relatórios existentes nos sistema de OLTP da UNOESC Chapecó, identificou-se a **área acadêmica** como carente de informações gerenciais, conforme já abordado no Capítulo 4. Com base nas informações coletadas, dá-se início à modelagem propriamente dita.

5.2.2 Processo de Negócio a ser modelado

Uma vez identificada a área de negócio, faz-se necessário definir qual o processo de negócio a ser modelado. No caso deste protótipo, tomou-se por base o processo de **matricula**, a saber:

- No início de cada semestre letivo, os acadêmicos realizam o processo de matrícula, ou seja, é ofertada uma série de disciplinas para serem cursadas, sempre observando o cumprimento de pré-requisitos e co-requisitos, bem como choque de horários. Com base na escolha das disciplinas, podem ser compiladas várias informações de caráter gerencial, que ao longo do tempo sofrem mutações que impactam na tomada de decisão. Um exemplo é um discente que, no início do semestre, matriculou-se em um número determinado de créditos e no mês seguinte realizou trancamento total ou parcial e, posteriormente, obteve aproveitamento de alguns créditos. Isso ocorre com frequência e muitas vezes contribui para grandes discrepâncias no fluxo de caixa da Instituição.

Existe uma necessidade da Pró-reitoria de Administração saber ao longo do tempo como será o impacto dessas mutações no processo de tomada de decisão. O sistema acadêmico atual, cuja arquitetura é OLTP, não consegue responder a tal pergunta, visto que o mesmo apenas mantém a situação corrente, para efeitos de controle do processo de matrículas aos cursos.

5.2.3 Levantamento de requisitos

É importante sabermos que, ao se implementar sistemas cuja finalidade é análise, o refinamento dos requisitos necessários deve ser realizado de forma incremental. O ponto de partida para as demais etapas desse projeto são os requisitos levantados inicialmente, que deverão ser revistos à medida que o protótipo for implementado e apresentado. Devido à própria natureza exploratória do trabalho de pesquisa e da própria natureza dos sistemas que se procura atender no ambiente DW, em um primeiro momento não se obtém uma especificação completa de todos os requisitos. Esta se forma com o desenrolar da implementação e validação do protótipo.

Através de reuniões com os usuários potenciais do Data Mart, foram levantados alguns objetivos e, também, identificado a necessidade de informação a partir de diversas situações descritas a seguir:

Os dirigentes querem saber, ao longo do tempo:

- Índice de desistências, trancamentos, cancelamentos e aproveitamentos por curso.
- Índice de Reprovação/Aprovação por curso.
- Índice de Evasão por curso.
- Índices relativos por sexo e idade (desistência, reprovação, evasão) Previsão de vagas x Aproveitamento Final (excluindo evasão, inadimplência e desistência) x (Total de Alunos no início do período).
- Evolução dos índices de reprovação/aprovação/cancelamento ao longo do tempo em cada curso.
- Distribuição espacial e temporal dos conceitos na UNOESC, por curso, disciplina e centro.

Kimball et. al. (1998) consideram o levantamento de requisitos o centro do Data Warehouse. É este levantamento que determina que dados devem estar disponíveis no DW. Como este se organiza e qual frequência é atualizado. Neste trabalho, o levantamento é resultado dos estudos apresentados no Capítulo 4 e de entrevistas junto a dirigentes da área acadêmica. A Tabela 9 apresenta os indicadores solicitados na área acadêmica e sua importância para os dirigentes.

Tabela 9 – Indicadores solicitados pela alta administração da área acadêmica.

Indicador	Relevância para Dirigente Acadêmico
Índice de desistências por curso	Permite obter o nível de desistência, possibilitando políticas específicas por curso na UNOESC.
Índice de trancamento por curso	Permite obter o nível de trancamento, possibilitando políticas específicas por curso.
Índice de cancelamentos por curso	Permite obter o nível de cancelamentos, possibilitando políticas específicas por curso.
Índice de aproveitamento Por curso	Permite obter o nível de aproveitamentos, possibilitando políticas específicas por curso.

5.2.4 Medidas Envolvidas

Uma vez definidos os objetivos, fa-ze necessário elencar as medidas que serão envolvidas para atingir os objetivos propostos. As principais medidas envolvidas no protótipo são:

- Quantidade de Créditos Matriculados
- Quantidade de Créditos Aproveitados
- Quantidade de Créditos Dispensados
- Quantidade de Créditos Cancelados
- Quantidade de Créditos Trancados
- Conceito Final

5.2.5 Granularidade

Com base nas medidas envolvidas, é importante estabelecermos qual será o nível de granularidade adequado, conforme abordado no Capítulo 3. Para esse protótipo, será adotado o **modelo levemente resumido**, visto que o conteúdo dos dados sofre modificações em relação ao dados correntes e

serão armazenados em **séries mensais**. As medidas envolvidas serão agregadas por mês, por aluno e por curso.

5.2.6 Modelagem Dimensional

A UNOESC Chapecó tem necessidade declarada por informações estratégicas, mas freqüentemente não dispõe de mecanismos adequados para tratar seus dados e gerar tais informações. A questão está no fato do volume de dados ser irrelevante, se ele não estiver organizado de forma a agregar valores. Agregar valores em uma Instituição significa transformar dados em informações úteis. Como visto no capítulo 3, a modelagem dimensional é uma abordagem que pode dar sentido a esta montanha de dados. O modelo dimensional representa os dados como matriz na qual cada dimensão é um tema ou assunto de negócio que será objeto da análise e o tempo é sempre uma das dimensões consideradas.

a) Esquema Estrela

A modelagem dimensional do protótipo é efetivada segundo o esquema estrela. Conforme já abordado no Capítulo 3, o esquema estrela é formado por dois elementos principais: tabelas de fatos e tabelas de dimensão. As tabelas de fato, também chamadas de tabelas principais, contêm dados *quantitativos* sobre o processo de matrícula. Já as tabelas de dimensão, chamadas de tabelas secundárias, contêm classes que descrevem as medidas contidas nas tabelas de fatos. O objetivo do protótipo é apresentar aos tomadores de decisão da UNOESC, em um curto espaço de tempo, o que vem a ser um sistema de Data Mart e como esse sistema poderá melhorar a qualidade das decisões.

Tabela 10 – Elementos do Modelo Multidimensional Desenvolvido.

Elemento	Conceito	No Protótipo
Tabela de Fatos	Tabela primária no modelo dimensional, onde cada fato representa uma medida de negócio da organização (Kimball et. al. 1998).	Dados quantitativos sobre o processo de matrícula <ul style="list-style-type: none"> ▪ MATRICULA
Tabelas de Dimensão	Tabelas secundárias no modelo dimensional, que guardam um conjunto de tabelas relacionadas a uma tabela de fato (Kimball et. al. 1998).	Dados relativos a registro de <ul style="list-style-type: none"> ▪ TEMPO ▪ ALUNO ▪ CURSO ▪ DISCIPLINA

A Tabela 10 apresenta os elementos do modelo multidimensional desenvolvido para o Data Mart acadêmico da UNOESC. A relação entre estes elementos foi representada segundo o esquema estrela da Figura 18.

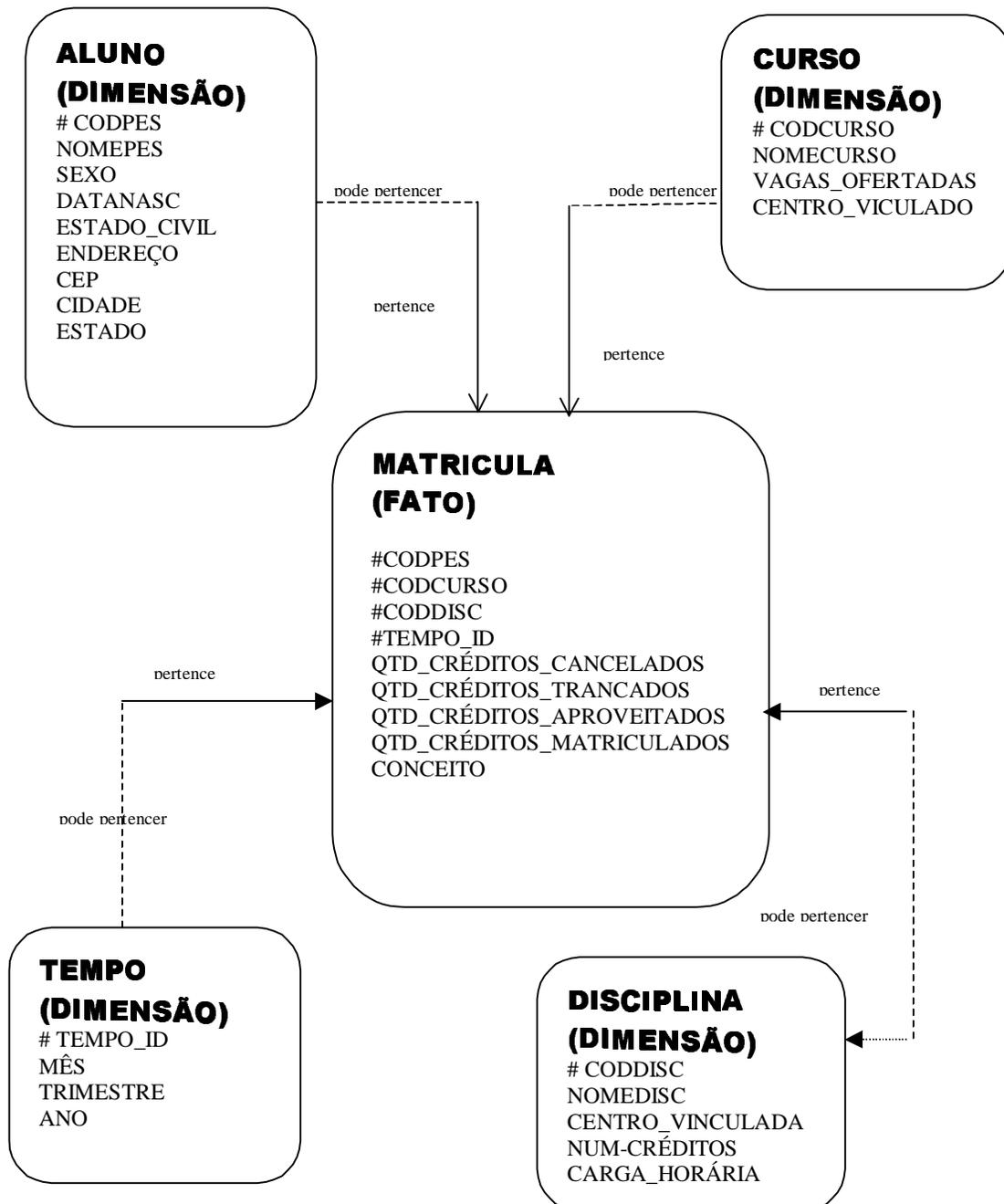


Figura 18 – Esquema Estrela do Protótipo

O esquema-estrela do protótipo contém uma tabela de fato e 4 tabelas de dimensão, a saber:

- ✓ Tabela de fato *matrícula*: contém os atributos *qtd_créditos_cancelados*, *qtd_créditos_trancados*, *qtd_créditos_aproveitados*, *qtd_créditos_matriculados* e *conceito*. A chave primária consiste nas quatro chaves estrangeiras: *codpes*, *codcurso*, *coddisc* e *tempo_id*, as quais ainda não são representadas neste nível mais alto do projeto, mas aparecerão durante a implementação.
- ✓ A tabela de dimensão *tempo* possui sua chave primária, cujo atributo é um identificador de tempo que na fase de implementação é subdividido em *ano*, *mês*, *trimestre*.
- ✓ A tabela de dimensão *aluno* contém oito elementos, além da chave primária *nome do aluno*, *sexo*, *data de nascimento*, *estado civil*, *endereço*, *código de endereçamento postal*, *cidade*, *unidade da federação* e tem como chave primária o *código do aluno*.
- ✓ A tabela de dimensão *curso* tem como chave primária o código do curso e abriga mais três atributos: nome do curso, as vagas ofertadas, o centro vinculado.
- ✓ A tabela de dimensão *disciplina* é composta pela chave primária, código da disciplina e pelos atributos: nome da disciplina, centro de vínculo da disciplina, número de créditos e carga horária.

5.3 Implementação

Após análise exaustiva junto aos usuários do protótipo e completado o projeto do banco de dados, é hora de passar para o próximo passo, ou seja, para implementação. Neste tópico serão abordadas também as características das ferramentas que foram usadas na implementação do Data Mart e o modo como foram utilizadas.

5.3.1 Ferramentas de Implementação

Para a implementação do protótipo, utilizou-se a ferramenta *Oracle Data Mart Suíte®*. Há várias razões pela escolha desta ferramenta:

- âmbito do projeto é departamental com enfoque em dados acadêmicos.
- *Oracle Data Mart Suíte®* proporciona um rápido começo no projeto de Data Mart e os resultados já podem ser obtidos em algumas semanas ao invés de meses de projeto. Esta característica deve ser observada com especial atenção gerentes de projetos em DW em IES. Períodos longos sem retorno aos dirigentes podem implicar em diminuição no grau de comprometimento com o projeto, dado os custos de implantação.
- É uma ferramenta bastante fácil de aprender (incorpora manuais on-line). No caso específico deste protótipo, foram dedicados menos de três meses para aprendizado da ferramenta e implementação.
- O *Oracle Data Mart Suíte®* suporta do ciclo de vida do Data Mart, desde o projeto até a análise dos resultados pelo usuário final.
- O ambiente Oracle já se encontra incorporado dentro da UNOESC, o que proporcionou uma certa familiaridade com a ferramenta. O projetista do ambiente DW deve estar atento à infra-estrutura disponível na organização e deve procurar informar-se sobre as soluções DW que os fornecedores atuais da IES possuem. Isto pode reduzir custos e facilitará a disseminação da tecnologia DW nas demais áreas de suporte da universidade.
- O pacote *Oracle Data Mart Suíte®* provê todo o software que foi preciso para construir esse protótipo e tem incorporado as seguintes ferramentas:
 - ✓ *Oracle Data Mart Designer®*: ferramenta utilizada para projetar o Data Mart.
 - ✓ *Oracle Data Mart Builder®*: Para extrair e transformar dados do sistema operacional.
 - ✓ *Oracle8 Enterprise Server®*: O servidor da base de dados.
 - ✓ *Oracle WebServer®*: Para a intranet ter acesso ao Data Mart.

- ✓ *Oracle Discoverer®*, ferramenta para consultas, relatórios e análise de dados.
- ✓ *Oracle Reports®*, usada para desenvolvimento e entrega de relatórios.

Para a implementação do protótipo foram utilizadas algumas das ferramentas acima, isto é, o *Oracle Data Mart Designer®*, *Oracle8 Enterprise Server®* e *Oracle Discoverer®*. Não foi necessário usar o Oracle Data Mart Builder, porque os dados não irão residir no mesmo sistema operacional. Da mesma forma o *Oracle WebServer®* foi substituído pela ferramenta o Oracle WebDB®, pela sua praticidade e a ferramenta Oracle Reports® não foi necessária nesse primeiro momento.

5.3.2 Passos da Implementação

Nesta seção serão descritos os passos que foram necessários para implementação do Data Mart com o uso do *Oracle Data Mart Suíte®*.

a) Representação do esquema-estrela no Oracle Data Mart Designer®

A ferramenta Data Mart Designer® permite projetar parte do ciclo de vida do data mart e possibilitou projetar o Data Mart desde sua concepção inicial. Após completado o projeto do esquema-estrela, foi inserido de forma visual o modelo de Entidade Relacionamento. Nesse momento foram inseridos os nomes das tabelas, seus atributos e a propriedade de cada atributo, conforme Anexo B. Posteriormente foram especificadas as relações que existem entre essas tabelas e a respectiva cardinalidade. A Figura 19 exemplifica o diagrama E-R obtido nessa fase.

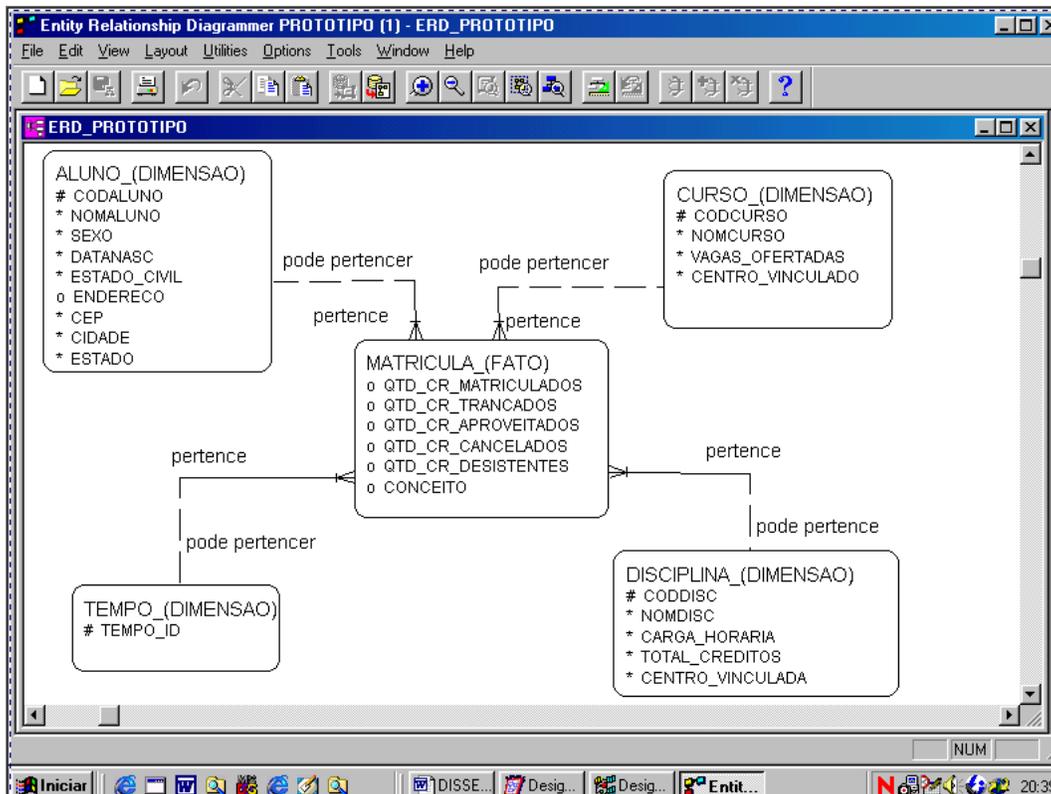


Figura 19 – Diagrama Entidade Relacionamento

Deve-se notar a semelhança entre o diagrama E-R na Figura 19 e o esquema estrela da Figura 18. Isto indica que o modelo multidimensional é o ponto de partida da especificação do DM e que a representação E-R é o resultado do recurso disponível na ferramenta adotada. O projetista do ambiente DW da IES deverá estar atento para as características da ferramenta adotada e deve procurar manter o modelo multidimensional como referência de seu Data Mart.

b) Projeto do banco de dados para o Oracle® Enterprise Server®

Nesse passo foi usado o *Oracle Data Mart Designer®* para a criação do esquema físico do Data Mart. Esta ferramenta tem a funcionalidade de criar o esquema físico no modelo servidor de banco de dados, bem como criar a integridade referencial, *constraints* e chaves estrangeiras. O resultado

desse passo pode ser visto na Figura 19 e foi utilizado para implementar o modelo original proposto.

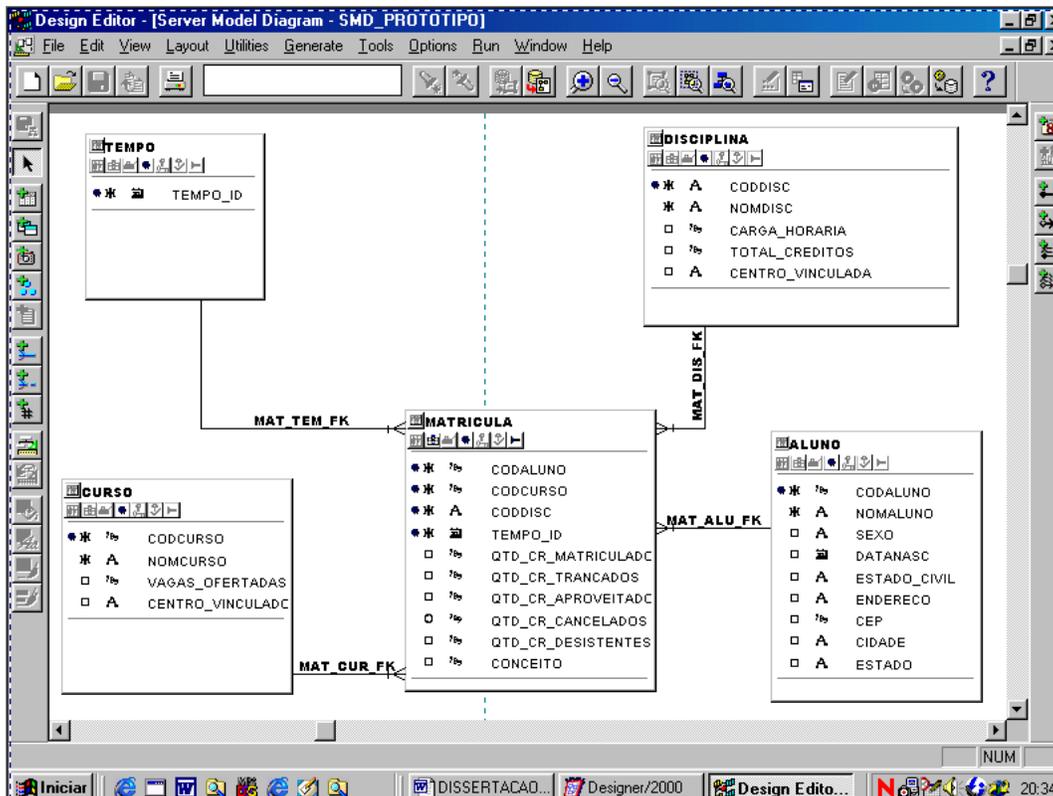


Figura 20 – Esquema Físico obtido a partir do modelo E-R

c) Geração do esquema para o database (SQL DDL)

O *Oracle Data Mart Suite®* tem a habilidade para gerar SQL DDL a partir do modelo servidor gerado no passo anterior. Essa DDL é típica, apresenta comentários e é útil para propósitos de documentação. Fazendo uso das características do *Oracle Data Mart Suite®*, foram gerados os *scripts* do esquema da base de dados, que podem ser vistos nos Anexos D e E.

d) Criação das tabelas no Banco de Dados Oracle® Enterprise Server®

Uma vez concluído o esquema, basta somente rodar os *scripts* de SQL para criar as tabelas. No caso do protótipo da UNOESC, utilizou-se a função específica do *Oracle Data Mart Suíte®* que permite executar os *scripts* diretamente no esquema específico no servidor de banco de dados.

e) Carga (Povoamento do DM)

O processo de carga é formado por uma série de procedimentos de banco de dados e estruturas intermediárias, necessárias para *popular* o DM proposto. Nesse protótipo, a carga não foi automatizada. Uma vez validado o esquema, deverá ser implementado o processo de carga automatizado através do uso de *schedules*, para viabilizar a carga no esquema estrela, bem como prover a eliminação de dados que expirarem, ou seja, dados anteriores ao período determinado para ser mantido no Data Mart.

Os procedimentos de carga geralmente são implementados através de processos intermediários; a carga não ocorre diretamente dos sistemas OLTP (bancos em produção) para o Data Mart, mas sim via *área de estagiamento* de dado. Como visto no capítulo 3, trata-se de estrutura de armazenamento intermediária que contempla o isolamento do DM em relação aos sistemas de OLTP. A carga somente será inserida no Data Mart, se todas as estruturas intermediárias tiverem sido constituídas adequadamente, o que evita que o Data Mart se torne indisponível por alguma inconsistência no banco de dados em produção. É importante lembrar que o processo de construção das dimensões geralmente exige diversas junções de tabelas e que as medidas na tabela de fatos implicam em cálculos e agregações. Essas etapas são caras e podem comprometer o banco de produção. Os passos intermediários podem implementar algumas destas junções, a serem realizadas em momentos em que haja recursos ociosos. A Figura 21 representa o processo de carga implementado no protótipo.

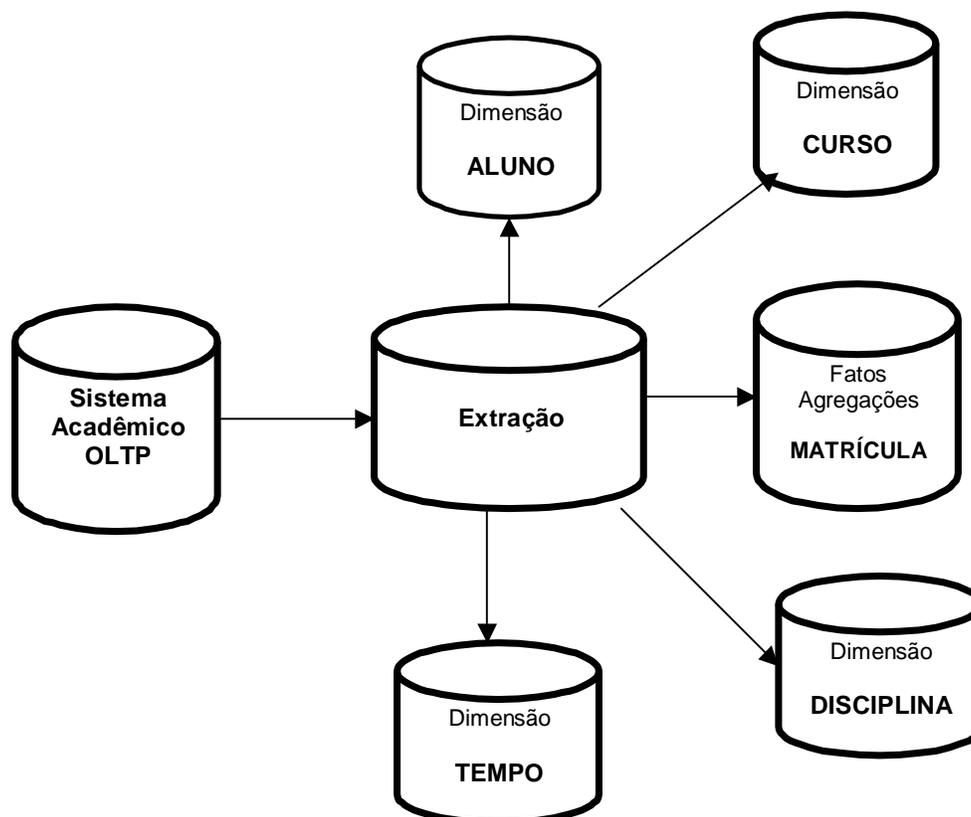


Figura 21 - Processo de Carga implementado no protótipo

Os objetivos da etapa intermediária são preservar a atomicidade da carga do Data Mart. Caso a carga tenha de ser abortada, o conteúdo do Data Mart é preservado. A carga só ocorre se as etapas intermediárias forem bem sucedidas. Deve-se isolar os cálculos das agregações que geram fatos e junções para formar as dimensões do OLTP.

f) Aplicações para o Usuário Final

Uma vez realizada a população do Data Mart, foi utilizado o *Oracle Discoverer®* como ferramenta *front-end* para suporte, análise e consulta relacionadas ao ciclo de vida de construção do DM. É importante lembrar que o repositório do Data Mart pode ser acessado através de diferentes ferramentas *front-end*. No caso específico desse protótipo serão

centrados sob o *Oracle Discoverer®*, com objetivo de validar o protótipo. Outrossim, todo e qualquer projeto de um data mart ou data warehouse, a *priori*, não necessita especificar em um primeiro momento a ferramenta ideal para o usuário final. Somente após o processo de validação do protótipo, com a equipe envolvida, é que se deve chegar à conclusão de qual é a ferramenta ideal para extração dos dados do DM.

No caso das IES, o projetista também deverá estar ciente dos fornecedores com quem a universidade trabalha e procurar racionalizar recursos físicos e humanos. O fato mais importante, no entanto, deverá ser o perfil dos analistas de informação que utilizarão as aplicações desenvolvidas sobre o ambiente DW. É muito provável que os dirigentes tenham que contar com técnicos em informação, devido ao pouco tempo que podem dispende nas análises on-line ou mesmo à pouca experiência com sistemas de análise como os que se necessita em um ambiente DW.

No protótipo da UNOESC, foram utilizados os seguintes componentes da ferramenta Discoverer:

- ✓ *Oracle Discoverer for Users®*: Utilizado para construir consultas, gerar relatórios e publicar informações no dicionário de dados. No Anexo A é apresentada a interface gráfica.
- ✓ *Oracle Discoverer Administration®*: Para administradores de dados criarem, dar manutenção e administrarem a interação dos dados dos usuários com os dados do banco de dados. O administrador prepara a área de trabalho do usuário com itens de dados do seu negócio. O Anexo C apresenta o módulo de administração.
- ✓ *End User Layer®*: É a interface entre o banco de dados e as áreas de estagiamento(trabalho) criada pelo administrador;
- ✓ BD: Realiza a coleta e organiza as informações. Em um banco de dados as informações estão armazenadas em um servidor e são acessadas por vários usuários.

- ✓ **BD Relacionais:** As informações são armazenadas em tabelas que contêm linhas e colunas. O BD gerencia o relacionamento destas tabelas. Para acessar os dados, utiliza-se de uma linguagem de programação SQL. O *Oracle discovere®* gera os programas em SQL. O usuário final não precisa conhecer SQL.

5.3.3 Extração e Análise dos Resultados

Com base no levantamento de requisitos, a seguir são elencados alguns exemplos com o objetivo de servir de apoio ao processo decisório dos gestores acadêmicos. Para efeito de aplicabilidade, serão tomadas como base as oito grandes áreas do conhecimento, a fim de preservar os dados privativos a cada curso.

As ferramentas utilizadas para a geração destes exemplos foram o *Discoverer Administration Edition®* e o *Discoverer User Edition®*, ambos da ORACLE. O *Discoverer Administration Edition®* foi utilizado para a preparação e administração do ambiente de trabalho para os usuários finais. O *Discoverer User Edition®* foi utilizado para a geração das aplicações desenvolvidas e extrações de informações. A geração obedece a alguns procedimentos, como seleção das tabelas a serem utilizadas, seleção de atributos, definição das operações sobre os atributos, condições pré-determinadas e cálculos.

Exemplo 1: Índice de Trancamento por Grande Área do Conhecimento

Apresenta os percentuais de trancamento de créditos por área do conhecimento no primeiro semestre do ano de 2000 na UNOESC, campus de Chapecó. A Figura 22 apresenta a planilha gerada pelo *Discoverer User Edition®*. Procedendo-se uma análise, observa-se que algumas grandes áreas do conhecimento apresentam índices elevados de trancamento de créditos. Já a Figura 23, na forma de gráfico de barras, deixa claro essa diferença. Com base nesses dados, o gestor tem informações importantes para proceder

correções e/ou ajustes que se fizerem necessários nas políticas acadêmicas da universidade.

Oracle Discoverer - [Índice de Créditos Trancados por Área do Conhecimento no 1 sem_2000.DIS]

Índice de Créditos Trancados por Área do Conhecimento no Primeiro Semestre de 2000.

	Nome do Curso	Percentual
1	CIENCIAS AGRARIAS	5,72
2	CIENCIAS BIOLÓGICAS	3,29
3	CIENCIAS DA SAUDE	2,57
4	CIENCIAS EXATAS E DA TERRA	9,42
5	CIENCIAS HUMANAS	2,76
6	CIENCIAS SOCIAIS E APLICADAS	3,66
7	ENGENHARIAS	7,86
8	LINGUISTICA, LETRAS E ARTES	4,87

Figura 22 - Índices de trancamentos por área do conhecimento

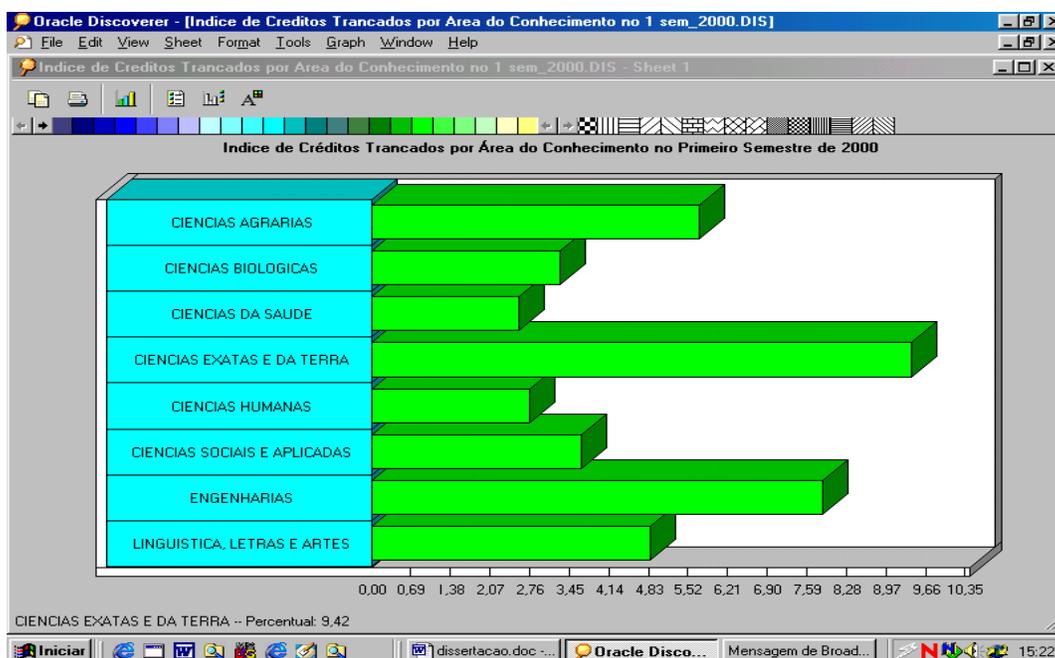
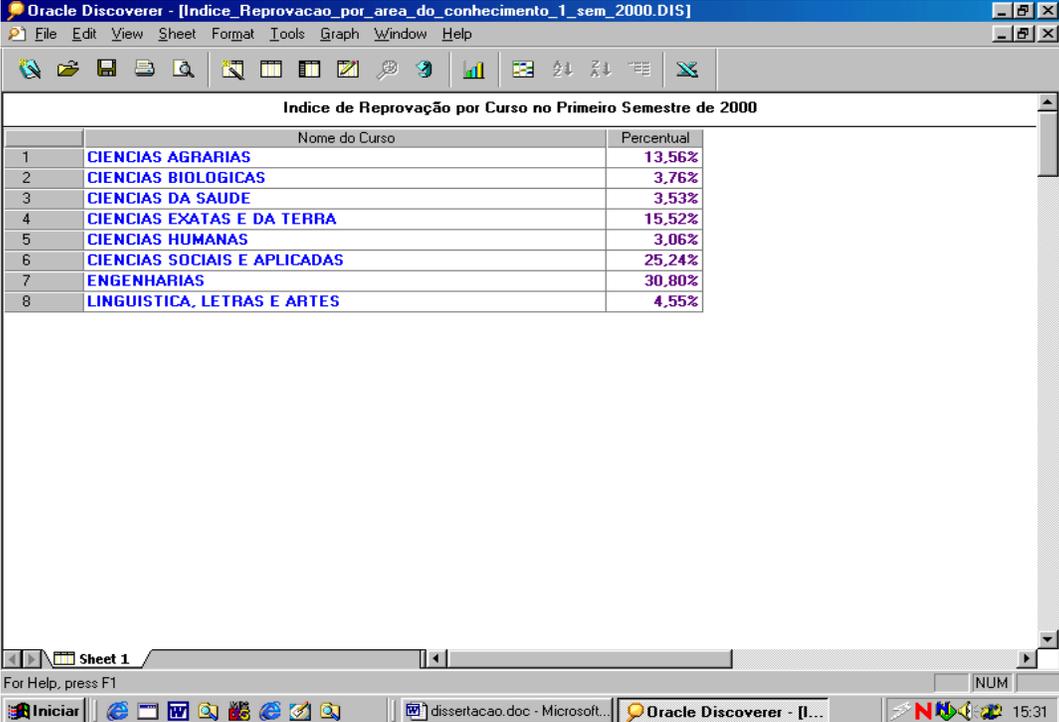


Figura 23 - Representação gráfica dos índices de trancamentos

Exemplo 2: Índice de Reprovação por Grande Área do Conhecimento

Ao analisar a Figura 24 observa-se o índice de reprovação por grande área do conhecimento no primeiro semestre de 2000. Já a Figura 25 apresenta esse índice em forma gráfica. Com base nessas informações, os gestores têm informações sumarizadas para utilizá-las no processo de tomada de decisão sobre o estágio atual de avaliação final dos discentes nas disciplinas.

O estudo mostra uma diferença de até dez vezes entre o menor índice de reprovação (Ciências Humanas), e o maior índice (Engenharias). É importante comparar estas análises com indicadores racionais e verificar a natureza destas informações em termos comparativos.



Oracle Discoverer - [Índice_Reprovacao_por_area_do_conhecimento_1_sem_2000.D15]

File Edit View Sheet Format Tools Graph Window Help

Índice de Reprovação por Curso no Primeiro Semestre de 2000

	Nome do Curso	Percentual
1	CIENCIAS AGRARIAS	13,56%
2	CIENCIAS BIOLÓGICAS	3,76%
3	CIENCIAS DA SAUDE	3,53%
4	CIENCIAS EXATAS E DA TERRA	15,52%
5	CIENCIAS HUMANAS	3,06%
6	CIENCIAS SOCIAIS E APLICADAS	25,24%
7	ENGENHARIAS	30,80%
8	LINGUISTICA, LETRAS E ARTES	4,55%

Sheet 1

For Help, press F1

NUM

Iniciator | Microsoft... | Oracle Discoverer - [I... | 15:31

Figura 24 - Índice de Reprovação por Grande Área do Conhecimento

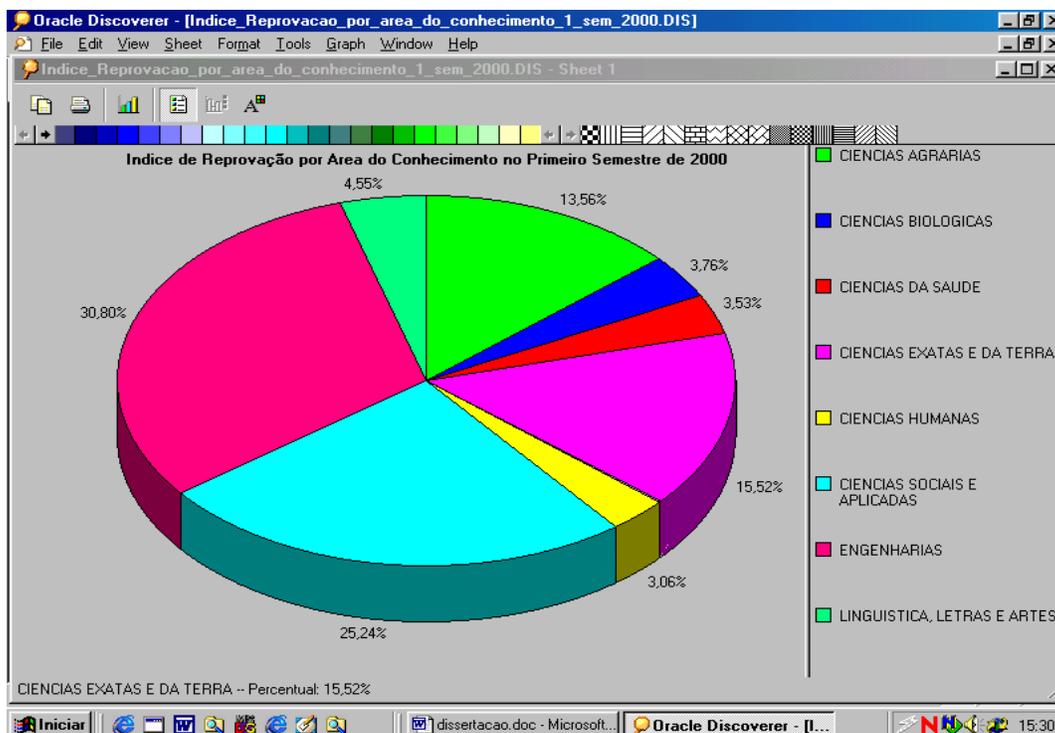


Figura 25 - Gráfico com o Índice de Reprovação

Exemplo 3: Índice de Trancamentos por Sexo

Ao observar a Figura 26, o gestor tem o índice de trancamento por sexo no primeiro semestre de 2000. Baseado nessas informações pode-se fazer um levantamento minucioso, como, por exemplo, ao clicar em uma região específica do gráfico é possível recuperar em detalhes quais são os alunos que foram aquele índice.

A análise do nível de trancamentos nos cursos, por sexo do estudante, confirma a intuição dos dirigentes. Estudantes do sexo feminino de forma geral trancam com mais frequência do que alunos do sexo masculino. Uma das causas prováveis é a maternidade. Mas somente uma análise com mais variáveis (ex. estado civil) poderia indicar as reais razões. Este é um exemplo das situações com que o projetista do DW da IES deverá se defrontar ao longo do ciclo de vida de cada Data Mart. Os requisitos iniciais podem mudar ou necessitar maior detalhamento, o que exige o ajuste e ampliação do

modelo inicial, desde a modelagem multidimensional, até a análise, passando pelo repovoamento do Data Mart.

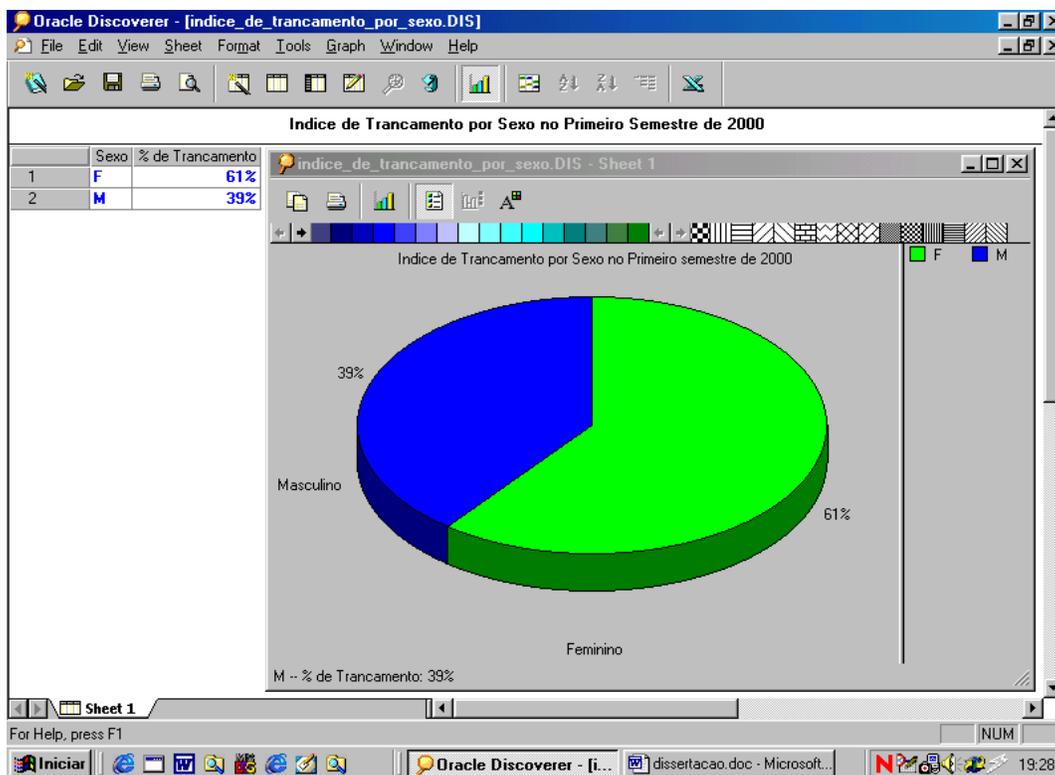


Figura 26 - Índice de Trancamentos por Sexo

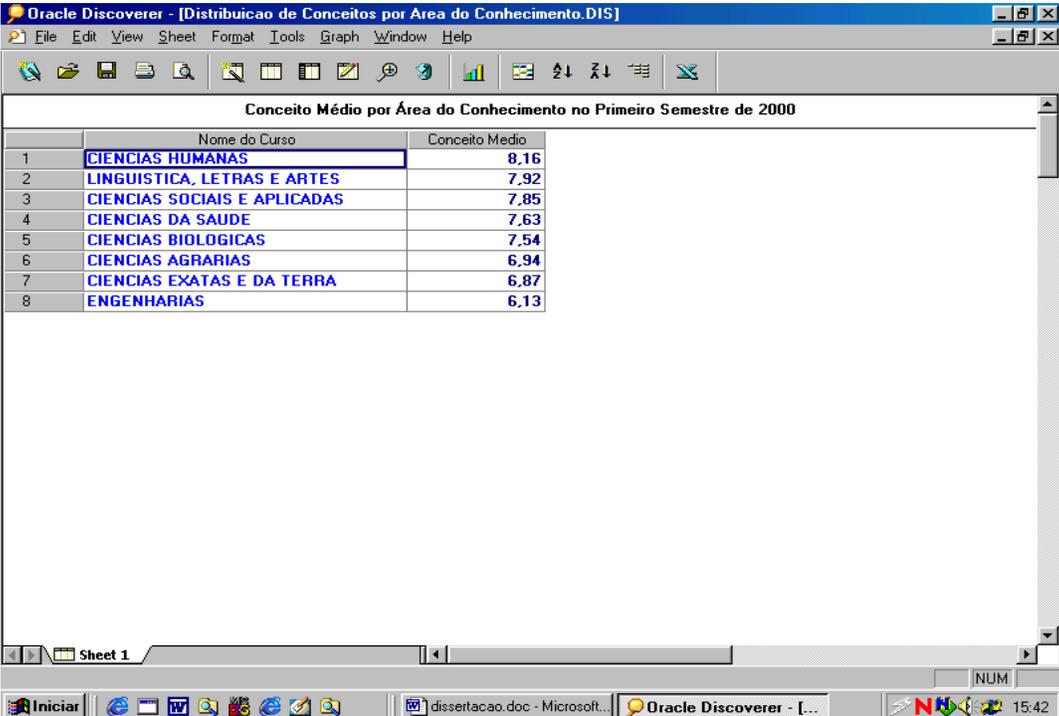
Exemplo 4: Conceito Médio por Grande Área do Conhecimento

O gestor acadêmico, ao proceder uma análise na Figura 27, tem o conceito médio por grande área do conhecimento, no primeiro semestre de 2000. Isso fica mais claro através do gráfico apresentado na Figura 28. Com base nessas informações, é possível tirar algumas conclusões.

O conceito médio das disciplinas da UNOESC varia de 6,13 (Engenharia) a 8,16 (Ciências Humanas). Deve-se observar que estes valores são inversamente proporcionais aos índices de reprovação, como era de se esperar. Os cursos com maior reprovação têm menor média de conceito. A mesma relação não ocorre de forma tão direta quando o conceito médio (ou a reprovação) é comparado com o índice de trancamento, em cada grande área.

O maior índice de trancamento está nas Ciências Exatas e da Terra, enquanto o menor está nas Ciências Humanas. Portanto, parece haver uma relação parcial entre a avaliação e o trancamento, como se nota na grande área de maior conceito e menor índice de trancamento.

Estas comparações serão muito provavelmente solicitadas pelos dirigentes acadêmicos, o que deverá levar o projetista do ambiente DW a retomar as especificações do Data Mart e recomeçar o ciclo de desenvolvimento, agora sob uma nova ótica de análise.



Oracle Discoverer - [Distribuicao de Conceitos por Area do Conhecimento.DIS]

File Edit View Sheet Format Tools Graph Window Help

Conceito Médio por Área do Conhecimento no Primeiro Semestre de 2000

	Nome do Curso	Conceito Medio
1	CIENCIAS HUMANAS	8,16
2	LINGUISTICA, LETRAS E ARTES	7,92
3	CIENCIAS SOCIAIS E APLICADAS	7,85
4	CIENCIAS DA SAUDE	7,63
5	CIENCIAS BIOLOGICAS	7,54
6	CIENCIAS AGRARIAS	6,94
7	CIENCIAS EXATAS E DA TERRA	6,87
8	ENGENHARIAS	6,13

Sheet 1

NUM

Iniciat | Microsoft Word | dissertacao.doc - Microsoft... | Oracle Discoverer - [...] | 15:42

Figura 27 - Conceito Médio por Grande Área do Conhecimento

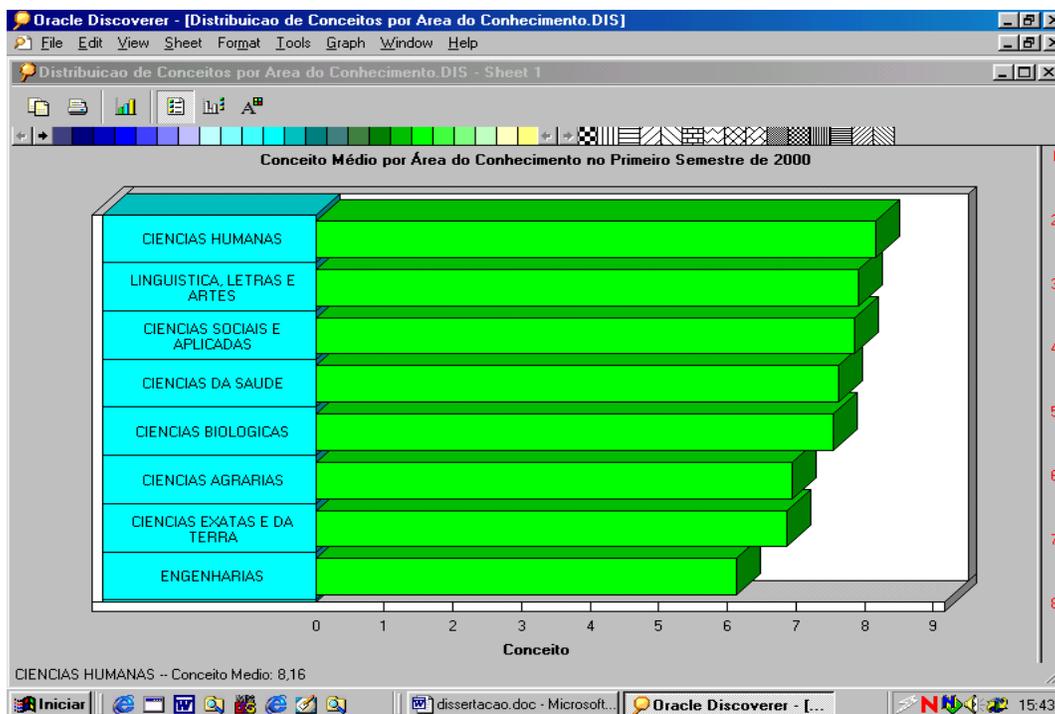


Figura 28 - Gráfico de Conceito Médio por Grande Área do Conhecimento

5.3.4 Considerações Finais

Este Capítulo apresentou o processo de implementação do modelo de ambiente DW proposto para a Universidade do Oeste de Santa Catarina – UNOESC – Chapecó, segundo as especificações do Capítulo 4. Para tal, foram estabelecidos os passos necessários à construção do primeiro Data Mart da universidade, desde sua definição (Data Mart Acadêmico) à sua utilização em quatro casos de estudo e análise viabilizadas pelo DM desenvolvido:

As análises desenvolvidas resultaram tanto em informações já intuitivas aos dirigentes (ex. cursos que reprovam mais, têm médias de conceitos menores). Como deduções que levaram os dirigentes acadêmicos à reflexão (ex. há uma relação indireta entre avaliação no curso e índice de trancamento, mas esta só ocorre em cursos de determinadas áreas).

O ambiente projetado para a UNOESC permite tanto a construção de novos Data Marts para a universidade, como a ampliação dos sistemas *front-end* para análise do DM acadêmico. Para tal, a adoção das ferramentas ORACLE facilita tais desenvolvimentos, tanto pela natureza dos aplicativos utilizados como por sua disponibilidade na universidade.

6 CONCLUSÕES E RECOMENDAÇÕES

Este trabalho descreve como foi realizada a concepção, projeto e implementação de um Data Mart em uma Instituição de Ensino Superior, através de um estudo de caso junto à UNOESC - Universidade do Oeste de Santa Catarina - Campi de Chapecó. Com a finalidade de atingir os objetivos propostos, várias etapas foram percorridas. No Capítulo 2, ao estudar as estruturas e organizações universitárias, foi possível entender os modelos de organizações universitárias, bem como a hierarquia e fluxo de tomada de decisão nessas instituições. Ficou claro, nesse estudo que a UNOESC se encaixa no modelo burocrático e colegiado, e processo decisório assemelha-se ao modelo baseado em áreas afins (Centros), uma vez que a figura do departamento é substituída diretamente pelo curso vinculado ao centro. No Capítulo 3, foi realizado um estudo da tecnologia de Data Warehouse - Data Mart, onde foram abordadas as principais arquiteturas e modelos de DW. Ficou claro que o modelo organizacional das IES aponta para a criação de um Data Mart departamental, uma vez que o modelo de dados defendido por INMON (Inmon, 1997) é pouco adequado para Universidades. O modelo mais adequado é o de KIMBALL (kimball et. al. 1998) por ser **incremental** por departamento, e prever sempre uma integração.

No Capítulo 4, é proposta a implantação de um Data Mart com enfoque na gestão acadêmica. Para isso, foi necessário um estudo da estrutura organizacional da UNOESC e o fluxo das informações para a tomada de decisão na gestão acadêmica. Para tanto, foram realizados levantamentos nos sistemas existentes, e ficou clara a existência de várias "*ilhas de informação*", conseqüentemente o isolamento de dados entre os setores. Por outro lado, constatou-se que a cada dia essas ilhas estão sendo integradas. Com base nesse estudo, foi proposto o DM para o gestor acadêmico.

No Capítulo 5, é construído um protótipo de parte do modelo proposto com o objetivo de validá-lo, onde foi possível percorrer, na prática todas as fases para criação de um Data Mart e também verificado o potencial das ferramentas ORACLE, no processo de construção de DM.

Diante do referido estudo, deve-se ressaltar que o processo de gestão é um processo puramente humano, por mais informatizado que seja. As informações colocadas diante dos tomadores de decisão são indicadores para melhor visualização do contexto que o rodeia.

A vantagem de implantação de um DM em uma IES é propiciar aos gestores acadêmicos uma significativa economia de tempo e esforço no processo de decisão. As decisões continuaram a ser tomadas como de praxe, o que muda é o grau de certeza que antecede à tomada de decisão e, por consequência, sua probabilidade de acerto e precisão.

Faz-se à natureza exploratória do trabalho, foi necessário delimitar o escopo de atuação. Sendo assim, gostaria de deixar, a título de sugestões futuras, a exploração de um estudo teórico sobre a comparação das metodologias de construção de DW, bem como um estudo detalhado de como a tecnologia de descoberta, em base de dados (*data mining*), poderá vir agregar valor ao Data Warehouse.

7 FONTES BIBLIOGRÁFICAS

- ANJOS, Sara Joana Gadotti. **Uma Contribuição para a Arquitetura de Informações Estratégicas (AIE) para Setores de Pesquisa em Universidades Brasileiras**. 1999. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-graduação em Engenharia de Produção, UFSC, Florianópolis.
- ALVIM, Paulo C. R. de Carvalho. O papel da informação no processo de capacitação tecnológica das micro e pequenas empresas. **Ciência da Informação**, Brasília, v.27 n.1,p.28-35, Jan./abr. 1998.
- BARQUINI, Ramon. **Planning and designing the Warehouse**. New Jersey, Prentice-Hall, 1996. 311p.
- BATINI, C., LENZERINI, M. **Comparative Analysis Of Methodologies For Database Schema Integration, ACM Computing Surveys**. New York, v.18, nº 4, pág.323 - 364, Dez/1986.
- BERRY, Michael J. A. **Data Mining Techniques: For Marketing, Sales, and Customer Support**. John Willey & sons, Inc., New York. 1997.
- BOHN, K. **Converging Data for Warehouses**. DBMS, Califórnia, Junho 1997.
- BORGES, Mônica E. N. A informação como recurso gerencial das organizações na sociedade do conhecimento. **Ciência da Informação**, São Paulo, v.24 n.2,1995.
- BRACKETT, Michael H. **The Data Warehouse Challenge**. Taming Data Chaos. John Wiley & Sons, Inc - New York - USA, 1996.
- CAMPOS, Maria Luiza e ROCHA, Arnaldo V. **Características do Data Warehouse**. Disponível em <http://genesis.nce.ufrj.br/dataware/tutorial/cbasicos.html#tg_I5>. Acesso em: 15 de jan. 2001.
- CAMPOS, Maria Luiza e ROCHA, Arnaldo V. **Data Warehouse XVII Congresso da Sociedade Brasileira de Computação, XVI Jornada de Atualização em Informática**, Rio de Janeiro, 1997.
- CHAUDHURI, S. e DAYAL, U. **An Overview of Data Warehousing and Olap**

- Tecnology**, SIGMOD Record, New York, v.26, nº 1, Mar/1997.
- CRUB, Conselho de Reitores das Universidades Brasileiras. **IES dados de Universidades filiadas ao CRUB**. Brasília [s.n], 1998. 35p.
- FINGER, Almeri Paulo. **Gestão de Universidades: novas abordagens**, Champagnat, Curitiba, 1997.
- FIRESTONE, Joseph M. **Architectural Evolution in DataWarehousing and Distributed Knowledge Management Architecture**. Disponível em: <<http://www.dkms.com/ARCHEV.htm>>. Acesso em: 02 de fev. 2001.
- HARJINDER, G. e RAO, P. C. **The Official Guide to Data Warehousing**. Que Corporation, 1996.
- HARDY, Cynthia. **Gestão de Estratégica na universidade brasileira: teoria e casos**, Ed. Universidade/UFRGS, Porto Alegre, 1996.
- INMON, William H. **Como Construir o Data Warehouse**. Campus, Rio de Janeiro, 1997.
- INMON, William & WELCH, J.D., Glassey, Katherine I. **Gerenciando Data Warehouse**. Makron Books, 1999.
- KIMBALL, Ralph. **The Data Warehouse Toolkit**. John Wiley & Sons Inc., New York, 1996.
- KIMBALL, Ralph; REEVES Laura; ROSS Margy; THORNTHWAITE Warren. **The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses**. John Wiley & Sons Inc., New York, 1998.
- MORAES, Rodrigo Leal – Página com informações sobre os **Sistemas de Data Warehouse**. Disponível em: <<http://www.inf.ufrgs.br/~rlmoraes/dw.html>>. Acesso em: 15 ago.1999.
- DAVID, Marco. **Managing Metadata Today and into the 21st Century**. Disponível em: <<http://www.dbmsmag.com/9708d16.html>>. Acesso em: 29 set.1999.
- ONEIL, B. **Oracle Data Warehousing**. Indianapolis, Sams Publishing, 1997.
- PROENS. **Guia do Universitário 1999**. UNOESC - Chapecó, 1999.
- RIBEIRO, Nelson Figueiredo. **Administração acadêmica universitária: a teoria, o método de Clóvis C. da Gama Malcher**, Livros Técnicos e

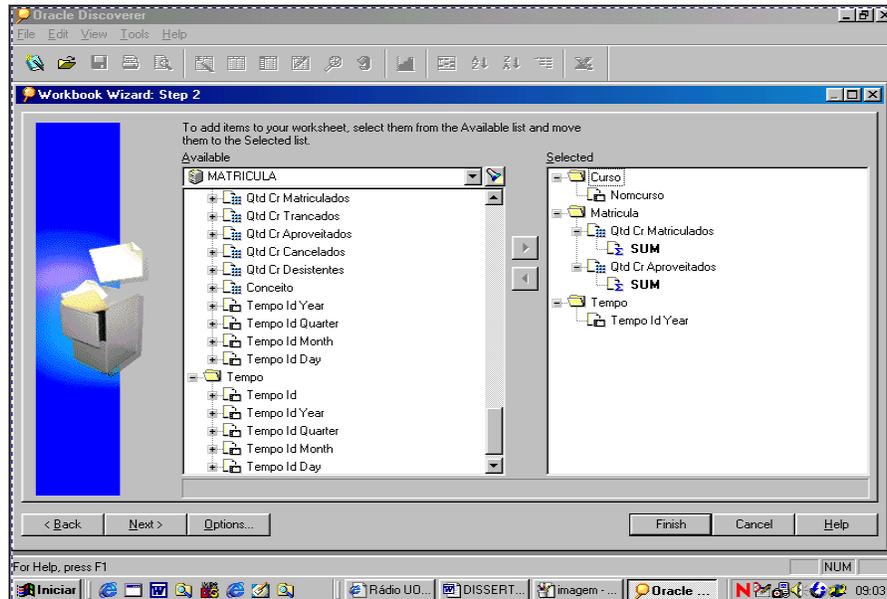
Científicos, Rio de Janeiro, 1977.

RIBEIRO, C. **Bancos de Dados Heterogêneos** - Mapeamento dos Esquemas Conceituais em um modelo Orientado a Objetos, Tese de Doutorado, Porto Alegre, CPGCC da UFRGS, 1995.

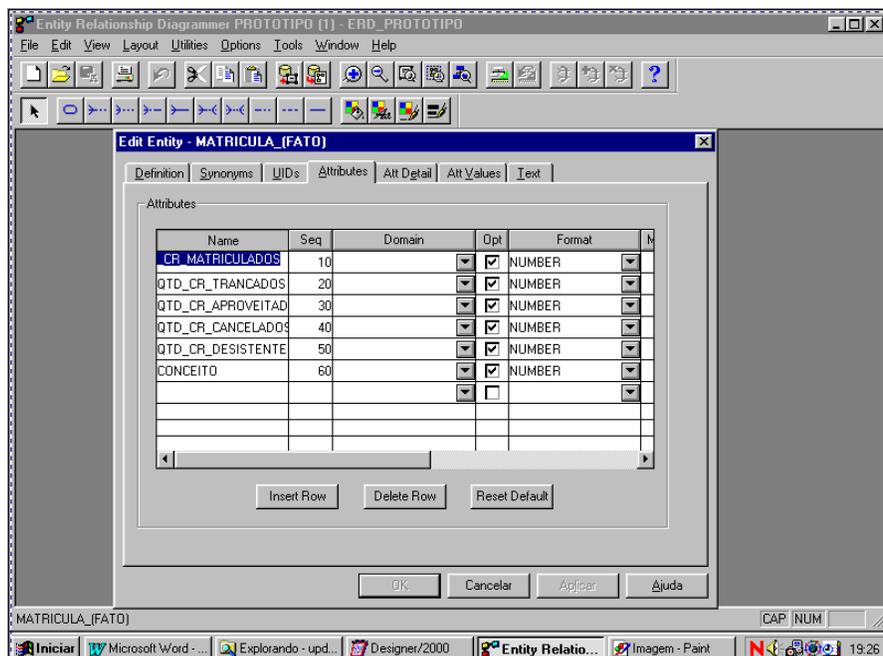
VALENTE, Daphnis Lopes. **Estudo sobre Armazém de Dados** CPGCC da UFRGS, Porto Alegre, 1996, 56 pág.

8 ANEXOS

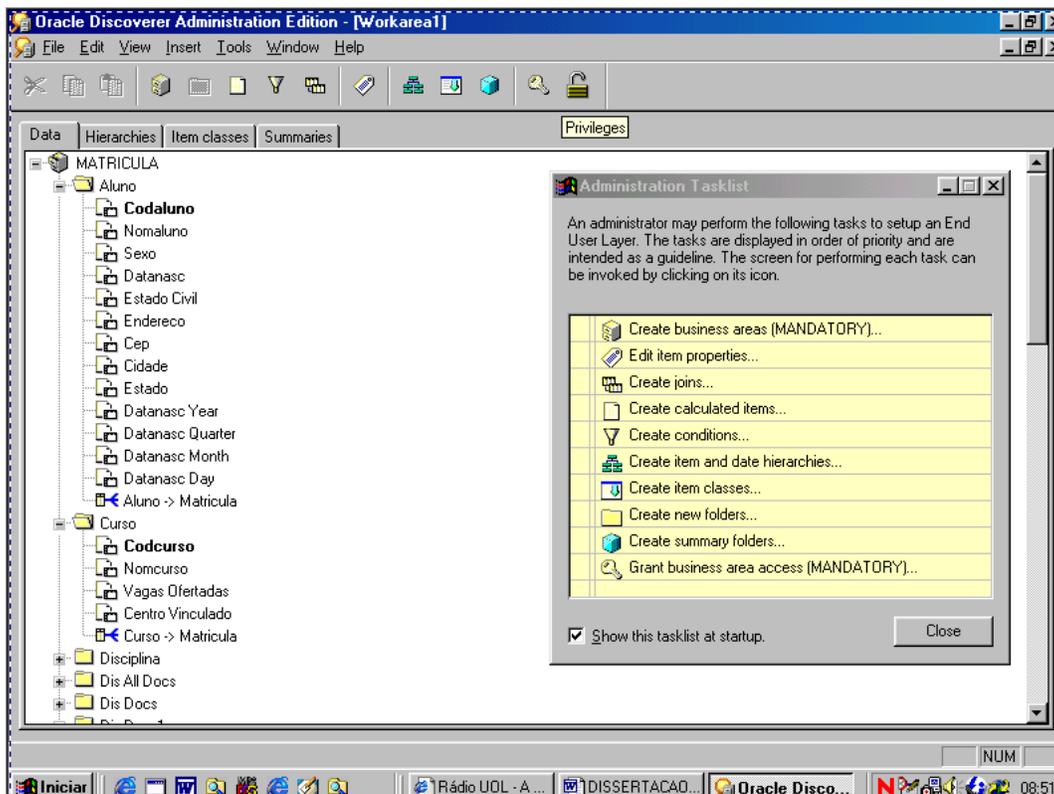
8.1 Anexo A: Interface do Oracle Discoverer for Users®



8.2 Anexo B: Definição dos Atributos das tabelas do DM



8.3 Anexo C: Interface do Oracle Discoverer Administration®



8.4 Anexo D: Scripts do Esquema do BD gerado para o DM

```
-- Generated for Oracle 8 on Wed Sep 20 15:24:14 2000 by Server Generator 2.1.19.5.0
PROMPT Creating Table 'DISCIPLINA'
CREATE TABLE DISCIPLINA
(CODDISC VARCHAR2(10) NOT NULL
,NOMDISC VARCHAR2(80) NOT NULL
,AREA_CONHECIMENTO VARCHAR2(50)
,CARGA_HORARIA NUMBER(3,0) DEFAULT 0 NOT NULL
,TOTAL_CREDITOS NUMBER(4,0) DEFAULT 0 NOT NULL
)
COMMENT ON COLUMN DISCIPLINA.CODDISC IS 'Código da Disciplina'
COMMENT ON COLUMN DISCIPLINA.NOMDISC IS 'Nome da Disciplina'
COMMENT ON COLUMN DISCIPLINA.AREA_CONHECIMENTO IS 'Area do Conhecimento'
COMMENT ON COLUMN DISCIPLINA.CARGA_HORARIA IS 'Carga Horária da Disciplina'
COMMENT ON COLUMN DISCIPLINA.TOTAL_CREDITOS IS 'Total de Creditos da Disciplina'
PROMPT Creating Table 'TEMPO'
CREATE TABLE TEMPO
(TEMPO_ID DATE DEFAULT '0' NOT NULL
)
COMMENT ON COLUMN TEMPO.TEMPO_ID IS 'Identificador de Tempo'
PROMPT Creating Table 'ALUNO'
CREATE TABLE ALUNO
(CODALUNO NUMBER(12,0) DEFAULT 0 NOT NULL
,NOMALUNO VARCHAR2(80) NOT NULL
,SEXO VARCHAR2(15) NOT NULL
,DATANASC DATE NOT NULL
```

```

,ESTADO_CIVIL VARCHAR2(20) NOT NULL
,ENDERECO VARCHAR2(100)
,CEP NUMBER(8,0) DEFAULT 0 NOT NULL
,CIDADE VARCHAR2(80) NOT NULL
,ESTADO VARCHAR2(1) NOT NULL
)
COMMENT ON COLUMN ALUNO.CODALUNO IS 'Código do Academico'
COMMENT ON COLUMN ALUNO.NOMALUNO IS 'Nome do Academico'
COMMENT ON COLUMN ALUNO.SEXO IS 'Sexo do Academico'
COMMENT ON COLUMN ALUNO.DATANASC IS 'Data de Nascimento'
COMMENT ON COLUMN ALUNO.ESTADO_CIVIL IS 'Estado Civil'
COMMENT ON COLUMN ALUNO.ENDERECO IS 'Endereco'
COMMENT ON COLUMN ALUNO.CEP IS 'CEP'
COMMENT ON COLUMN ALUNO.CIDADE IS 'Cidade'
COMMENT ON COLUMN ALUNO.ESTADO IS 'Estado'
PROMPT Creating Table 'CURSO'
CREATE TABLE CURSO
(CODCURSO NUMBER(8,0) DEFAULT 0 NOT NULL
,NOMCURSO VARCHAR2(80) NOT NULL
,VAGAS_OFERTADAS NUMBER(3,0) DEFAULT 0 NOT NULL
,CENTRO_VINCULADO VARCHAR2(80) NOT NULL
)
COMMENT ON COLUMN CURSO.CODCURSO IS 'Código do Curso'
COMMENT ON COLUMN CURSO.NOMCURSO IS 'Nome do Curso'
COMMENT ON COLUMN CURSO.VAGAS_OFERTADAS IS 'Vagas Ofertadas'
COMMENT ON COLUMN CURSO.CENTRO_VINCULADO IS 'Centro Vinculado'
PROMPT Creating Table 'MATRICULA'
CREATE TABLE MATRICULA
(CODALUNO NUMBER(12,0) NOT NULL
,CODCURSO NUMBER(8,0) NOT NULL
,CODDISC VARCHAR2(10) NOT NULL
,TEMPO_ID DATE NOT NULL
,QTD_CR_MATRICULADOS NUMBER(12,2) DEFAULT 0
,QTD_CR_TRANCADOS NUMBER(12,2) DEFAULT 0
,QTD_CR_APROVEITADOS NUMBER(12,2) DEFAULT 0
,QTD_CR_CANCELADOS NUMBER(12,2) DEFAULT 0
)
COMMENT ON COLUMN MATRICULA.CODALUNO IS 'Código do Academico'
COMMENT ON COLUMN MATRICULA.CODCURSO IS 'Código do Curso'
COMMENT ON COLUMN MATRICULA.CODDISC IS 'Código da Disciplina'
COMMENT ON COLUMN MATRICULA.TEMPO_ID IS 'Identificador de Tempo'
COMMENT ON COLUMN MATRICULA.QTD_CR_MATRICULADOS IS 'Quantidade de Creditos
Matriculados'
COMMENT ON COLUMN MATRICULA.QTD_CR_TRANCADOS IS 'Quantidade de Créditos Cancelados'
COMMENT ON COLUMN MATRICULA.QTD_CR_APROVEITADOS IS 'Quantidade de Creditos
Aproveitados'
COMMENT ON COLUMN MATRICULA.QTD_CR_CANCELADOS IS 'Quantidade de Creditos
Cancelados'

```

8.5 Anexo E: Scripts de Criação Primary Key, Foreign Key e Constraint

-- Generated for Oracle 8 on Wed Sep 20 15:24:14 2000 by Server Generator 2.1.19.5.0

```

PROMPT Creating Primary Key on 'DISCIPLINA'
ALTER TABLE DISCIPLINA
ADD CONSTRAINT DIS_PK PRIMARY KEY
(CODDISC)
/

```

```

PROMPT Creating Primary Key on 'TEMPO'
ALTER TABLE TEMPO
ADD CONSTRAINT TEM_PK PRIMARY KEY

```

```
(TEMPO_ID)  
/
```

```
PROMPT Creating Primary Key on 'ALUNO'  
ALTER TABLE ALUNO  
ADD CONSTRAINT ALU_PK PRIMARY KEY  
(CODALUNO)  
/
```

```
PROMPT Creating Primary Key on 'CURSO'  
ALTER TABLE CORSO  
ADD CONSTRAINT CUR_PK PRIMARY KEY  
(CODCURSO)  
/
```

```
PROMPT Creating Primary Key on 'MATRICULA'  
ALTER TABLE MATRICULA  
ADD CONSTRAINT MAT_PK PRIMARY KEY  
(CODALUNO  
,CODCURSO  
,TEMPO_ID  
,CODDISC)  
/
```

```
PROMPT Creating Foreign Keys on 'MATRICULA'  
ALTER TABLE MATRICULA ADD CONSTRAINT  
MAT_ALU_FK FOREIGN KEY  
(CODALUNO) REFERENCES ALUNO  
(CODALUNO) ADD CONSTRAINT  
MAT_CUR_FK FOREIGN KEY  
(CODCURSO) REFERENCES CORSO  
(CODCURSO) ADD CONSTRAINT  
MAT_TEM_FK FOREIGN KEY  
(TEMPO_ID) REFERENCES TEMPO  
(TEMPO_ID) ADD CONSTRAINT  
MAT_DIS_FK FOREIGN KEY  
(CODDISC) REFERENCES DISCIPLINA  
(CODDISC)  
/
```