

Data Warehouse

Cursos: Ciência da Computação e
Informática

Disciplina: Banco de Dados II
Profa. Dra. Maria Madalena Dias

1

Tópicos

- Introdução
- Evolução Histórica
- Objetivo do DW
- Vantagens do DW
- Características da Tecnologia de DW
- Diferenças entre OLTP e DW
- Granularidade
- Data Mart
- Modelagem Dimensional
- Considerações Gerais
- Bibliografia

2

Introdução

DW surgiu no final da década de 70.

DW é o produto da evolução de:

- Banco de dados cliente/servidor;
- Sistemas operacionais de rede;
- Discos rígidos de grande capacidade;
- Interface gráfica;
- Computadores de alto desempenho;
- Processamento paralelo;
- OLAP;
- Data Mining (mineração de dados).

3

Introdução

- Um data warehouse é uma coleção de dados orientada por assuntos, integrada, variante no tempo, e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão.”
(Inmon, 1992)

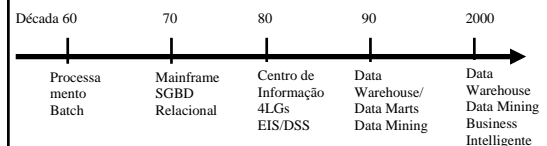
4

Introdução

- O DW contém informações históricas e resumos de qualquer época;
- Fornece informações para SAD;
- Permite uma visão geral do negócio;
- Facilita a análise de tendências de mercado;
- Permite identificar relacionamentos desconhecidos.

5

Evolução Histórica



Baseado em
(Badijeiro, 1996)

6

Década de 60

- Processamento Batch

7

Década de 70

- Mainframes com pouca capacidade de armazenamento e processamento dos dados
- SGBD relacional – promessa de ser capaz de analisar milhões de dados históricos
- Teoria da normalização como uma teoria e metodologia para estruturação de dados para aplicações OLTP
- DW como um simples armazém de dados

8

Década de 80

- Surgimento das 4GLs, dos Centros de Informações e das tecnologias DSS e EIS
- Limitações da teoria da normalização na análise dimensional para suporte à decisão
- Introdução do esquema estrela
- DSS eram desenvolvidos tendo como base os dados operacionais

9

Década de 80

- Tomada de decisão: arte pura - tentativas e erros
- Aumento do uso de computadores nos níveis operacionais, enquanto o alto escalão distanciava-se da TI
- Surgimento de algoritmos *data mining*
- Estruturas de dados desnormalizadas e super-dimensionadas na *data mining*

10

Década de 90

- Primeiro Data Warehouse: RedBrick Warehouse VPT da Red Brick Systems
- Gerações de modelos DSS:
 - Primeira geração: tecnologia mainframe suportando DSSs e sistemas de produção, provou ser impraticável
 - Segunda geração: impraticável, SIs corporativos consolidam dados em sistemas únicos e grandes
 - Terceira geração: papel corporativo do data warehouse e data marts, e o tempo limitado para construir este modelo híbrido DSS

11

Década de 90

- Diferenças arquitetônicas nas bases de dados para data warehouse e para OLTP
- DW agora tem uma missão: gerenciar os dados que analisam o negócio da empresa
- Anúncio ou lançamento de novos produtos sobre modelos de DSS
- A metodologia de esquema estrela ganha aceitação difundida para data warehousing

12

Década de 90

- Tecnologias especializadas tanto para ambientes OLTP quanto para DW
- Necessidade de suporte a grandes quantidades de dados, centenas de gigabytes e dezenas de terabytes
- DWMS (Sistemas de Gerenciamento de DW), que contempla três tipos de dados (Temporal, Relacional e Multidimensional)

13

Década de 90

- Critérios de avaliação de tecnologias DW:
 - capacidade de armazenamento
 - carga e desempenho de indexação
 - integridade operacional, confiabilidade e capacidade de gerenciamento
 - conectividade cliente/servidor
 - desempenho de processamento de query

14

Década de 90

- Novas ferramentas de desenvolvimento Web que permitem browsers Web para acessar data warehouses
- Data marts distribuídos substituindo uma única base de dados de alta manutenção
- Ferramentas automáticas para manter data marts atuais e consistentes
- Aumento de atividades de pesquisa, com aplicações em suporte a decisão, OLAP, otimização de query e replicação

15

Década de 90

- Abordagem BOTTOM-UP
 - Metodologia na qual primeiro constrói-se um **data mart** e então expande-se gradualmente seu escopo e tamanho
- Custo e complexidade são duas razões para data marts serem tão populares

16

Década de 90

- Business Intelligence
 - Conjunto de conceitos e metodologias que visa ao apoio à tomada de decisões nos negócios a partir da transformação do dado em informação e da informação em conhecimento (Côrtes, 2002).
 - Objetiva usar os dados da organização para apoiar decisões bem informadas, facilitando o acesso e a análise de dados e possibilitando a descoberta de novas oportunidades (Almeida et al., 1999).
 - O foco do BI é facilitar o entendimento do negócio das organizações, fornecendo a todos os níveis informações relevantes sobre suas operações internas e o ambiente externo, incluindo clientes e competidores, parceiros e fornecedores. (Moss; Atre, 2003; Waltz, 2003).

17

Objetivos do DW

- O principal objetivo do DW é identificar novas oportunidades e melhorar o foco do negócio.
- Exemplos:
 - A venda de cerveja no fim de semana está relacionado com a venda de fraldas.
- CRM (Customer Relationship Management)
 - Identificar cliente
 - Diferenciar cliente
 - Interagir com o cliente
 - Personalizar o contato com o cliente

18

Objetivos do DW

- Deve fazer com que informações de uma empresa possam ser facilmente acessadas
- Deve apresentar as informações da empresa de modo consistente
- Deve ser adaptável e flexível a mudanças
- Deve ser um armazém de dados seguro que protege as informações
- Deve funcionar como a base para uma melhor tomada de decisão
- Para que um DW seja considerado de sucesso, a comunidade de negócio deve aceitá-lo.

19

Vantagens do DW

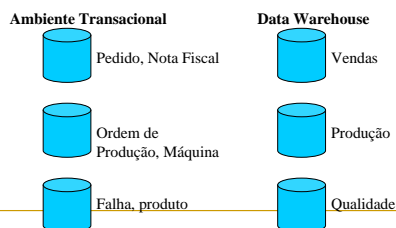
- permitem que sistemas mais antigos continuem em operação;
- consolidam dados inconsistentes dos sistemas mais antigos em conjuntos coerentes;
- extraem benefícios de novas informações oriundas das operações correntes;
- provêm ambiente para o planejamento e arquitetura de novos sistemas de cunho operacional.

20

Características da Tecnologia de DW

➢ Orientação por assunto:

- Um DW armazena as informações agrupadas por assunto de interesse da empresa que são mais importantes.



21

Características da Tecnologia de DW

➢ Variação de Tempo

- Os dados são um "snapshot" (instantâneo);
- Não necessita mudanças diárias.

➢ Não Volátil

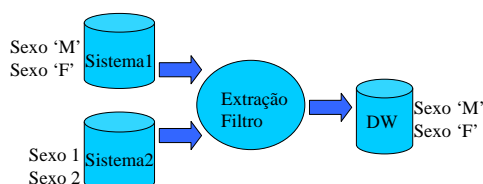
- Inclusão e consulta;
- Dados são filtrados, limpos e transformados;
- Não existe problemas como *deadlock*.

22

Características da Tecnologia de DW

➢ Integração

- Necessidade de ter uma convenção de nome e valores entre diferentes base de dados;



23

Diferenças entre OLTP e DW

| | OLTP | DW |
|---------------------|-----------------------|--------------------------------------|
| Propósito | Operações Cotidianas | Recuperação de informações e análise |
| Foco | Registros Individuais | Milhões de Registros |
| Tipos de dados | Que administram | Que analisam |
| Modelo Dado | Normalizado | Multidimensional/desnormalizado |
| Modelo Processo | Customizado | Padrão |
| Acesso à Informação | SQL | SQL + extensões de análise de dados |
| Dados | Inseridos/Atualizados | Inseridos/Incrementados |

24

Diferenças entre OLTP e DW

- Banco de dados para OLTP inapropriado para DW processamento de transações (alta capacidade)
 - muitas tabelas
 - constantes mudanças
 - consultas rápidas
 - entradas duplicadas
 - dados ausentes
 - não bloquear os usuários, etc.
- Data Warehouse
 - processamento de consultas
 - fatos e dados históricos
 - análises comerciais

25

Granularidade

- nível de detalhamento das unidades de dados tempo:
 - segundos
 - minutos
 - horas
 - dias, meses, anos, décadas,...
- maior nível de detalhe
 - menor granularidade
 - maior volume de dados
 - maior tempo de processamento
 - maior complexidade de consultas

26

Granularidade

- Estimativa bruta
 - Identificar todas as tabelas que serão criadas
 - Estimar o tamanho da linha de cada tabela
 - Calcular, para o horizonte de um ano, o número máximo e o número mínimo de linhas da tabela
 - Repetir o processo de estimativa para cinco anos.
- Quais serão os níveis de granularidade
 - O método normal para a escolha dos níveis de granularidade consiste em usar bom senso, criar uma pequena parte do warehouse e deixar o usuário acessar os dados e dar o feedback.

27

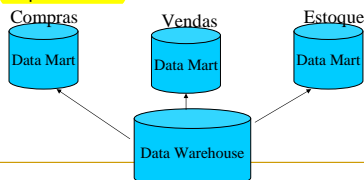
Data Mart

- Representa um subconjunto de dados do DW
- Os dados do DM são direcionados a um departamento ou a uma área específica do negócio
- Custo mais baixo e um tempo menor de implementação

28

Data Mart

- Diferenças entre DW e DM:
 - volume de dados
 - complexidade de carga
 - DM são voltados somente para uma determinada área, já o DW é voltado para os assuntos da empresa toda.



29

Modelagem Dimensional

- Modelagem dimensional é o nome da técnica de design lógica freqüentemente usada para DW e permite alto desempenho de acesso
- Todo modelo dimensional é composto de uma tabela com uma chave composta, chamada de tabela de fatos e um conjunto de tabelas menores chamadas de tabelas de dimensões. Cada tabela de dimensão tem uma única chave primária que corresponde exatamente um dos componentes da chave composta da tabela de fatos.

30

Modelagem Dimensional

- Fato é tudo aquilo que reflete a evolução dos negócios do dia-a-dia de uma organização.
- Um fato é uma coleção de dados. Cada fato representa um item de negócio, uma transação ou um evento de negócio, e são os fatos que são utilizados para fazer a análise sobre a empresa, ou a instituição.
- O segredo para obter a flexibilidade das consultas é criar as tabelas de fatos no nível mais granular possível.

31

Modelagem Dimensional

- Em uma tabela de fatos, uma linha corresponde a uma medição. Todas as medições em uma tabela de fatos devem estar alinhadas na mesma granularidade.
- Os fatos mais úteis em uma tabela de fatos são numéricos e aditivos.
- As tabelas de fatos expressam relações de muitos-para-muitos em modelos dimensionais.

32

Modelagem Dimensional

- Dimensões são os elementos que participam de um fato, como por exemplo:
 - Tempo;
 - Localização;
 - Cliente, etc..
- A dimensão pode ser organizada de maneira hierárquica, sendo constituída de vários níveis. Por exemplo, a dimensão tempo pode ser constituída de ano, trimestre, mês e dia.

33

Modelagem Dimensional

Dimensionalidade do DW

Qual a tendência em termos de volume de vendas para o ano que vem dos modelos Fiesta verde ?

Vendas X modelos

Vendas X cor

Vendas X fabricante

Vendas X período de tempo

Dimensões:

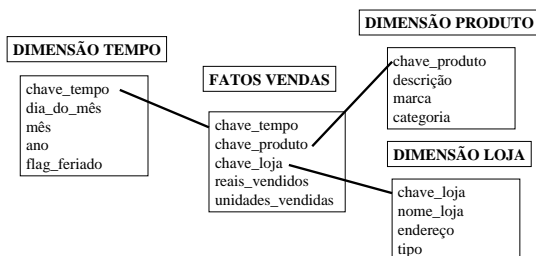
diferentes perspectivas (campos não numéricos)

Medidas: campos numéricos

34

Modelagem Dimensional

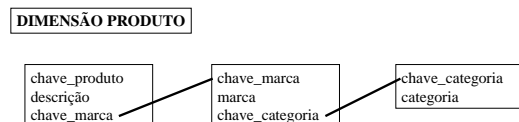
Modelo estrela



35

Modelagem Dimensional

Modelo Snowflake



36

Metadados

- Todas as fases de um projeto de Data Warehouse, desde a modelagem até a visualização da informação, geram metadados.
- Neles estarão contidas informações como atributos das tabelas, agregadas utilizadas, cálculos necessários, descrições, periodicidade das cargas, histórico de mudanças etc..
- Os metadados mantêm informações sobre "o que está e onde," no DW.

37

Metadados

- São dados sobre os dados ou sobre a estrutura dos dados;
- Funcionam também como documentação
- Podem ser divididos em dois conjuntos :
 - Integração de dados
 - Transformação

38

Metadados

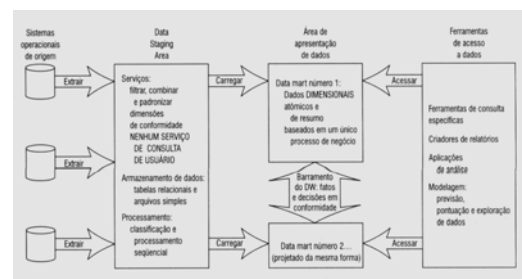
Tipicamente, os aspectos sobre os quais os metadados mantêm informações são:

- A estrutura dos dados, segundo a visão do programador;
- A estrutura dos dados, segundo a visão dos analistas;
- A fonte de dados que alimenta o DW;
- As transformações sofridas pelos dados no momento de sua migração para o DW;
- O modelo de dados;
- O relacionamento entre o modelo de dados e o DW;
- O histórico das extrações de dados.

39

Arquitetura de DW

Componentes de um DW

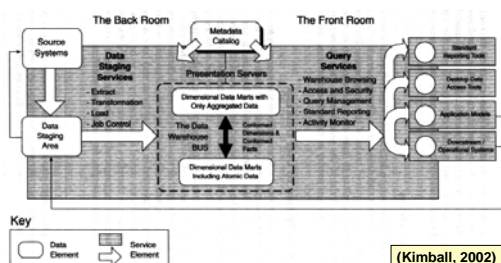


(Kimball, 1998)

40

Arquitetura de DW

Modelo da Arquitetura Técnica de Alto Nível



41

Arquitetura de DW

Uma Estrutura Arquitetural

| Nível de Detalhe | Dados (o que?) | Técnica Back Room | Front Room (como?) | Infraestrutura (Onde?) |
|---------------------------------|----------------|-------------------|--------------------|------------------------|
| Requisitos de Negócio | | | | |
| Modelos e Doctos da Arquitetura | | | | |
| Modelos Detalhados | | | | |
| Implementação | | | | |

42

Arquitetura de DW

A Área de Arquitetura de Dados

■ Requisitos de negócio

- ❑ Qual informação nós necessitamos para tomar melhores decisões de negócio?
- ❑ Inclui os conteúdos do DW – a lista de dados que é importante ao negócio, os depósitos de dados de todo o ambiente do DW e as fontes que os alimentam

■ Modelos e documentos da arquitetura

- ❑ Modelo Dimensional:
 - Quais são as melhores entidades (fatos e dimensões) que constituem esta informação e como se relacionam?
 - Como deveriam essas entidades serem estruturadas?

43

Arquitetura de DW

A Área de Arquitetura de Dados

■ Modelos Detalhados

■ Os modelos lógico e físico

- Quais são os elementos individuais, suas definições, domínios e regras para derivação?
- Quais são as fontes e como são mapeadas para os objetivos?

■ Implementação

- ❑ Criar as bases de dados, índices, backup, etc.

44

Arquitetura de DW

A Área de Arquitetura de Técnica

- Cobre os processos e ferramentas que nós aplicamos aos dados

- Responde a questão “como” - como nós buscamos os dados de suas fontes, colocamos eles na forma que encontrem os requisitos de negócio e os movemos para um lugar acessível?

45

Arquitetura de DW

A Área de Arquitetura de Técnica

Back Room

- Requisitos de negócio
 - ❑ Como nós buscamos os dados, transformamos e os tornamos disponíveis aos usuários?
 - ❑ Como isto é feito hoje?

Front Room

- Requisitos de negócio
 - ❑ Quais são as maiores questões de negócio?
 - ❑ Como nós medimos essas questões?
 - ❑ Como nós queremos analisar os dados?

46

Arquitetura de DW

A Área de Arquitetura de Técnica

Back Room

■ Modelos e docs. da arquitetura

- ❑ Quais são as capacidades específicas que nós necessitaremos para buscar dados, de forma usual, nas localizações desejadas, nos tempos apropriados?
- ❑ Quais são os maiores armazéns de dados e onde eles deveriam estar localizados?

Front Room

■ Modelos e docs. da arquitetura

- ❑ Quais usuários necessitarão buscar a informação de forma usual?
- ❑ Quais as maiores classes de análise e de apresentação nós necessitamos fornecer e quais são as prioridades?

47

Arquitetura de DW

A Área de Arquitetura de Técnica

Back Room

■ Modelos detalhados

- ❑ Quais padrões e produtos fornecem as capacidades necessárias?
- ❑ Como eles podem ser presos juntos?
- ❑ Quais são nossos padrões de desenvolvimento para gerenciamento de código, nomeação, etc.?

Front Room

■ Modelos detalhados

- ❑ Quais são as especificações para os templates de relatório, incluindo linhas, colunas, seções, cabeçalhos, filtros, etc.?
- ❑ Quem necessita deles?
- ❑ Com qual frequência?
- ❑ Como nós os distribuiremos?

48

Arquitetura de DW

A Área de Arquitetura de Técnica

- | Back Room | Front Room |
|--|---|
| ■ Implementação <ul style="list-style-type: none">□ Escrever as extrações e cargas□ Automatizar o processo□ Documentar | ■ Implementação <ul style="list-style-type: none">□ Implementar o relatório e ambiente de análise, construir o conjunto de relatório inicial e treinar os usuários□ Documentar |

49

Arquitetura de DW

A Área de Arquitetura de Infraestrutura

- Requisitos de negócio
 - Quais capacidades em nível de hardware e de sistema nós necessitamos?
 - O que nós temos atualmente?
- Modelos e documentos da arquitetura
 - De onde vêm e para onde vão os dados?
 - Nós temos capacidade suficiente de cálculo e armazenagem?
 - Quais são as capacidades específicas que nós estamos contando? Elas existem? Quem é responsável por elas?

50

Arquitetura de DW

A Área de Arquitetura de Infraestrutura

- Modelos detalhados
 - Como nós interagimos com essas capacidades?
 - Quais são as utilidades do sistema, chamadas, APIs, etc.?
- Implementação
 - Instalar e testar novos componentes da infraestrutura
 - Conectar as fontes aos objetivos do *desktop*
 - Documentar.

51

Arquitetura de DW

A Arquitetura Bus

- O DW é planejado como uma arquitetura de dados que tem várias metas específicas e bem estabelecidas
- O objetivo é produzir um conjunto mestre de dimensões e padronizar as definições de fatos
- Os *data marts* são implementados separados passo-a-passo, por diferentes grupos em diferentes momentos
- A fase de arquitetura de dados produz linhas guias específicas que as equipes de desenvolvimento dos *data marts* podem seguir, de forma independente e assíncrona

52

Arquitetura de DW

A Arquitetura Bus

- A maior tarefa na construção do DW é projetar o sistema de extração que transfere os dados de um sistema legado específico para uma área de estágio onde eles podem ser transformados para dar origem à base de dados de consulta
- O sistema de extração deve ser projetado para cada *data mart*

53

Arquitetura de DW

A Arquitetura Bus

- Um bus em seu computador é uma interface padrão que permite que muitos diferentes tipos de dispositivos sejam conectados ao seu computador
- A arquitetura bus possibilita a inclusão de um novo *data mart*, porque o barramento oferece uma cadeia de dimensões padrão e fatos paralelos que aumenta a expressividade do novo *data mart*.

54

Arquitetura de DW

A Arquitetura Bus

■ Matriz de barramento do DW

- É usada para criar, documentar e comunicar a arquitetura de barramento.
- As linhas da matriz de barramento correspondem a *data marts*. Você deve criar linhas de matriz separadas se as origens forem diferentes, os processos forem diferentes ou a linha da matriz representar mais do que o que realmente pode ser representado em uma iteração de uma única implementação.

55

Arquitetura de DW

A Arquitetura Bus

■ Matriz de barramento do DW

- As colunas da matriz representam as dimensões comuns usadas na empresa. Normalmente, é útil criar uma lista completa de dimensões antes de preencher a matriz.
- A criação da matriz de barramento do DW é um dos produtos genéricos preliminares mais importantes de uma implementação de DW. É um recurso híbrido que faz parte da ferramenta de design técnico, da ferramenta de gerenciamento de projeto e da ferramenta de comunicação.

56

Arquitetura de DW

A Arquitetura Bus

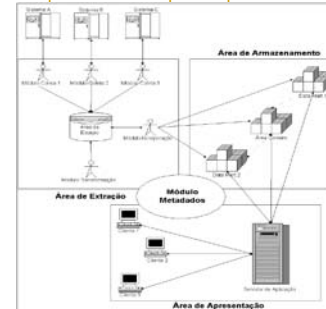
■ Matriz de barramento do DW

| PROCESSOS DE NEGÓCIO | DIMENSÕES COMUNS | | | | | | | |
|-----------------------|------------------|---------|------|----------|-----------|------------|----------|----------------|
| | Data | Produto | Loja | Promoção | Warehouse | Fornecedor | Contrato | Transportadora |
| Vendas no varejo | X | X | X | X | | | | |
| Estoque no varejo | X | X | X | | | | | |
| Entregas no varejo | X | X | X | | | | | |
| Estoque do warehouse | X | X | | | X | X | | |
| Entregas no warehouse | X | X | | | X | X | | |
| Ordens de compra | X | X | | | X | X | X | X |

57

Arquitetura de DW

A Arquitetura Proposta por Sell

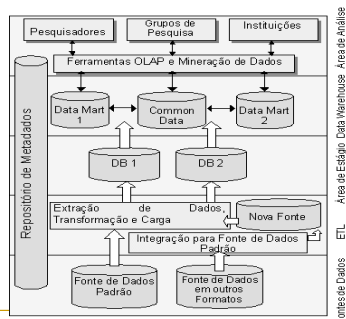


(Sell, 1998)

58

Arquitetura de DW

Uma Arquitetura para C&T



(Menolli, 2004)

59

Considerações Gerais

- Em geral, pessoas são resistentes a mudanças, o data warehouse representa uma quantidade significativa de mudanças.
- Na Emory University em Atlanta, departamentos resistiram ao DW por causa de questões sobre quem poderia controlar os dados, como os dados poderiam ser compartilhados e segurança dos dados.

60

Considerações Gerais

- Gerenciar expectativas do usuário pode ser o maior desafio na implementação de um DW.
- Necessidade de implantação imediata de data mart para alguma área emergente, considerando tempo, investimento inicial e retorno de investimento.
- Necessita de ferramentas de extração e conversão.

61

Considerações Gerais

- Necessita de ferramentas inteligentes de prospecção e análise de dados.
- Necessita de ferramentas de administração e gerenciamento.
- DW melhora a eficácia e a disponibilidade dos dados.
- DW reduz o tempo de tomada de decisões de negócio e a quantidade de recursos para a construção e manutenção de relatórios.

62

Bibliografia

- DIAS, M.M.; MATTOS, M.M.; ROMÃO, W.; TODESCO, J.L.; PACHECO, R.C.S. DIAS, M.M. et al. Data warehouse – presente e futuro. In: *Revista Tecnológica*, No. 7, 1998, p. 59-73.
- INMON, W.H. Como construir o Data Warehouse. Editora Campus. 1997.
- KIMBALL, R.; Reeves, L.; Ross, M.; Thornthwaite, W. The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses, John Wiley & Sons Inc., New York, United States of America, 1998.
- KIMBALL, R.; Ross, M. The Data Warehouse Toolkit. Editora Campus, 2002.

63

Bibliografia

- MACHADO, F. N. R. Projeto de Data Warehouse – Uma visão Multidimensional. Érica, 2000.
- MENOLLI, A. L. A. Definição de uma Arquitetura de Data Warehouse para Gestão em Ciência e Tecnologia no Brasil. Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Estadual de Maringá, 2004.
- SELL, D. Uma Arquitetura para Distribuição de Componentes Tecnológicos de Sistemas de Informações Baseadas em Data Warehouse. Programa de Pós-Graduação em Engenharia de Produção, Florianópolis – SC. Dissertação de Mestrado, UFSC, Florianópolis, 2001.

64