

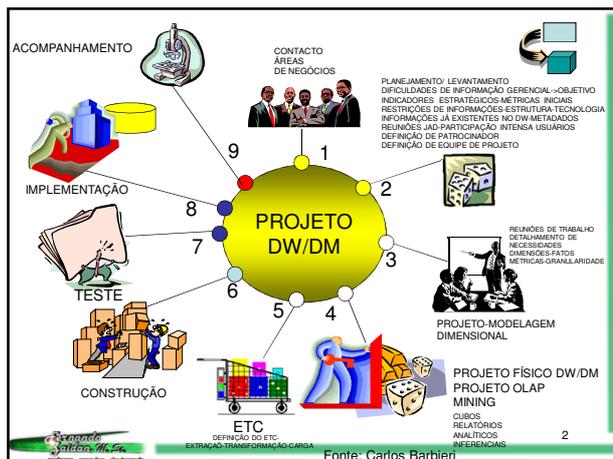
CEFET-MG

PÓS-GRADUAÇÃO LATO SENSU

Curso: Banco de Dados

Disciplina: Data Warehouse e Business Intelligence
Professor: Fernando Zaidan

Unidade 2
2012

LEVANTAMENTO DE NECESSIDADES

- **DEFINIR AS NECESSIDADES DE INFORMAÇÃO PARA O NEGÓCIO:**
 - DIFICULDADES-RESTRITÇÕES SUGEREM OBJETIVOS
 - INDICADORES,MÉTRICAS,COMPARATIVOS
 - SEMPRE OBSERVAR "DIMENSÕES"/COMBINAÇÕES
 - MODELO DE DADOS EXISTENTES:
 - ENTIDADES-RELACIONAMENTOS-ATRIBUTOS-(KEYS-ORIGINAIS-DERIVADOS)-PROPORÇÕES
 - ARQUIVOS, DADOS MANUAIS EXISTENTES



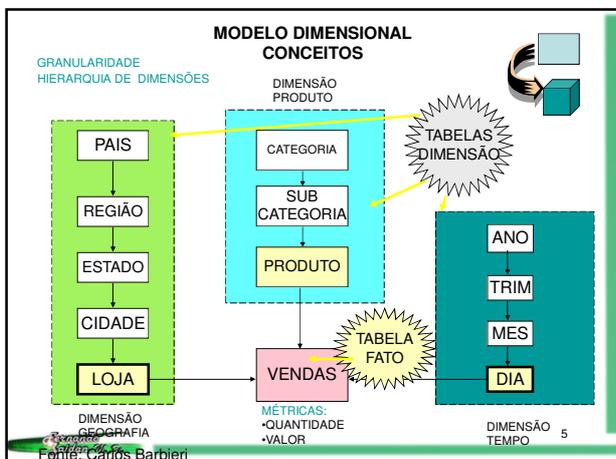
3

MODELAGEM DIMENSIONAL

- **TABELAS FATOS**
 - MÉTRICAS E VALORES
- **TABELAS DIMENSÃO**
 - TEXTOS, CAMPOS DIVERSOS



4

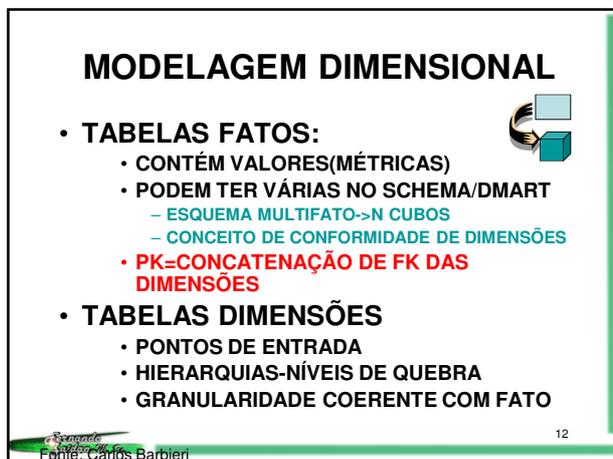
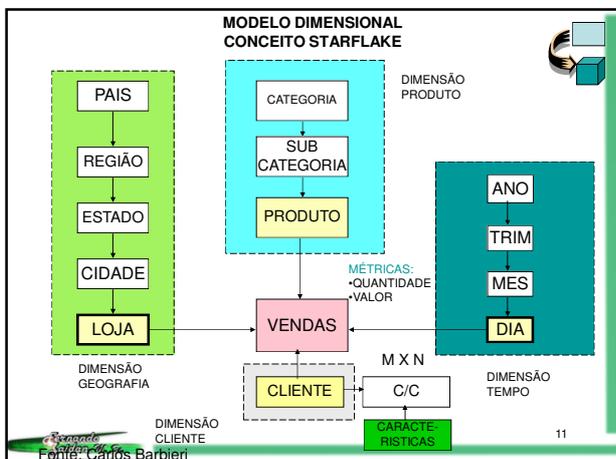
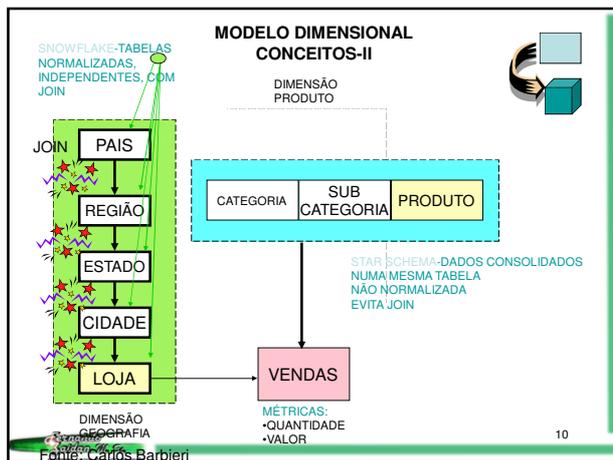
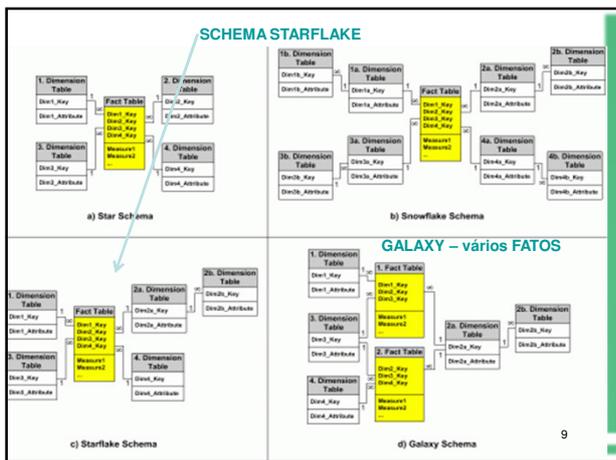
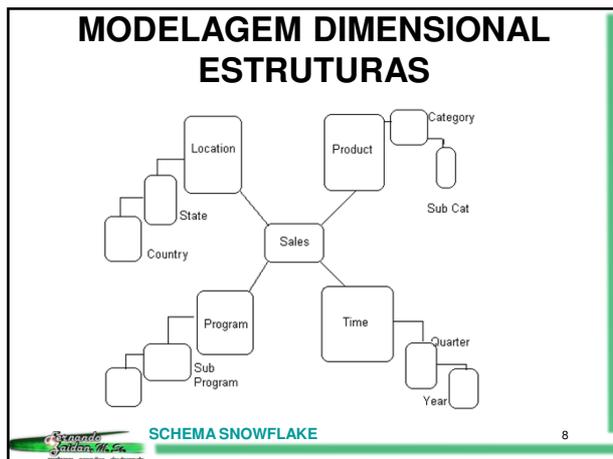
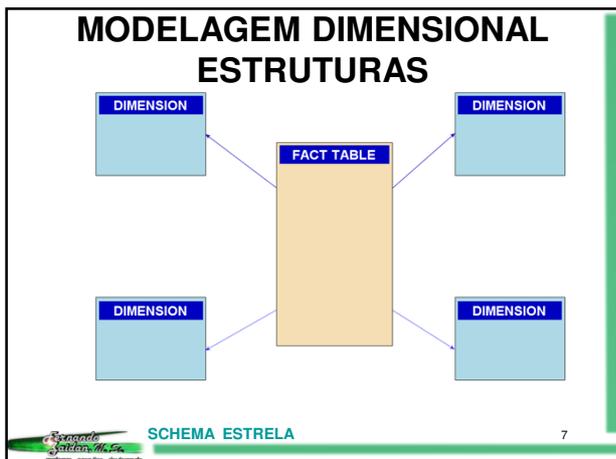


MODELAGEM DIMENSIONAL ESTRUTURAS

- **SCHEMA ESTRELA:**
 - DIMENSÕES DESNORMALIZADAS
 - VOLTADO PARA ACESSOS C/ PERFORMANCE
 - HIERARQUIAS ACHATADAS
- **SCHEMA SNOWFLAKE (flocos de neve):**
 - TRADICIONAL+ E/R(TABELAS EM CASCATA)
 - NORMALIZADO
 - HIERAQUIAS MANTIDAS
 - MUITAS TABELAS-->MUITAS JUNÇÕES-1:N
- **SCHEMA STARFLAKE**
 - COMBINAÇÃO DAS DUAS
 - DIMENSÕES COM M X N COM OUTRAS TABELAS



6

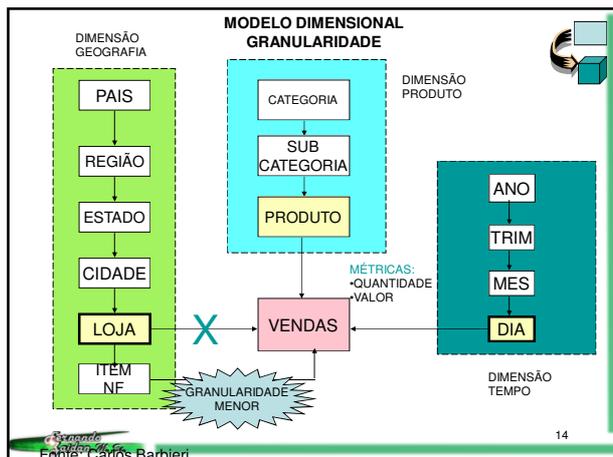


GRANULARIDADE DE FATOS E DADOS

- NÍVEL ATÔMICO DE DADOS NA(S) ENTIDADE(S)/TABELA(S) FATO/DIMENSÃO
- POSSIBILIDADES:
 - NÍVEL DE TRANSAÇÕES(DOCUMENTO-NF)
 - NÍVEL DE ÍTEM DE UM DOCUMENTO (NF, OC, OEXPEDIÇÃO, APÓLICE)
 - TEMPO:
 - NÍVEL DIÁRIO
 - NÍVEL SEMANAL
 - NÍVEL MENSAL, ETC
 - CONSIDERAÇÕES:
 - VOLUMES DE DADOS
 - NECESSIDADE DE INFORMAÇÕES P/ NEGÓCIO
 - DISPONIBILIDADE DO DADO FONTE

13

Fonte: Carlos Barbieri



DIMENSÕES

- PONTOS DE ENTRADAS DA ESTRUTURA
- DIMENSÕES E SEUS ATRIBUTOS SERVM TAMBÉM COMO FILTROS E COMO **HEADER** DOS RELATÓRIOS
- DIMENSÕES TÍPICAS:
 - PRODUTO/SERVIÇO-O QUE VENDEO
 - CLIENTE-QUEM COMPRA
 - TEMPO-QUANDO FOI FEITO A COMPRA
 - LOCAL(ARMAZÉM,LOJA,ETC)-ONDE
 - STATUS, PROMOÇÕES-CONDIÇÕES DA COMPRA
- DESCREVER TODOS OS ATRIBUTOS DAS DIMENSÕES
- DEVEM SER ATRIBUTOS DESCRITIVOS SEM CAMPOS NULOS
- NORMALMENTE UM DM TEM ENTRE 4-15 DIMENSÕES
 - MENOS=FALTOU OBSERVAÇÃO(TEMPO-ESPAÇO-TIPO)
 - MAIS-DIMENSÕES SUPÉRFLUAS
- SÃO OS DESCRITORES DAS TFATOS
- **CONCEITO DE SK** (surrogate Key – chave sequencial) - INDEPENDÊNCIA

15

Fonte: Carlos Barbieri

DIMENSÕES EM HIERARQUIAS

- DIMENSÕES NORMALMENTE TEM HIERARQUIAS
- HIERARQUIAS TEM NÍVEIS
- NÍVEIS TEM MEMBROS(MEMBERS)
- TIPOS DE HIERARQUIA-RELACIONAMENTOS 1:N
 - **BALANCEADA:** N DIFERENTE DE ZERO EM TODOS OS NÍVEIS-
 - EX: ANO->MÊS->DIA
 - **DESBALANCEADA:** N PODE SER ZERO
 - EX:ÓRGÃO->DIVISÃO(PODE TER ÓRGÃO SEM DIVISÃO)
 - **RAGGED:** UM DO NÍVEIS PODE NÃO TER MEMBROS
 - EX: PAIS-ESTADO-CIDADE-EM ISRAEL NÃO TEM ESTADO. EXISTE SOMENTE CIDADE E PAIS
 - ESTADO: ASSUME CHAVE DO PAIS OU BRANCO

16

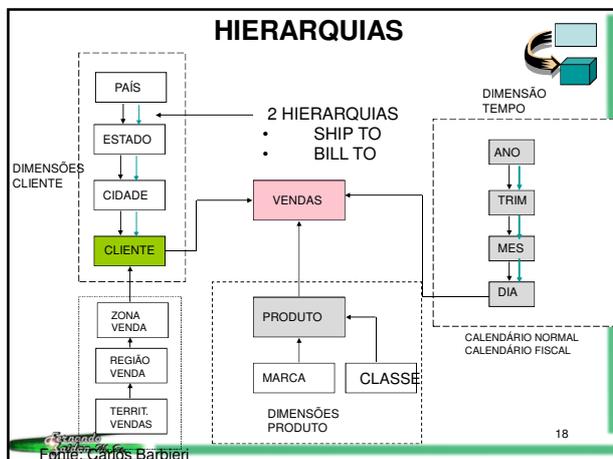
Fonte: Carlos Barbieri

DIMENSÕES EM HIERARQUIAS

- DIMENSÕES ESPECIAIS PODEM TER MÚLTIPLAS HIERARQUIAS. EXEMPLO: TEMPO
- **TEMPO CALENDÁRIO NORMAL**
 - ANO->TRIMESTRE->MÊS->DIA
 - COMEÇA EM JANEIRO
- **TEMPO CALENDÁRIO FISCAL**
 - ANO->TRIMESTRE->MÊS->DIA
 - COMEÇA EM ABRIL
- OS SERVIDORES OLAP TRATAM A DIMENSÃO TEMPO COMO ESPECIAL
 - PODEM SER OBTIDAS DIRETAMENTE DE UMA FONTE SIMPLES-CAMPO DATA DE UMA TABELA
 - PODEM SER OBTIDAS DE UMA TABELA FONTE-DIMENSÃO TEMPO BEM PROJETADA-COM DIA, FERIADOS,TAGS DE FIM DE SEMANA, ETC
- NORMALMENTE DEFINE-SE TEMPO COMO UMA DIMENSÃO A SER COMPARTILHADA COM OS CUBOS DO DMART

17

Fonte: Carlos Barbieri



DIMENSÕES COMPARTILHADAS



- A DIMENSÃO É COMPARTILHADA ENTRE VÁRIOS PROJETOS DE DM/DW
- FUNDAMENTAL PARA A INTEGRAÇÃO ENTRE OS VÁRIOS "DMARTS"
- AS DIMENSÕES NORMALMENTE SÃO DESENVOLVIDAS EM SUA **MAIOR** GRANULARIDADE
 - TEMPO=> ANO-SEMESTRE-TRIMESTRE-MÊS-DIA
 - CLIENTE=> TIPO-CLIENTE
 - GEOGRAFIA=>PAIS-REGIÃO-ESTADO-CIDADE-LOJA
- AS DIMENSÕES PODEM SER COMPARTILHADAS EM HIERARQUIAS PARCIAIS. POR EX: CATEGORIA->SUBCATEGORIA->PRODUTO. SOMENTE VOU COMPARTILHAR NO MEU CUBO CATEGORIA, OU CATEGORIA->SUBCATEGORIA. DESABILITO O NÍVEL INDESEJÁVEL(AUTOMATICAMENTE DESABILITAM OS NÍVEIS MENORES)

Fonte: Carlos Barbieri

19

DIMENSÃO TEMPO

- (QUASE) SEMPRE PRESENTE NOS MODELOS DIMENSIONAIS
- TABELA DIMENSÃO TEMPO **PADRÃO**:
 - CHAVE DE DATA(PK)
 - DATA-COMPLETA(01-01-2010)
 - DIA-SEMANA(6A FEIRA)
 - NÚMERO-DIA-MÊS(01)
 - NÚMERO-DIA-GERAL(CORRIDO NO ANO)(01 a 365)
 - NÚMERO-SEMANA-ANO(01 a 52)
 - NÚMERO-SEMANA-GERAL(CORRIDO)
 - MÊS
 - NÚMERO-MÊS-GERAL(CORRIDO)
 - TRIMESTRE
 - PERÍODO-FISCAL
 - TAG-DIA-SEMANA
 - TAG-ÚLTIMO-DIA-MÊS

Fonte: Carlos Barbieri

20

DIMENSÃO CLIENTE

- MAIOR DETALHAMENTO POSSÍVEL, COM MODELAGEM DOS ATRIBUTOS COM ALTA INDEPENDÊNCIA ENTRE ELES
- TABELA DIMENSÃO CLIENTE **PADRÃO**:
 - CHAVE DE DATA(PK)
 - SAUDAÇÃO EX:MR
 - ESTILO DE SAUDAÇÃO EX:PROFISSIONAL
 - PRENOME E MEIO-NOME EX:R. JAMES
 - SOBRENOME EX: WOOD
 - SUFIXO EX:JR
 - ETNIA DO NOME EX:INGLÊS
 - GÊNERO EX:MASCULINO
 - TÍTULO EX:ADVOGADO
 - RELACIONAMENTO EX:REPRESENTANTE DE JOHN DOE
 - ORGANIZAÇÃO EX:ABC GENERIC POWER
 - SUB-ORGANIZAÇÃO EX:DEPARTAMENTO JURÍDICO

Fonte: Carlos Barbieri

21

CHAVES DE DIMENSÕES E DE FATOS

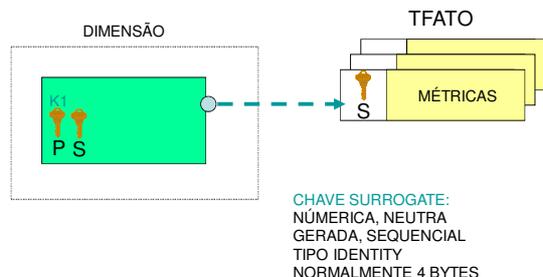


- RECOMENDAÇÃO - USAR **SURROGATE KEY(SK)**
- SK=CHAVE SEQUENCIAL, SEM SENTIDO EMBUTIDO
- CRIA MAIOR ESTABILIDADE
- EVITA CONFLITO DE MUDANÇAS DE CHAVES E DE SUAS SEMÂNTICAS
- EVITAR/CUIDADO COM O USO DE **SMART KEY**(CHAVES COM SEMÂNTICA EMBUTIDA)
- 4 BYTES: 2 BILHÕES DE OCORRÊNCIAS DE SK

Fonte: Carlos Barbieri

22

CHAVES SURROGATE



Fonte: Carlos Barbieri

23

FATOS E DADOS



- ESCOLHER PARA CADA **TFATO** OS ATRIBUTOS NUMÉRICOS E ADITIVOS
- TÍPICOS:
 - QUANTIDADE VENDIDA
 - VALOR VENDIDO
 - CUSTO DO PRODUTO (VENDIDO)
 - LUCRO
 - CONSUMO
- MANTER CONFORMIDADE/COERÊNCIA TAMBÉM ENTRE FATOS E AS MEDIDAS/VALORES, COM O MESMO SENTIDO, FÓRMULAS DE CÁLCULOS, ETC
- GRANULARIDADE DA TFATO ESTA DIRETAMENTE RELACIONADA COM A DAS TDIM
- LEMBRE-SE PORÉM: AS TFATOS SÃO GIGANTESCAS(ALTO VOLUME) E ISSO REQUER COMPROMISSOS NA ESCOLHA DE SEUS CAMPOS

Fonte: Carlos Barbieri

24

FATOS E DADOS

- ESCOLHER COM CUIDADO OS CAMPOS, DEVIDO AO TAMANHO EXPONENCIAL DAS TFATOS
- ETERNO COMPROMISSO ENTRE PERFORMANCE E ARMAZENAMENTO
- CAMPOS CANDIDATOS A REMOÇÃO:
 - CAMPOS USADOS POR POUCOS USUÁRIOS
 - CAMPOS POTENCIALMENTE DERIVADOS
 - EX: VALOR UNITÁRIO E QUANTIDADE DO ITEM
 - ARMAZENO O VALOR TOTAL DA VENDA DO ITEM????(V.UNITÁRIO*QUANTIDADE)
 - CAMPOS QUE NÃO TENHAM VALOR DE NEGÓCIO
 - CAMPOS DE DIMENSÕES DEGENERADAS, COMO NÚMERO DE ORDEM/PEDIDOS, CASO A GRANULARIDADE SEJA O ITEM DESSAS ENTIDADES
- ANALISAR O TAMANHO DE CADA CAMPO
 - USE CHAVES SURROGATE QUANDO POSSÍVEL(CHAVE DEFINIDA PELO PROJETO, SEM SIGNIFICADO INTRINSECO)

25

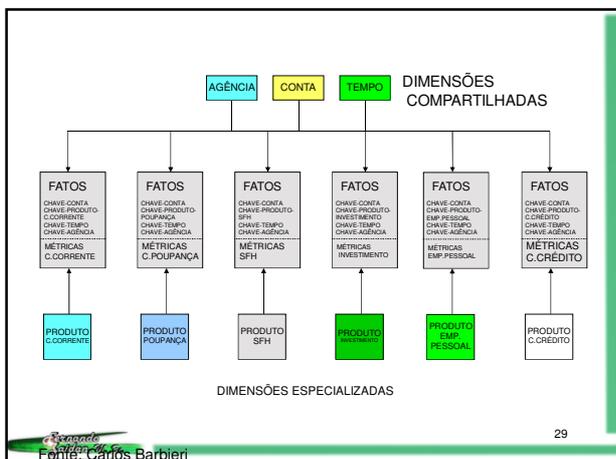
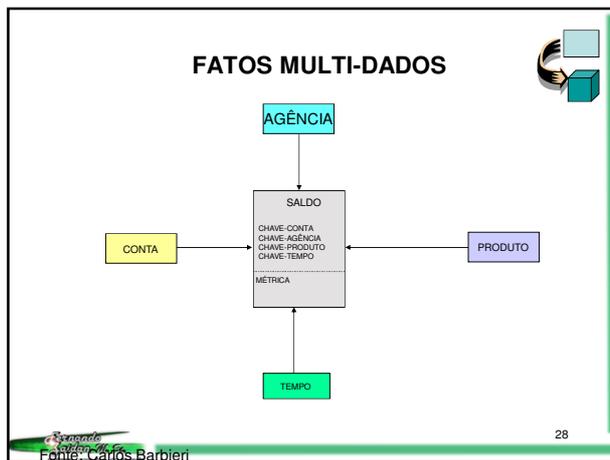
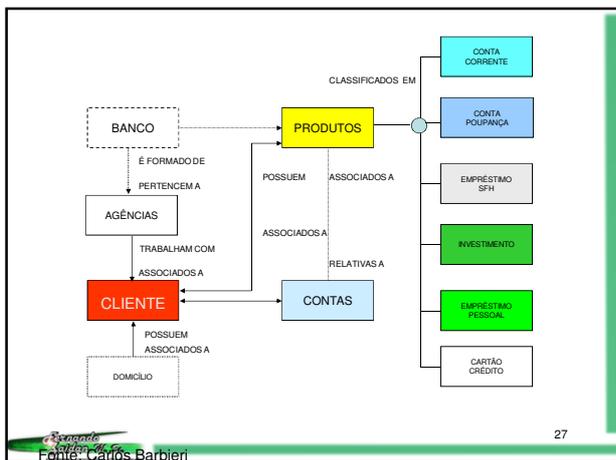
Fonte: Carlos Barbieri

FATOS E DADOS HETEROGENEIDADE

- PRODUTOS HETEROGÊNEOS
- INDÚSTRIA FINANCEIRA
- CONTA CORRENTE, SEGURO, EMPRÉSTIMO, POUPANÇA, HABITAÇÃO, ETC
- DIFERENTES FATOS E DADOS PARA CADA LINHA DE NEGÓCIO
- DIMENSÕES COMUNS(CLIENTES, AGÊNCIAS)

26

Fonte: Carlos Barbieri



DADOS E FATOS

- ALGUMAS FERRAMENTAS CONSIDERAM A TFATO COMO MAIS UMA DIMENSÃO
- FACILITA O USO DE EXPRESSÕES
- OS VALORES DA TFATO ESTÃO SEMPRE ASSOCIADOS A ELEMENTOS FOLHA DAS DIMENSÕES
- LEMBRAR QUE AS FATOS PODERÃO SER PROCESSADAS POR VÁRIOS TIPOS DE ELABORAÇÃO:
 - SOMA (MAIS COMUM)
 - VALOR MÁXIMO, MÍNIMO, CONTADOR, CONTADOR(DISTINTO)
- AS CÉLULAS DA TFATO PODEM SER CALCULADAS EM FUNÇÃO DE VALORES DE OUTRAS CÉLULAS DA MESMA TFATO-SÃO OS MEMBROS CALCULADOS
 - EX: MÉDIA=TOTAL VENDAS/TOTAL UNIDADES VENDIDAS

30

Fonte: Carlos Barbieri

Bibliografia

BARBIERI, Carlos. BI - Business Intelligence: Modelagem e tecnologia. Rio de Janeiro, Axcel Books, 2001.

CAMPOS, M. L. Data Ware Housing. UFRJ, 2007.

COME, Gilberto de. **Contribuição ao Estudo da Implementação de Data Warehousing: um caso no setor de telecomunicações** – São Paulo : FEA/USP, 2001. 133 p

FANTAUZZI, F. A. C.; ROCHA, Rogério Morais. **Directório de Softwares para Inteligência Competitiva** Monografia apresentada ao Departamento de Ciência da informação como requisito para a conclusão do curso de especialização em Gestão Estratégica da Informação da Universidade Federal de Minas Gerais - UFMG, Belo Horizonte, ano de 2006.

FARIA, João Marcos Bonadio de. **Artefatos da Semiótica Organizacional na Elicitação de Requisitos para Soluções de Data Warehouse** Trabalho final (mestrado profissional) - Universidade Estadual de Campinas, Instituto de Computação, fevereiro de 2006.

Bibliografia

INMON, William. **What is Data Warehouse ?** UNJobs, acessado em 19 de abril de 2009, disponível em < <http://unjobs.org/authors/w.-h.-inmon>>

SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. Sistema de banco de dados. Tradução da 5ª Edição. São Paulo: Campus, 2006.

TERESKO, John. **Information Rich, Knowledge Poor ?** IndustryWeek.com, acessado em 19 de abril de 2009, disponível em < <http://www.industryweek.com/PrintArticle.aspx?ArticleID=245> >

 Fernando Zaidan M.Sc.
Instituto de Computação - UNICAMP

Obrigado e bom trabalho,

Zaidan

www.fernandozaidan.com.br
fhzaidan@gmail.com

"Aí está o mérito do êxito de meus projetos: sempre fui muito exigente e rigoroso com procedimentos que aparentemente não faziam muito sentido na época. Mais tarde viu-se que esse rigor fez a diferença entre afundar ou não, concluir ou não um projeto".

Amyr Klunk

 Fernando Zaidan M.Sc.
Instituto de Computação - UNICAMP