

Modelagem para Data Warehouse

A serviço do suporte decisório nas organizações.

FIM

Prof. Grimaldo Oliveira
www.aprendavirtual.ninehub.com
www.bicomvatapa.blogspot.com
<http://br.linkedin.com/in/grimaldo>

Programa

- Planejamento Estratégico
- Evitando a Desinformação
- Business Intelligence
- Data Warehouse
- Processo de Construção de Um Data Warehouse
 - Desenvolvimento do Data Warehouse
- Demonstração Ferramenta Olap
- Conclusão
- Bibliografia

Planejamento Estratégico



Começando a planejar...



- Em que situação pretendo que minha empresa esteja a médio e longo prazo ?
- Qual é a razão de ser da minha empresa ? Por que e para que ela existe ?
- Qual a situação atual em que ela se encontra ?
- Quais são e de que forma posso classificar a situação atual dos meus processos de negócio ?
- Quais ações devem ser elaboradas para que meus processos de negócio sejam potencializados ?

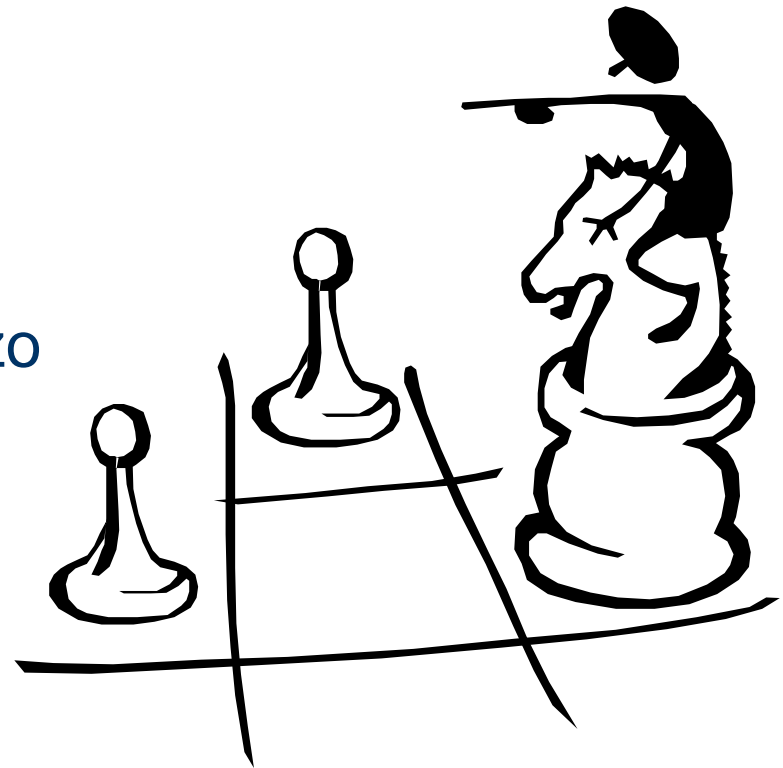
Planejamento Estratégico nas Organizações



- Para que a empresa se torne competitiva no mercado, deve aperfeiçoar continuamente seus processos, procurando adaptar-se a realidade de constantes mudanças e incertezas.
- O planejamento estratégico auxilia os dirigentes a tomarem decisões, a se anteciparem às mudanças ou mesmo a se prepararem para tal.
- Envolve a determinação de metas, a construção de estruturas organizacionais adequadas e a mobilização dos recursos necessários para o alcance dos objetivos traçados

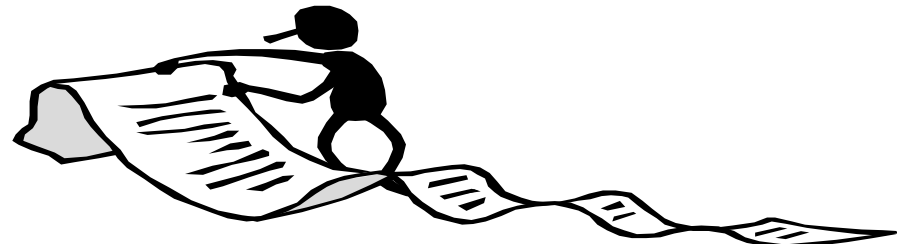
Planejamento Estratégico nas Organizações

- Produtos Finais :
 - Missão da empresa
 - Objetivos a médio e longo prazo
 - Fatores Críticos de Sucesso



O processo decisório nas empresas

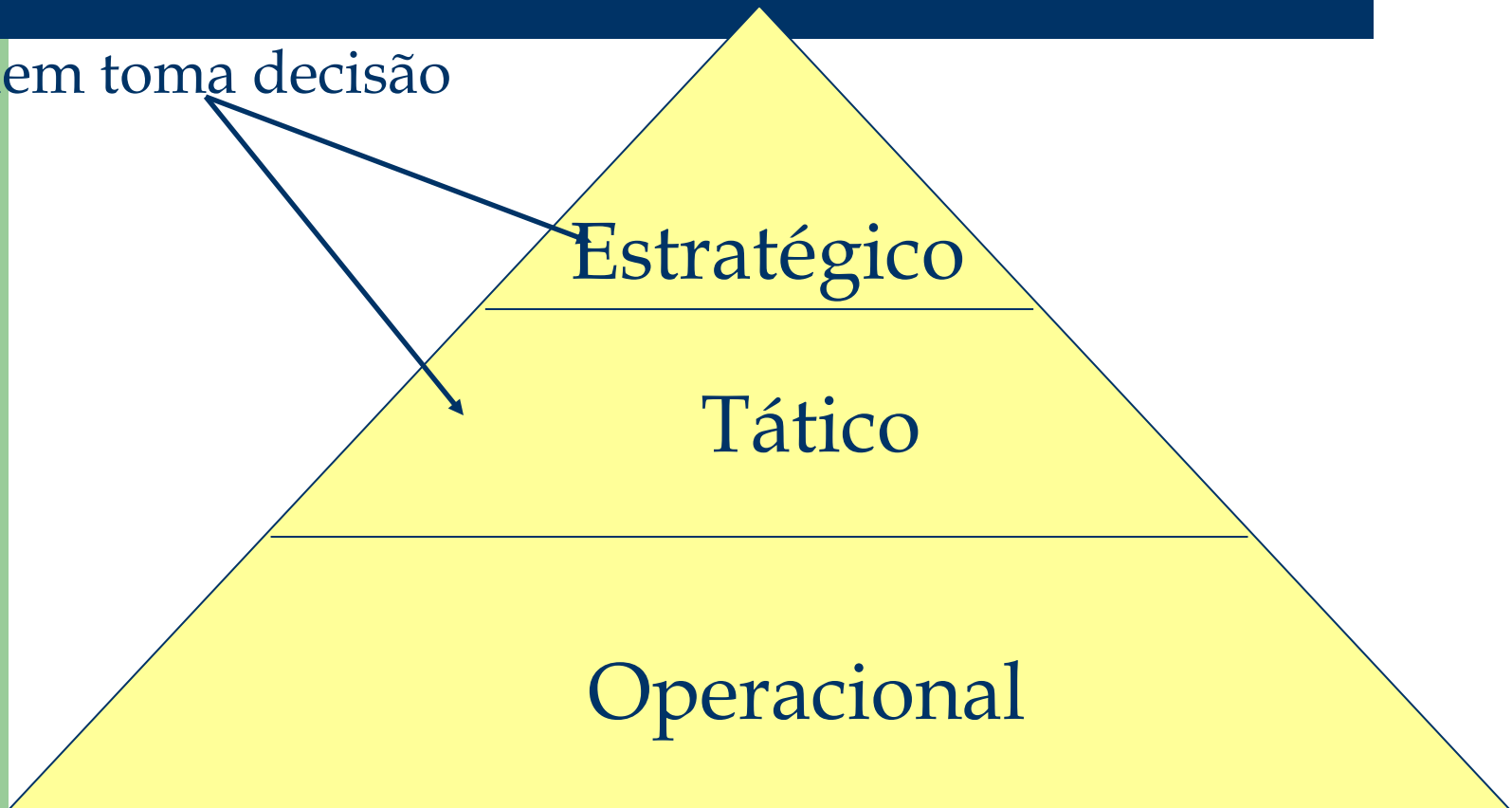
- Qual é o suporte de informações para o executivo ?
 - Documentos impressos
 - Planilhas Eletrônicas



- Contexto Atual : Sobram dados e falta informação

Processo Decisório

Quem toma decisão



Nível da tomada de decisão gerencial

- **Administração Estratégica** – principais executivos, desenvolvem as metas globais, estratégias, políticas e objetivos organizacionais como parte de um processo de planejamento estratégico.
 - **Administração Tática** – gerentes de operações. Desenvolvem planejamentos de médio e curto prazos, orçamentos, e definição da distribuição de recursos ao ambiente operacional.
 - **Administração Operacional** – equipes autogeridas e os gerentes de operações, planos de curto prazo. Desenvolvem planos de curto prazo. Resolução de problemas restritos ao operacional da empresa.
- Nível da tomada de decisão gerencial

O processo decisório nas empresas



- Qual é a qualidade das informações para o executivo ?
 - Conflito de Conceitos
 - Falta de Integridade
 - Excesso de Papéis
 - Perda de tempo para composição das informações
- Contexto Atual : As informações disponibilizadas aos executivos não possuem confiabilidade

Efeitos colaterais da Desinformação

?

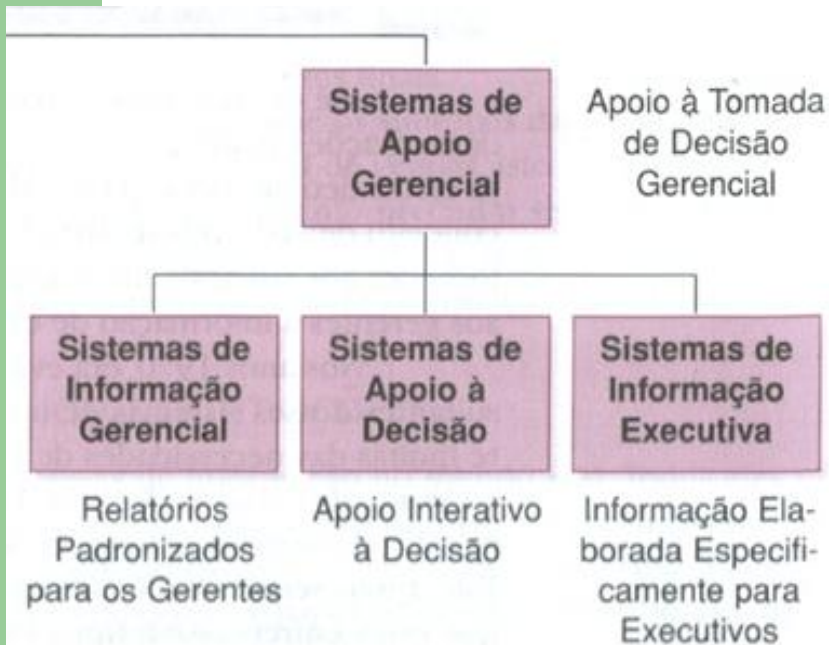
- Aumento da pressão sobre o executivo
- Incerteza na tomada de decisão
- Perda de oportunidade de negócios



Evitando a Desinformação



Sistemas de Apoio Gerencial



- Fornecer informação e apoio para a tomada de decisão eficaz pelos gerentes.
- Apóiam as necessidades de tomada de decisão da administração estratégica (principal), administração tática (média) e administração de operação (supervisora).

Histórico

- E.I.S : Executive Information System
 - Respostas Prontas, Modelo Rígido, fácil uso.
 - Uso tipo “point and click”
 - Novos relatórios demandam programação
 - O usuário não pode definir novas agregações, quebras ou criar dados calculados

Histórico

- D.S.S : Decision Suport System
 - Relatórios configuráveis pelo usuário
 - O modelo rígido ainda não permite que quebras e agrupamentos sejam realizados diferentes dos programados
 - Criação de dados calculados

Histórico

- O.L.A.P. : On Line Analysys Process
 - Múltiplas respostas, Modelo Flexível
 - Quaisquer combinações dos dados podem ser feitas pelo próprio usuário
 - Criação de dados calculados
 - Respostas rápidas
 - Enfoque no modelo multidimensional

Histórico

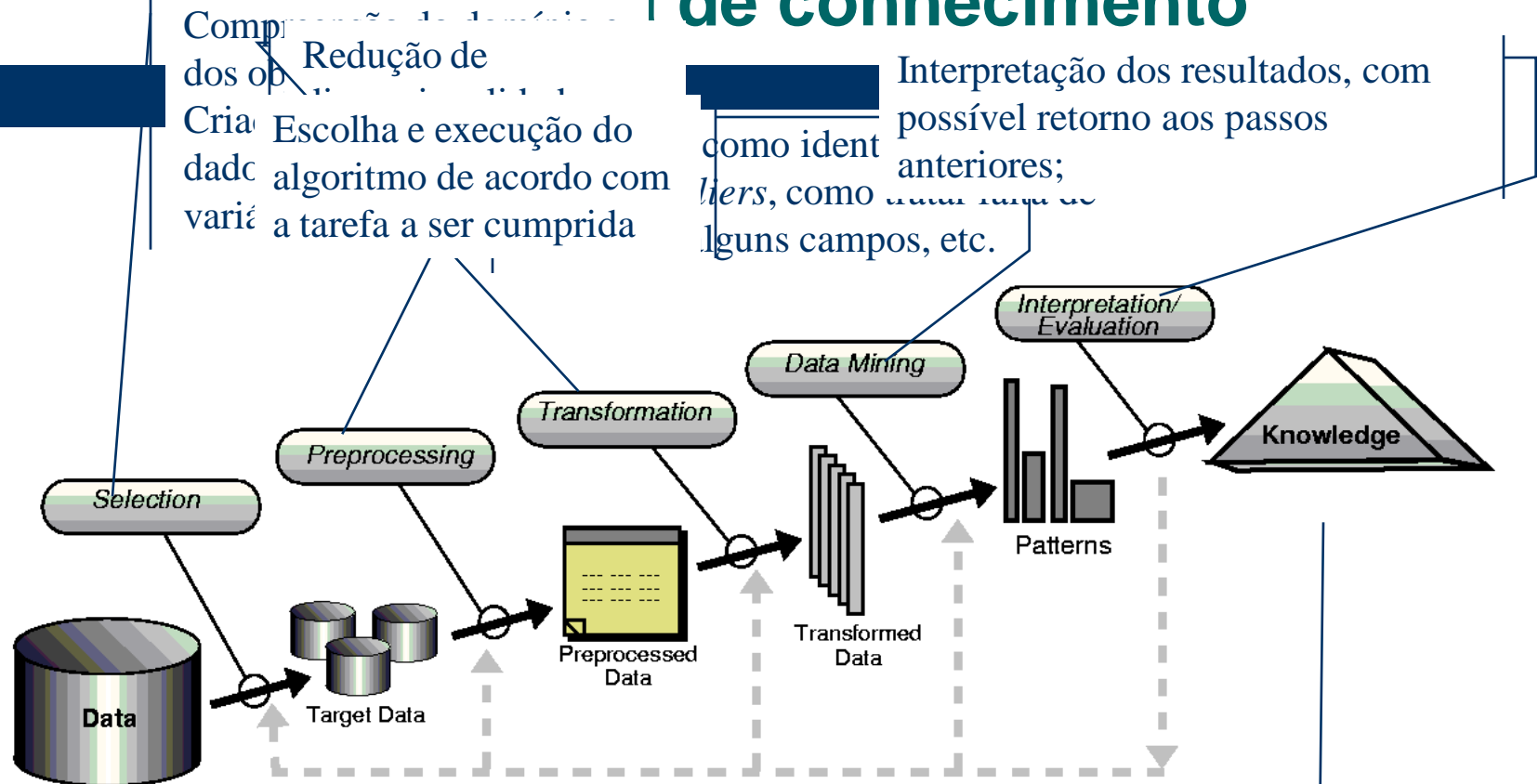
- Data Warehouse

- Permite que a história dos dados sejam registrada.
- Visão integrada
- Não Volátil
- É a fonte de dados para responder questões de tomada de decisão

Business Intelligence



Processo mínimo de descoberta de conhecimento



Comp dos op
Cria
dados
variá

Redução de
Escolha e execução do
algoritmo de acordo com
a tarefa a ser cumprida

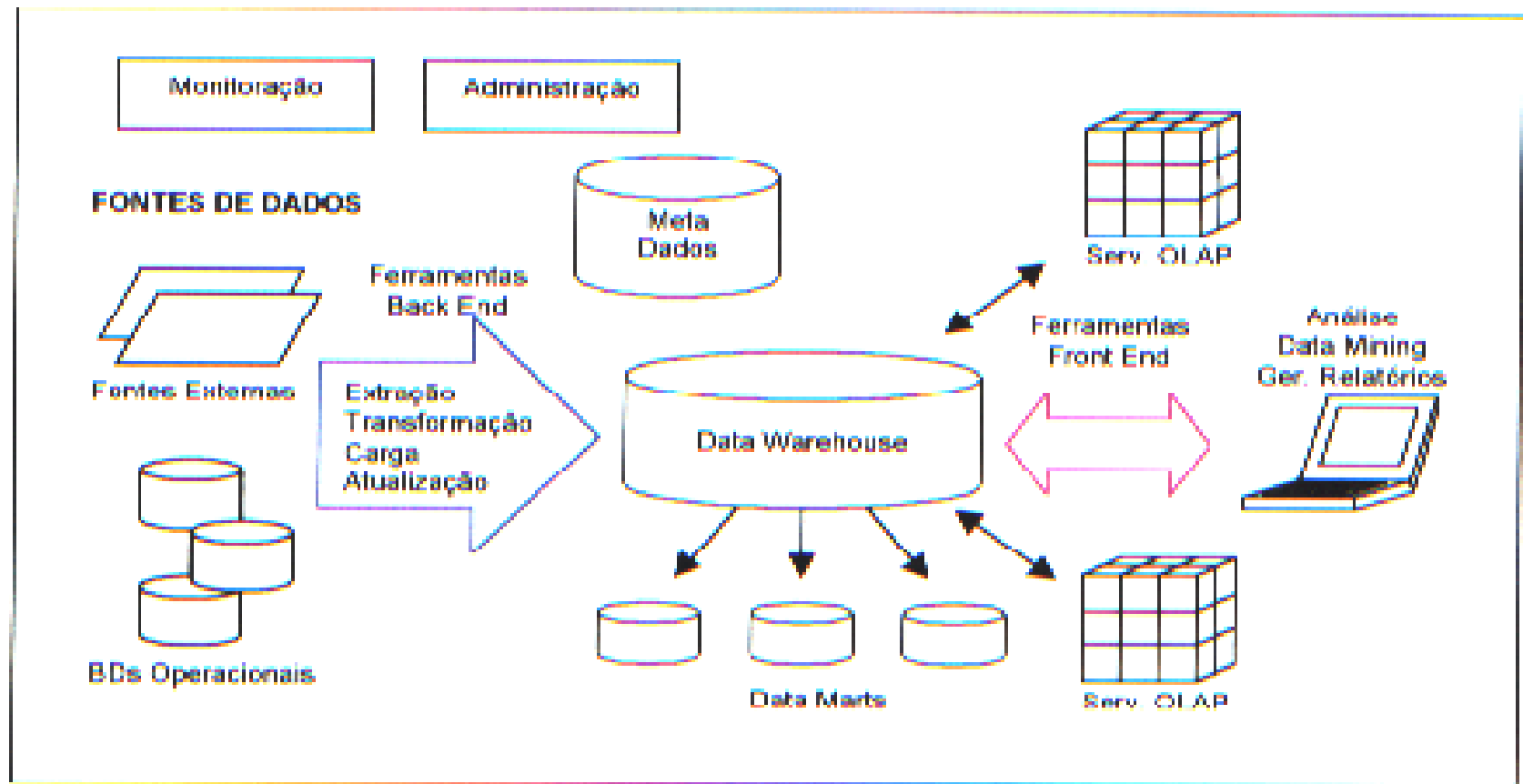
Interpretação dos resultados, com
possível retorno aos passos
anteriores;
como ident
liers, como
alguns campos, etc.

Consolidação: incorporação e documentação do
conhecimento e comunicação aos interessados;

Business Intelligence - Conceito

- Coleta de dados a partir de diversas aplicações operacionais, integrando-as em áreas lógicas de processos de negócios, armazenando-as em um repositório central e disponibilizando as informações aos tomadores de decisão através de uma ferramenta de visualização de forma rápida, fácil e intuitiva.

Ambiente BI



Business Intelligence - Objetivos

- Prover suporte decisório de qualidade nas organizações disponibilizando informações sobre clientes, mercado, negócios e processos com as seguintes características :
 - Informações confiáveis, padronizadas, unificadas, com fácil e rápido acesso.
 - Composição de análises diferenciadas
 - Visualização intuitiva das informações



Preparação dos Dados

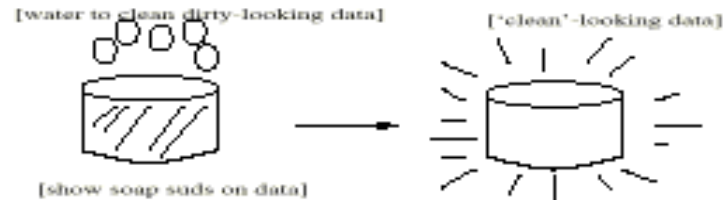
Os dados no mundo real estão “sujos”:

- Incompletos
 - ausência de atributos de interesse
 - apenas dados agregados
 - ausência de valores
- Ruidosos
 - erros aleatórios
 - valores aberrantes (outliers)
- Inconsistentes
 - discrepâncias nas codificações ou nos nomes

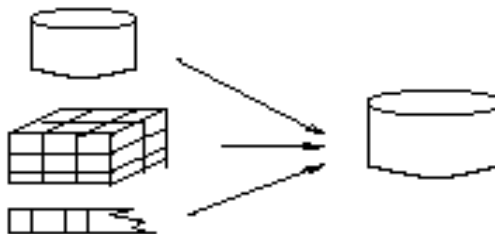
Sem dados de boa qualidade o resultado da mineração é pobre

Pré-processamento dos Dados

Data Cleaning



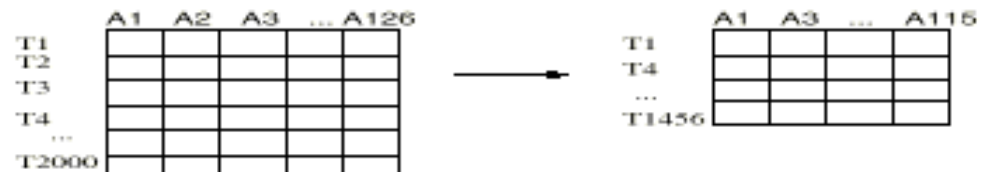
Data Integration



Data Transformation

-2, 32, 100, 59, 48 → -0.02, 0.32, 1.00, 0.59, 0.48

Data Reduction

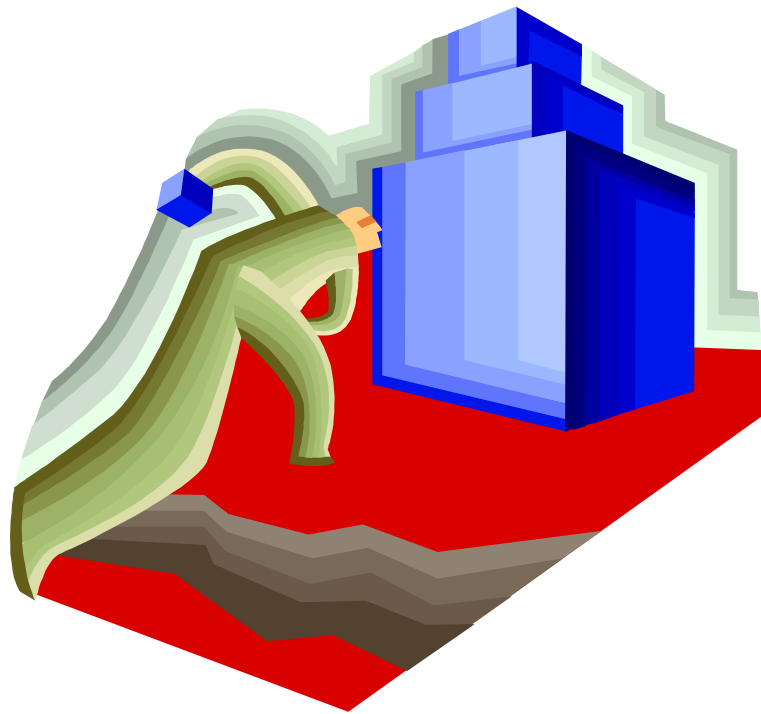


Limpeza dos Dados

Em que consiste a “limpeza” dos dados?

- preencher dados ausentes
- identificar valores aberrantes -> substituir/retirar
- Identificar inconsistências -> substituir/retirar
- etc

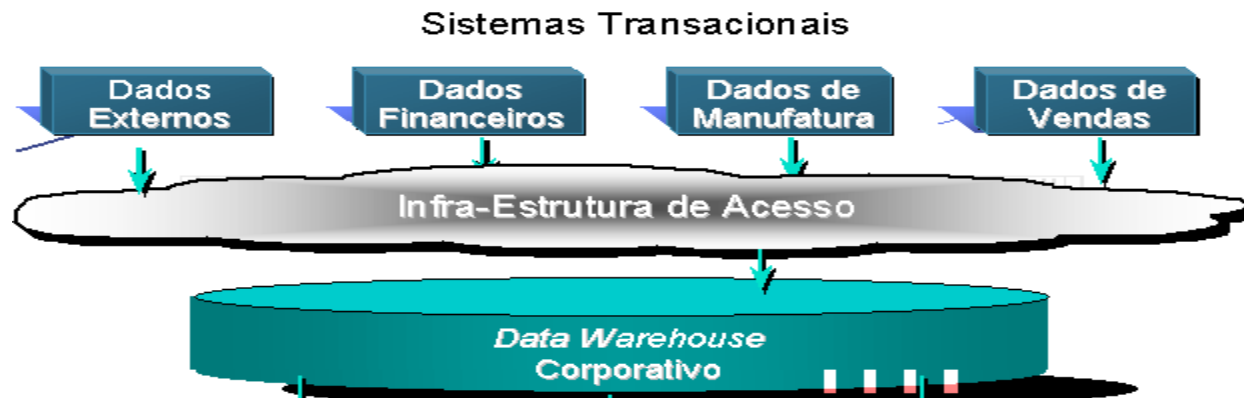
Data WareHouse



Data Warehouse - Conceito

É o repositório central dos dados da organização com o objetivo de prover suporte à decisão.

- Orientado por assunto: contêm informações sobre os processos de negócio da empresa.
- Não volátil: permite apenas a carga de novos dados e consultas.
- Variável no tempo: contém dados não atualizáveis que se referem a algum momento específico.
- Integrado: contêm dados em um estado uniforme, ou seja, existe uma consistência entre nomes, unidades de medida e etc.



Data Warehouse - Objetivos

- Tornar a informação mais acessível e consistente para toda a organização.
- Ser uma fonte segura para proteger a informação da empresa.
- Deve ser a base para a tomada de decisão.



Data Warehouse - Benefícios

- Elimina as redundâncias
- Padroniza as terminologias
- Democratiza a informação
- Agiliza o processo de decisão
- Libera tempo do executivo



Data Warehouse – Fatores Críticos para o sucesso

- Alta Direção como Patrocinadora do Projeto
- Escolha de Consultoria adequada e experiente
- Criação de equipe interna efetiva
- Utilização de Campanhas culturais diversas
- Documentação
- Escolha de ferramentas adequadas



Processo de Construção do Data Warehouse



Evolução dos SGBD's

- **Bancos Pré-relacionais**
 - coleções de arquivos
- **Bancos Relacionais – anos 70**
 - coleções de tabelas
 - integridade referencial
 - independência de dados
 - álgebra relacional – linguagem SQL
- **Bancos Pós-relacionais**
 - extensões ao modelo relacional (ORDBMS) ou
 - abordagens totalmente distintas do modelo relacional

OLTP

- *On Line Transactional Processing* :
Processamento 'on line' de Transações.
- Processamentos que executam as operações do dia-a-dia da organização
- Ênfase ao suporte do negócio, através de um processamento rápido, acurado e eficiente de dados
 - Ex: movimento bancário

Modelo Relacional

Carro	Cor	Modelo
Gol	Azul	3
Gol	Cinza	2
Gol	Preto	5
Gol	Branco	6
Santana	Azul	4
Santana	Cinza	2
<i>Santana</i>	<i>Preto</i>	<i>3</i>
Santana	Branco	4
Polo	Azul	5
Polo	Cinza	1
Polo	Preto	2
Polo	Branco	3

Suportado
pelos bancos
de dados
relacionais:
Oracle, SQL
Server, DB2,
etc.

OLAP

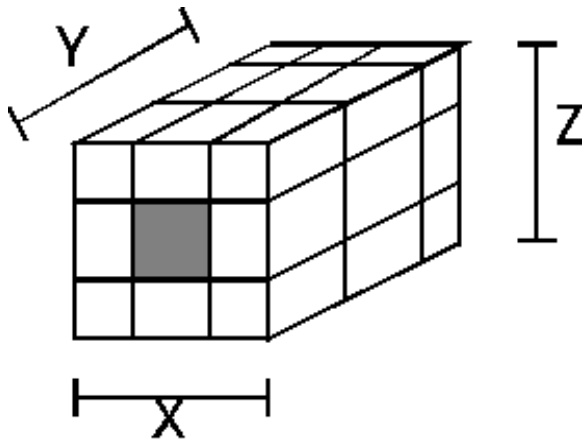
- *On Line Analytical Processing* :
Processamento 'on line' de análise.
 - Múltiplas respostas, Modelo Flexível.
 - Processamentos que suportam a tomada de decisões
- Permite analisar tendências e padrões em grande quantidades de dados
 - ao longo do tempo (histórico)
 - e em diferentes localizações (geográficos)

Modelo Multidimensional

		Gol CL	Santana	Polo
1996				
1997				
Azul	3	4	5	
Cinza	2	2	1	
Preto	5	3	2	
Branco	6	4	3	

E há algum
tipo de banco
de dados que
suporte esse
tipo de
modelo ?

Modelo Multidimensional



armazenados em um formato que assemelha-se a um cubo.

- Cada célula do cubo pode conter outro cubo (Meta Cubo)
- Histórico dos Dados

OLTP

x OLAP

dados orientados à aplicação
última versão dos dados
dados atualizáveis
desempenho é fator crítico
acesso orientado a linha
dados voláteis
alta disponibilidade
ausência de redundância

vs.

vs.

vs.

vs.

vs.

vs.

vs.

vs.

orientados ao assunto

snapshots

somente para leitura

não tão crítico

orientado a conjunto

dados históricos

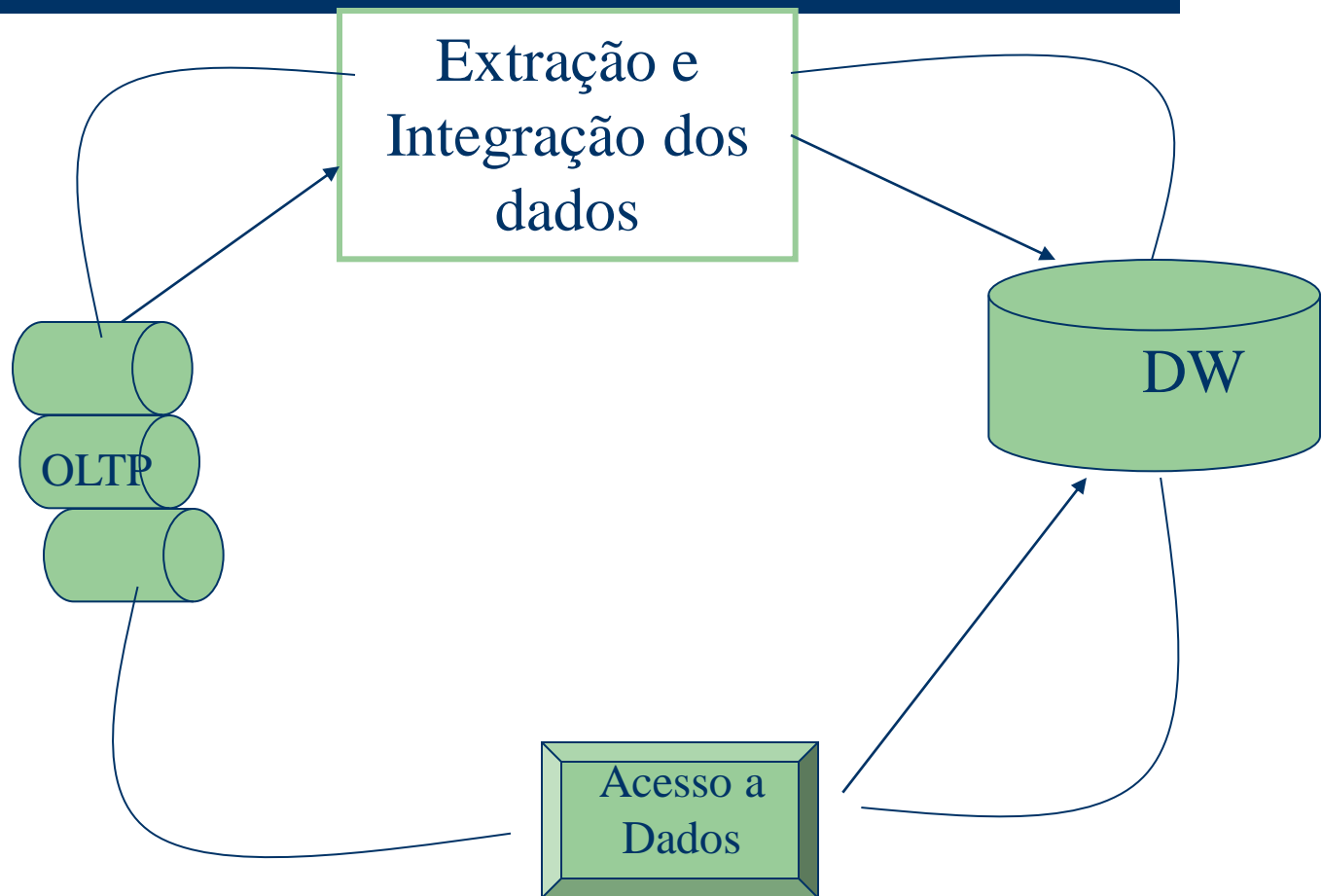
não tão alta

redundância gerenciada

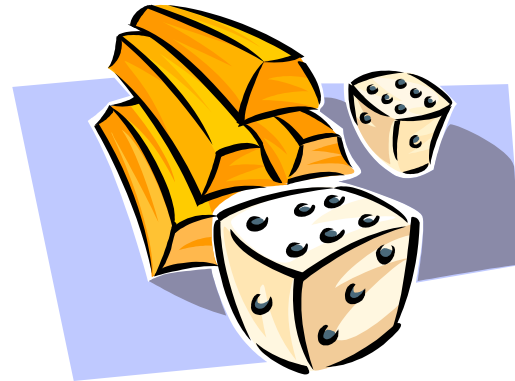
Bancos de Dados Multidimensionais

- Permitem a criação, gerenciamento e consulta de dados no formato puramente multidimensional.
- Têm excelente performance para consulta.
- Os dados são gravados via cargas batch.
- Via de regra, não há atualizações por parte dos usuários.
- Exemplo: SQL Analysis Server, EssBase, Oracle Express.

Ambiente do DW

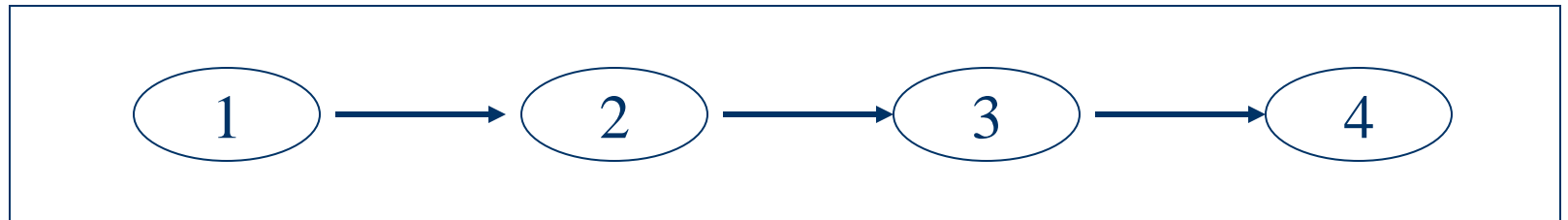


Desenvolvimento do Data WareHouse



Data Warehouse – Principais Etapas do Desenvolvimento

- Identificação dos indicadores
- Modelagem Dimensional
- ETL + Carga
- Criação dos Relatórios (Ferramenta OLAP)
- Pós-Implantação



Data Warehouse – Identificação dos Indicadores

- Através do planejamento estratégico da organização, todas as informações de caráter estratégico e tático necessárias para apoio a tomada de decisão são identificadas.
- A existência de um planejamento estratégico na organização agiliza este processo de identificação dos indicadores, uma vez que já estão elaborados e são conhecidos por toda organização.



Data Warehouse – Modelagem Dimensional

- O modelo dimensional de um DW tem como objetivo ser intuitivo para um administrador do negócio além de realizar consultas com alta performance.
- Dimensão : informações descritivas relacionadas aos processos de negócio. Ex : Dados de empresa, cliente, produto, fornecedor.
- Tabelas fato : medidas dos processos de negócio que devem ser analisadas. Ex : vendas, faturamento, despesa, estoque.

Visão Multidimensional

Região	Trimestre	Linha de Produtos	Vendas
Sul	T1	Linha Branca	R\$ 250M
		Outros	R\$ 127M
	T2	Linha Branca	R\$ 225M
		Outros	R\$ 143M
	T3	Linha Branca	R\$ 275M
		Outros	R\$ 148M
	T4	Linha Branca	R\$ 253M
		Outros	R\$ 131M
Nordeste	T1	Linha Branca	R\$ 280M
		Outros	R\$ 147M
	T2	Linha Branca	R\$ 255M
		Outros	R\$ 163M
	T3	Linha Branca	R\$ 305M
		Outros	R\$ 148M
	T4	Linha Branca	R\$ 283M
		Outros	R\$ 151M

Dimensões

Fatos

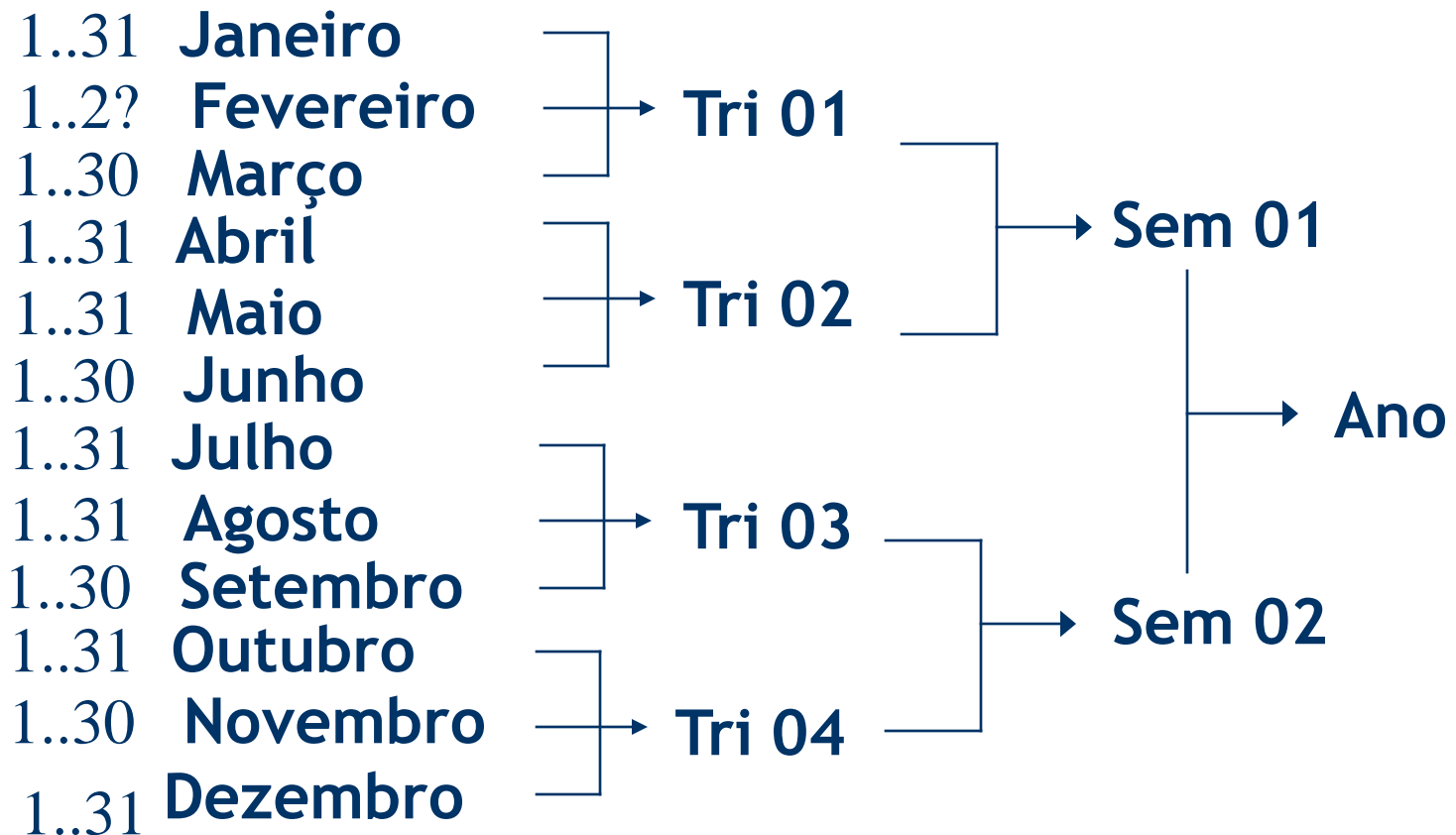
Dados Multidimensionais

- Apenas relacionamentos 1:N ou N:N devem ser trazidos para o mundo multidimensional
- Relacionamentos 1:1 geram esparsividade
- Exemplo : Cliente e Sexo. Não pode ter em duas dimensões separadas, pois um cliente só tem um sexo.

Inteligência de Tempo

- Um MDDB reconhece e gerencia perfeitamente os diversos agrupamentos de tempo :
 - Dia, Mês , Ano
 - Dia, Mês, Trimestre, Semestre, Ano
 - Dia, Mês, Estação , Ano
 - Dia, Semana
- O gerenciamento dos dados é automático ao lidarmos com o tempo.

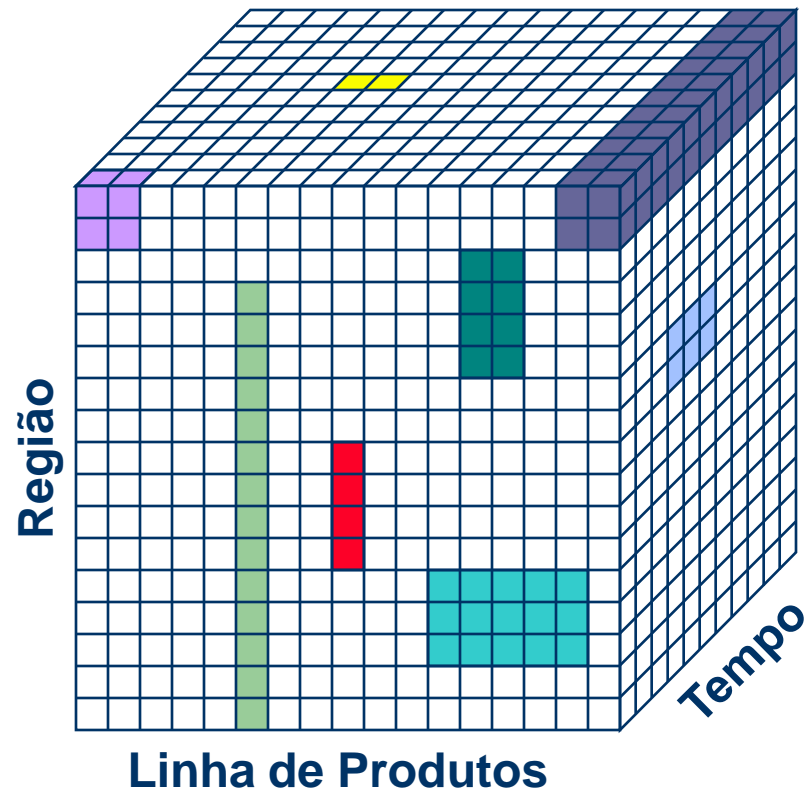
Inteligência de Tempo



Operações OLAP

- Slice and Dice

- Ranging
- Ranking
- Drilling



Ranging (montado a consulta)

- COR
- TEMPO
- CARRO
- LOJA
- QUANTIDADE

PÁGINAS

COLUMNAS

L
I
N
H
A
S

CÉLULAS

Exemplo

Ano : 1999

Loja : Cobape

	<i>Gol</i>	<i>Parati</i>	<i>Polo</i>	<i>Golf</i>
<i>Azul</i>	23	43	23	93
<i>Prata</i>	36	98	76	76
<i>Branco</i>	87	74	86	21

Loja : Sanave

	<i>Gol</i>	<i>Parati</i>	<i>Polo</i>	<i>Golf</i>
<i>Azul</i>	23	43	23	93
<i>Prata</i>	36	98	76	76
<i>Branco</i>	87	74	86	21

Ranking

- Permite a classificação de uma dimensão através de um fato.
- Aplica-se para saber os maiores ou os menores.
- Exemplo : TOP (Contribuinte, ICMS, 5)
 - Retorna os 5 maiores contribuintes com valor de ICMS

Drilling

D
O
W
N



A
C
R
O
S
S



<i>Loja</i>	<i>Produto</i>	<i>Mês</i>	<i>Valor</i>
Bahia	Big Mac	Jan/2000	\$5340,00

<i>Loja</i>	<i>Produto</i>	<i>Mês</i>	<i>Valor</i>
Barra	BigMac	Jan/2000	\$2200,00
Igua01	Big Mac	Jan/2000	\$3140,00

<i>Bebida</i>	<i>Produto</i>	<i>Mês</i>	<i>Valor</i>
Coca	Big Mac	Jan/2000	\$1340,00

U
P



Slice and Dice

- Filosofia das ferramentas OLAP que permite ao usuário acessar todas as operações OLAP através da interface gráfica, sem precisar recorrer à linguagens de comando.

(fatiar e girar)

Staging Area

- A staging area ou área auxiliar serve como ponto único para a carga efetiva no data warehouse.
- A cada carga seu conteúdo é limpo
- Fornece unicidade e performance para a carga
- Evita acesso à produção em caso de recarga durante o dia

Fundamentos Básicos: Modelagem Multidimensional

- Desnormalização: anti-forma normal.
- Dimensões: Chaves Artificiais e Histórico
- Hierarquias
- Fatos
- Tempo
- Granularidade do Fato
- Agregados

Desnormalização

- Só a 1ºFN deve ser respeitada. As demais FN obrigatoriamente devem ser desrespeitadas.
- Vantagens: Excelente tempo de 'query response'.
- Exemplo:

CLIENTE

Cod_cliente

Nom_cliente

Cod_tipo_cliente

TIPO CLIENTE

Cod_tipo_cliente

Des_tipo_cliente

Cod_condicao_tipo_cliente

@Cod_cliente

Nom_Cliente

Cod_tipo_cliente

Des_tipo_cliente

Cod_condicao_tipo_cliente

Des_condicao_tipo_cliente

Considerações sobre espaço

- A desnormalização simplifica o modelo, pois quem vai gerar as queries são as ferramentas OLAP e não uma pessoa, mas causa excesso de uso de disco.
- Exemplo : Clientes do Bank Money

	<u>Normalizado</u>	<u>Desnormalizado</u>
Tamanho	400Mb	2Gb
Qtde Campos	~70	~180
Query Estatística	> 10hs	<15 min

Dimensões

- Contém os descritores textuais do negócio
- Representam objetos físicos do mundo real, locais, conceitos, etc.
- São os eixos do cubo
- Exemplo : Tempo , Cliente, Produto, Tipo de Embalagem, Situação, etc.

Chaves Artificiais

- Permite que o controle de histórico dos dados seja facilmente implementado
- Gera independência de relacionamento com outras tabelas.
- Devem ser apenas números e não carregar em si nenhum significado.
- Exemplo : 1,32453
- Contra Exemplo : Jan/99 , IFZ01

Tipos de Dimensões

- Slow Changing Dimensions
- Fast Changing Monster Dimensions
- Junk Dimensions
- Dimensões Degeneradas

Slow Changing Dimensions

Tipo 1 : Sobrescrever os Dados


- O novo registro substitui o registro original. Só existe um registro no banco de dados - os dados atuais.
- Não mantém histórico

Tipo1 : Controle de Versão

- Exemplo seria o de uma tabela de banco de dados que mantém as informações do fornecedor.

Supplier_key	Supplier_Name	Supplier_State
001	Phlogistical Sociedade de Abastecimento	CA

Supplier_key	Supplier_Name	Supplier_State
001	Phlogistical Sociedade de Abastecimento	IL



Slow Changing Dimensions

- Tipo 2 : Controle de Versão
 - Mantém histórico
 - Um novo registro é adicionado na tabela de dimensão. Dois registros existentes no banco de dados - os dados atuais e dados da história anterior
 - É recomendável para 99% dos casos

Tipo 2 : Controle de Versão

- Exemplo seria o de uma tabela de banco de dados que mantém as informações do fornecedor.

Supplier_key	Supplier_Code	Supplier_Name	Supplier_State	Data_inicial	Data_final
001	ABC	Phlogistical Sociedade de Abastecimento	CA	01 de janeiro-2000	21-Dez-2004
002	ABC	Phlogistical Sociedade de Abastecimento	IL	22-Dez-2004	

Slow Changing Dimensions

- Tipo 3 : Criação de Campos
 - Criar campos dinamicamente na tabela (em tempo de carga) para que os valores anteriores sejam guardados
 - Os dados originais são modificados para incluir novos dados. Um registro existe no banco de dados novas informações são unidas com informações antigas na mesma linha.
 - Aumenta o custo de manutenção
 - Totalmente desaconselhável

Tipo 3 : Criação de Campos

- Exemplo seria o de uma tabela de banco de dados que mantém as informações do fornecedor.

Supplier_key	Supplier_Name	Original_Supplier_State	Effective_Date	Current_Supplier_State
001	Phlogistical Sociedade de Abastecimento	CA	22-Dez-2004	IL

Fast Changing Monster Dimensions

- Algumas dimensões que possuem grande volume de registros e muitos campos, crescem rapidamente, explodindo o espaço físico de armazenamento.
- Solução : Colocar campos que trocam de valores mais rapidamente em outra tabela, sem alterar a versão do registro.

Dimensões Degeneradas

- Existem apenas como referência a uma entidade
- Exemplo : Considere que uma tabela de fato com os itens das notas fiscais de uma empresa varejista. O que fazer com o número da nota fiscal ? Ele em si, não representa nada. Apenas serve para agrupar os itens de uma mesma compra. Não existe fisicamente uma dimensão nota fiscal, embora exista uma coluna na tabela de fatos com o número da nota propriamente dito.

Junk Dimensions

- São agrupamentos de campos que não se encaixam em nenhuma dimensão
- Pode ser um conjunto de Flags. A combinação de valores de cada flag é um elemento da dimensão
- Um certo número de dimensões muito pequenas podem ser agrupados para formar uma única dimensão, uma dimensão lixo - os atributos não são estreitamente relacionadas.

Hierarquias

- Agrupamentos de conceitos correlatos em uma única dimensão
- Aplica-se às dimensões para que os fatos sejam agrupados em níveis de detalhamento.
- Tipos de Hierarquia :
 - Simples : um nó tem somente um pai
 - Múltipla : um nó pode ter vários pais
- Transparência de navegação para o usuário

Hierarquias

- Exemplo : País → Região → UF → Cidade
- As tabelas de País, Região, UF e Cidade são armazenadas separadamente no sistema fonte (normalizado). No Data Warehouse, elas compõem uma única tabela, a dimensão geografia. Cada nível da hierarquia deve ser representado individualmente.
- Usada para Permitir Drill :
 - Drill Down : Detalha a informação
 - Drill Up : Sumariza a informação
 - Drill Across : Muda de dimensão, mantém fato.
 - Drill Through : Vê registros do ambiente transacional que originaram aquele fato

Exemplo de Implementação

Seq	País	Região	Estado	Cidade	Loja	Nível
1	BR	SE	RJ	Rio	L1	1
2	BR	SE	RJ	Rio	L2	1
3	BR	SE	RJ	Rio		2
4	BR	SE	RJ			3
5	BR	SE	SP			3
6	BR	SE				4
7	BR					5

Atenção !!!

- Em dimensões hierarquizadas, devemos ter um cuidado especial para realizar o controle de versão :
 - Ao atualizar um nó de uma hierarquia devemos, automaticamente no processo de carga, gerar para esse novo nó, todos os filhos do nó que foi alterado.

Exemplo : DIM_CLIENTE

Seq_cliente

Cod_cliente

Nom_cliente

Cod_tipo_cliente

Des_tipo_cliente

Dtc_inicio

Dtc_fim

Sts_corrente

Em preto: campos de controle utilizados pela modelagem multidimensional

Em vermelho: campos do sistema fonte.

Exemplo

Seq_cliente : 1
Cod_cliente : X1
Nom_cliente : Fulano de Tal
Cod_tipo_cliente : 1
Des_tipo_cliente : Ativo
Dtc_inicio : 01/03/1998
Dtc_fim : <null>
Sts_corrente : S

E se o
cliente
vier a
falecer ?

Exemplo

Seq_cliente : 3435
Cod_cliente : X1
Nom_cliente : Fulano de Tal
Cod_tipo_cliente : 5
Des_tipo_cliente : Falecido
Dtc_inicio : 01/03/2002
Dtc_fim : <null>
Sts_corrente : S

Um ***novo***
registro é
inserido na
base de
dados!!!

E o outro
registro ?

Exemplo

Seq_cliente : 1
Cod_cliente : X1
Nom_cliente : Fulano de Tal
Cod_tipo_cliente : 1
Des_tipo_cliente : Ativo
Dtc_inicio : 01/03/1998
Dtc_fim : 01/03/2002
Sts_corrente : N

Ambos os
registros
estão no
banco de
dados!!!

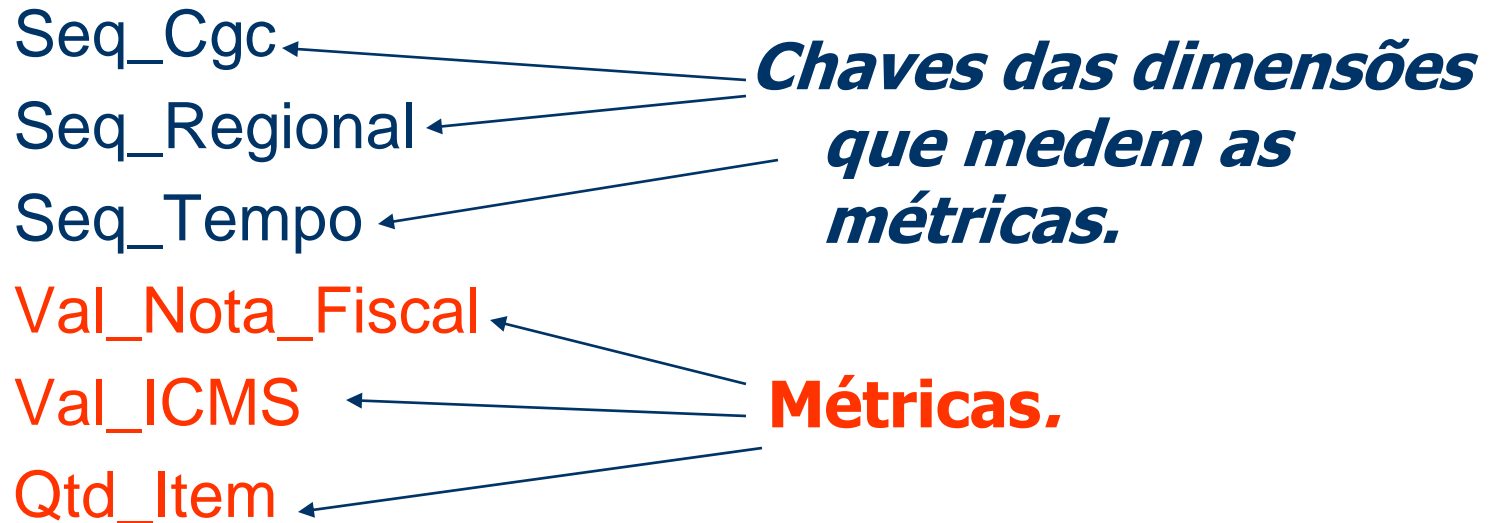
Considerações

- O uso de chaves artificiais permite o controle do histórico e facilita a criação das tabelas de fatos.
- O campos de controle de data permitem saber *quando* aquele registro representava a verdade para o dado.
- O campo *Sts_Corrente* permite a fácil identificação do registro que contém os dados mais recentes.

Fatos

- Termo utilizado para a medição do negócio.
- Representam valores ou ocorrências de eventos.
- São as células do cubo, ou seja, a interseção dos eixos.
- Exemplo: quantidade de produtos vendidos, preço de compra, preço de venda, lucro, etc.

Exemplo: FATO_Arrecadacao



Grão

- Conceito que identifica a unidade de medida das métricas.
- Nível de detalhe dos dados.
- Mais detalhe -> Mais dados->Análise mais longa ->Informação mais detalhada.
- Menos detalhe -> Menos dados
->Análise mais rápida ->Informação menos detalhada.

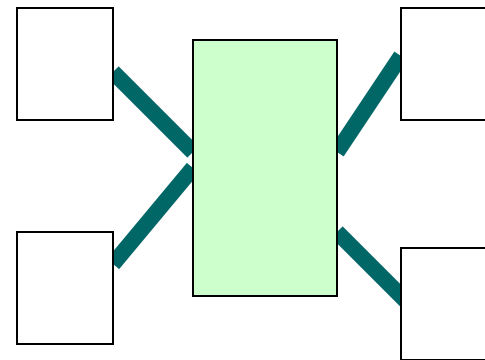
Agregados

- São agrupamentos dos fatos
- Geralmente, diminui-se o grão da tabela de fatos
- Pode-se criar agregados por qualquer atributo da dimensão
- Deve-se avaliar o tempo de carga e o espaço ocupado em disco
- Beneficiar o maior número possível de usuários
- Possibilitar o ganho de performance no maior número de queries possível.

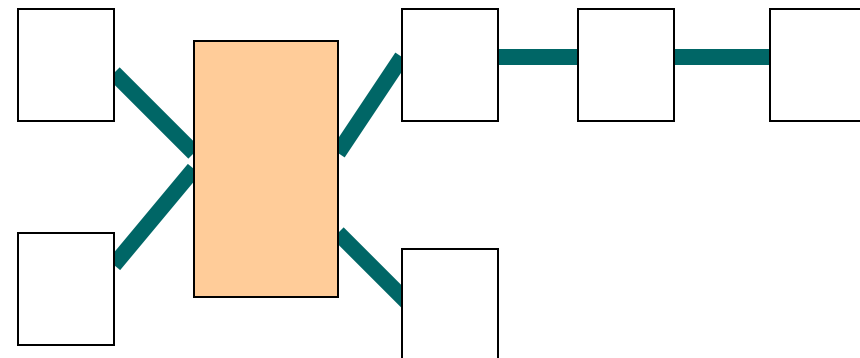
Modelagem Multidimensional

- Técnica de modelagem que visa dar simplicidade e performance de consulta.

- Star Schema



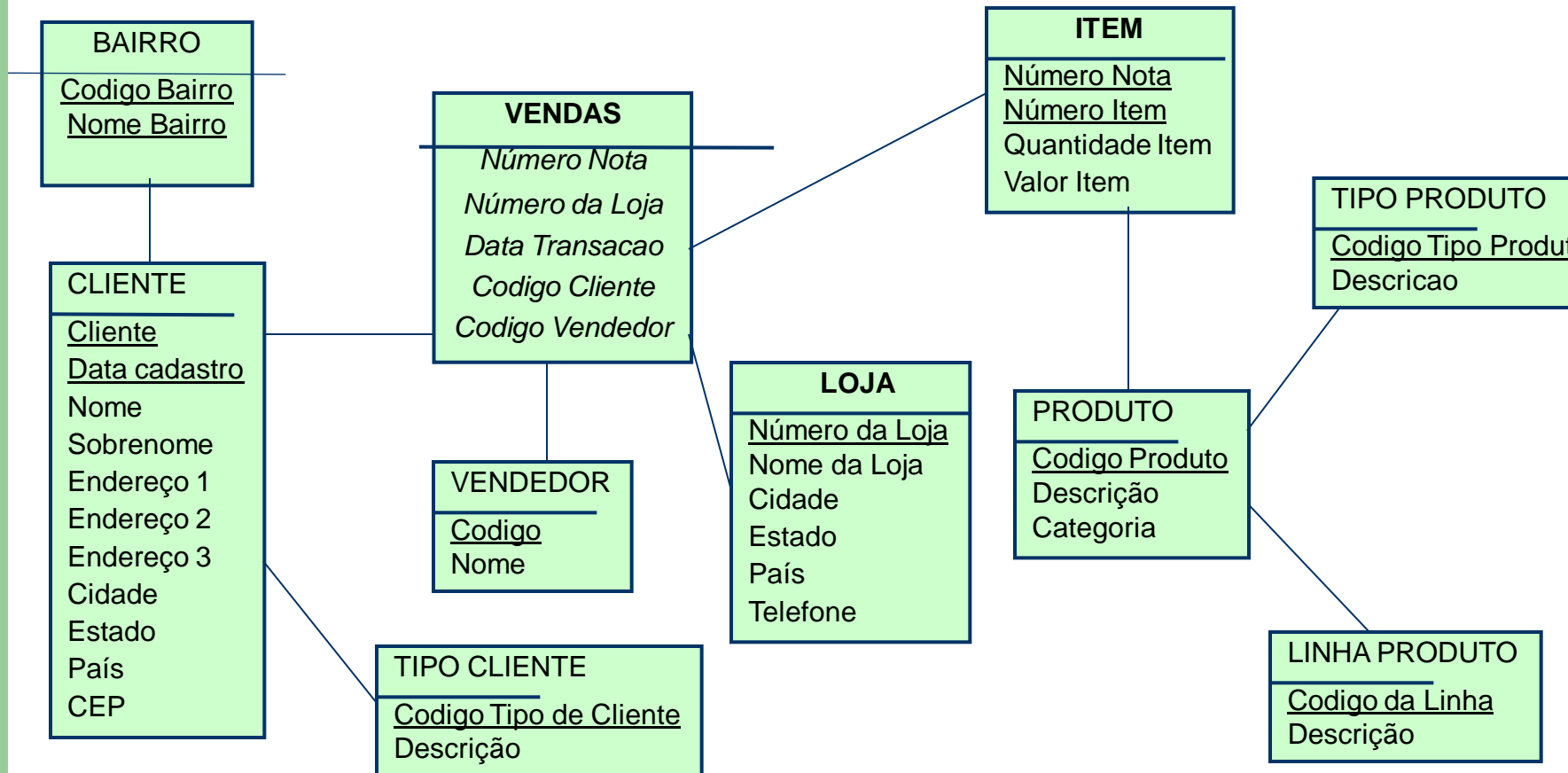
- Snow Flake



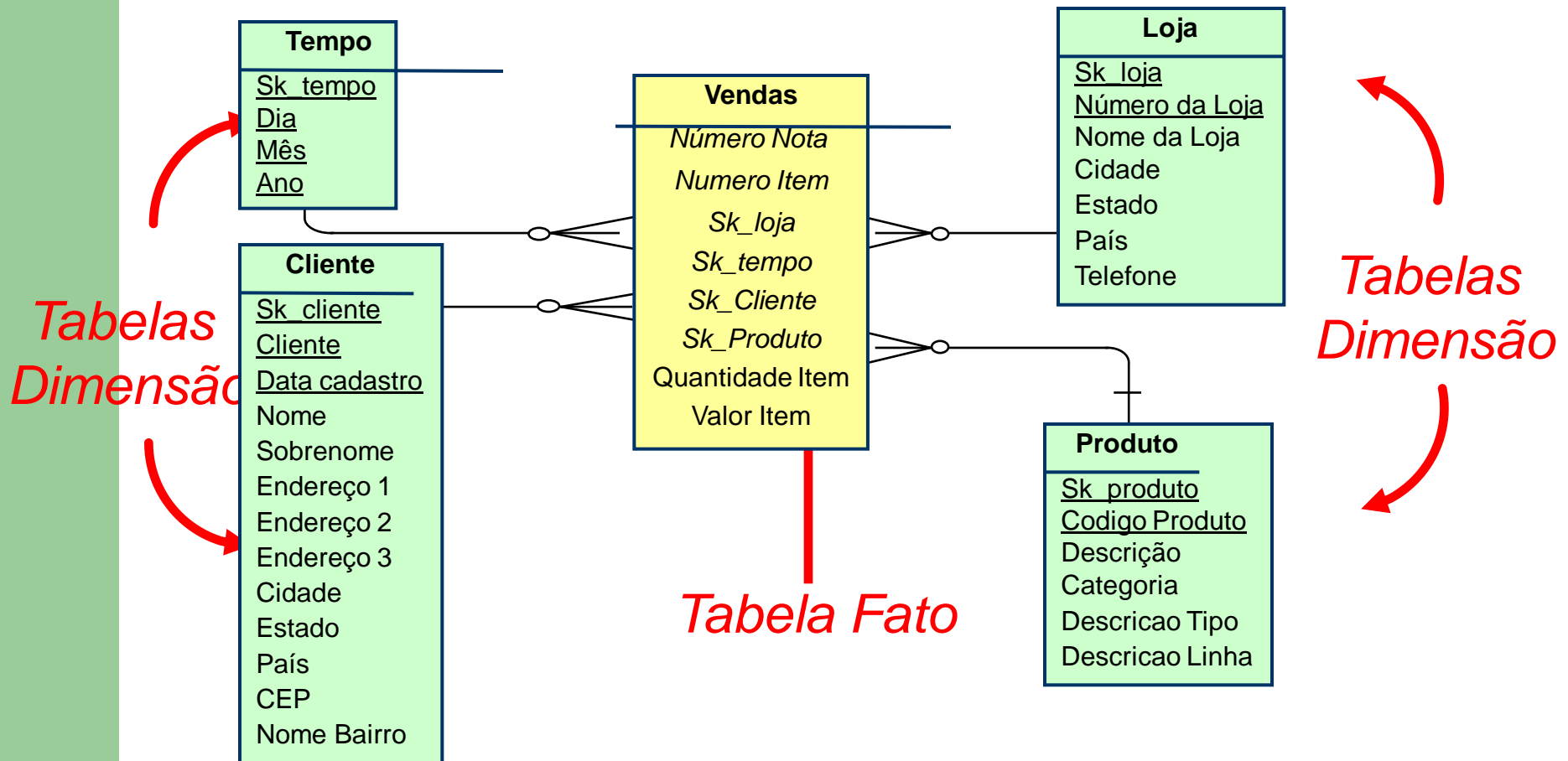
Data Warehouse – Modelagem Dimensional

- Existem 2 formas de implementação do modelo dimensional :
 - Modelo Star-Schema : Dimensões desnormalizadas (Alta performance, porém com requisitos de espaço de armazenamento em disco maior do que o modelo Snow-Flake)
 - Modelo Snow-Flake : Dimensões normalizadas (Performance menor porém com requisitos de espaço de armazenamento em disco menor do que o modelo Estrela)

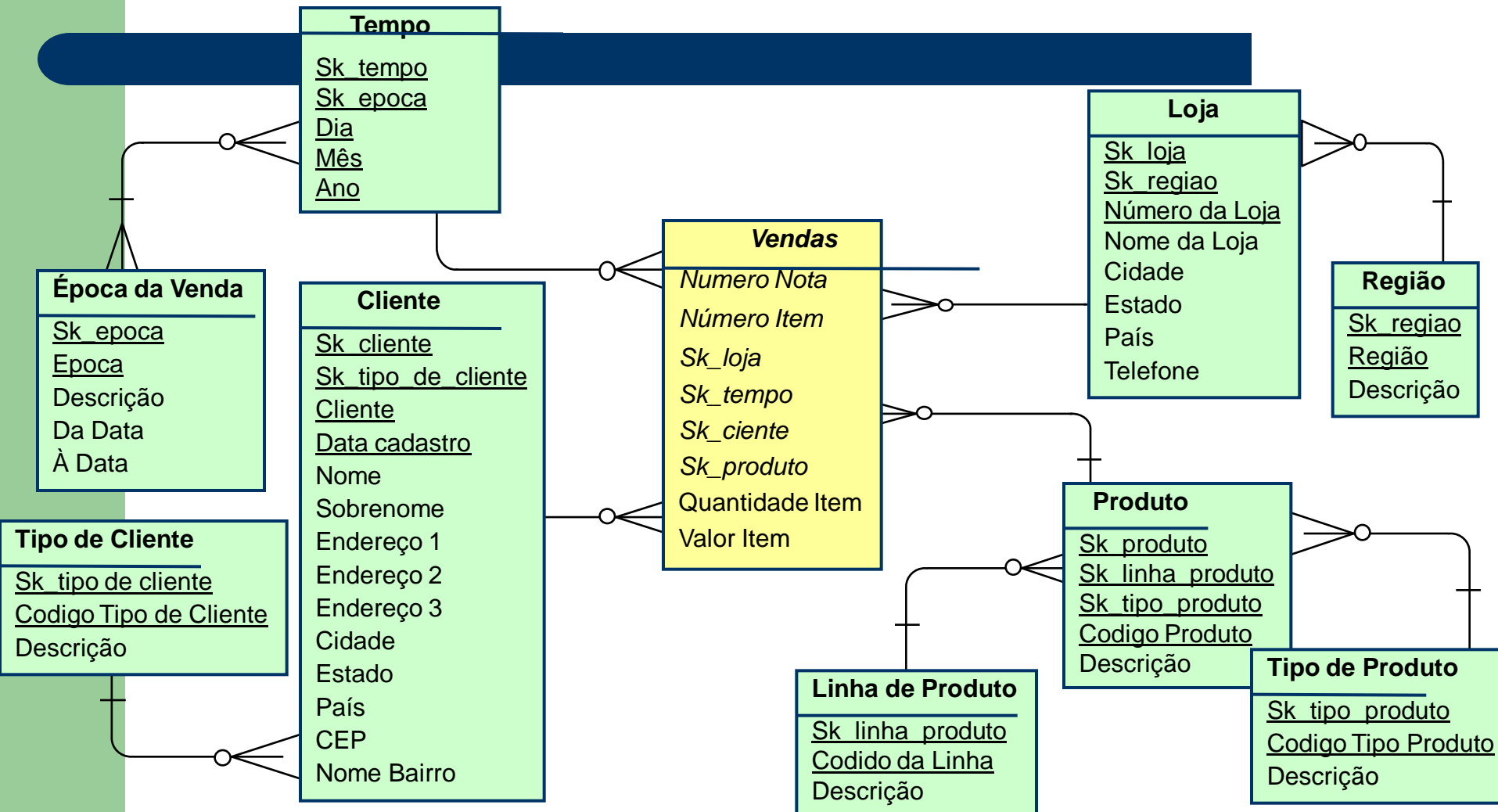
Tabelas Operacionais Vendas



Data Warehouse – Star Schema



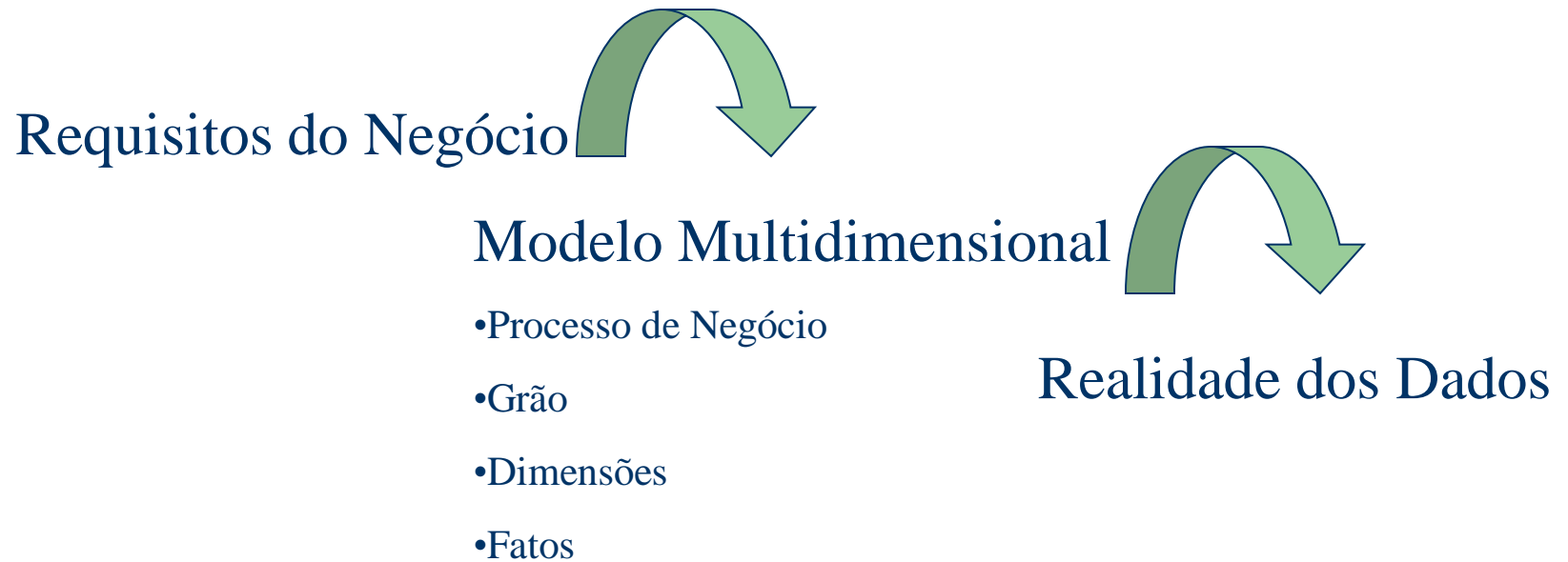
Data Warehouse – Snow-Flake



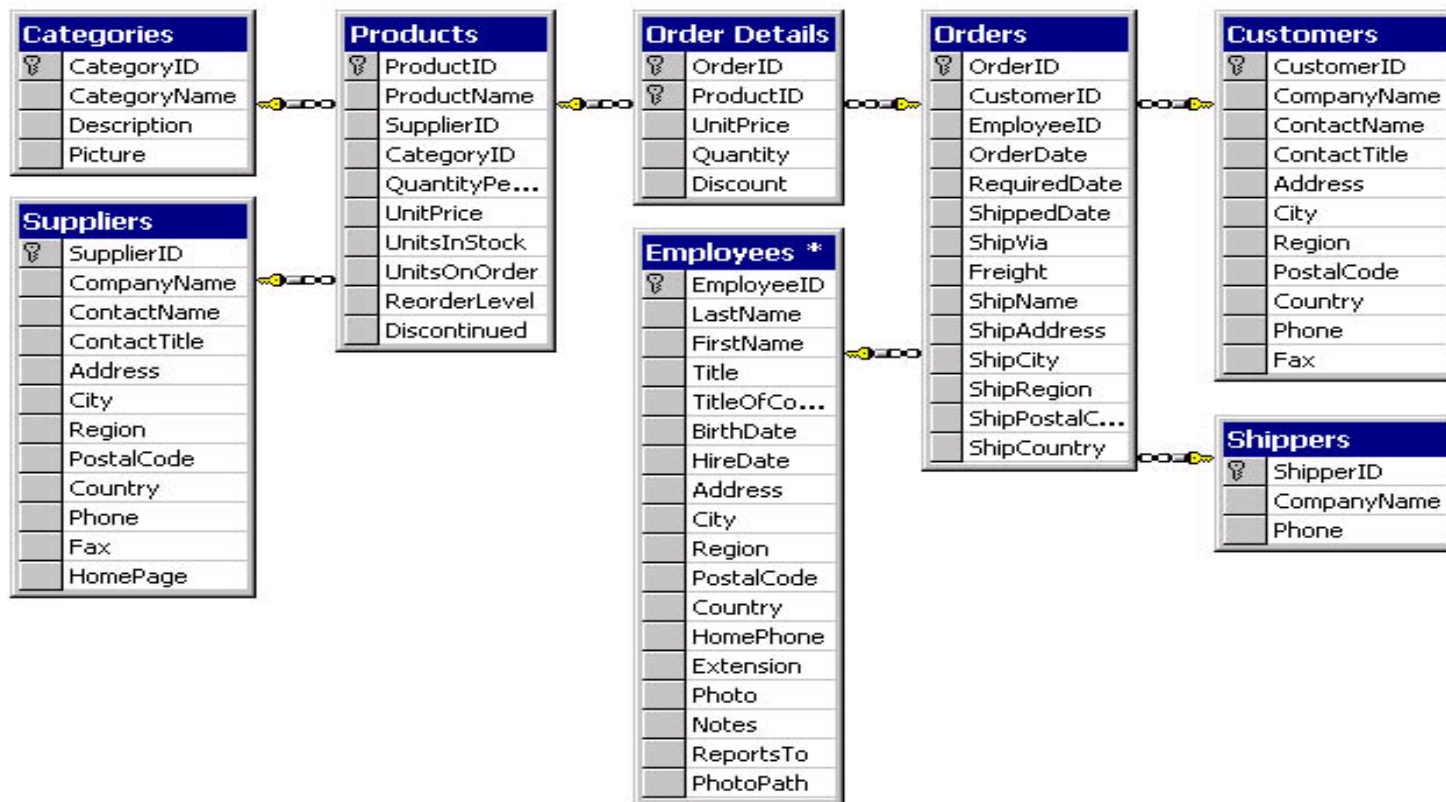
Passo a Passo

- Defina o assunto
 - Bom entendimento do negócio
 - Informações desejadas no nível de tomada de decisão
- Defina a granularidade da tabela fato
 - Identifique as dimensões
 - Identifique o nível hierárquico de cada dimensão
 - O grão influencia no tamanho da tabela.
- Identifique as métricas
- Identifique a periodicidade de carga, bem como o volume de dados a cada carga.
- Identifique possíveis agregados

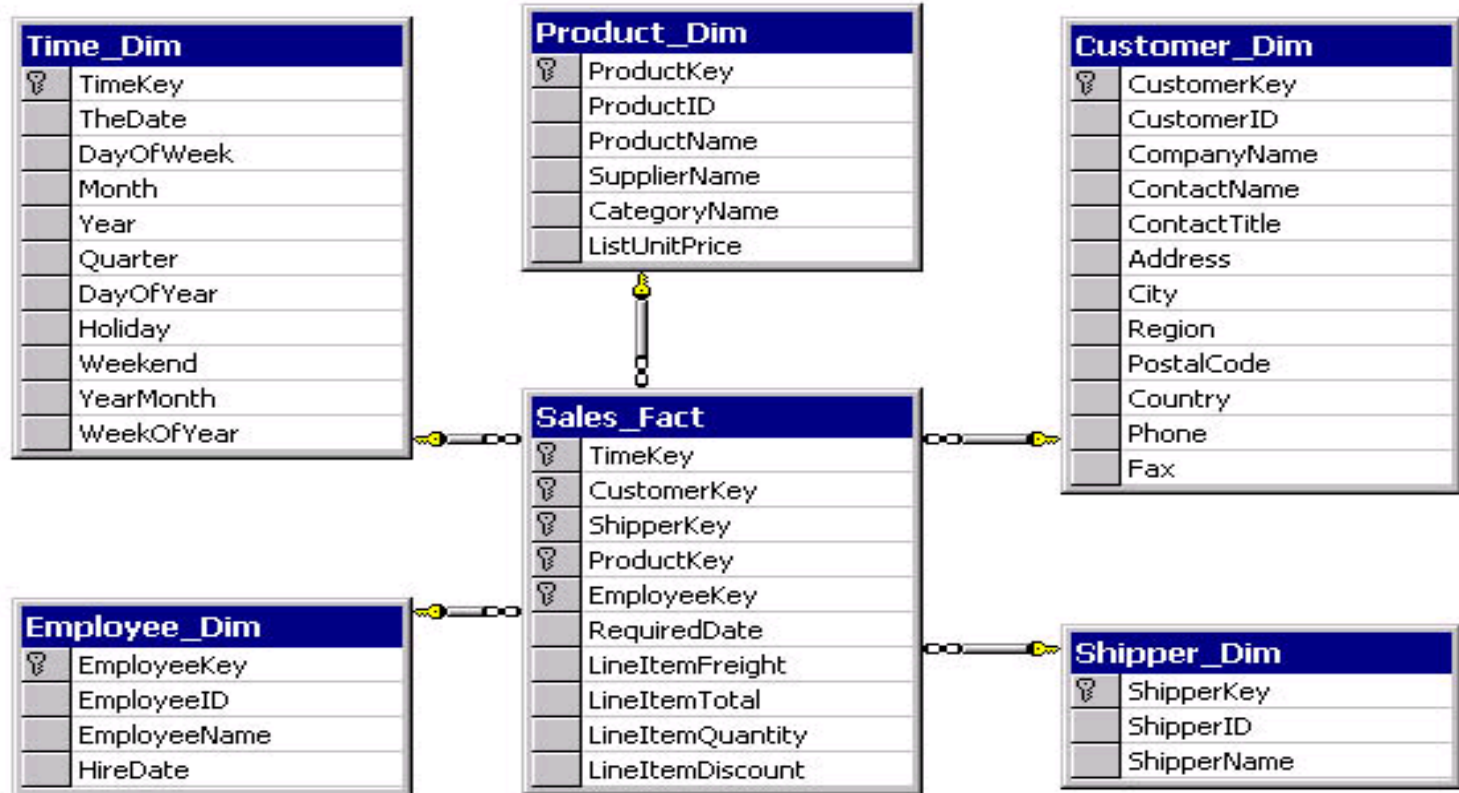
Processo de Criação Multidimensional



Exemplo Transaccional



Exemplo Multidimensional



Matriz Fatos x Dimensões

- Diagrama usado para documentar o data warehouse
- Nas colunas coloca-se as dimensões e nas linhas as tabelas de fatos.
- Assinala-se um “X” quando a tabela de fato possuir a dimensão em seu grão.
- Útil para avaliar o impacto de alterações nas dimensões e nos fatos.

Matriz Fatos x Dimensões

DIMENSÃO	D A T A	F O R N E C E D O R	P E Ç A	C O B E R T U R A	S E G U R A D O
Tabela de Fatos					
TRANSACOES DA APOLICE	X	X		X	X
PAGAMENTO DO SEGURADO	X				X
CLIENTE	X	X			
OPERACOES NA WEB	X			X	X

Também deve
ser feita uma
matriz

Fato x
Dimensão
e

Dimensão x
Fatos
Agregado.

Meta Dados

- Refere-se a todas as informações no ambiente de data warehouse (o que significa, com que dados se relaciona, domínio, etc.)
- Também responsável pela documentação do processo de carga, guardando a ordem de execução, relacionamento entre objeto (que tabelas são usadas por quais procedures, etc.)

Meta Dados

- Data Warehouse sem meta-dados significa o caos.
- Data Warehouse com meta-dados significa um caos ordenado e documentado.
- Deve ser alimentado a cada carga.
- O usuário final pode, e deve, acessá-lo.

Previendo o espaço gasto

Vendas diárias:

Dimensão Tempo: 730 dias (2 anos)

Dimensão Loja: 300 lojas

Dimensão Produto: 20.000 produtos (média de 10 por compra)

Dimensão Cliente: 100.000 clientes (média de 500 por dia)

Número de registros: $730 \times 300 \times 10 \times 500 = 1.095.000.000$

Tamanho médio do registro = 24 bytes

Tamanho da tabela fatos =

$1,095 \text{ milhões} \times 24 \text{ bytes} = 26 \text{ GB}$

Data Warehouse – ETL

- A partir da modelagem dimensional, cada campo de tabela deve ter sua origem de dados identificada
- Após a extração dos dados, podem ser necessários aplicação de processos de padronização, limpeza e consistência nos dados antes da carga no DW.
- Geralmente, este processo é realizado por Ferramentas ETL, que são ferramentas próprias para acessar dados em múltiplas plataformas, SGBDs e em diversos formatos.



Considerações

- SGBDMd são rápidos para consulta, mas apresentam restrições de volume, quantidade de campos.
- SGDBR são mais lentos para consultas, mas permitem grandes volumes de dados, quantidade ilimitada de campos.

Implementações do cubo

- MOLAP :
 - Modelo Multidimensional puro. Todas as informações são gravadas em cubos de dados
- ROLAP :
 - Modelo Multidimensional totalmente gravado em relacional
- HOLAP :
 - Modelo Híbrido. Fatos são gravados em relacional e agregações são gravadas em cubos

Data Warehouse – Criação dos Relatórios

- Esta etapa tem como atividade a construção das análises e consultas que proverão informações de suporte ao processo decisório, utilizando-se de todos os recursos das ferramentas OLAP, tais como : tabelas dinâmicas, gráficos, drills, e outros.
- Podem ser acessadas via modelo cliente-servidor ou através da Web – USO DE PORTAIS.



Data Warehouse – Pós-Implantação

- Acompanhamento das rotinas de atualizações e sua performance
- Acompanhamento do nível de utilização das informações pelos executivos
- Caso esteja aquém das expectativas, podem ser criados mecanismos para aumento do nível de utilização das informações pelos executivos como, por exemplo, envio de emails e outras formas de notificação.



Bibliografia Recomendada

- KIMBALL, Ralph - Data Warehouse Toolkit. Makron Books, São Paulo, 2004
- INMON, W.H. – Como Construir o Data Warehouse. Campus, Rio de Janeiro, 2005
- COREY, M. - Oracle 10g Data Warehouse. Campus, Rio de Janeiro, 2004

Bibliografia na Internet

- www.dwbrasil.com.br
- www.datawarehouse.com
- www.ralphkimball.com