

MODELAGEM PARA DATA WAREHOUSE

A serviço do suporte decisório
nas organizações

GRIMALDO OLIVEIRA

Programa

- Introdução
 - Panorama Geral
 - Planejamento Estratégico
 - Prática : Construção da Matriz de Necessidades
 - Prática : Construção da Fonte de Dados
 - Demonstração : Ferramentas para Tomada de Decisão
 - Compreendendo Business Intelligence
 - Compreendendo Data Warehouse (Armazém de Dados)
- Processo de Construção do Data Warehouse
 - Conceitos
 - Definir : o que são Dimensões
 - Definir : o que são Fatos

Produtos que Aprenderemos

Matriz de Necessidades – Fonte de Dados – Modelo Multidimensional

Finalidade: Coleta de dados com os gestores para a Construção do BI.

| Área | Função | Requisito | Detalhe | Observação |
|------------------|-------------------|--------------------------------|---------------------------------------|------------|
| Financeiro | Contabilidade | Relatório de vendas | Por cliente, por produto, por período | |
| Operacional | Produção | Relatório de produção | Por produto, por linha, por período | |
| Marketing | Marketing Digital | Relatório de marketing digital | Por canal, por produto, por período | |
| Logística | Logística | Relatório de estoque | Por produto, por local, por período | |
| Recursos Humanos | Recursos Humanos | Relatório de pessoal | Por departamento, por período | |
| Administrativo | Administrativo | Relatório de custos | Por departamento, por período | |
| Total | Total | Total | Total | Total |

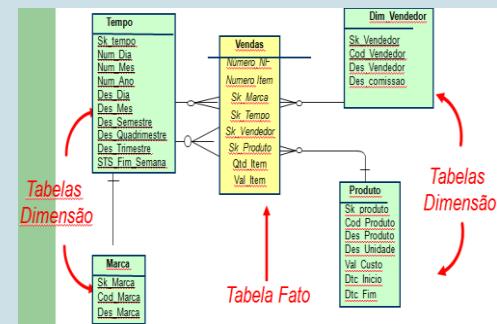
Matriz de
Necessidades

Finalidade: Levantamento dos relacionamentos, objetos,etc. Que armazenam os dados da empresa.

| ITEMS | DATA | Detalhe | |
|---------|----------------|----------|----------------|
| | | Venda | Detalhe |
| ITEMS | Código_produto | Chave_FK | |
| DETALHE | Descrição | | PPV/Devolvidos |
| ITEMS | PRODUTO | | PPV/Devolvidos |
| DETALHE | Nome | | |
| ITEMS | CLIENTE | | |
| DETALHE | Nome | | |
| ITEMS | Código_produto | Chave_FK | |
| DETALHE | Nome | | Chave_2 |
| ITEMS | PRODUTO | | Chave_2 |
| DETALHE | Nome | | Nome_Produto |
| ITEMS | CLIENTE | | Chave_2 |
| DETALHE | Nome | | Nome_Cliente |
| ITEMS | Número_NF | Chave_FK | |
| DETALHE | Numero_NF | | Chave_2 |
| ITEMS | PRODUTO | | Chave_2 |
| DETALHE | Nome | | Nome_Produto |
| ITEMS | CLIENTE | | Chave_2 |
| DETALHE | Nome | | Nome_Cliente |
| ITEMS | Data | | Chave_2 |
| DETALHE | Nome | | |

Fonte de Dados

Finalidade: Modelo adequado para realizar as consultas as bases que servirão ao BI



Modelagem
Multidimensional

Panorama Geral Business Intelligence



Panorama Geral - Business Intelligence



Sabe qual é a pergunta
mais comum entre
Diretores e Gerentes?

eis a pergunta...

**Como tomar decisões
de forma mais rápida?**



Panorama Geral - Business Intelligence

eis a resposta...



Use os dados
ao seu redor.

Panorama Geral - Business Intelligence

SIRH

PLANILHAS

COM
PESSOAS

COM
PARCEIROS

WEB
SITES

SISTEMAS
LEGADOS

COM
TERCEIROS

MEUS
SISTEMAS

onde estão
os dados?

Panorama Geral - Business Intelligence

**AS RESPOSTAS
ESTÃO NOS DADOS
QUE A SUA
ORGANIZAÇÃO
JÁ POSSUI...**



**...E QUE PRECISARÃO
SER ANALISADOS
PARA QUE SE
POSSA CONSTRUIR
INFORMAÇÕES
DE VALOR.**

Panorama Geral - Business Intelligence

MAS COMO POSSO TRANSFORMAR ESSES DADOS
EM INFORMAÇÃO PARA O MEU DIA-A-DIA?

Panorama Geral - Business Intelligence

QUAL O
FATURAMENTO
DO ULTIMO SEMESTRE?

QUANTO ESTAMOS
GASTANDO MENSALMENTE
COM MANUTENÇÃO DE
VEÍCULOS?

QUAL O PESO DESSE
AUMENTO SOBRE
A FOLHA DE PAGAMENTO?

QUANTOS COLABORADORES
IRÃO SE APOSENTAR
EM JANEIRO?

QUAL A NOSSA DESPESA
COM TAXI
SEMANALMENTE?

QUAL RODOVIA
REQUER MAIS
REPAROS DURANTE
O ANO?

QUAL O ALUNO
COM MELHOR
DESEMPENHO
ANUAL?

QUAL A ESCOLA
COM MAIOR
EVASÃO DE
ALUNOS?

COM PERGUNTAS QUE TE AUXILIEM NA TOMADA DE DECISÃO

QUAL A ESCOLA
COM MENOR
EVASÃO DE
ALUNOS?

QUAIS OS MUNICÍPIOS
QUE MAIS MELHORARAM
O SEU IDH?

QUAIS OS MUNICÍPIOS
COM MAIOR
ARRECADAÇÃO
BRUTA DE ICMS E IPVA?

QUAL A REGIÃO
ESTADUAL MAIS
CARENTE DE
HOSPITAIS?

QUAL MUNICÍPIO É
O MAIOR PRODUTOR
DE PESCADOS
NO ESTADO?

QUAL O MUNICÍPIO
QUE MAIS SE
DESENVOLVEU
APÓS O PROGRAMA
BOLSA FAMILIA?

QUAL BAIRRO
DA CAPITAL
É MENOS FAVORECIDO
DE ESCOLAS?

Panorama Geral - Business Intelligence

AGORA VOCÊ DEVE ESTAR DIZENDO...

Os sistemas daqui não se comunicam e as informações que preciso, demoram meses a chegar... Você está dizendo que existe uma solução chamada:

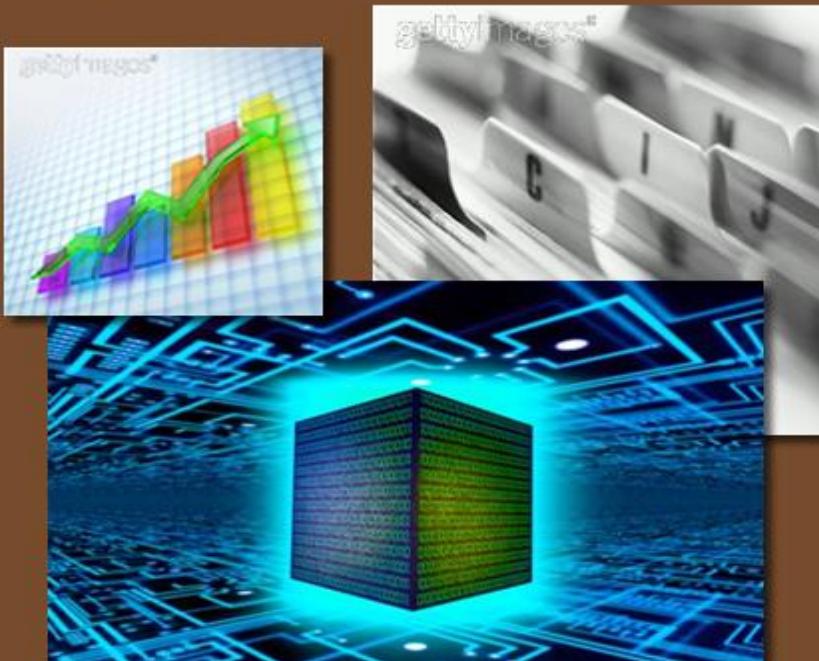
Integração Milagrosa de Sistemas!!!



Panorama Geral - Business Intelligence

...e como fazemos isso?

Montando um armazém de dados organizados e qualificados, com dados extraídos das fontes das respostas dadas às perguntas utilizadas como base de apoio à tomada de decisão.



Panorama Geral - Business Intelligence

Finalmente... mostre sua habilidade!

Construa de forma analítica, gráficos, relatórios e painéis de gestão com o simples ato de arrastar e soltar.

Os dados agora passam a ser informações valiosas.



Panorama Geral - Business Intelligence

tranquile-se...

Queremos apenas libertar sua habilidade em transformar dados e informações em conhecimento de negócio, para que você possa otimizar o seu processo corporativo de tomada de decisão.



O Que Necessito para trabalhar com BI?

Curioso

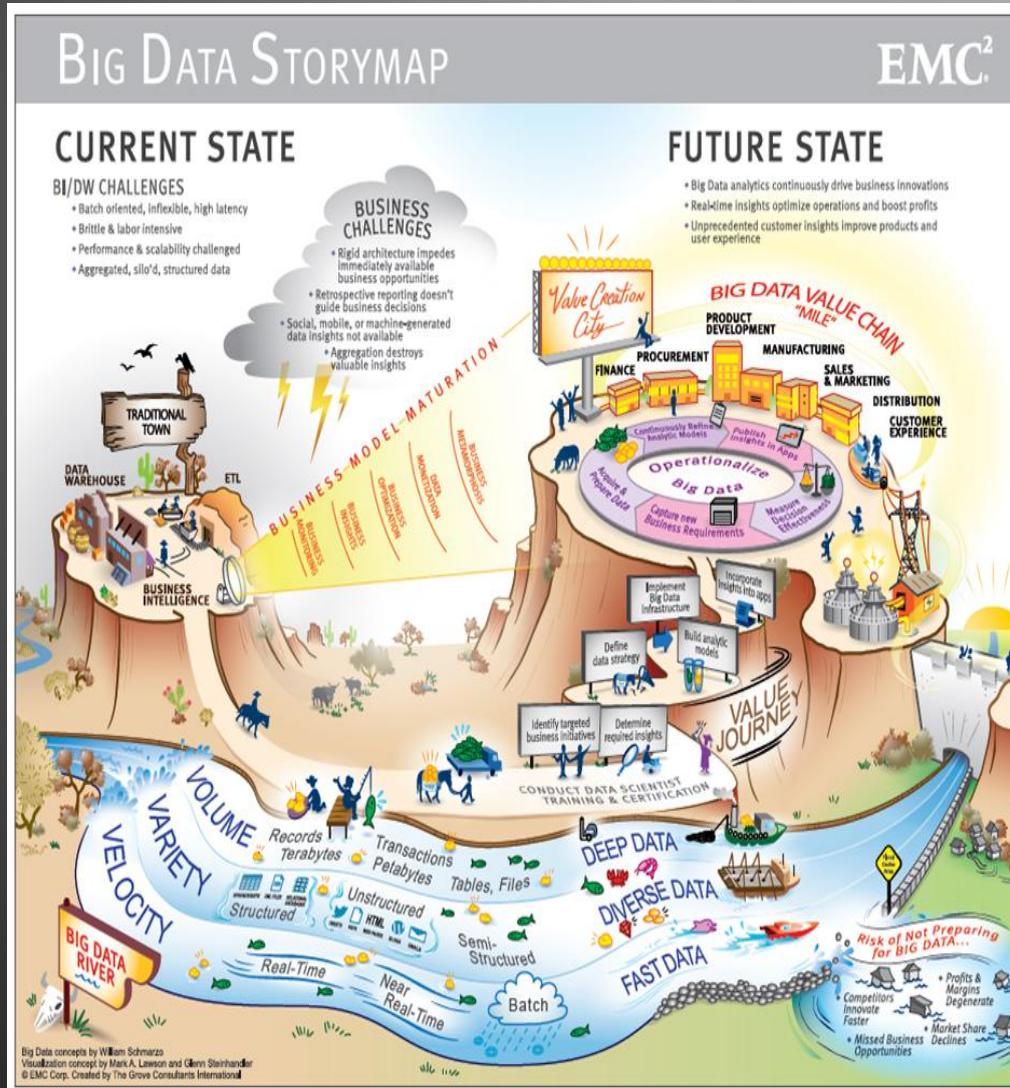
Interdisciplinaridade (negócio)

Banco de Dados, Modelagem de Dados,
Ferramentas OLAP, Ferramentas Data Discovery.

Conhecer Tecnologias



O Hoje e o amanhã da nossa área



Como Diferenciar DW X Big Data e Data Mining

Pergunta: Armazenar todos os bons e maus pagadores do Banco HSBC no Mundo



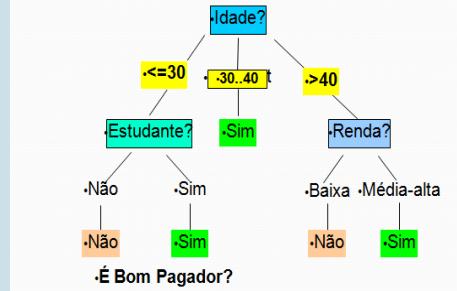
Big Data

Pergunta: Gerar Análise com os maus pagadores apenas nas capitais

| Nome | Idade | Renda | Profissão | Bom Pagador |
|--------|-----------|------------|------------|-------------|
| Daniel | ≤ 30 | Média | Estudante | Sim |
| João | 31..50 | Média-Alta | Professor | Sim |
| Carlos | 31..50 | Média-Alta | Engenheiro | Sim |
| Maria | 41..50 | Baixa | Vendedora | Não |
| Paulo | ≤ 30 | Baixa | Porteiro | Não |
| Otávio | > 60 | Baixa | Aposentado | Não |

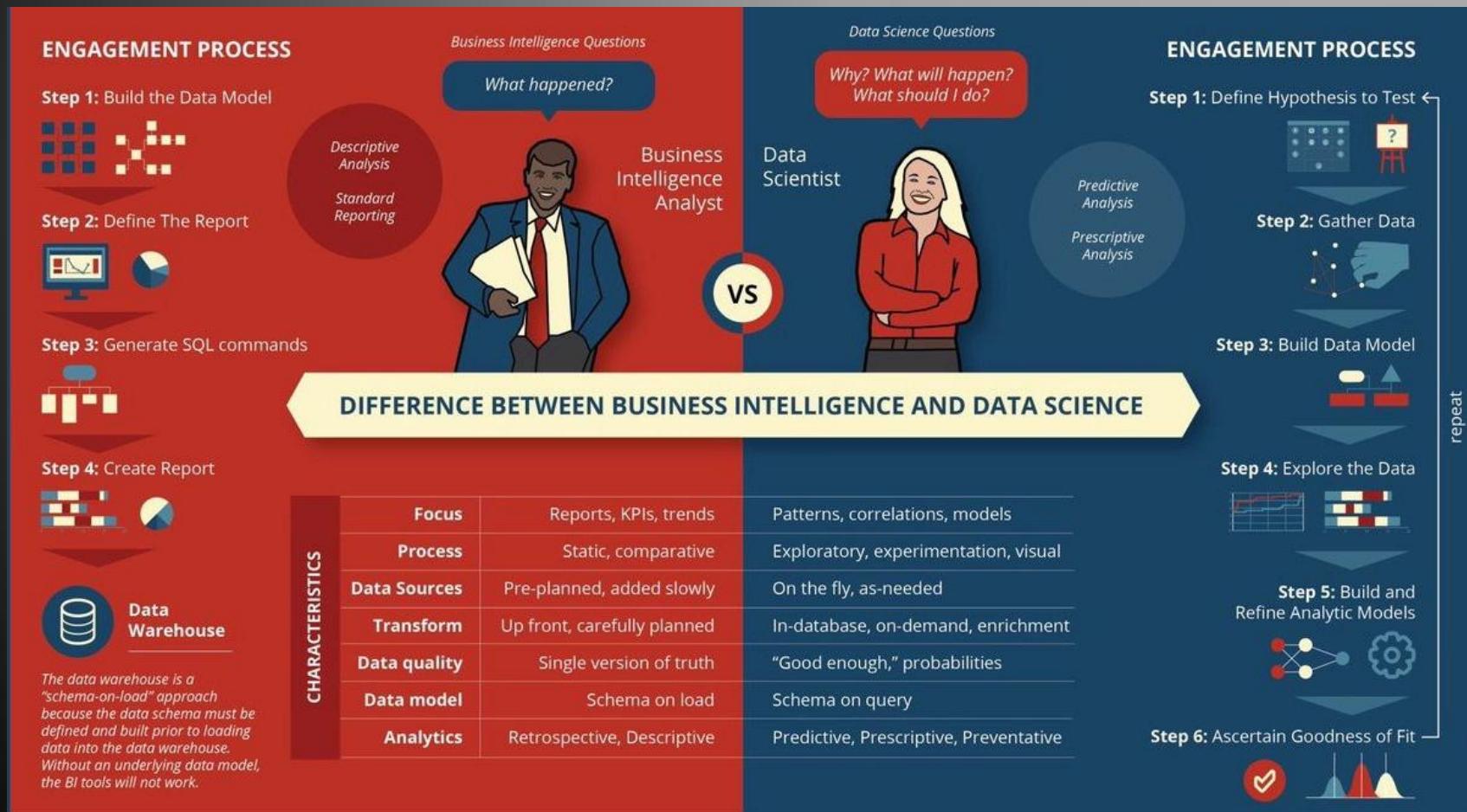
DW

Pergunta: João dos Santos pediu crédito em São Paulo, ele será um bom pagador?



Mineração de Dados

Como Diferenciar BI x Data Science



Planejamento Estratégico



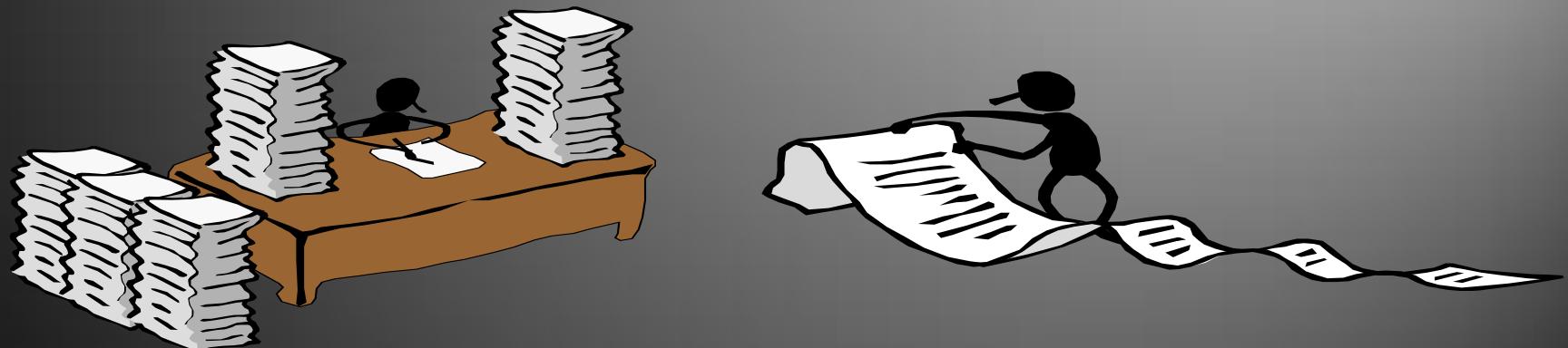
Começando a planejar...



- ▶ 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 ?

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



O processo decisório nas empresas

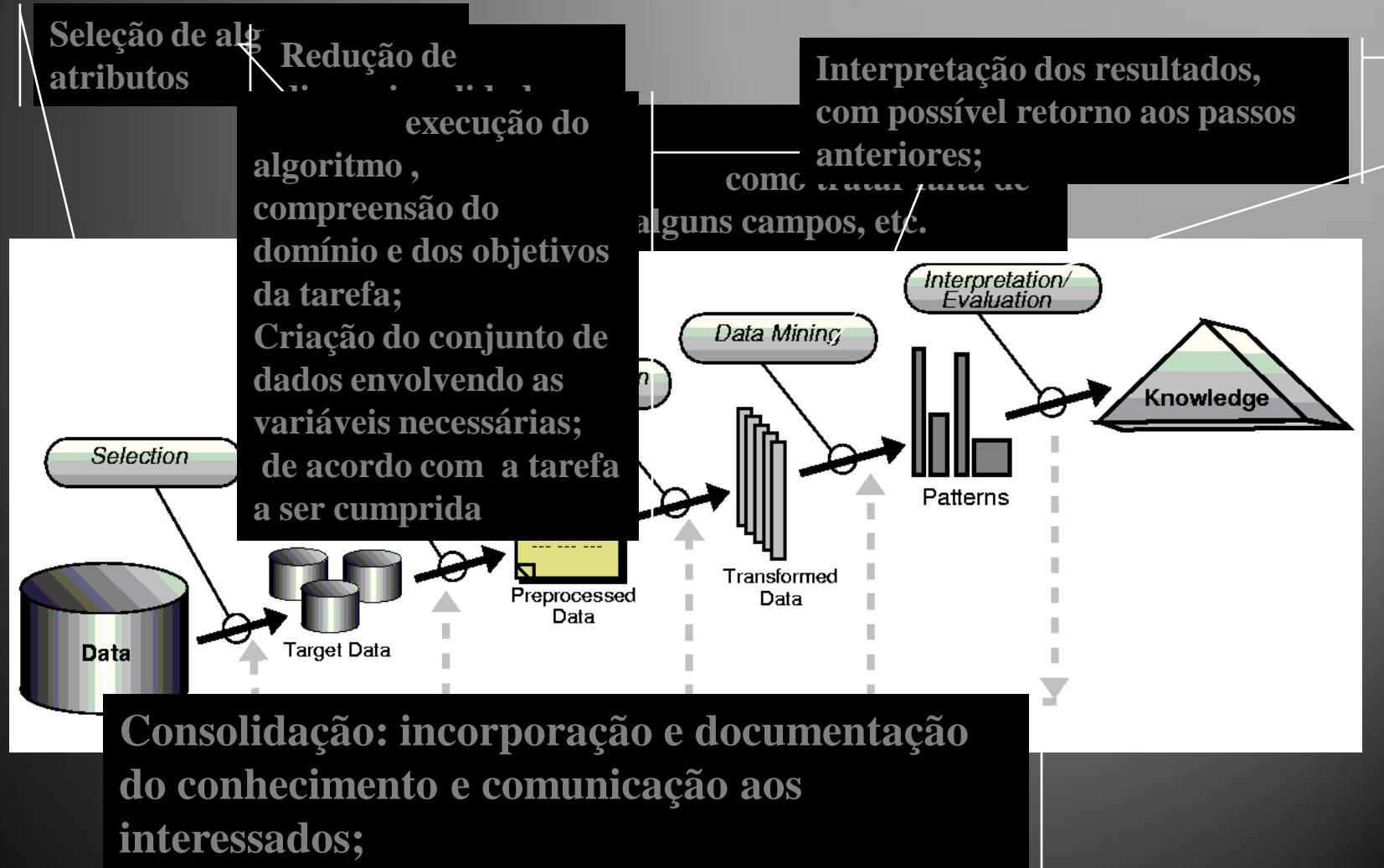


- ▶ Qual é a qualidade das informações para o executivo ?
 - Conflito de Conceitos
 - Falta de Integridade
 - Excesso de Papéis
 - Perda de tempo para composição das informações
- ▶ Contexto Atual : As informações disponibilizadas aos executivos não possuem confiabilidade

Comprendendo Business Intelligence



Processo mínimo de descoberta de conhecimento



Qualidade dos Dados

Os dados no mundo real estão “sujos”:

Incompletos

- ausência de atributos de interesse
- apenas dados agregados
- ausência de valores

Ruidosos

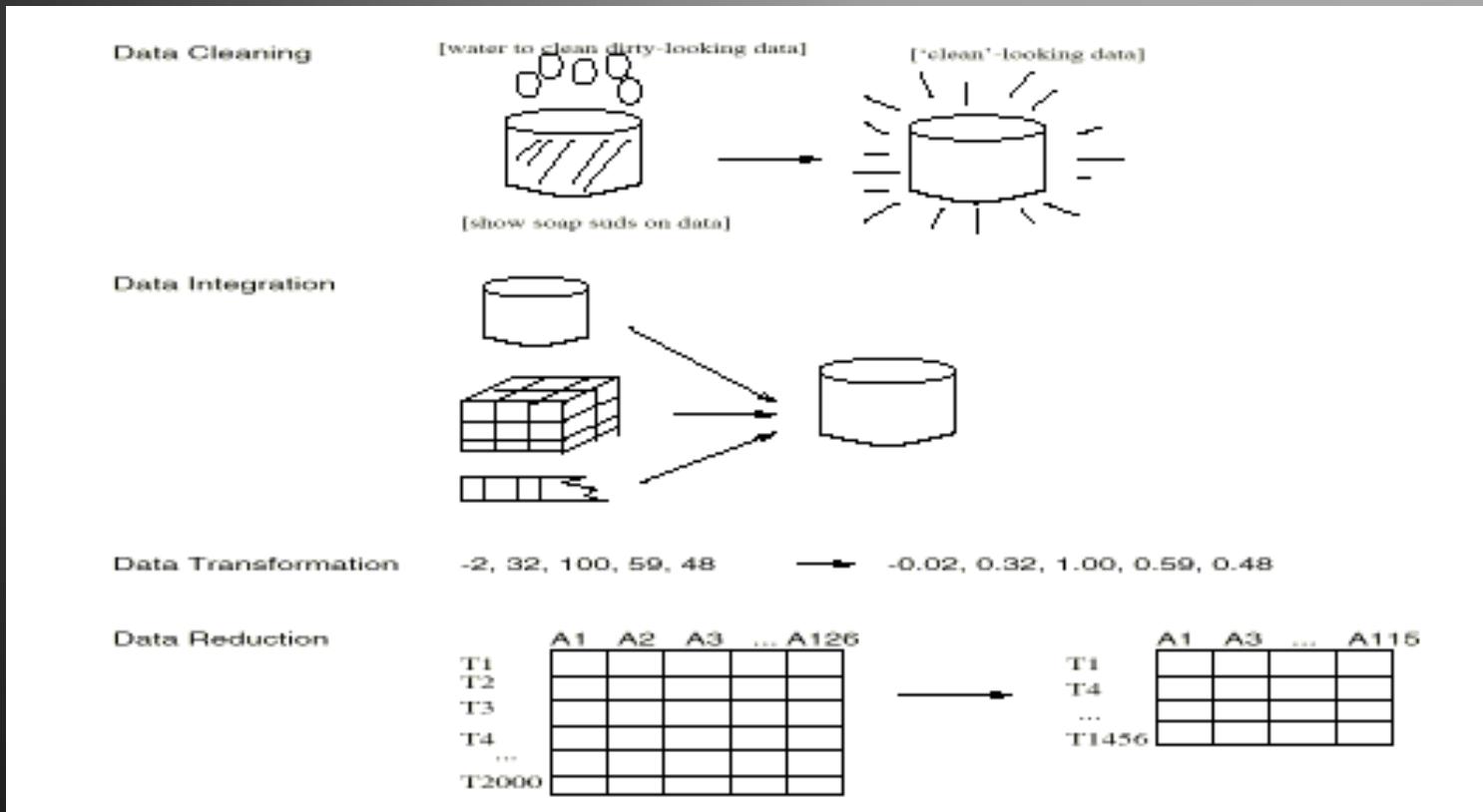
- erros aleatórios
- valores aberrantes (outliers)

Inconsistentes

- discrepâncias nas codificações ou nos nomes

Sem dados de boa qualidade o resultado é pobre

Pré-processamento dos Dados



Limpeza dos Dados

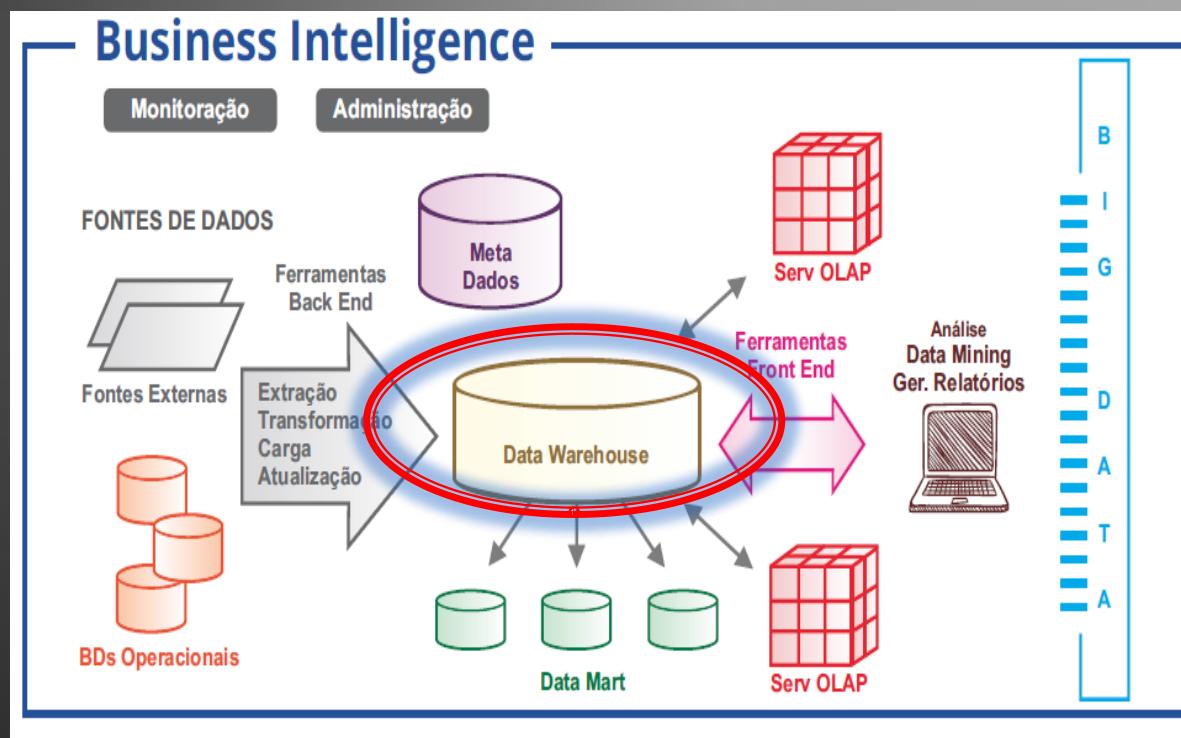
Em que consiste a “limpeza” dos dados?

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

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

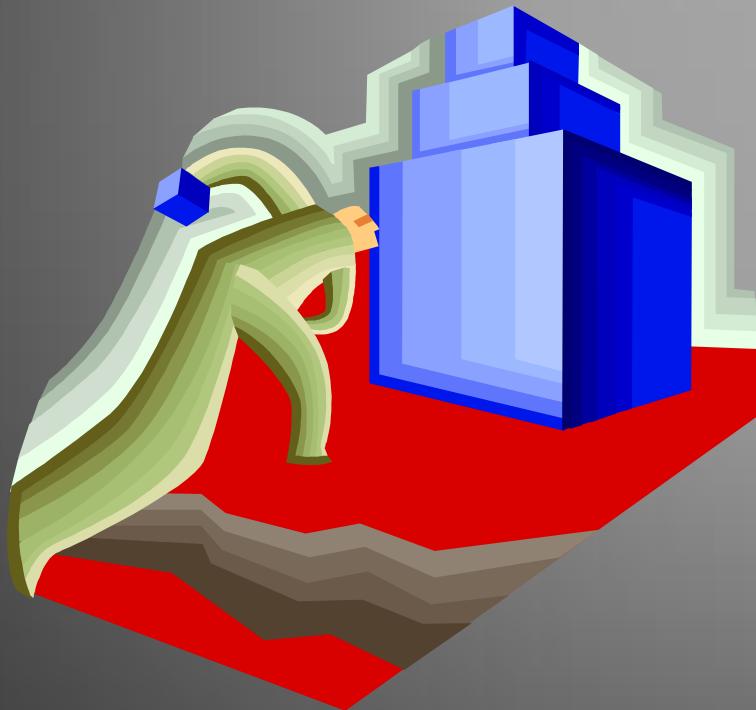


Business Intelligence – Objetivos

- ▶ 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



Compreendendo Data WareHouse



Data Warehouse – Conceito

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

- ▶ Orientado por assunto: contém 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: contém 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 – 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



Típica consulta em um BI (Data Warehouse)

A tabela apresenta informações do total do mês e a comparação percentual com o mês anterior das vendas de um produto.

Este relatório é uma summarização dos dados do DW Vendas.

| Produto | Filial | Vendas no Mês | Compração com o Mês Anterior |
|----------|----------------|---------------|------------------------------|
| Papel A4 | Salvador | 20.000 | 12% |
| Papel A4 | Campina Grande | 3.000 | -5% |
| Total | | 23.000 | 7% |

O que isto sugere?

Típica consulta em um BI (Data Warehouse)

Neste contexto, o executivo deseja saber a razão para os 5% negativos . Assim, ele executa uma operação acrescentando o atributo “Tamanho da embalagem” do produto.

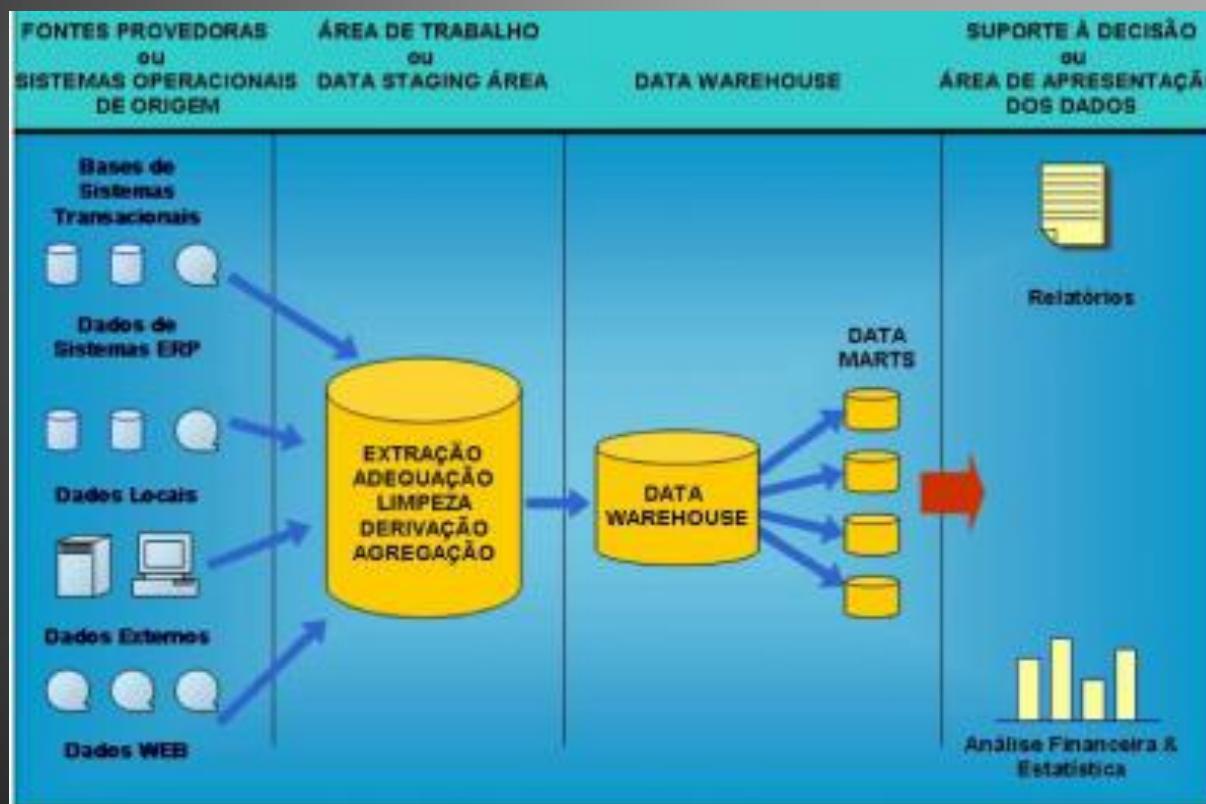
| Produto | Filial | Tamanho | Vendas no Mês | Comparação Com o Mês Anterior |
|--------------|----------------|---------|---------------|-------------------------------|
| Papel A4 | Salvador | A | 10.000 | 11% |
| Papel A4 | Salvador | B | 10.000 | 12% |
| Papel A4 | Total | | 20.000 | 12% |
| Papel A4 | Campina Grande | A | 1.000 | -4 |
| Papel A4 | Campina Grande | B | 1000 | -5 |
| Papel A4 | Campina Grande | C | 1.000 | -4 |
| Papel A4 | Total | | 3.000 | -5 % |
| <i>Total</i> | | | 23.000 | 7% |

Típica consulta em um BI(Data Warehouse)

O executivo com o resultado percebe que necessita investigar equipes de vendas, com isso descobre uma notável inconsistência, podendo agora tomar providências realmente efetivas.

| Produto | Filial | Equipe de Vendas | Vendas no Mês | Comparação Com o Mês Anterior |
|--------------|----------------|------------------|---------------|-------------------------------|
| Papel A4 | Salvador | X1 | 2.000 | 5% |
| Papel A4 | Salvador | Y1 | 18.000 | 19% |
| Papel A4 | Total | | 20.000 | 12% |
| Papel A4 | Campina Grande | X2 | 2.000 | 4% |
| Papel A4 | Campina Grande | Y2 | 800 | 1% |
| Papel A4 | Campina Grande | Z2 | 200 | -20% |
| Papel A4 | Total | | 3.000 | -5 % |
| <i>Total</i> | | | 23.000 | 7% |

O que desejamos atingir!!!



Consultas OLAP e Data Discovery

The collage consists of four windows:

- Processor BI - Centro Universitário - Acadêmico**: A screenshot of a web-based BI application showing a bar chart titled "Store Sales by Gender by Product by Education Level". The chart displays sales volume across different gender, product, and education level categories. Below the chart is a detailed hierarchical tree view of product categories like Meat/Poultry, Produce, and specific items such as Longlife Tofu and Manjimup Dried Apples, further broken down by country (Argentina, Austria, Brazil, Canada, Denmark, France, Germany, Ireland) and state/province (RJ, SP).
- Sum(Num)**: A dashboard featuring a horizontal bar chart with three segments (green, yellow, red) representing values from 0.0 to 10.0. Below it is a circular gauge with a green needle pointing towards the center. To the right is a large red wedge-shaped gauge with a scale from 0.0 to 8.0.
- Mapa de Entrega do Dia - Google Maps**: A screenshot of a Google Map showing delivery locations marked with red and pink pins across a city area. The map includes street names and various landmarks.
- Processor BI - Centro Universitário - Acadêmico**: A second screenshot of the same BI application interface, showing a table titled "Price Quantity Unit" with data rows corresponding to the products listed in the first window.

Ferramentas para Tomada de Decisão



Ferramentas de Mercado

- ▶ Open Source

- Jmagallanes
- Pentaho
- Talend (ETL)

- ▶ Proprietárias

- Power BI
- Tableau
- Microstrategy
- Qlik View e Qlik Sense
- Cognos
- SAP BW
- ODI – ORACLE (ETL)
- OBIEE

Processo de Construção de Data Warehouse



Necessário Saber!!

- ▶ Diferenciar os termos OLTP e OLAP
- ▶ Perguntas que desejam respostas!!!!
 - Como irei transferir os dados da base operacional para a base gerencial (Data Warehouse)?
 - É apenas copiar os dados ou necessito compreender o modelo de dados que serão armazenados?
 - O modelo de dados do Data warehouse é como se estivéssemos modelando para base de dados 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

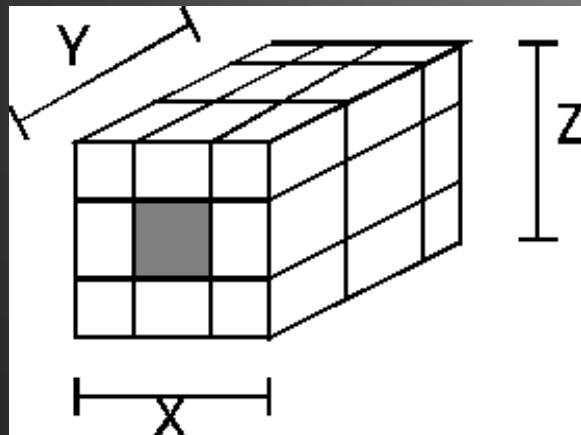
- ▶ *On Line Analytical Processing :*
Processamento ‘on line’ de análise.
 - Múltiplas respostas, Modelo Flexível.
 - Processamentos que suportam a tomada de decisões
 - Permite analisar tendências e padrões em grande quantidades de dados
 - ao longo do tempo (histórico)
 - e em diferentes localizações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



- ▶ Dados são 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

vs. orientados ao assunto

última versão dos dados

vs. *snapshots*

dados atualizáveis

vs. somente para leitura

desempenho é fator crítico

vs. não tão crítico

acesso orientado a linha

vs. orientado a conjunto

dados voláteis

vs. dados históricos

alta disponibilidade

vs. não tão alta

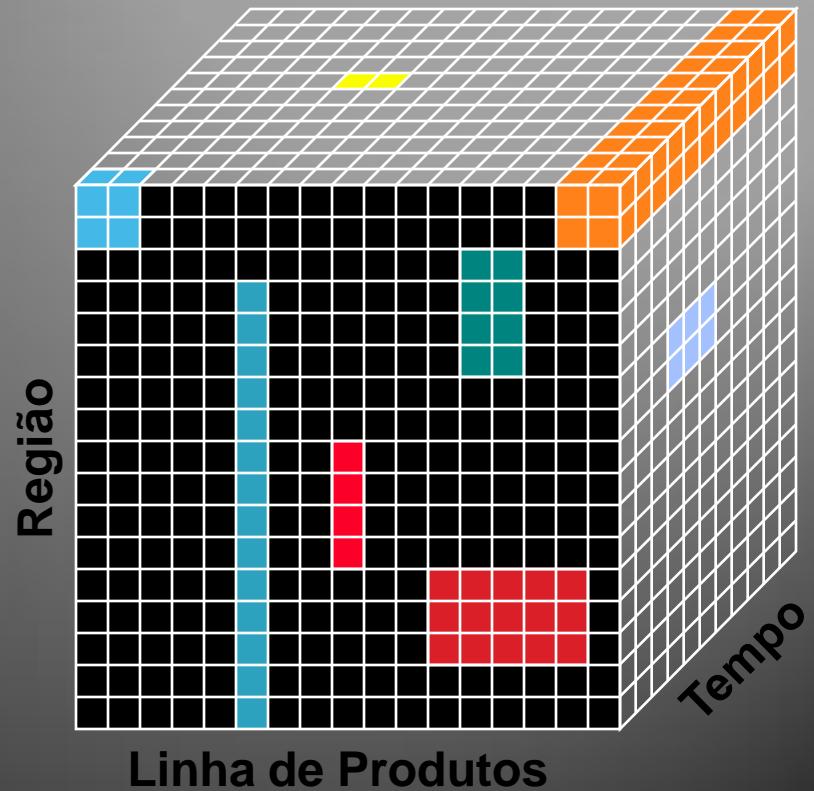
ausência de redundância

vs. redundância gerenciada

Operações OLAP

▶ Slice and Dice

- Ranging
- Ranking
- Drilling



Ranging (montado a consulta)

COR

TEMPO

CARRO

LOJA

QUANTIDADE

PÁGINAS

COLUNAS

L
I
N
H
A
S

CÉLULAS

Exemplo

Ano : 2010

Loja : STARCAR

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

Loja : CARPLUS

| | <i>Gol</i> | <i>Amarok</i> | <i>Polo</i> | <i>Golf</i> |
|---------------|------------|---------------|-------------|-------------|
| Azul | 23 | 43 | 23 | 93 |
| Prata | 36 | 98 | 76 | 76 |
| Branco | 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 |

U
P

| <i>Loja</i> | <i>Produto</i> | <i>Mês</i> | <i>Valor</i> |
|-------------|----------------|------------|--------------|
| Barra | BigMac | Jan/2000 | \$2200,00 |
| Iguaraci | 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 |

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)

Levantamento dos Dados



Produtos que Aprenderemos

Matriz de Necessidades – Fonte de Dados – Modelo Multidimensional

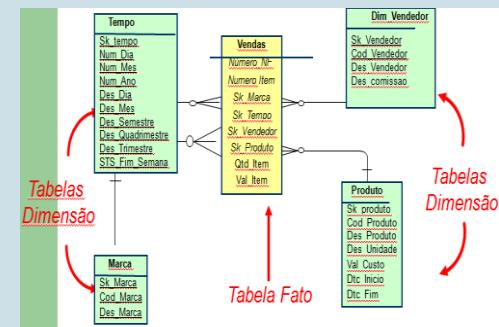
Finalidade: Coleta de dados com os gestores para a Construção do BI.

Matriz de
Necessidades

Finalidade: Levantamento dos relacionamentos, objetos,etc. Que armazenam os dados da empresa.

Fonte de Dados

Finalidade: Modelo adequado para realizar as consultas as bases que servirão ao BI



Modelagem
Multidimensional

Matriz Métricas x Descritores

- Diagrama usado para documentar o data warehouse
- Nas linhas coloca-se os descritores do negócio e nas colunas as métricas.
- Assinala-se um “X” quando a métrica for referente ao descriptor.
- Útil para avaliar o impacto de alterações do seu DW.

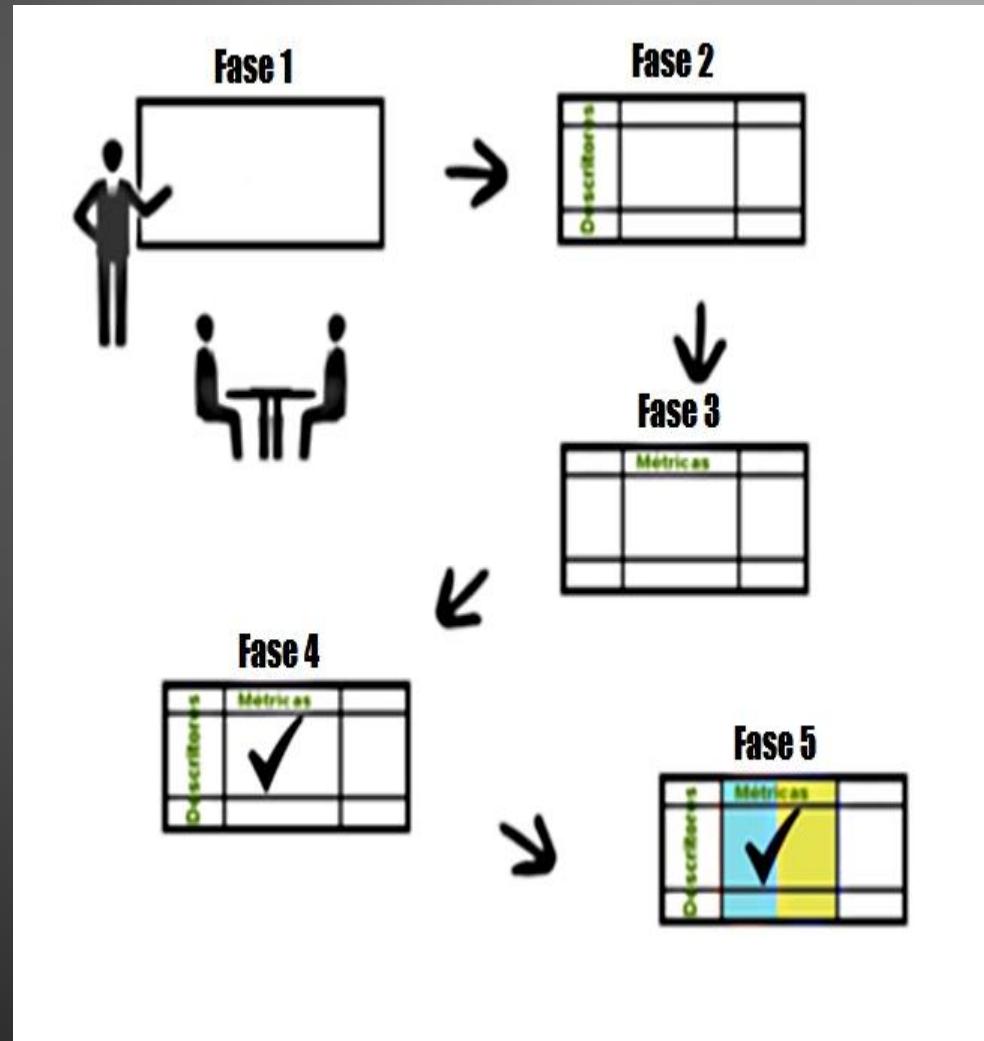
Matriz Métricas x Descritores

| Descritores do Negócio | | Métrica → | | | | |
|------------------------|------------|------------|-----------|-------|------------------|----------|
| | | TRANSAÇÕES | PAGAMENTO | TAXAS | OPERAÇÕES NA WEB | SEGURADO |
| ↓ | TEMPO | X | X | X | X | X |
| | FORNECEDOR | X | | | | |
| | PEÇA | X | X | | | |
| | CLIENTE | X | | | X | X |

Levantamento dos dados com os Gestores

- **Passo a Passo**
 - **Fase 1**– Convoque uma reunião com todos os gestores que utilizarão o BI.
 - **Fase 2**– Com um projetor e uma planilha eletrônica faça uma abertura explicando como será preenchida a Matriz de Necessidades.
 - **Fase 3**– Solicite que os gestores expliquem sobre as principais necessidades na análise de dados, indagando: quais métricas e descritores (valores e cruzamentos) vocês desejam analisar?
 - **Fase 4**– Comece a desenvolver a Matriz de Necessidades, completando com os dados informados pelos gestores.
 - **Fase 5**– Ao completar a Matriz de Necessidades com todas as informações discutidas na reunião, verifique se as perguntas dos gestores podem ser respondidas pelo documento criado.

Dinâmica de Construção da Matriz de Necessidades



Exercício

Construção da Matriz de Necessidades

- ▶ De acordo com o modelo de dados do Hotel Dallas, vamos montar uma Matriz Métricas X Descritores para o desenvolvimento do seu DW
- ▶ Assunto: Faturamento e Diárias do Hotel Dallas

Pense nas perguntas para a tomada de decisão

Matriz de Necessidades

- Perguntas dos gestores:
 - 1) Gostaria de analisar o **Total do Valor do Bar**, gasto pelo hóspede (**nome e cidade de origem**).
 - 2) Quero analisar o **Total de Diárias**, gasto pelo hóspede da **classe econômica**.
 - 3) Identifique os **hóspedes** do hotel Dallas, que vieram do **país 'Brasil'**.
 - 4) Solicito analisar a **Média do valor da hospedagem** de todos os hóspedes que são oriundos da **cidade de 'Salvador'** e que saíram do **aeroporto de 'Porto Seguro'**.
 - 5) Gostaria de realizar uma agregação de todos os valores passados ao governo da região com o **valor turismo**, separados por **país de origem** do hóspede e **cidade de origem** do hóspede.
 - 6) Gostaria que fosse exibido uma análise dos **dias da semana** que as pessoas mais se hospedam (diária) no Hotel Dallas, separados por **mês, bimestre e semestre**.
 - 7) Quero criar uma fórmula $\text{Receita} = \text{valor da hospedagem} + \text{valor do bar} - \text{valor do turismo}$ por **nome de hóspede**.

Fonte de Dados

- Diagrama usado para documentar os campos levantados na Matriz de Necessidades oriundas das bases transacionais da organização;
- Nas linhas coloca-se os descritores do negócio e os relacionamentos que existem no modelo, as métricas são identificadas e destacadas;
- Você deve ficar atento as chaves (PK) para mapear os dados corretamente;
- Útil para avaliar quantos descritores e métricas você terá no seu DW.

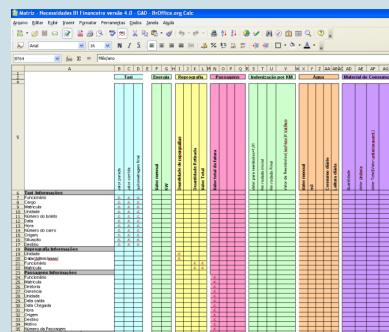
Fonte de Dados

| A | B | C | D | E |
|----|--------------------------|---|--|-------------|
| 1 | | | | |
| 2 | CLIENTE | FATO BENEFICIO | Obs: Campos sim/não se estiverem nulos não informado | |
| 3 | | | | |
| 4 | Dimensões | | Base Original | |
| 5 | | DDM | Atributo | Relac. |
| 6 | Informações Beneficiário | | | Domínio |
| 7 | | REGISTROATIVIDADEOPERACIONALSISTEMAPRODUCAO | idRegistroAtividadeOperacional | CHAVE-1 |
| 8 | | REGISTROATIVIDADEOPERACIONALBENEFICIADOS | idRegistroAtividadeOperacional | CHAVE-1 |
| 9 | | REGISTROATIVIDADEOPERACIONALBENEFICIADOS | idBeneficiado | CHAVE-2 |
| 10 | | BENEFICIADO | idBeneficiado | CHAVE-2 |
| 11 | CPF/CNPJ | BENEFICIADO | CPF | |
| 12 | Nome | BENEFICIADO | NOME | |
| 13 | Data Cadastro | BENEFICIADO | data_criacao | |
| 14 | Data nascimento | BENEFICIADO | dataNascimento | |
| 15 | Idade | BENEFICIADO | idade | Calculado |
| 16 | | BENEFICIADO | idCidadeNaturalidade | CHAVE-3 |
| 17 | | CIDADE | idCidade | CHAVE-3 |
| 18 | Naturalidade (município) | CIDADE | nome | UPPER(nome) |
| 19 | | CIDADE | idEstado | CHAVE-3-4 |
| 20 | | ESTADO | idEstado | CHAVE-3-4 |
| 21 | Naturalidade (Estado) | ESTADO | nome | |
| 22 | Renda Extra | | | |
| 23 | | REGISTROATIVIDADEOPERACIONALSISTEMAPRODUCAO | idRegistroAtividadeOperacional | CHAVE-1 |
| 24 | | REGISTROATIVIDADEOPERACIONALBENEFICIADOS | idRegistroAtividadeOperacional | CHAVE-1 |
| 25 | | REGISTROATIVIDADEOPERACIONALBENEFICIADOS | idBeneficiado | CHAVE-2 |
| 26 | CPF/CNPJ | BENEFICIADO | idBeneficiado | CHAVE-2 |
| 27 | | BENEFICIADORENDAAEXTRA | idBeneficiado | CHAVE-2-3 |
| 28 | | BENEFICIADORENDAAEXTRA | idTipoRendaExtra | CHAVE-2-3 |
| 29 | % de Renda | BENEFICIADORENDAAEXTRA | percentual renda | |
| 30 | Tempo | | | |
| 31 | Ano de Referência | | PERIODO_ANO | CHAVE |

Produtos que Aprenderemos

Matriz de Necessidades – Fonte de Dados– Modelo Multidimensional

Finalidade: Coleta de dados com os gestores para a Construção do BI.

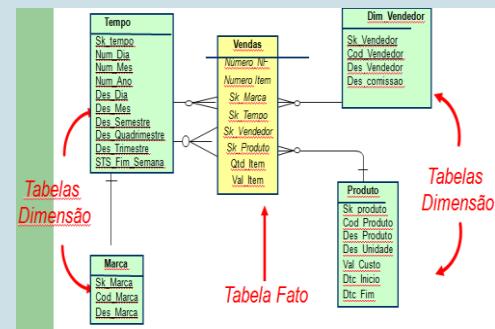


Matriz de Necessidades

Finalidade: Levantamento dos relacionamentos, objetos,etc. Que armazenam os dados da empresa.

Fonte de Dados

Finalidade: Modelo adequado para realizar as consultas as bases que servirão ao BI



Modelagem Multidimensional

Exercício

Construção da Fonte de Dados

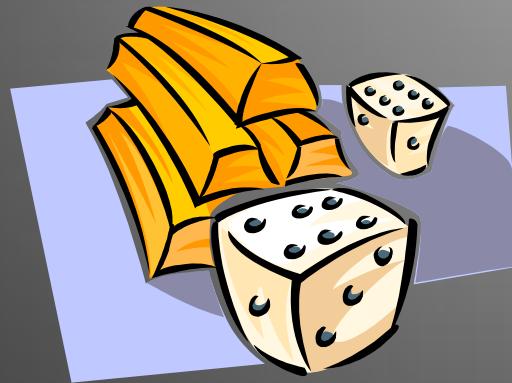
- ▶ De acordo com o modelo de dados do Hotel Dallas, vamos montar a Fonte de Dados para o desenvolvimento do seu DW;
- ▶ Assunto: Faturamento e Diárias do Hotel Dallas.

MODELO DE DADOS – HOTEL DALLAS

| <u>Hospede</u> | <u>Aeroporto Saída</u> | <u>Diaria</u> | <u>Tipo quarto</u> |
|----------------------|---------------------------|------------------------------|--------------------|
| Cod_hospede | Cod_aeroporto | Cod_diaria | Cod_Tipo_quarto |
| Nom_hospede | Des_aeroporto | Cod_hospede | Des_Tipo_quarto |
| Des_endereco | Nom_Localidade | Cod_tipo_quarto | |
| Cod_cidade | | Cod_classe_quarto | |
| Cod_aeroporto | | Val_diaria | |
| DtcCadastro | <u>Tipo classe quarto</u> | Qtd_diaria | <u>Faturamento</u> |
| | Cod_classe_quarto | Qtd_pessoas | Cod_faturamento |
| | Des_classe_quarto | Dtc_registro_primeira_diaria | Cod_hospede |
| <u>Cidade Origem</u> | | | |
| Cod_cidade | | | Dtc_entrada |
| Nom_cidade | | | Dtc_saida |
| Cod_pais | | | Val_hospedagem |
| <u>Pais Origem</u> | | | Val_turismo |
| | | | Val_bar |
| | | | |

Desenvolvimento do Armazém de Dados

Data WareHouse



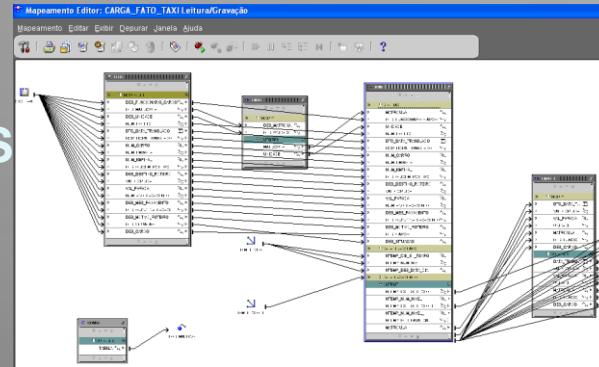
COMO CRIAMOS O DW

CONSOLIDA OS
DADOS

MATRIZ DE
NECESSIDADES

Matriz - Necessidades BI Financeiro versão 4.0 - GAD - BrOffice.org Calc

| | Taxi | Externa | Respiratoria | Passeiros | Identificação por KM | Aux | Material de Consumo |
|----------------------------------|-------------------------------|--------------------|--------------|-----------|----------------------|---------|---------------------|
| 1. Inf Informações | pequeno pequeno pequeno | pequeno pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 2. Carga | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 3. Materiais | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 4. Unidade | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 5. Utilizado no Toilette | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 6. Data | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 7. Endereço | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 8. Número do carro | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 9. Origem | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 10. Situação | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 11. Destino | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 12. Reservadas Informações | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 13. Descrição (Passageiro) | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 14. Materiais | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 15. Passeiros Informações | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 16. Localização | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 17. Materiais | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 18. Passeiros | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 19. Identificação | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 20. Material | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 21. Local | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 22. Endereço | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 23. Data saída | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 24. Hora Chegada | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 25. Origem | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 26. Destino | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 27. Atividade | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 28. Número da Passageiro | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |
| 29. Identificação km Informações | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno | pequeno |

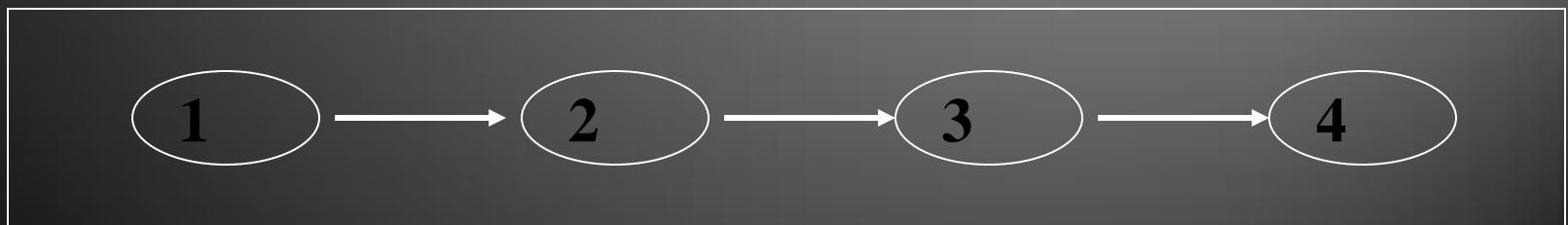


FERRAMENTA OLAP e
DATA DISCOVERY



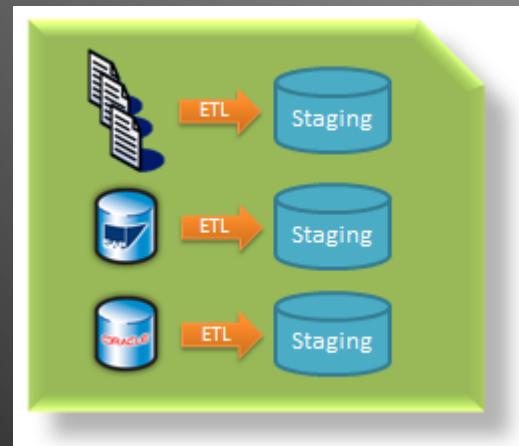
Data Warehouse – Principais Etapas do Desenvolvimento

- ▶ Identificação dos indicadores
- ▶ Transferir dados do Operacional
- ▶ Modelagem Dimensional
- ▶ ETL (Carga)
- ▶ Criação dos Relatórios (Ferramenta OLAP e Data Discovery)
- ▶ Pós-Implantação



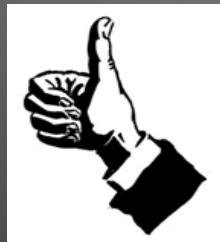
Transferência dos dados do operacional : Staging Area

- 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.



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.
- ▶ Montagem da Matriz de Necessidades e Fonte de Dados.



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 : métricas dos processos de negócio que devem ser analisadas. Ex : vendas, faturamento, despesa, estoque.

Dados Multidimensionais

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

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

Dimensões

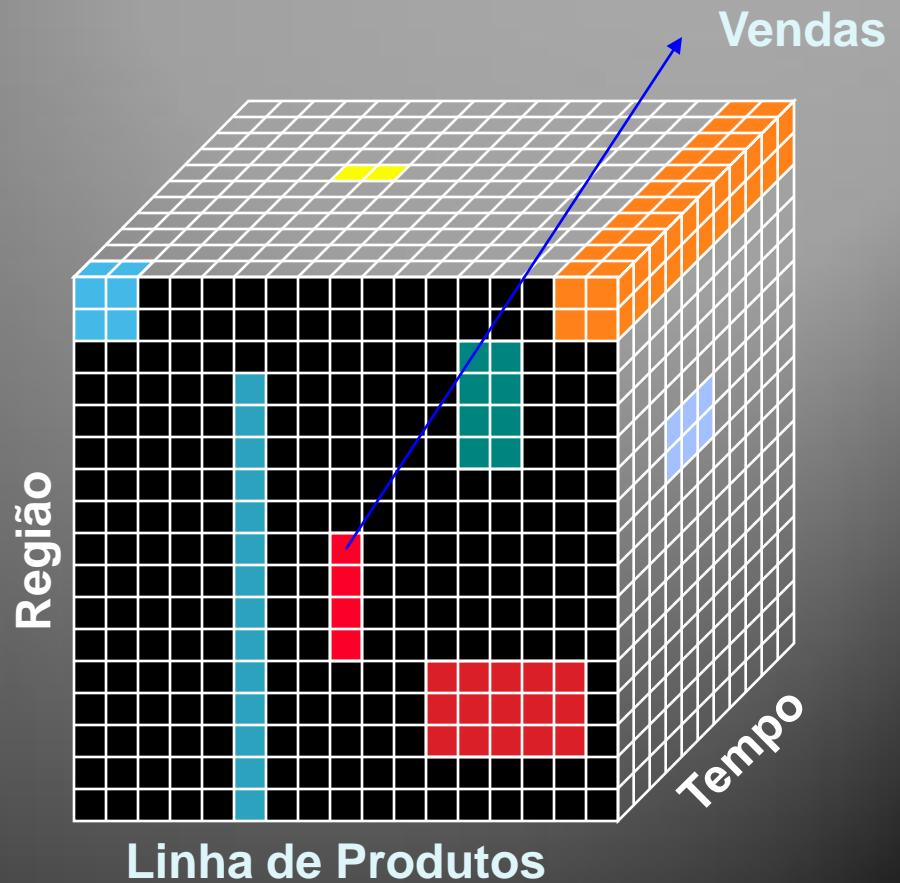
- Contém os descritores textuais do negócio
- Representam objetos físicos do mundo real, locais, conceitos, etc.
- São os eixos do cubo
- Exemplo : Tempo , Cliente, Produto, Tipo de Embalagem, Situação, etc.

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.

Identificando Dimensões e Fatos

- ▶ Dimensões
 - Região
 - Linhas de Produto
 - Tempo (obrigatória)
- ▶ Fato
 - Vendas



Fundamentos Básicos: Modelagem Multidimensional

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

Desnormalização

- ▶ Só a 1°FN 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

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 |

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

Tipo 1 : Sobrescrever os Dados

- Exemplo seria o de uma tabela de banco de dados que substitui 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 | |

Exemplo : DIM_CLIENTE

Seq_cliente

Cod_cliente

Nom_cliente

Cod_tipo_cliente

Des_tipo_cliente

Dtc_inicio

Dtc_fim

Sts_corrente

Em azul: 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.

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 acrescenta campos 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

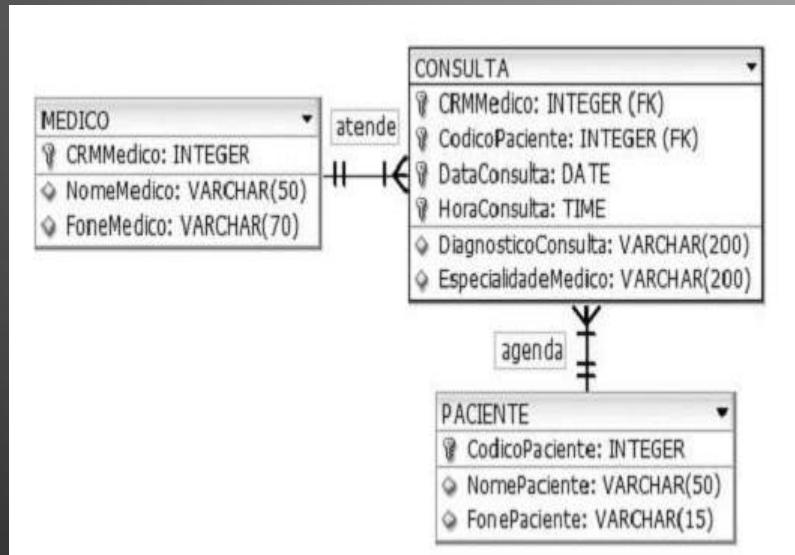
- Existem apenas como referência a uma entidade
- Exemplo : Considere que uma tabela de fato com os itens das notas fiscais de uma empresa varejista. O que fazer com o número da nota fiscal ? Ele em si, não representa nada. Apenas serve para agrupar os itens de uma mesma compra. Não existe fisicamente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

Dimensão Ponte (Bridge Table)

- Uma tabela com chave composta capturando um relacionamento muitos–para–muitos que não possa ser acomodado pela granularidade natural de uma tabela de fatos ou tabela de dimensão. Serve como uma ponte entre a tabela de fatos e a tabela de dimensão de forma a permitir dimensões multivaloradas.



Outro Exemplo: Correntista e conta corrente de um banco

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 compõe 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

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



Exemplo: FATO_Arrecadacao



*Chaves das
dimensões que
medem as métricas.*

Métricas.

Tabelas Fato

Fato Aditivas : Todas as dimensões podem ser utilizadas para summarizações.

Fato Semi-Aditivas : Algumas dimensões podem ser utilizadas para summarizações.

Ex.:Saldo Bancário – Faz sentido somar o seu saldo caso ele tenha conta em mais de um banco, mas não faz sentido somar seu saldo todos os dias.

Fato Não-Aditivas : Nenhuma dimensão pode ser utilizada para summarização.

Ex.:% margem de lucro

Tabelas Fato sem fatos

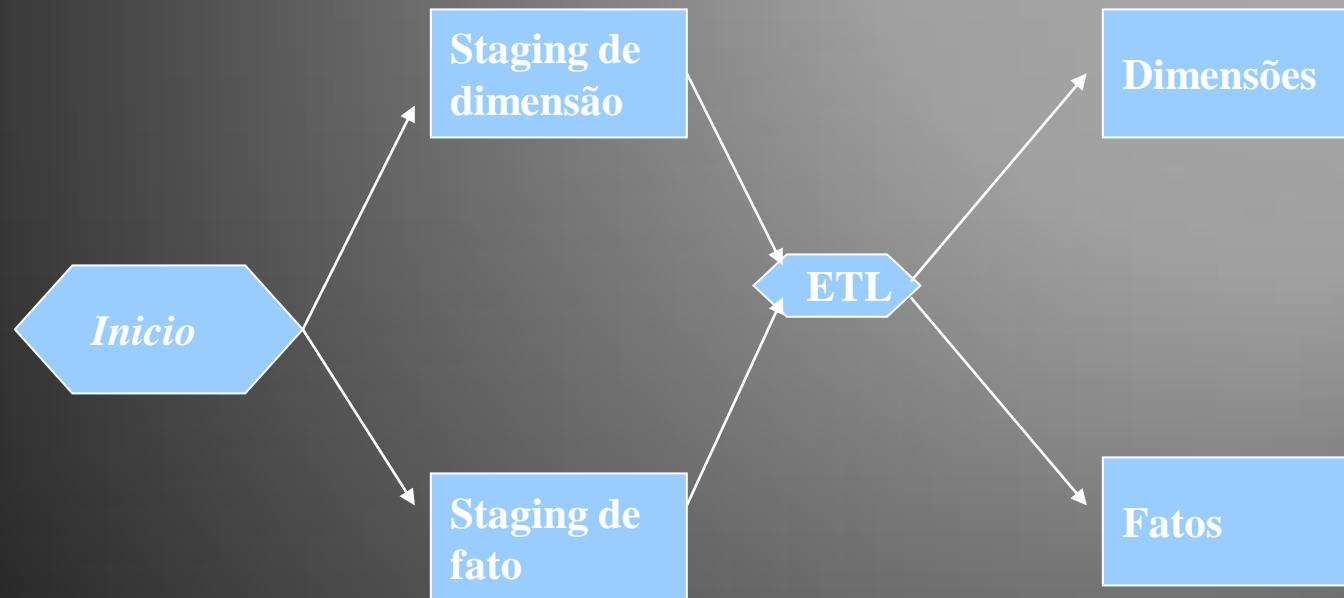
Uma tabela de fatos que não tem fatos captura alguns relacionamentos muitos-para-muitos entre chaves de dimensões.

Ex: Uma tabela de fatos, tipicamente sem fatos, que registra todos os produtos que estão em promoção numa determinada loja, independentemente de ser vendidos ou não. Consulta: Quais produtos estavam em promoção mas não venderam?

Grão

- ▶ Conceito que identifica a unidade de medida das métricas.
- ▶ Nível de detalhe dos dados.
- ▶ Menor Grão: Mais detalhe -> Mais dados->Análise mais longa ->Informação mais detalhada-> Mais “grão” de dados.
- ▶ Maior Grão: Menos detalhe -> Menos dados ->Análise mais rápida ->Informação menos detalhada-> Menos “grão” de dados.

Seqüência obrigatória para execução das cargas



Produtos que Aprenderemos

Matriz de Necessidades – Fonte de Dados – Modelo Multidimensional

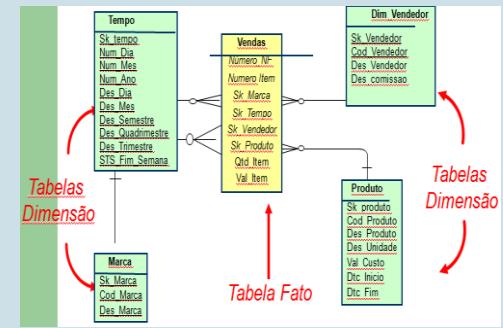
Finalidade: Coleta de dados com os gestores para a Construção do BI.

Matriz de
Necessidades

Finalidade: Levantamento dos relacionamentos, objetos,etc. Que armazenam os dados da empresa.

Fonte de Dados

Finalidade: Modelo adequado para realizar as consultas as bases que servirão ao BI

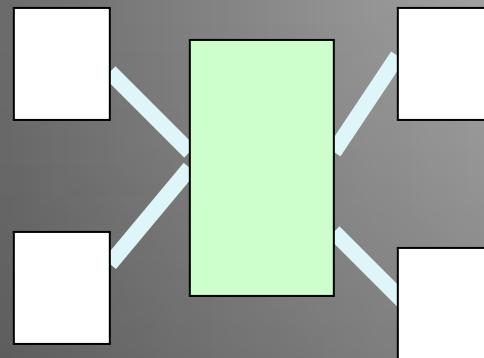


Modelagem
Multidimensional

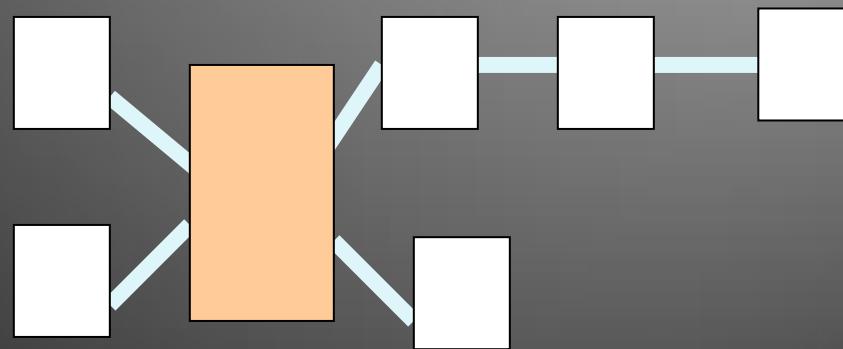
Modelagem Multidimensional

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

- Star Schema



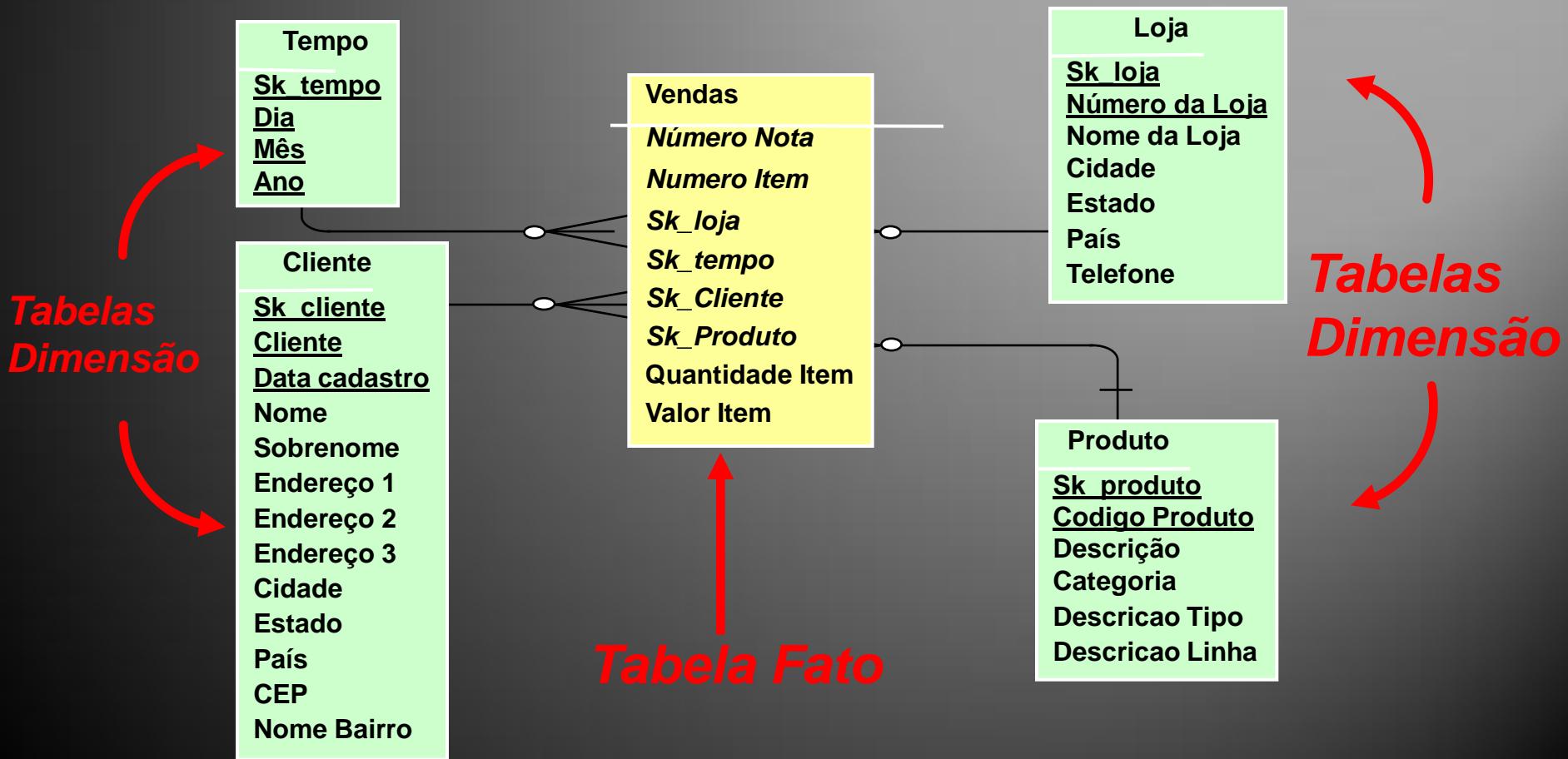
- SnowFlake



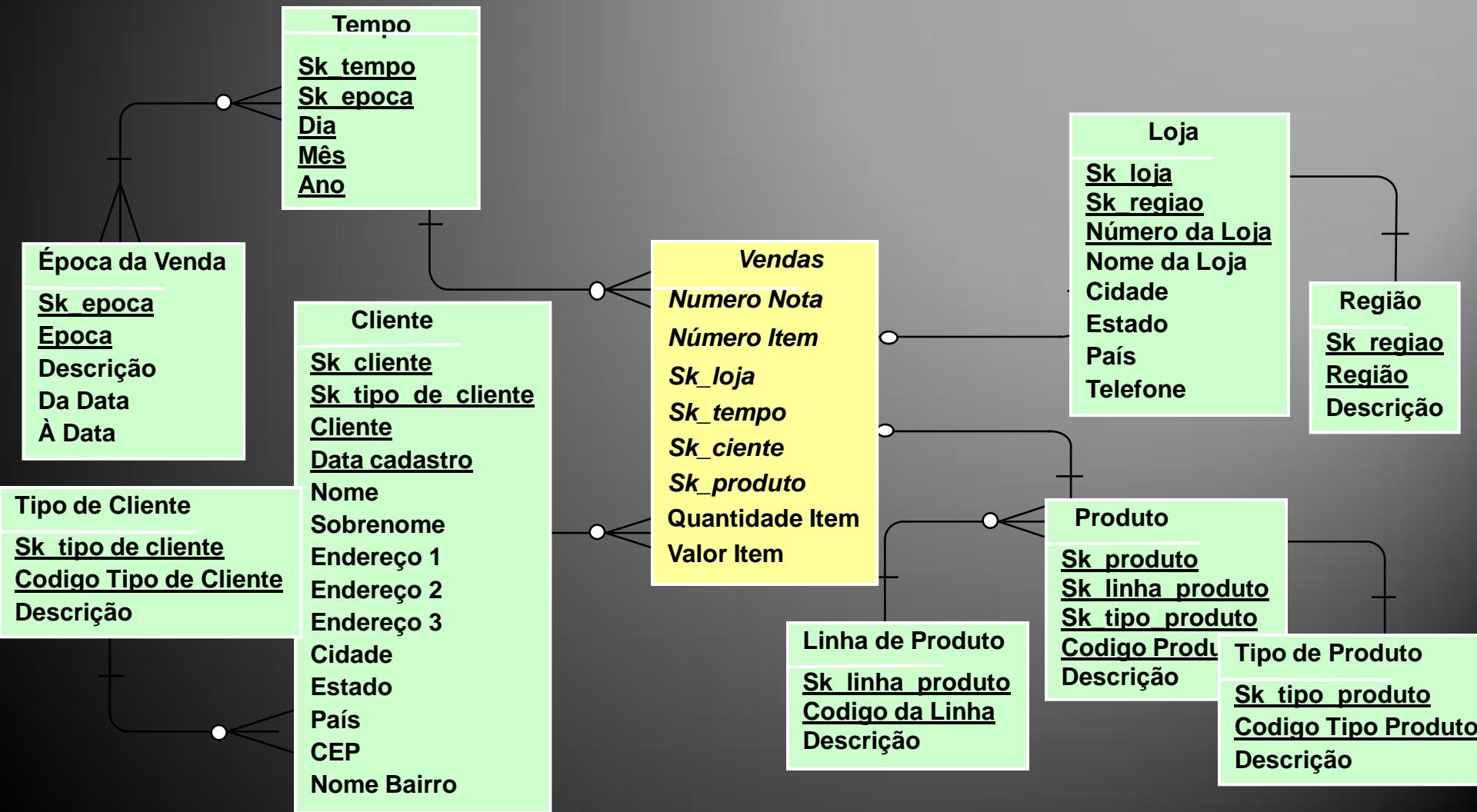
Data Warehouse - Modelagem Dimensional

- ▶ 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 SnowFlake)
 - Modelo SnowFlake : Dimensões normalizadas (Performance menor porém com requisitos de espaço de armazenamento em disco menor do que o modelo Estrela)

Data Warehouse - Star Schema



Data Warehouse - SnowFlake



Modelagem Dimensional Hotel Dallas

- ▶ Abrir o documento da matriz de necessidades
- ▶ Abrir o documento de fonte de dados
- ▶ Identifique as dimensões
- ▶ Identifique as fatos
- ▶ Criar um modelo Star Schema

OS TRÊS PRODUTOS - Ex: Hotel Dallas

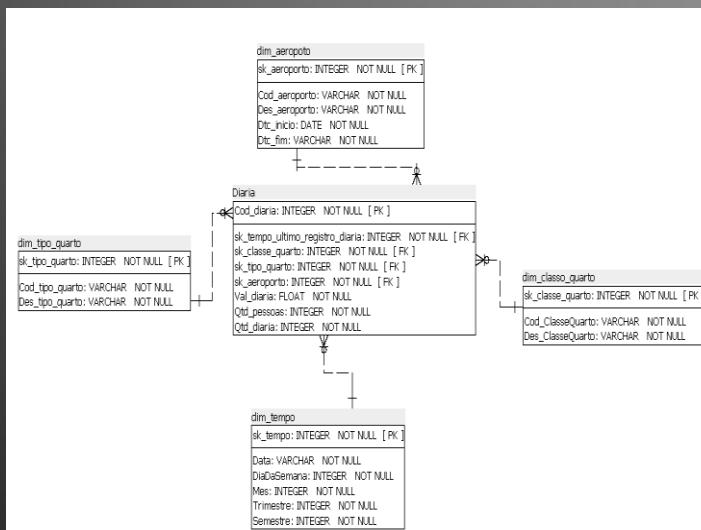
MATRIZ DE NECESSIDADES

| INDICADORES | FATURAMENTO | | | | | DIÁRIA | | | | | INDICADORES | |
|---------------------------------------|---|---|----------------|---|---|-------------|---|---------|---|------------|--------------------------|--|
| | DADOS LEVANTADOS | | VAL_HOSPEDAGEM | | | VAL_TURISMO | | VAL_BAR | | VAL_DIARIA | | |
| HOSPEDE | | | | | | | | | | | | |
| COD.HOSPEDE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | COD.HOSPEDE | |
| NOM.HOSPEDE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | NOM.HOSPEDE | |
| TIPO QUARTO | TIPO QUARTO | | | | | | | | | | | |
| DES.TIPO_QUARTO | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | DES.TIPO_QUARTO | |
| TIPO_CLASSE_QUARTO | TIPO_CLASSE_QUARTO | | | | | | | | | | | |
| DES.CLASSE_QUARTO | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | DES.CLASSE_QUARTO | |
| LOCALIDADE | LOCALIDADE | | | | | | | | | | | |
| NOM.CIDADE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | NOM.CIDADE | |
| NOM.PAIS | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | NOM.PAIS | |
| AÉROPORTO | AÉROPORTO | | | | | | | | | | | |
| DES.AEROPORTO | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | DES.AEROPORTO | |
| NOM.LOCALIDADE_AEROPORTO | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | NOM.LOCALIDADE_AEROPORTO | |
| TEMPO - DTC_ENTRADA, DTC_SAIDA | TEMPO - DTC_REGISTRO_PRIMEIRA_DIARIA | | | | | | | | | | | |
| Descrição ano mês | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | Descrição ano mês | |
| Descrição bimestre | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | Descrição bimestre | |
| Descrição data dia | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | Descrição data dia | |
| Descrição do dia | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | Descrição do dia | |

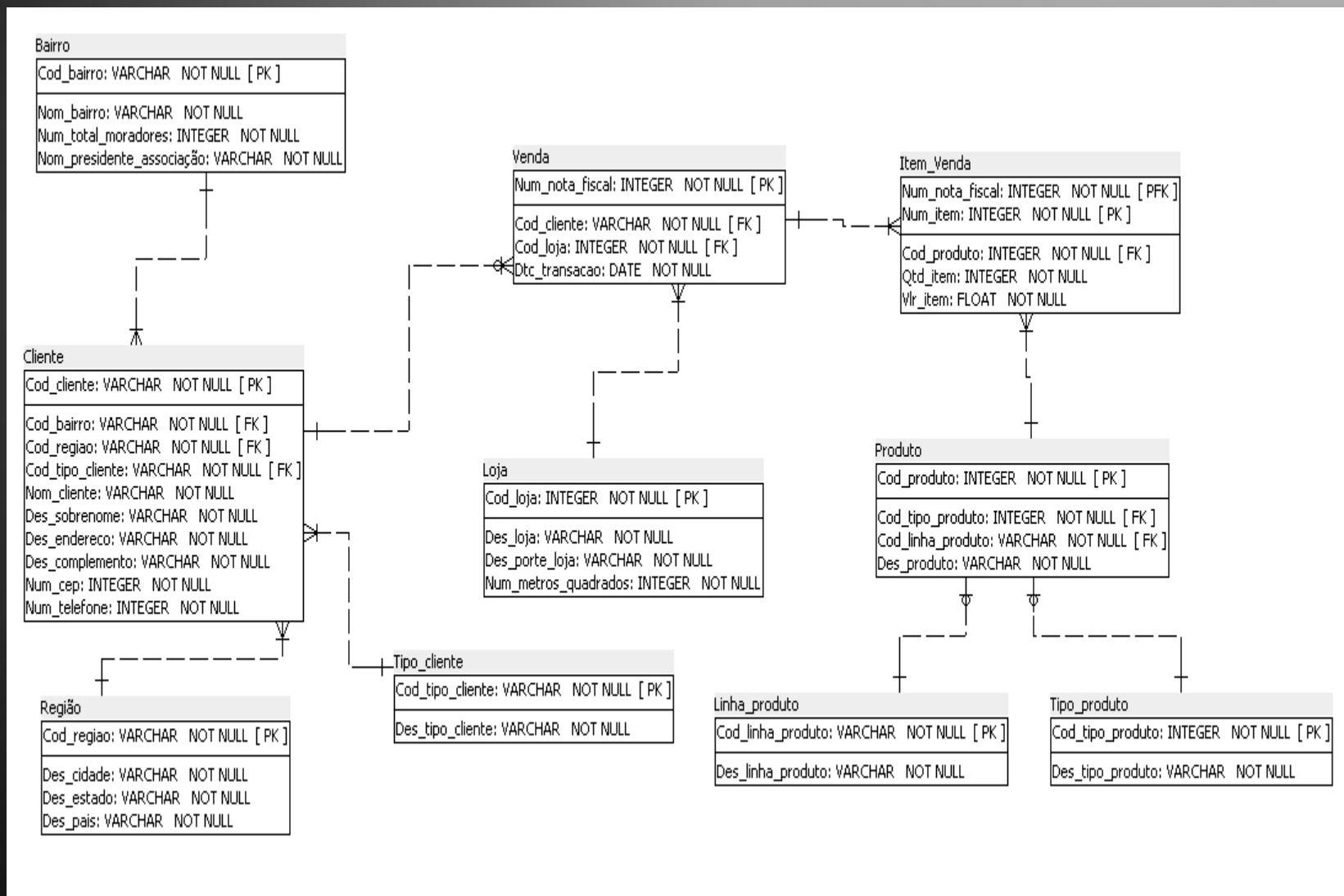
FONTE DE DADOS

| A | B | C | D |
|----|--------------------------|-----------------|-----------------|
| 1 | | | |
| 2 | HOTEL DALLAS | FATO DIARIA | |
| 3 | | | |
| 4 | Dimensões | DOM | Base Original |
| 5 | | | Attributo |
| 6 | HOSPEDE | | Relac |
| 7 | DIARIA | COD.HOSPEDE | CHAVE-1 |
| 8 | HOSPEDE | COD.HOSPEDE | CHAVE-1 |
| 9 | COD.HOSPEDE | COD.HOSPEDE | |
| 10 | NOM.HOSPEDE | NOM.HOSPEDE | |
| 11 | LOCALIDADE | | |
| 12 | DIARIA | COD.HOSPEDE | CHAVE-1 |
| 13 | HOSPEDE | COD.HOSPEDE | CHAVE-1 |
| 14 | HOSPEDE | COD.CIDADE | CHAVE-2 |
| 15 | CIDADE.ORIGEM | COD.CIDADE | CHAVE-2 |
| 16 | NOM.CIDADE | CIDADE.ORIGEM | NOM.CIDADE |
| 17 | CIDADE.ORIGEM | COD.PAIS | CHAVE-3 |
| 18 | PAYS.ORIGEM | COD.PAIS | CHAVE-3 |
| 19 | NOM.PAIS | PAYS.ORIGEM | NOM.PAIS |
| 20 | AEROPORTO | | |
| 21 | DIARIA | COD.HOSPEDE | CHAVE-1 |
| 22 | HOSPEDE | COD.HOSPEDE | CHAVE-1 |
| 23 | HOSPEDE | COD.AEROPORTO | CHAVE-4 |
| 24 | AEROPORTO.SAIDA | COD.AEROPORTO | CHAVE-4 |
| 25 | DES.AEROPORTO | AEROPORTO.SAIDA | DES.AEROPORTO |
| 26 | NOM.LOCALIDADE.AEROPORTO | AEROPORTO.SAIDA | NOM.LOCALIDADE |
| 27 | TIPO QUARTO | | |
| 28 | DIARIA | COD.TIPO_QUARTO | CHAVE-1 |
| 29 | TIPO QUARTO | COD.TIPO_QUARTO | CHAVE-1 |
| 30 | DES.TIPO_QUARTO | TIPO QUARTO | DES.TIPO_QUARTO |

MODELO DIMENSIONAL



Tabelas Operacionais Vendas



Tabelas Operacionais

**ALUGUEL DE
VEICULOS**

**INFRAÇÕES DE
TRÂNSITO**

LOGÍSTICA

NOTAS FISCAIS

Construção do Data Warehouse

▶ Criação Dimensões e Carga Staging



▶ Carga Dimensões



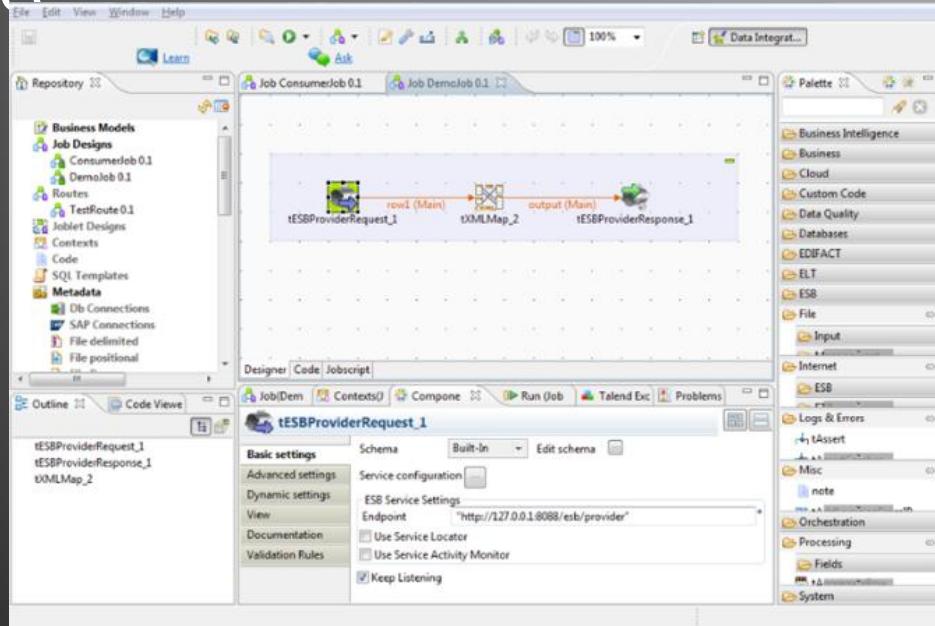
▶ Criação da Fato



▶ Carga da Fato



- UTILIZAR O TALEND
- É uma ferramenta de ETL, que oferece um ambiente gráfico para construir, gerenciar e manter processos de integração de dados em sistemas de Business Intelligence



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

**Requisitos do
Negócio**



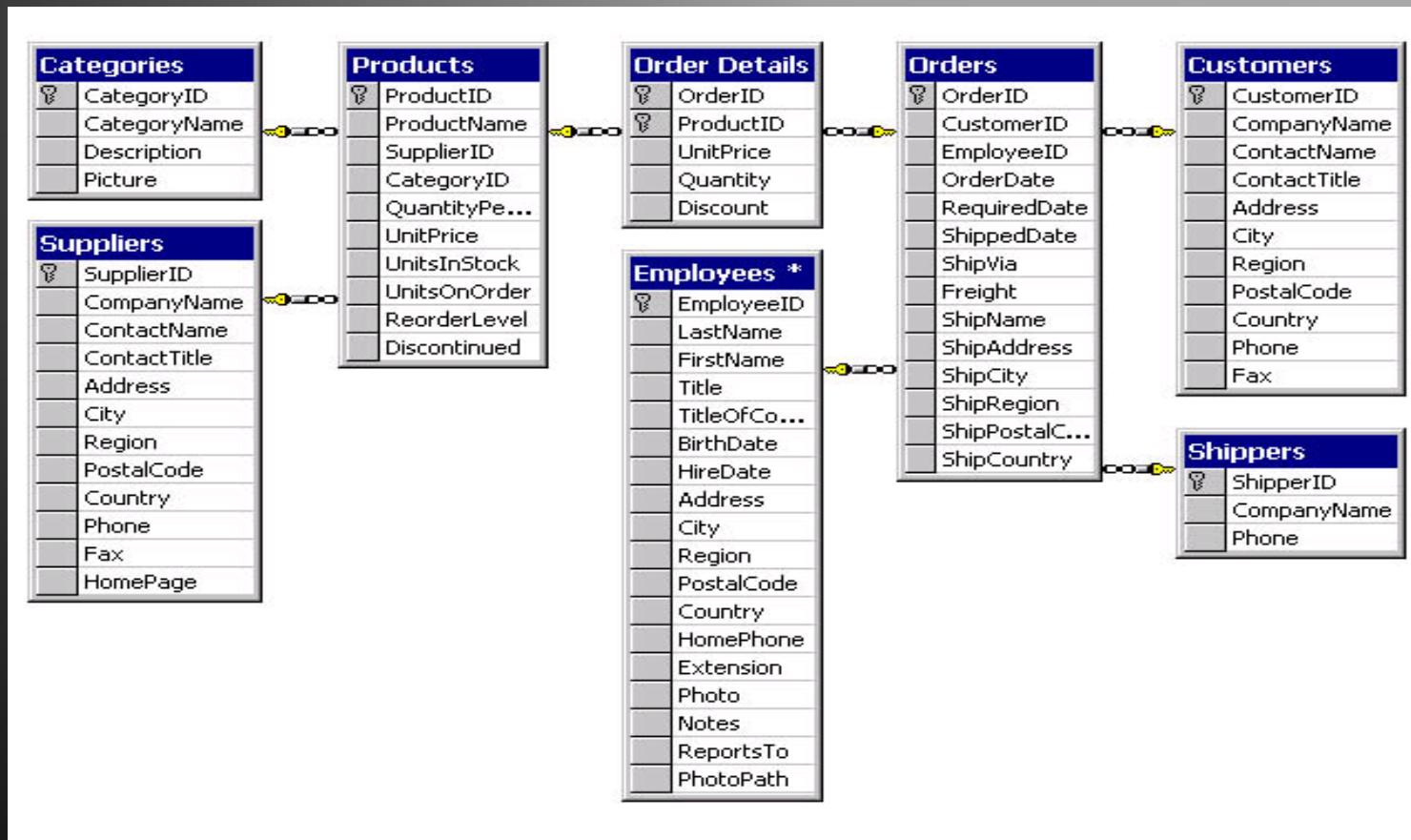
**Modelo
Multidimensional**

Processo de Negócio
Grão
Dimensões
Fatos

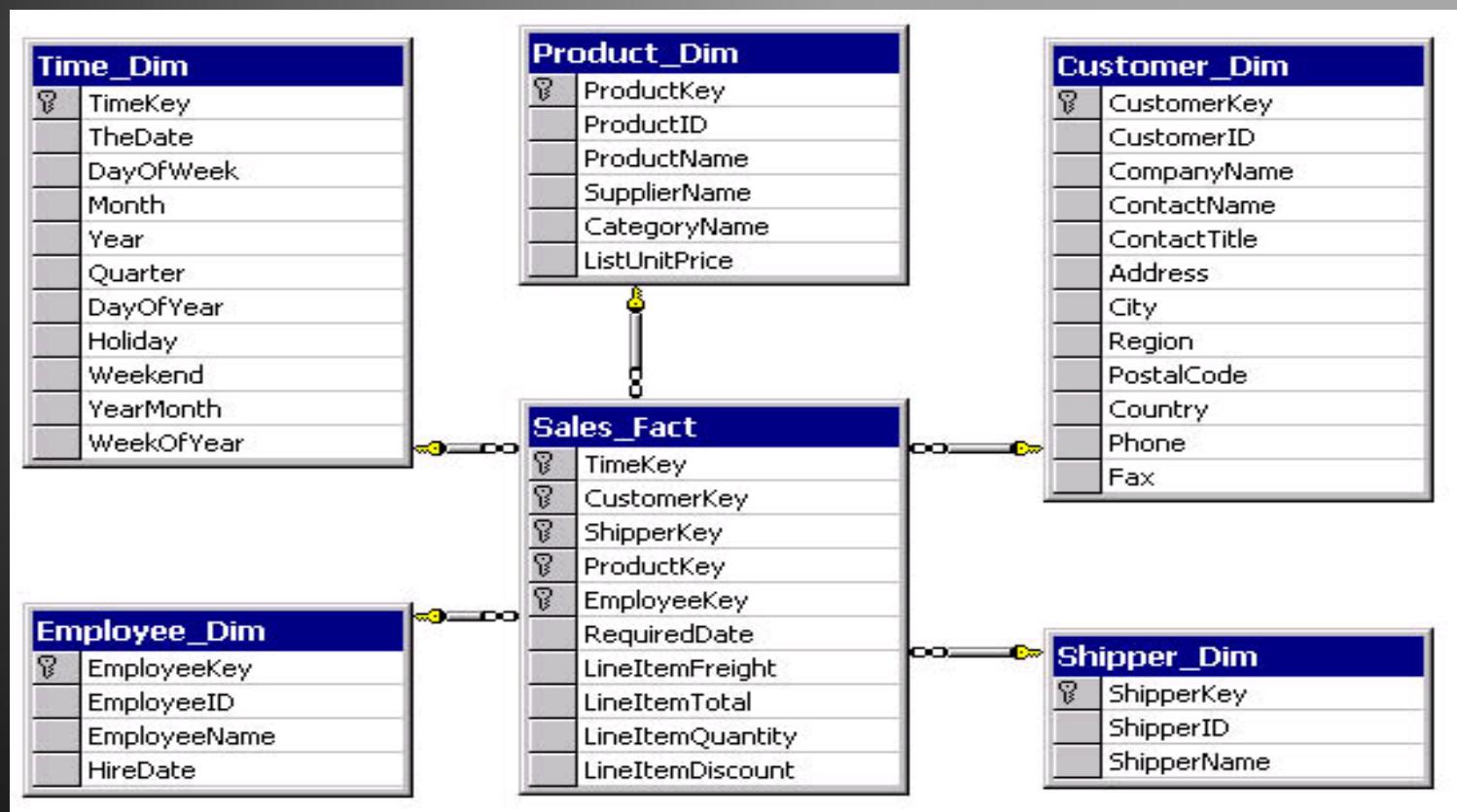


Realidade dos Dados

Exemplo Transacional



Exemplo Multidimensional



Arquitetura de um DW

- ROLAP (“Relacional OLAP”) - suporte a consultas OLAP por meio de Sistemas de Gerência de Bancos de Dados Relacionais — SGBDR;
- MOLAP (“Multidimensional OLAP”) - suporte a consultas OLAP baseadas em arrays multidimensionais.
- HOLAP = ROLAP + MOLAP

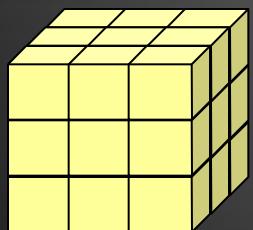
ROLAP

| Vantagens | Desvantagens |
|--|--|
| <ul style="list-style-type: none">✓ Utilização de uma tecnologia consolidada e robusta;✓ Arquitetura mais facilmente escalável;✓ Tecnologia e cultura já disponíveis nas organizações. | <ul style="list-style-type: none">✓ Tecnologia relacional (SQL) não é natural para consultas analíticas (OLAP);✓ Performance na manipulação de registros em caso de grandes volumes de dados; |



MOLAP

| Vantagens | Desvantagens |
|---|---|
| <ul style="list-style-type: none">✓ Utilização de tecnologia mais próxima da consulta analítica (OLAP)✓ Performance. | <ul style="list-style-type: none">✓ Tecnologia e cultura não muito difundidas e consolidadas nas organizações;✓ Limitações para armazenamento de grande volume de dados. |



DW/DM
Banco de Dados
Multidimensional



Ferramenta de
Consulta OLAP



HOLAP

| Vantagens | Desvantagens |
|---|---|
| <ul style="list-style-type: none">✓ Combina as vantagens das arquiteturas ROLAP e MOLAP;✓ Permite o armazenamento de informações transacionais detalhadas sem causar impacto na performance das consultas. | <ul style="list-style-type: none">✓ Ambiente híbrido;✓ Necessidade de ferramentas para armazenar dados em duas tecnologias diferentes. |



MetaDados

- ▶ 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.)

Metadados

- ▶ Preparado para detalhar tudo que foi construído no Data Warehouse
- ▶ Data Warehouse sem metadados significa o caos.
- ▶ Data Warehouse com metadados significa um caos ordenado e documentado.
- ▶ Deve ser alimentado sempre que novas implementações surjam como carga, tabelas, campos, etc.
- ▶ O usuário final pode, e deve, acessá-lo.

Prevendo 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 milhões x 24 bytes = 26 GB

Considerações

- ▶ SGBMD 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.

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 desktopvou 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
- ▶ Oliveira, Grimaldo, Oliveira Diego – BI Como Deve Ser – O Guia Definitivo. Salvador, 2016

Bibliografia na Internet

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

OBRIGADO!

Grimaldo Lopes de Oliveira
grimaldo_lopes@hotmail.com