

BD II

Professor Msc. Aparecido Vilela Junior
`aparecido.vilela@unicesumar.edu.br`

- Histórico
- Arquitetura do Data Warehouse
- Data Warehouse
- Modelagem Multidimensional
- OLTP e OLAP
- Processo e Procedimentos de Carga - ETL

A história do **BI** teve início na década de 70.

Alguns produtos de **BI** foram fornecidos para os analistas de negócio.

Os **principais problemas** desses produtos:

- Exigiam intensa e exaustiva programação;

- Não disponibilizavam informação em tempo hábil nem de forma flexível;

- Tinham alto custo de implantação.

Com o surgimento dos **BDs relacionais** e das **interfaces gráficas**:

Começaram a surgir os primeiros produtos de **BI** direcionados aos analistas de negócios.

Os novos produtos passaram a ser mais rápidos e apresentaram maior flexibilidade de análise.

BI pode ser entendido com a utilização de variadas **fontes de informação** para se definir **estratégias de competitividade** nos negócios de uma empresa.

O que é Business Intelligence ?

Conjunto de tecnologias que permitem o **cruzamento de informações** e suportam a **análise dos indicadores de performance** de um negócio.

Principal ferramenta de suporte à tomada de decisão.

Crucial para o **bom gerenciamento** de uma empresa na **Era Digital**.

**BI É UM CONJUNTO DE TÉCNICAS E
PROCESSOS, NÃO UMA
FERRAMENTA**

Case: SBT

No projeto de BI do SBT, eles usaram por muito tempo um projeto piloto em Qlikview. Mas para o que eles queriam fazer, o Qlikview começou a travar e ficar lento.

E isso quer dizer que o Qlikview é ruim? Não. Quer dizer que cada ferramenta tem uma melhor aplicabilidade do processo. E no caso deles, o Qlikview não era a escolha certa.

Eles queriam utilizar as técnicas de **balanced scorecard e KPI**, vendo isso em tempo real.

Quem conhece essa técnica, entenderia o que eles estavam pedindo, mas a pessoa que começou esse projeto era um ferramenteiro, ele entendia de **Qlikview**, não de **BI**.

E o que estavam fazendo, eram métricas com gráficos, não KPIs.

Processo

BI é toda a técnica e processo para a coleta, organização, análise, compartilhamento e monitoramento dos dados para suportar a tomada de decisão. E como todo bom processo, ele segue uma sequência de passos, que são:

passo 1: ele **coleta** os dados de uma fonte de dados, como um banco de dados legado, uma planilha, e outras fontes de dados;

passo 2: depois **organiza** todos esses dados coletados em um lugar, que nesse caso nós chamamos de **Data Warehouse**;

passo 3: depois disso, a gente pode **analisar** e **compartilhar** esses dados;

passo 4: e por fim se toma uma **decisão** com base nos dados que foram coletados, organizados e analisados;

passo 5: depois a gente **monitora** todo esse processo para entender se a nossa tomada de decisão foi **coerente** ou se **piorou** a situação.

Por quê BI

“Conheça a si mesmo e ao inimigo e, em cem batalhas, você nunca correrá perigo.

Conheça a si mesmo, mas desconheça seu inimigo, e suas chances de ganhar e perder são iguais.

Desconheça a si mesmo e ao inimigo e você sempre correrá perigo. ”

Livro Arte da Guerra - Sun Tzu

BI - Perguntas -> Respostas

O BI utiliza métricas e KPIs para medir o desempenho passado e poder orientar o planejamento futuro.

Ou seja, olhamos o que já aconteceu, que pode ser um dia atrás, uma semana, no mês passado, no ano passado, 10 anos ou um minuto atrás.

E existem 3 perguntas que o BI vai responder para suportar a tomada de decisão;

#1: O que aconteceu?

Por exemplo:

aconteceu uma venda;

um funcionário foi demitido;

novos itens chegaram no estoque;

as contas foram pagas.

#1: O que aconteceu?



Nesse exemplo aqui, aconteceu venda de produtos em uma determinada região.

Vamos considerar que o produto seja café. E que a região seja Maringá.

#2: Quantas vezes isso aconteceu?

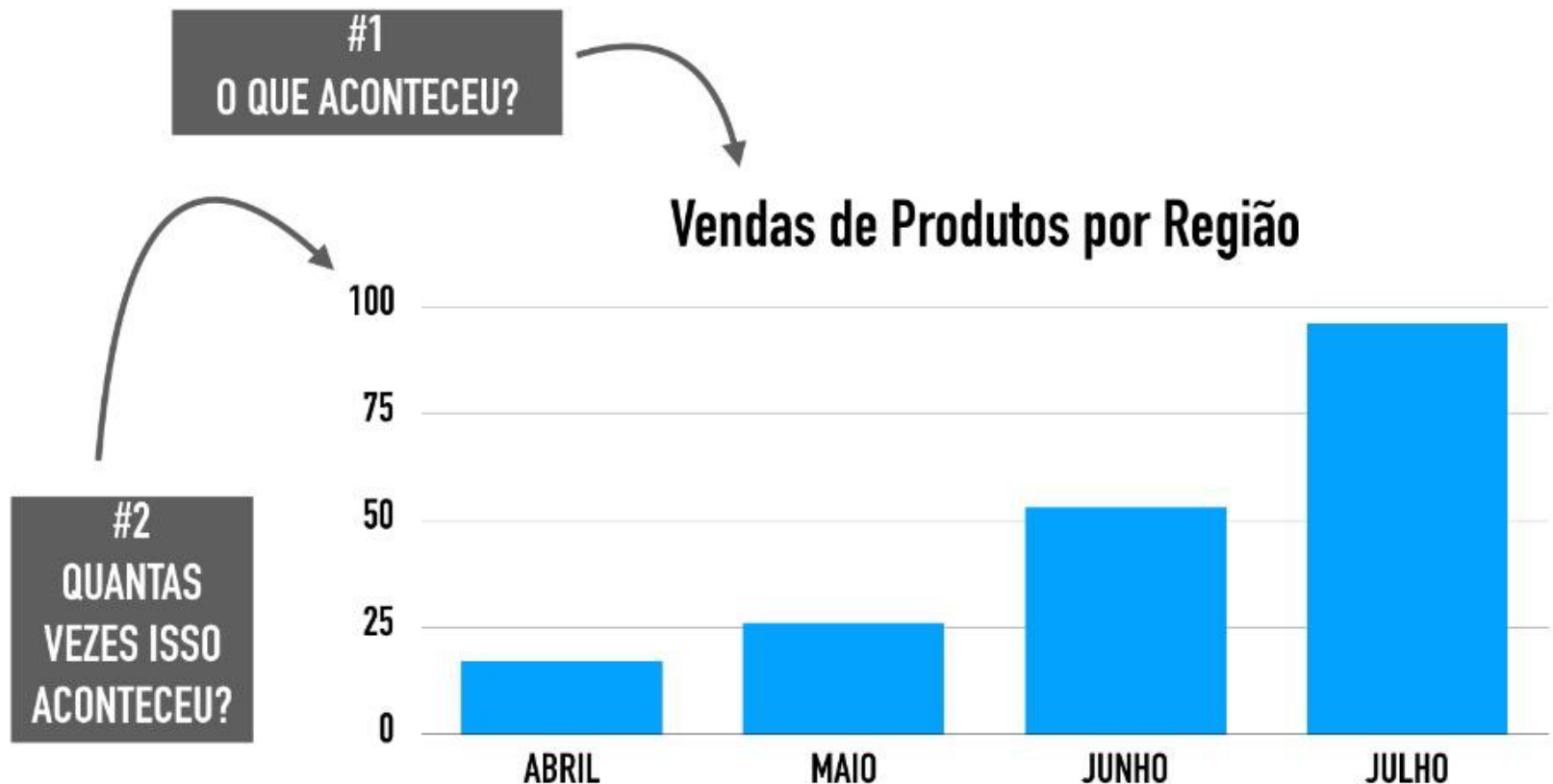
Já sabemos o que aconteceu. Agora, quantas vezes isso aconteceu?

Por exemplo:

Aconteceu uma venda. E quantas vezes isso aconteceu?

Se foram feitas 30 vendas daquele produto, essa venda aconteceu 30 vezes.

#2: Quantas vezes isso aconteceu?



Aqui a gente pode ver pela altura das barras, que em abril aconteceram 17 vendas, em maio aconteceram foram 26, 53 em junho e 96 em julho.

#3: Com que frequência isso acontece ou aconteceu?

Aqui tem relação com o tempo, com a frequência que aquele evento acontece. Por exemplo:

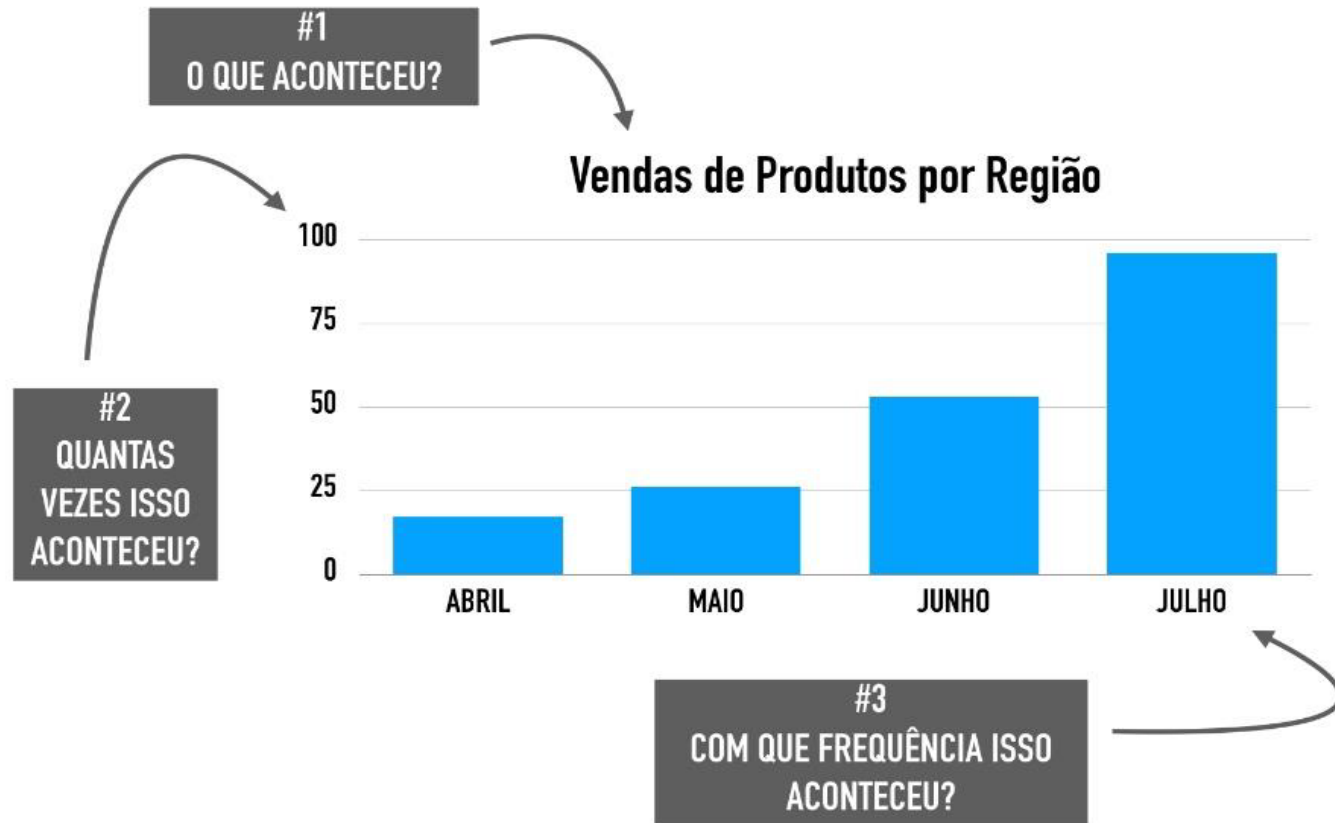
O que aconteceu? Contratação de colaborador.

Quantas vezes isso aconteceu? Uma vez.

Com que frequência? Em janeiro.

Você obviamente não tem como saber tudo que vai acontecer no futuro, mas o BI ajuda a gente a diminuir o erro.

#3: Com que frequência isso acontece ou aconteceu?



Nesse exemplo, a frequência é mês a mês. Podia ser ano a ano, semana a semana, hora a hora, de 5 em 5 minutos.

Lembrando:

As ferramentas de BI servem para nos auxiliar em uma determinada tarefa. Por exemplo:

Se eu tenho que fazer um buraco na parede, eu posso:

- usar uma furadeira;

- abrir com uma martelada;

- abrir com uma facinha que nem presidiário.

Todas são ferramentas para atingir um determinado objetivo.

No BI, esse objetivo é responder essas 3 perguntas.

Etapas da tomada de decisão

dados;

informação;

conhecimento;

sabedoria.

Esse é o ciclo de vida da tomada de decisão, ou ciclo de vida do conhecimento.

Os dados são uma parte pequena da informação, que sozinhos não fazem sentido.

São os ativos mais importantes de qualquer organização.

DADOS

Exemplo

| DADOS | INFORMAÇÃO | CONHECIMENTO | SABEDORIA |
|-------|------------|--------------|-----------|
| 40° | | | |
| 90° | | | |
| 20° | | | |

O que é isso para você? Pode ser qualquer coisa que trabalhe com graus. Por exemplo:

- coordenadas geográficas; temperatura; unidade de medida ótica; ângulo; etc.

Informação

Com base nesses dados, nós geramos a informação. São dados agrupados, organizados e lapidados. E é aqui que seu Data Warehouse vai entrar. Nele você vai fazer esse agrupamento e organização dos dados.

Há muitos dados pelo mundo para gerar informação, é aqui que entra o conceito de Big Data, dados estruturados e não estruturados.

Os dados são diversos, e para gerar informação, você precisa agrupar, organizar e limpar eles, deixar redondinho.



Informação

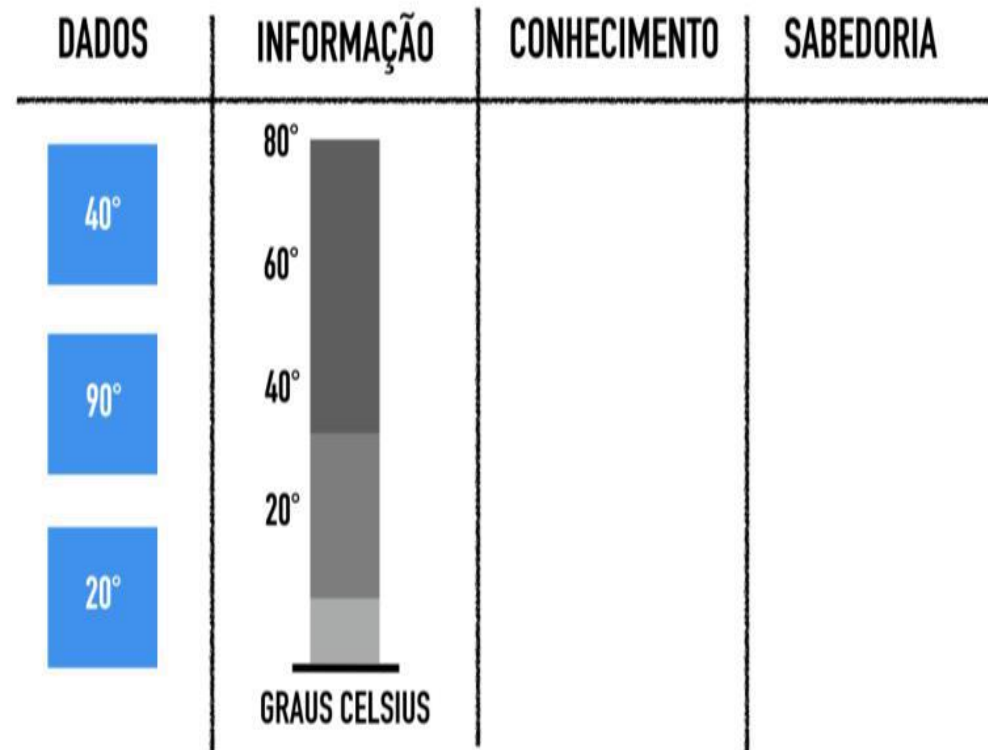
Nesse exemplo aqui é um termômetro, são graus Celsius.

Agora imagina que eu falo para você: “hoje está 40°.”

Você vai pensar que está quente demais para sair de casa. E como que você sabe que está quente demais? Para uma pessoa que vive no Rio de Janeiro, 40° é uma temperatura normal.

A interpretação depende do seu conhecimento

Então a próxima etapa desse ciclo da tomada de decisão é gerar o conhecimento.



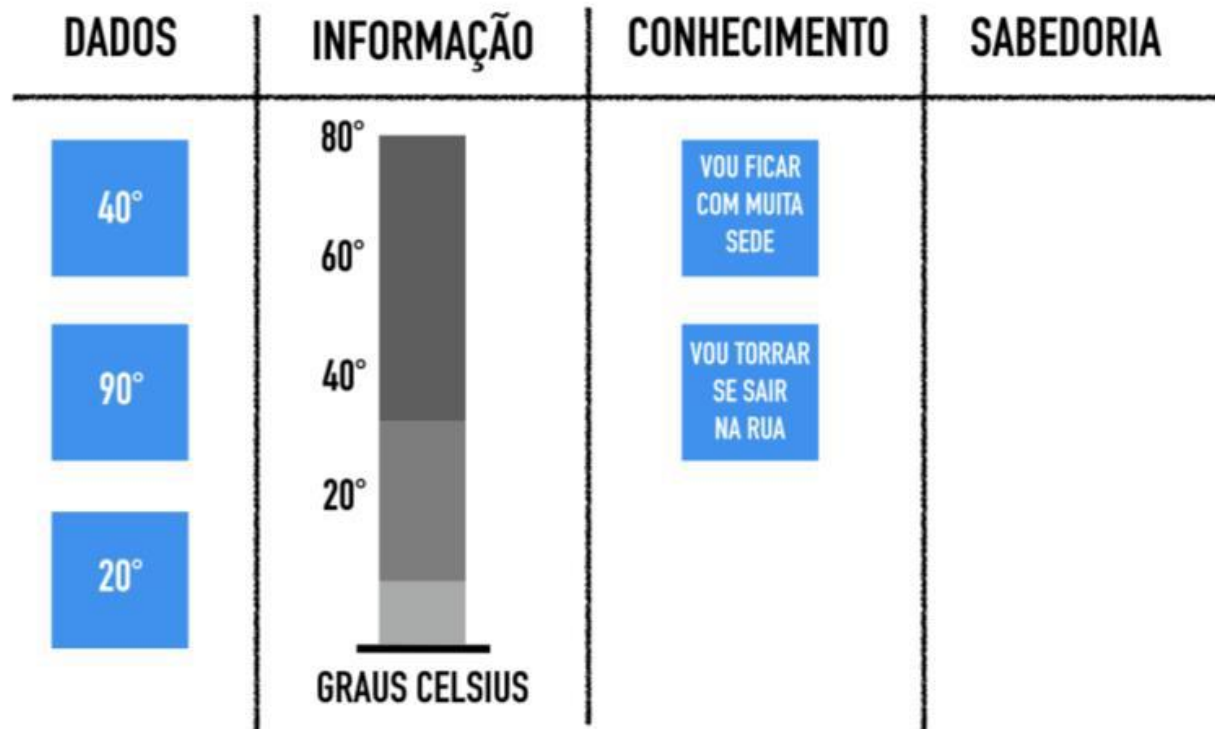
Conhecimento

Interpretando e aprendendo com as informações geradas, você vai gerar novos conhecimentos.

Depois, você combina seus conhecimentos atuais com os novos, para gerar um novo leque de conhecimento.

E com base nesse leque de conhecimentos, vem a parte da sabedoria.

Conhecimento



Essa é a grande sacada de ter o conhecimento.

E nós, como profissionais de BI, temos que entender isso para que a gente possa ajudar quem vai tomar a decisão a saber o que fazer.

Sabedoria

O momento da tomada de decisão.

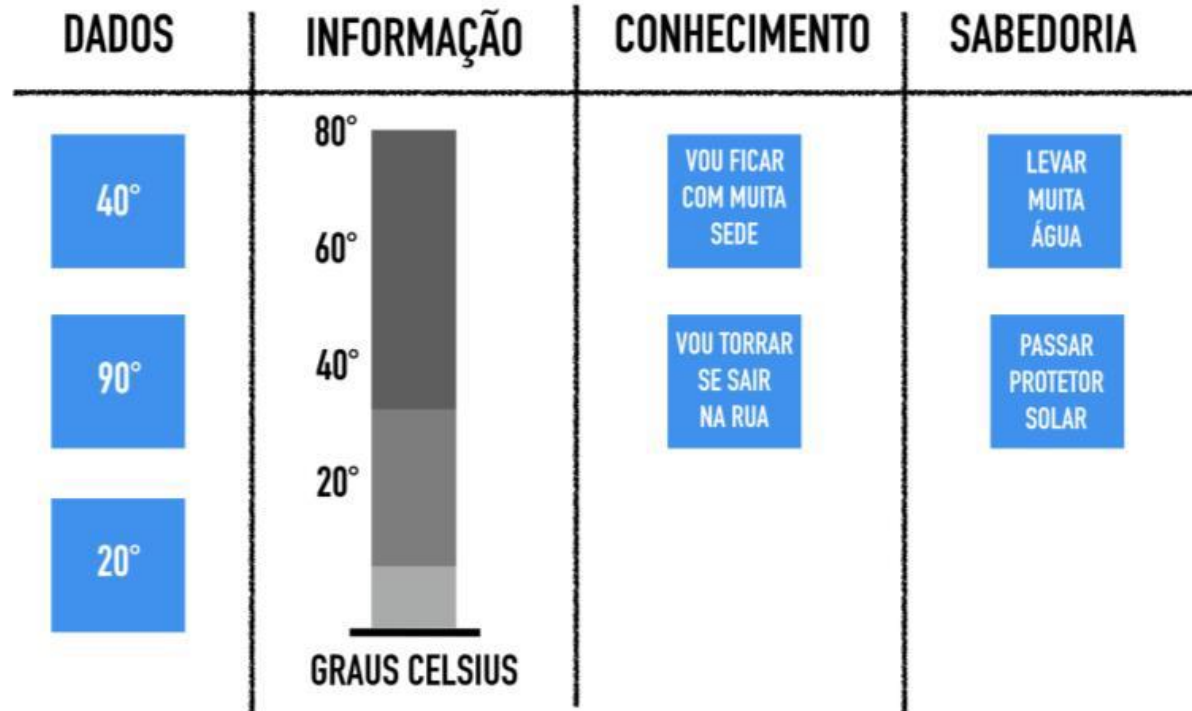
Com base no conhecimento que eu gerei, eu tenho a sabedoria de tomar uma decisão.

Se eu e você tivermos o mesmo problema, nós vamos tomar decisões diferentes se esse ciclo da tomada de decisão for diferente para você e para mim.

No momento que você organiza as suas informações, você pode entender e interpretar elas de um jeito, enquanto eu posso entender e interpretar de outro jeito.



Sabedoria



Esse processo que nós seguimos para a tomada de decisão é o mesmo no BI, no Data Warehouse, no Analytics, no Data Science, em qualquer coisa, porque é como nós, seres humanos, tomamos decisões.

Arquitetura do BI



Embora o Data Source não seja uma das etapas de uma solução de BI, gosto de representar ele como uma torre também porque é de onde os dados vão sair, é a origem deles.

Temos o ERP, CRM, arquivos CSV, arquivos TXT, etc. Importante: o BI trabalha com dados estruturados e semiestruturados. Ou seja, bancos de dados relacionais, algo que tenha uma estrutura implícita para a organização dos dados.

Data Integration



Depois de você ter identificado o Data Source, a primeira fase na arquitetura de BI é o Data Integration, que é onde ficam os robozinhos, é a parte do ETL ou EL-T.

Nessa etapa, é onde nós transformamos os dados em informação.

E = extract (extrair);

T = transform (transformar);

L = load (carregar).

Extract

O "E" vai extrair os dados de todos os Data Sources e colocar eles na staging área.

(assunto posterior)

Transform

Depois que os dados de todos os bancos foram tirados de lá e eu tenho todos juntos da staging area, o “T” vai unificar os dados, limpar eles e deixar pronto para aplicar o conhecimento.

Transform - Exemplo

Esses são os dados que eu tenho nos Data Sources:

ERP:

| ID | Nome do Cliente | Sexo do Cliente |
|----|-----------------|-----------------|
|----|-----------------|-----------------|

| | | |
|-----|-----------------|---|
| 333 | Vinicius Vilela | m |
|-----|-----------------|---|

Sistema Financeiro:

| ID | Nome do Cliente | Cidade do Cliente |
|----|-----------------|-------------------|
|----|-----------------|-------------------|

| | | |
|------|-----------------|---------|
| 1345 | Vinicius Vilela | Maringá |
|------|-----------------|---------|

CRM:

| ID | Nome do Cliente | Sexo do Cliente |
|----|-----------------|-----------------|
|----|-----------------|-----------------|

| | | |
|-----|-----------------|-----------|
| 102 | Vinicius Vilela | masculino |
|-----|-----------------|-----------|

As informações que temos sobre o cliente Vinicius Vilela nesses 3 sistemas é:

sexo: m;

cidade: Maringá;

sexo: masculino.

Depois, ele vai tirar as sujeiras e unificar tudo em uma informação só. Aqui ele tem duas informações sobre "sexo", o que ele vai fazer?

Vinicius Vilela:

sexo: masculino;

cidade: Maringá.

Nesse momento, a gente está unificando os dados.

Load

O "L" vai fazer a carga efetivamente.

Então você extraiu de algum lugar, fez a transformação, e depois vai pegar esses dados todos e vai carregar. Ou seja, vai inserir no Data Warehouse.

Esse processo do ETL vai acontecer com uma periodicidade, por exemplo, uma vez por dia, todo o dia a meia noite ele vai lá, pega os dados, faz a transformação e coloca no Data Warehouse.

A periodicidade vai depender da necessidade do negócio.

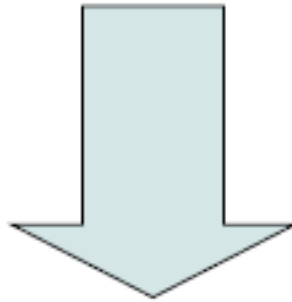
O que é Business Intelligence ?

Data Warehousing.

Ferramentas OLAP.

Data Mining.

Modelagem Analítica e Previsões



Base de acesso ao resultado do seu negócio.

Extrair e integrar dados de múltiplas fontes;

Fazer **uso da experiência**;

Analisar **dados contextualizados**;

Trabalhar com **hipóteses**;

Procurar **relações de causa e efeito**;

Transformar os registros obtidos em **informação útil** para o conhecimento empresarial.

Importante: Para um melhor desempenho dos sistemas, é recomendável que seja criada uma infra-estrutura específica para **BI**, separada do ambiente transacional.

Database Marketing;

Data Warehouse (**um dos principais pilares**);

OLAP;

Data Mining;

Webhouse.

O que essas **ferramentas têm em comum?**

Facilitam a transformação dos dados em informações, auxiliando os usuários das diferentes áreas da empresa a realizar análises e gerar relatórios.

Fidelização do Cliente

Otimização de alocação de recursos

Aumentar vendas cross-selling

Melhorar produtos e serviços

Fidelização do Cliente:

Registrando cada compra do seu cliente, é possível prever quando ele precisará do seu produto;

Essa antecipação geralmente garante um **atendimento diferenciado**.

O cliente sente-se especial e estabelece uma relação de parceria.

Aumentar vendas cross-selling:

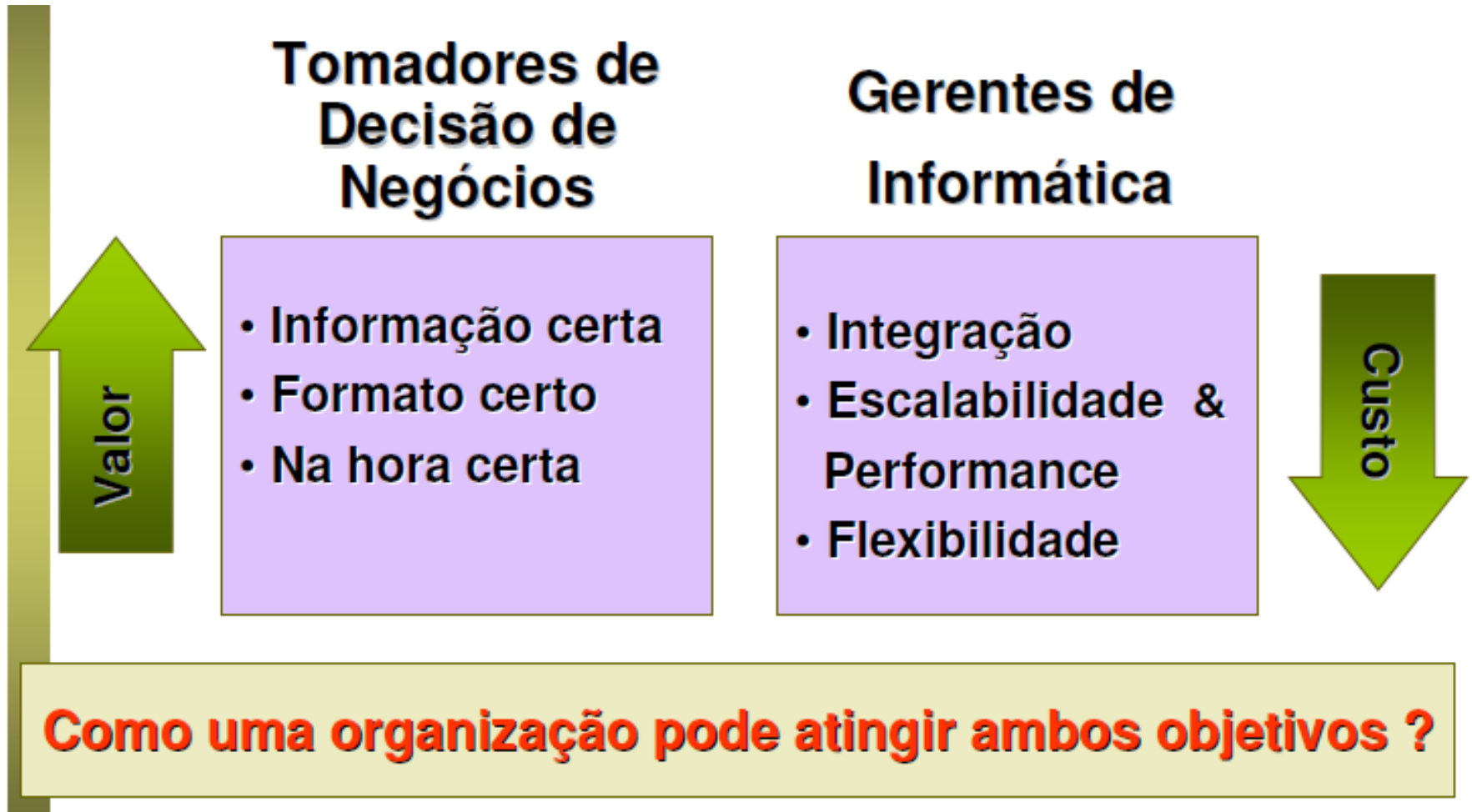
Traçar o **perfil de compra de um cliente** permite a criação de uma mensagem direcionada para a **promoção de vendas de produtos** complementares **gerando a demanda** que ainda não foi percebida pelo cliente.

Otimização de alocação de recursos:

Acompanhar quais **linhas de produtos** são mais lucrativas;

Acompanhar quais **estratégicas** permitem que o empresário **priorize seus investimentos**;

Controlar os **reflexos e consequências globais** de suas ações na empresa.



Antes de um **projeto BI**, é necessário um planejamento estratégico da informação.

Razão: Através dessa análise pode-se verificar se há necessidade de remodelar os processos ou apenas fazer ajustes para que os sistemas se enquadrem no **projeto de BI**.

Empresas de todos os tipos e portes podem e devem empregar **soluções de BI**:

Objetivo: injetar inteligência aos negócios.

Melhorar produtos e serviços:

65% dos Sistemas de Atendimento ao Consumidor no Brasil são somente usados de **forma reativa**;

As **informações colhidas** através desse serviço podem ser estudadas com o **suporte** de uma aplicação de **Business Intelligence** para a detecção de defeitos e necessidades do mercado decorrendo em melhorias no **produto** ou **serviço**.



A aplicação de data mining do Wal-mart descobriu que o perfil do consumidor de cervejas era semelhante ao de fraldas. Eram homens entre 25 e 30 anos, que compravam fraldas e cervejas às sextas-feiras no final da tarde no caminho do trabalho para casa.

Com base nestas informações - que não foram requeridas à aplicação - resolveu-se colocar os dois produtos em locais estratégicos, de forma que os compradores ao buscarem estes dois produtos, pudessem observar outros produtos em promoção que foram colocados entre os mesmos.

Resultado : o consumo cresceu 30 % às sextas-feiras.

“Essa pesquisa demorou 30 minutos, enquanto uma pessoa do marketing demoraria várias semanas para chegar a mesma conclusão”



O banco Itaú, pioneiro no uso de data warehouse no Brasil, costumava enviar mais de 1 milhão de malas diretas, para todos os correntistas.

No máximo 2% deles respondiam às promoções.

Hoje, o banco tem armazenada toda a movimentação financeira de seus milhões de clientes nos últimos 18 meses. A análise desses dados permite que correspondências/emails sejam enviadas apenas a quem tem maior chance de responder. A taxa de retorno subiu para 30%. A conta do correio foi reduzida a um quinto.

A Sprint, um dos líderes no mercado americano de telefonia de longa distância, desenvolveu, com base no seu armazém de dados, um método capaz de prever com 61% de segurança se um consumidor trocava de companhia telefônica dentro de um período de dois meses. Com um marketing agressivo, conseguiu evitar a deserção de 120 000 clientes e uma perda de 35 milhões de dólares em faturamento.

Outra empresa de telefonia detectou, ao implantar seu armazém de dados, que quatro grandes clientes empresariais eram responsáveis por mais da metade das chamadas de manutenção. Um deles estava prestes a abandonar os serviços. A telefônica fez reparos imediatos, convenceu o cliente a ficar e manteve uma receita anual de 150 milhões de dólares.

Apresentar:

Conceitos

Necessidade de um Data Warehouse

Características de um Data Warehouse

Algumas arquiteturas

Modelos de dados

Desenvolvimento do Data Warehouse

Povoamento do Data Warehouse

Extração de informações do Data Warehouse

Análise do uso da tecnologia Data Warehouse

Definição I:

“ É uma coleção de dados orientados por assuntos, integrados, variáveis no tempo e não voláteis, para dar suporte ao processo gerencial de tomada de decisão ” [Inmon]

Definição II:

“ É um processo em andamento que aglutina dados de fontes heterogêneas, incluindo dados históricos e dados externos para atender às necessidades de consultas estruturadas e *ad-hoc*, relatórios analíticos e de suporte a decisão ” [Harjinder]

Definição III:

“ É uma coleção de técnicas e tecnologias que juntas disponibilizam um enfoque pragmático e sistemático para tratar com o problema do usuário final de acessar informações que estão distribuídas em vários sistemas da organização ”

[Barquini]

Dados operacionais vs. Data Warehouse

| Características | BD Operacional | Data Warehouse |
|------------------------|-------------------------------|-----------------------|
| Objetivo | Operações diários do negócio | Analisar o negócio |
| Uso | Operacional | Informativo |
| Tipo de processamento | OLTP | OLAP |
| Unidade de trabalho | Inclusão, alteração, exclusão | Carga e consulta |
| Número de usuários | Milhares | Centenas |
| Tipo de usuário | Operadores | Comunidade gerencial |
| Interação do usuário | Somente pré-definida | Pré-definida e ad-hoc |
| Condições dos dados | Dados operacionais | Dados Analíticos |
| Volume | Megabytes - gigabytes | Gigabytes - terabytes |
| Histórico | 60 a 90 dias | 5 a 10 anos |

Dados operacionais vs. Data Warehouse

| Características | BD Operacional | Data Warehouse |
|------------------------|-----------------------|------------------------|
| Granularidade | Detalhados | Detalhados e resumidos |
| Redundância | Não ocorre | Ocorre |
| Estrutura | Estática | Variável |
| Manutenção desejada | Mínima | Constante |
| Acesso a registros | Dezenas | Milhares |
| Atualização | Contínua (tempo real) | Periódica (batch) |
| Integridade | Transação | A cada atualização |
| Número de índices | Poucos / simples | Muitos / complexos |
| Intenção dos índices | Localizar um registro | Aperfeiçoar consultas |

A credibilidade dos dados

É melhorada considerando a existência de uma única fonte de dados;

A produtividade

É melhorada visto que o trabalho efetuado para produzir o primeiro relatório prepara o ambiente para os demais;

A impossibilidade de transformar dados em informação

É atenuada graças a integração dos dados;

Data Warehouse



Data Analysis



Nessa etapa é onde nós vamos gerar o conhecimento, onde começamos a cruzar as informações disponibilizadas e começamos a analisar e identificar padrões passados e problemas futuros.

Aqui é onde vai entrar a parte de análise do negócio.

E tem algumas técnicas que você vai utilizar para fazer isso.

Métricas

A métrica é uma medida, ela serve para você metrificar algo, então não é nada incomum você tentar relacionar a uma fita métrica, daquelas de pedreiro, que começa a puxar e medir.

Métrica é também chamada de quantificador ou medida. Elas são utilizadas para metrificar algo e sempre são números, porque precisam ser contáveis.

Esses números são provenientes de transações da empresa.

Tudo que a empresa for mensurar é uma métrica. Na maioria das vezes, a métrica vai ser o que o usuário quer medir.

Normalmente vão responder: quero ver as vendas, os seguidores no Twitter, o fluxo de caixa etc....

Key Performance Indicator (KPI)

É um índice usado para mensurar percentualmente as variações que ocorrem na empresa, no produto, no setor, etc. Geralmente, um KPI tem uma meta.

Por exemplo:

Nossa meta de faturamento este ano é R\$200.000, e estamos em R\$100.000 de faturamento, então o KPI está em 50% do batimento da meta, é assim que se analisa.

KPI de Turnover (Rotatividade de Colaboradores)

Objetivo: Monitorar o giro de entradas e saídas de pessoal na empresa. Ou seja, todo funcionário que entrou e que saiu.

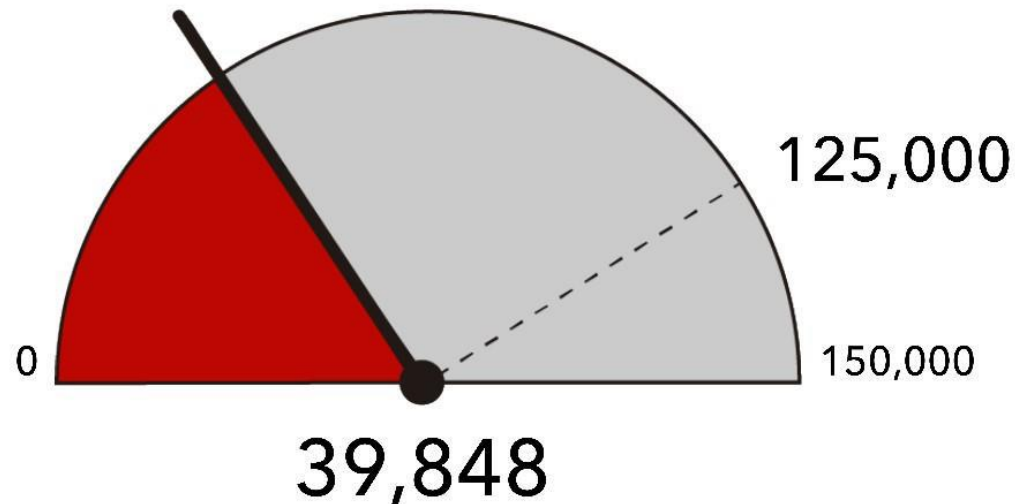
Fórmula: $(n^{\circ} \text{ de desligamentos} \times 100) / \text{média de empregados} = \% \text{ de Turnover}$.

$$(50 \times 100) / 100 = 50\% \text{ de Turnover}$$

Leitura: 50% dos colaboradores afastaram-se da empresa naquele determinado período.

Se isso é bom ou é ruim, depende da empresa. Eu acho 50% de turnover alto, mas para empresas de

Exemplo



O que acontece é que se em um determinado momento aquele ponteirinho deveria estar na linha pontilhada e no momento que a gente está analisando ele não está, ele vai estar vermelho.

Se estivesse no meio, estaria amarelo, e se estivesse no pontilhado estaria verde, porque estamos onde deveríamos para alcançar a meta a tempo.

Analysis OLAP (Cubos)



OLAP significa online analytical processing, ou seja, processamento analítico em tempo real.

Com ele, podemos combinar dados, pivotar e fazer diversos cruzamentos em tempo de execução, ou seja, com os cubos, a gente pode manipular e analisar um grande volume de dados em tempo real, por isso é online, e o mais importante, com múltiplas perspectivas e hipóteses.

Data Visualization



Depois de ter feito as análises, é hora de preparar a visualização delas.
Como essa informação vai ser entregue para o tomador de decisão?

Reporting

Uma delas são os relatórios de BI.

Quase todas as ferramentas de BI implementam esses relatórios. E qual a grande diferença deles para os demais relatórios?

A técnica de BI reporting consegue transformar os dados em informações compreensíveis, flexíveis e fáceis de serem analisadas pelo usuário final em relatórios pré-formatados.

Exemplo

STEELWHEELS

500 International Speedway, Daytona Beach FL 32114

(123) 456-7890 <http://www.steelwheels.com>

Qua Nov 27 17:26:56 GMT 2013

TO:

Australian Gift Network, Co
31 Duncan St. West End,
South Brisbane, Queensland 4101 Australia

INVOICE

Attn: Tony Calaghan

Invoice #: 10152

Sales Rep: 1611

Account Number: 333

Terms: Net 30 days

Date: Setembro 25, 2003

| SKU | Product Description | Price/Unit | Qty Ordered | Total Price |
|----------|--|------------|-------------|--------------------|
| S16_4027 | 1970 Triumph Spitfire | \$4,524.10 | 35 | \$4,524.10 |
| S32_3207 | 1950's Chicago Surface Lines Streetcar | \$1,681.35 | 33 | \$1,681.35 |
| S24_4048 | 1992 Porsche Cayenne Turbo Silver | \$2,802.09 | 23 | \$2,802.09 |
| S24_1444 | 1970 Dodge Coronet | \$1,632.75 | 25 | \$1,632.75 |
| | | | | <u>\$10,640.29</u> |

Payment History

| Date | Check# | Amount |
|----------|----------|--------------|
| 11-15-03 | HL209210 | \$ 27,098.80 |
| 10-17-03 | JK479662 | \$ 10,640.29 |
| 03-01-05 | NF959653 | \$ 21,730.03 |

Send Payment and Remittance Slip to:

Steel Wheels
500 International Speedway
Daytona Beach FL 32114

Thank you for your business!

Dashboards

O dashboard, também conhecido como painel, é uma técnica para visualização de dados, o famoso painel de controle da empresa.

Em uma única tela, você vai organizar análises, gráficos, métricas, KPIs e reports de toda a empresa ou de um ponto de vista do negócio ou departamento.

Exemplo



Levantamento de Dados

Matriz Métricas x Descritores

Diagrama usado para documentar o DW

Nas linhas coloca-se o descritores do negócio e nas colunas as métricas.

Assinala-se um “X” quando a métrica for referente ao descritor.

Útil para avaliar o impacto de alterações do seu DW

Matriz

Tem como origem a Bus Matrix, que estabelece a definição dos grupos de Dimensões e Fatos, estruturados de forma mais organizada, simples e intuitiva.

Nos livros de cabeceira e nas dicas de grandes mestres de DW, a ideia básica é:

Deve-se unir em uma mesma sala todos os gestores de negócios interessados na criação do projeto.

A ausência de um gestor importante para a construção do banco de dados consolidado irá fazer com que você ouça no final: “ESTÃO FALTANDO DADOS”.

Nas entrevistas, serão levantados os principais anseios dos gestores que deverão partir do questionamento:

Quais informações necessito para tomar decisões mais efetivas?

Ao final das “n” reuniões necessárias gera-se uma Matriz.

Matriz de necessidades

| INDICADORES | Faturamento | | | | Diária | INDICADORES |
|----------------------------|--------------|---------------------|---------------|--|-----------------------|----------------------------|
| DADOS LEVANTADOS | | | | | | DADOS LEVANTADOS |
| | Valor do Bar | Valor da Hospedagem | Valor Turismo | | Quantidade de Diárias | |
| Hospede | | | | | | Hospede |
| Nome do Hospede | ✓ | ✓ | ✓ | | ✓ | Nome do Hospede |
| | | | | | | |
| Regiao | | | | | | Regiao |
| Nome da Cidade | ✓ | ✓ | ✓ | | | Nome da Cidade |
| Pais | ✓ | ✓ | ✓ | | | Pais |
| | | | | | | |
| Classe Quarto | | | | | | Classe Quarto |
| Tipo da Classe | | | | | ✓ | Tipo da Classe |
| | | | | | | |
| Aeroporto | | | | | | Aeroporto |
| Nome do Aeroporto | ✓ | ✓ | ✓ | | | Nome do Aeroporto |
| | | | | | | |
| TEMPO | | | | | | TEMPO |
| Descrição ano mês | ✓ | ✓ | ✓ | | ✓ | Descrição ano mês |
| Descrição bimestre | ✓ | ✓ | ✓ | | ✓ | Descrição bimestre |
| Descrição data dia | ✓ | ✓ | ✓ | | ✓ | Descrição data dia |
| Descrição do dia | ✓ | ✓ | ✓ | | ✓ | Descrição do dia |
| Descrição do mês | ✓ | ✓ | ✓ | | ✓ | Descrição do mês |
| Descrição mês ano | ✓ | ✓ | ✓ | | ✓ | Descrição mês ano |
| Descrição mês ano numérico | ✓ | ✓ | ✓ | | ✓ | Descrição mês ano numérico |
| Descrição mês ano completo | ✓ | ✓ | ✓ | | ✓ | Descrição mês ano completo |
| Descrição quadrimestre | ✓ | ✓ | ✓ | | ✓ | Descrição quadrimestre |
| Descrição quinzena | ✓ | ✓ | ✓ | | ✓ | Descrição quinzena |
| Descrição semestre | ✓ | ✓ | ✓ | | ✓ | Descrição semestre |
| Descrição trimestre | ✓ | ✓ | ✓ | | ✓ | Descrição trimestre |
| Número ano | ✓ | ✓ | ✓ | | ✓ | Número ano |
| Número bimestre | ✓ | ✓ | ✓ | | ✓ | Número bimestre |
| Número dia | ✓ | ✓ | ✓ | | ✓ | Número dia |
| Número mês | ✓ | ✓ | ✓ | | ✓ | Número mês |
| Número nível | ✓ | ✓ | ✓ | | ✓ | Número nível |
| Número quadrimestre | ✓ | ✓ | ✓ | | ✓ | Número quadrimestre |
| Número quinzena | ✓ | ✓ | ✓ | | ✓ | Número quinzena |
| Número semestre | ✓ | ✓ | ✓ | | ✓ | Número semestre |

Fonte de Dados

| | A | B | C | D | E |
|----|--------------------------|--|---|-----------|-------------|
| 1 | | | | | |
| 2 | CLIENTE | FATO BENEFICIO | Obs: Campos simvão se estiverem nulos não informado | | |
| 3 | | | | | |
| 4 | Dimensões | | Base Original | | |
| 5 | | DOM | Atributo | Relaç. | Domínio |
| 6 | Informações Beneficiário | | | | |
| 7 | | REGISTROATIVIDADEOPERACIONAL SISTEMAPRODUCAO | idRegistroAtividadeOperacional | CHAVE-1 | |
| 8 | | REGISTROATIVIDADEOPERACIONALBENEFICIADOS | idRegistroAtividadeOperacional | CHAVE-1 | |
| 9 | | REGISTROATIVIDADEOPERACIONALBENEFICIADOS | idBeneficiado | CHAVE-2 | |
| 10 | | BENEFICIADO | idBeneficiado | CHAVE-2 | |
| 11 | CPF/CNPJ | BENEFICIADO | CPE | | |
| 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 | | REGISTROATIVIDADEOPERACIONAL SISTEMAPRODUCAO | idRegistroAtividadeOperacional | CHAVE-1 | |
| 24 | | REGISTROATIVIDADEOPERACIONALBENEFICIADOS | idRegistroAtividadeOperacional | CHAVE-1 | |
| 25 | | REGISTROATIVIDADEOPERACIONALBENEFICIADOS | idBeneficiado | CHAVE-2 | |
| 26 | CPF/CNPJ | BENEFICIADO | idBeneficiado | CHAVE-2 | |
| 27 | | BENEFICIADORENDAEXTRA | idBeneficiado | CHAVE-2.3 | |
| 28 | | BENEFICIADORENDAEXTRA | idTipoRendaExtra | CHAVE-2.3 | |
| 29 | % de Renda | BENEFICIADORENDAEXTRA | percentual renda | | |
| 30 | Tempo | | | | |
| 31 | Ano de Referência | | PERIODO_ANO | CHAVE | |

Modelagem de Dados

Por fim, é necessário após todo o levantamento, estruturar um modelo de dados que seja funcional e rápido para encontrar as respostas dos gestores, este modelo a ser criado é o conhecido modelo multidimensional.

O modelo é composto por uma tabela Fato central que contém todas as chaves artificiais (Surrogate Key) das tabelas de Dimensão além das métricas do projeto.

Estas métricas foram inicialmente identificadas na Matriz de Necessidades e encontradas no banco de dados relacional, planilhas eletrônicas, logs de dados, dentre outras fontes, e posteriormente informadas na Fonte de Dados.

Contém a única tabela normalizada do projeto, que é a tabela Fato. Existem diversas vantagens no uso deste tipo de modelo para a leitura dos dados de uma base de DW:

Levantamento dos dados com os Gestores

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 explanem 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 complementar a Matriz com todas as informações discutidas na reunião verifique se as perguntas dos gestores podem ser respondidas pelo documento criado.

Exemplo

De acordo com o modelo do Hotel FCV, vamos montar uma Matriz Métricas x Descritores para o desenvolvimento do DW.

Assunto: Faturamento e Diárias do Hotel.

Pense nas perguntas para a tomada de decisão.

Perguntas dos gestores

- 1) Gostaria de analisar o **Total do Valor do Restaurante**, gasto pelo hóspede (**nome e cidade de origem**).
- 2) Analisar o **Total das Diárias**, gasto pelo **hóspede da classe econômica**.
- 3) Identificar os **hóspedes** do Hotel, que vieram do **país “ Argentina”**;
- 4) Analisar a **Média do valor da hospedagem** de todos os hóspedes que são oriundos da **cidade** de “São Paulo” e que saíram do **aeroporto** de “Guarulhos”.
- 5) Realizar uma agregação de todos os valores passados ao governo da região com **o valor turismo**, separados por **país e cidade** de origem do hóspede
- 6) Análise dos **dias da semana** que as pessoas mais se hospedam (diária) no Hotel, separados por **mês, bimestre, trimestre e semestre**
- 7) Criar uma fórmula para a Receita = **valor da hospedagem + valor do restaurante – valor do turismo** por **nome** do hóspede.

Matriz

| INDICADORES | | Faturamento | | Diária | INDICADORES | |
|----------------------------|--|--------------|---------------------|---------------|-----------------------|----------------------------|
| DADOS LEVANTADOS | | | | | DADOS LEVANTADOS | |
| | | Valor do Bar | Valor da Hospedagem | Valor Turismo | Quantidade de Diárias | |
| Hospede | | | | | | Hospede |
| Nome do Hospede | | ✓ | ✓ | ✓ | ✓ | Nome do Hospede |
| | | | | | | |
| Regiao | | | | | | Regiao |
| Nome da Cidade | | ✓ | ✓ | ✓ | | Nome da Cidade |
| Pais | | ✓ | ✓ | ✓ | | Pais |
| | | | | | | |
| Classe Quarto | | | | | | Classe Quarto |
| Tipo da Classe | | | | | ✓ | Tipo da Classe |
| | | | | | | |
| Aeroporto | | | | | | Aeroporto |
| Nome do Aeroporto | | ✓ | ✓ | ✓ | | Nome do Aeroporto |
| | | | | | | |
| | | | | | | |
| TEMPO | | | | | | TEMPO |
| Descrição ano mês | | ✓ | ✓ | ✓ | ✓ | Descrição ano mês |
| Descrição bimestre | | ✓ | ✓ | ✓ | ✓ | Descrição bimestre |
| Descrição data dia | | ✓ | ✓ | ✓ | ✓ | Descrição data dia |
| Descrição do dia | | ✓ | ✓ | ✓ | ✓ | Descrição do dia |
| Descrição do mês | | ✓ | ✓ | ✓ | ✓ | Descrição do mês |
| Descrição mês ano | | ✓ | ✓ | ✓ | ✓ | Descrição mês ano |
| Descrição mês ano numérico | | ✓ | ✓ | ✓ | ✓ | Descrição mês ano numérico |
| Descrição mês ano completo | | ✓ | ✓ | ✓ | ✓ | Descrição mês ano completo |
| Descrição quadrimestre | | ✓ | ✓ | ✓ | ✓ | Descrição quadrimestre |
| Descrição quinzena | | ✓ | ✓ | ✓ | ✓ | Descrição quinzena |
| Descrição semestre | | ✓ | ✓ | ✓ | ✓ | Descrição semestre |
| Descrição trimestre | | ✓ | ✓ | ✓ | ✓ | Descrição trimestre |
| Número ano | | ✓ | ✓ | ✓ | ✓ | Número ano |
| Número bimestre | | ✓ | ✓ | ✓ | ✓ | Número bimestre |
| Número dia | | ✓ | ✓ | ✓ | ✓ | Número dia |
| Número mês | | ✓ | ✓ | ✓ | ✓ | Número mês |
| Número nível | | ✓ | ✓ | ✓ | ✓ | Número nível |
| Número quadrimestre | | ✓ | ✓ | ✓ | ✓ | Número quadrimestre |
| Número quinzena | | ✓ | ✓ | ✓ | ✓ | Número quinzena |
| Número semestre | | ✓ | ✓ | ✓ | ✓ | Número semestre |

Prática

Em grupos de 4 alunos, definir levantamento de dados para um DW.

Escolher entre os alunos um tema a ser abordado.

Apresentar as perguntas do(s) gestor(es).

Apresentar a Matriz Métricas x Descritores.

Modelagem de dados

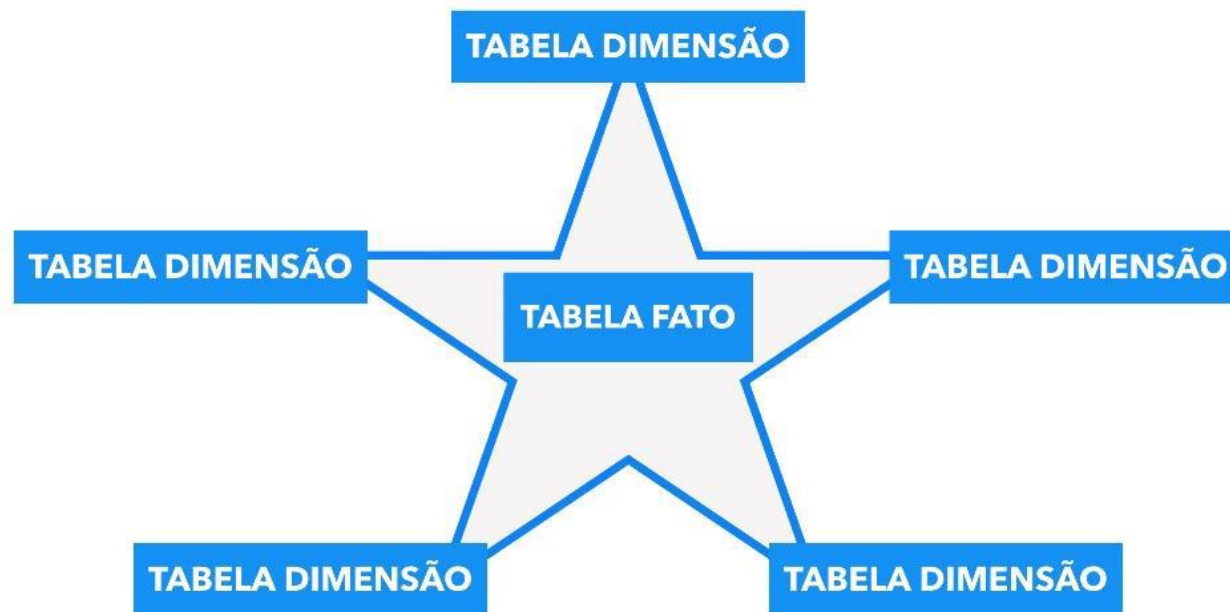
Estrutura – Modelo Estrela

O conceito de Star Schema, ou modelo estrela, foi idealizado por Ralph Kimball, um dos precursores do modelo Star Schema.

A ideia dele é propor uma visão para modelagem de dados que suporte a tomada de decisão orientada a dados, e a modelagem dimensional é uma das técnicas mais utilizadas para fazer isso, sendo o Star Schema o coração dessa modelagem dimensional.

Estrutura – Modelo Estrela

O Star Schema é composto no centro por uma tabela fato que é rodeada por tabelas de dimensão, ficando parecido com a forma de uma estrela (que é de onde vem o nome).



Modelos relacionais normalizados
(OLTP e ODS)

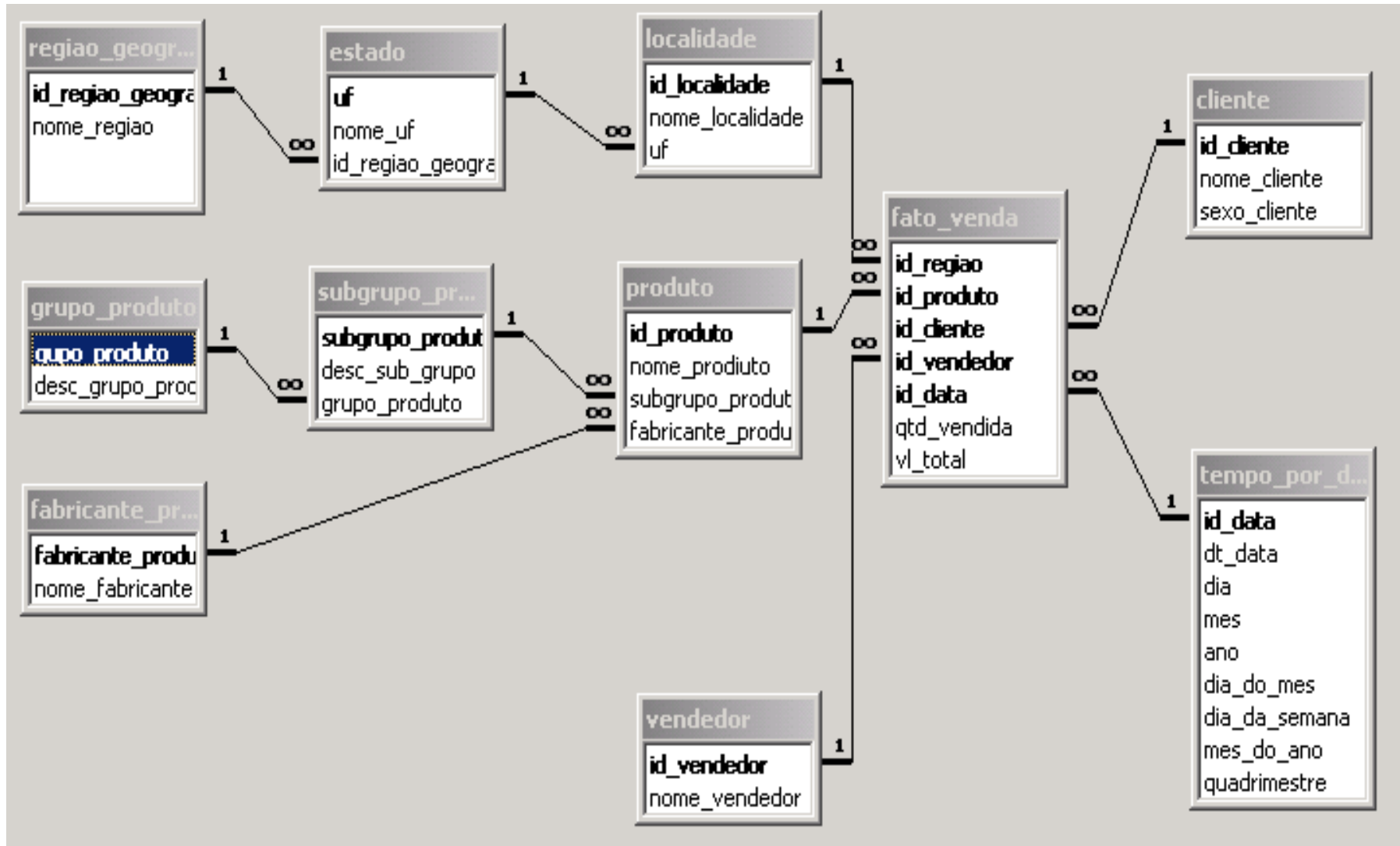
Modelos Snow Flakes

Visões ajustadas para o enfoque desejado

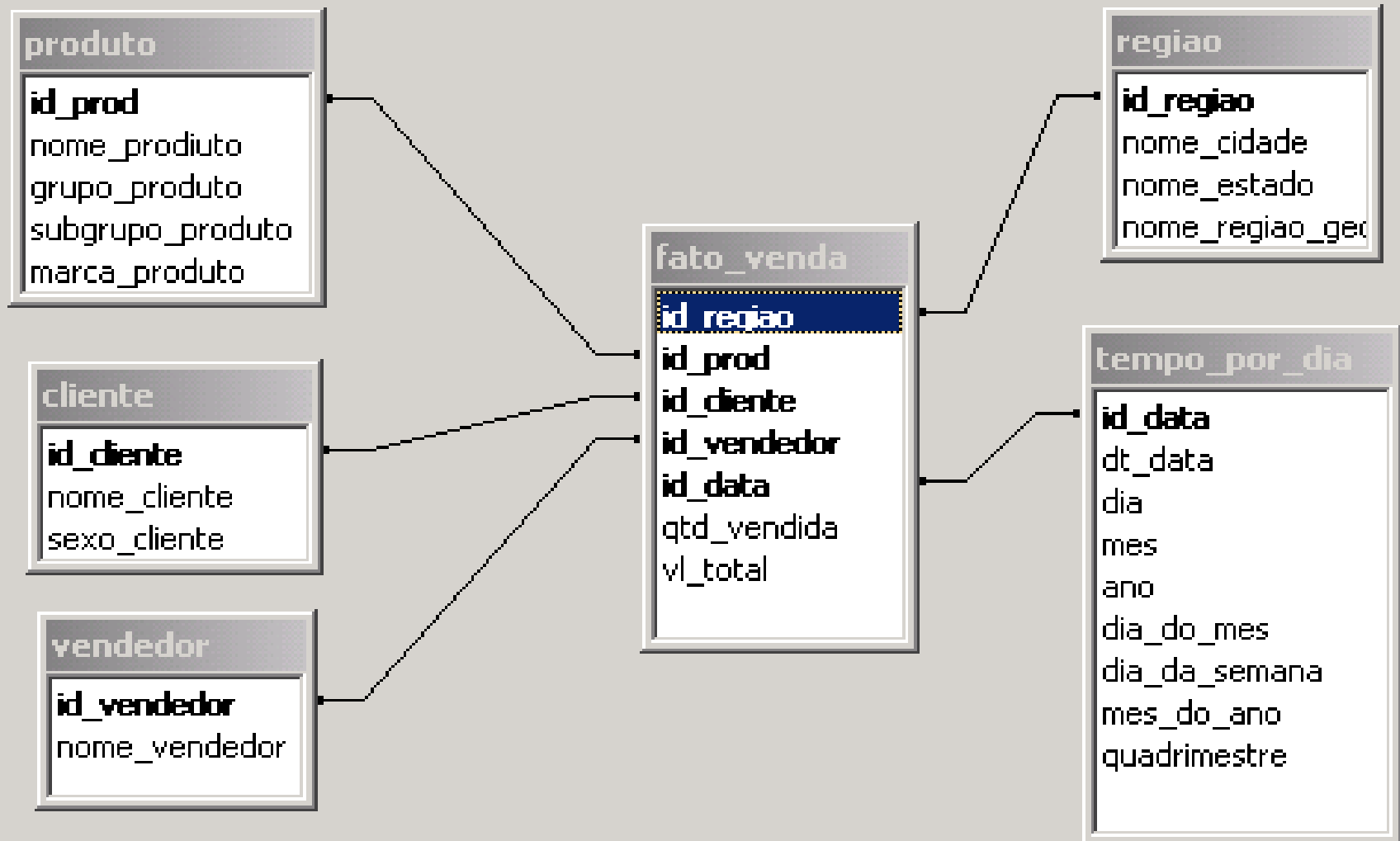
Modelos Star Schema
(Desnormalizado)

Modelos Multidimensionais
Cubos – Visa performance

Modelo Snow Flake



Modelo Star Schema



Dimensões

A dimensão, também chamada de qualificador, descreve o fato ocorrido, ela contém as características do evento.

Por exemplo, quando eu faço uma venda, quero saber por onde a venda foi feita, que produto foi vendido ou para quem.

Elas vão qualificar, classificar ou descrever os dados que estão nos fatos, ou seja, as métricas.

Para identificar as dimensões, normalmente perguntamos pelo que o usuário quer mensurar uma informação.

Elas vão ser toda e qualquer informação que qualifique uma métrica ou medida.

Dimensões

Por exemplo:

Preciso medir as vendas.

Ótimo, mas pelo quê?

Pela empresa / pelo produto / pelas filiais / por dia.

Você também pode usar a pergunta de quais filtros o usuário gostaria de ter naquela visão.

Por que filtros?

Dimensões

Imagina que você tem um dashboard com um menu e o usuário pode escolher: um relatório por produto.

Esses filtros, na maioria dos casos, vão ser as dimensões.

E embora nem todo filtro seja uma dimensão, podem ser dimensões candidatas, que ainda não foram colocadas no modelo, mas podem vir a ser colocadas.

Dimensões

Slow Changing Dimensions (Sobrescrever os dados)

O novo registro substitui o registro original. Só existe um registro no banco de dados – os dados atuais.

Não mantém histórico.

Dimensões

As dimensões armazenam 3 coisas:

a surrogate key;

a natural key;

os atributos.

TABELA DIMENSÃO

| Surrogate Key (PK) | Natural Key | Atributo 1 | Atributo 2 |
|-------------------------------------|-----------------------------|----------------------------|------------|
| Chave artificial autoincremental | PK da origem / legado | Qualificadores da dimensão | |

Dimensões

Alguns exemplos de dimensão seriam:

loja;

cliente;

produto;

data/tempo;

região/geografia;

funcionário;

fornecedor.

Dimensões

Em uma dimensão Loja, por exemplo, os atributos seriam:

nome;

endereço;

bairro;

cidade;

estado;

país.

Tipos de Chaves

Quanto às chaves do Data Warehouse, a surrogate key é uma primary key da dimensão. É uma chave artificial e autoincremental.

Ela é artificial porque não existe em lugar nenhum, não está no transacional, ela é criada no Data Warehouse.

Uma surrogate key nada mais é que um campo com as características de uma primary key, e é gerada automaticamente na hora da carga, quando você carrega a dimensão no ETL.

Tipos de Chaves

A foreign key, ou chave estrangeira, serve para você relacionar os dados entre duas tabelas.

Quando conectar a dimensão na fato, a surrogate key da dimensão vai ser referenciada na fato como uma foreign key, simples assim.

Já a natural key, ou business key, que é usada nas dimensões, é a primary key daquele dado na origem ou legado, é o que você vai usar para identificar de onde ele veio.

No Data Warehouse, ela vira uma simples natural key, não é esse código que vai controlar tudo e fazer o relacionamento, como era no transacional, aqui quem faz isso é a surrogate key.

Construção de um Modelo Estrela

Na construção de um Star Schema, a gente primeiro define a tabela fato, que possui:

foreign keys - servem para conectar com as primary keys das dimensões;

métricas - são dados numéricos.

Construção de um Modelo Estrela

TABELA FATO

| FK DIM 1 | FK DIM 2 | FK DIM 3 | FK DIM 4 | MÉTRICA |
|---|----------|----------|----------|---------------|
| FK da fato conecta com a PK da dimensão | | | | Dado numérico |
| | | | | |

O que tem em todas as dimensões que é padrão, não importando qual projeto está sendo desenvolvido:

surrogate key - uma primary key nas dimensões, uma chave artificial e autoincremental;

natural key - a primary key da origem ou legado;

atributos - qualificam os dados da dimensão.

Depois que a dimensão estiver definida, a gente faz um join dela com a tabela fato.

Construção de um Modelo Estrela

DIMENSÃO 1

| Surrogate Key (PK) | Natural Key | Atributo 1 | Atributo 2 |
|----------------------------------|-----------------------|----------------------------|------------|
| Chave artificial autoincremental | PK da origem / legado | Qualificadores da dimensão | |



TABELA FATO

| FK DIM 1 | FK DIM 2 | FK DIM 3 | FK DIM 4 | MÉTRICA |
|---|----------|----------|----------|---------------|
| FK da fato conecta com a PK da dimensão | | | | Dado numérico |

Construção de um Modelo Estrela

DIMENSÃO 1

| Surrogate Key (PK) | Natural Key | Atributo 1 | Atributo 2 |
|----------------------------------|-----------------------|----------------------------|------------|
| Chave artificial autoincremental | PK da origem / legado | Qualificadores da dimensão | |

DIMENSÃO 2

| Surrogate Key (PK) | Natural Key | Atributo 1 | Atributo 2 |
|----------------------------------|-----------------------|----------------------------|------------|
| Chave artificial autoincremental | PK da origem / legado | Qualificadores da dimensão | |

TABELA FATO

| FK DIM 1 | FK DIM 2 | FK DIM 3 | FK DIM 4 | MÉTRICA |
|---|----------|----------|----------|---------------|
| FK da fato conecta com a PK da dimensão | | | | Dado numérico |

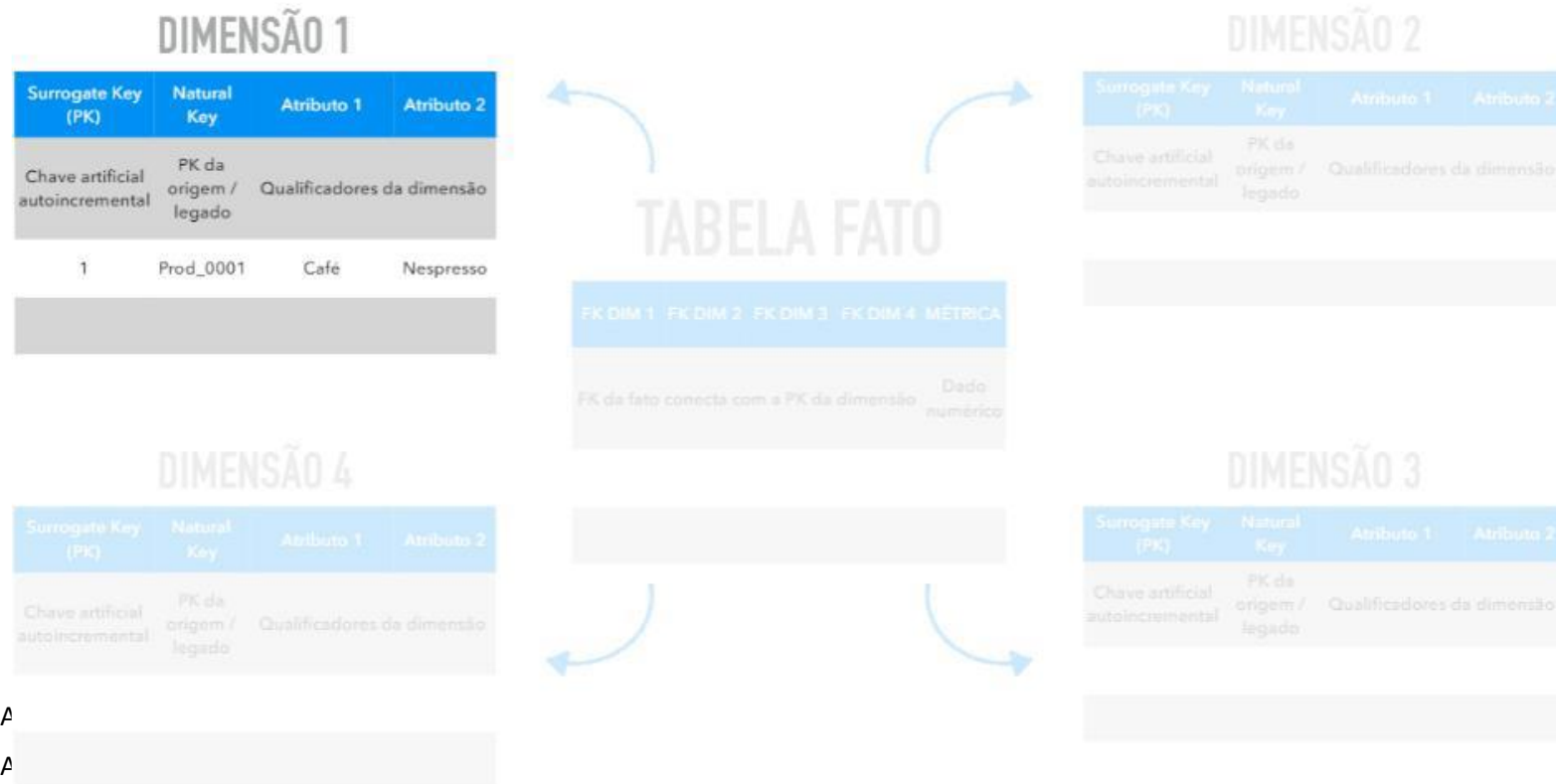
DIMENSÃO 4

| Surrogate Key (PK) | Natural Key | Atributo 1 | Atributo 2 |
|----------------------------------|-----------------------|----------------------------|------------|
| Chave artificial autoincremental | PK da origem / legado | Qualificadores da dimensão | |

DIMENSÃO 3

| Surrogate Key (PK) | Natural Key | Atributo 1 | Atributo 2 |
|----------------------------------|-----------------------|----------------------------|------------|
| Chave artificial autoincremental | PK da origem / legado | Qualificadores da dimensão | |

Construção de um Modelo Estrela



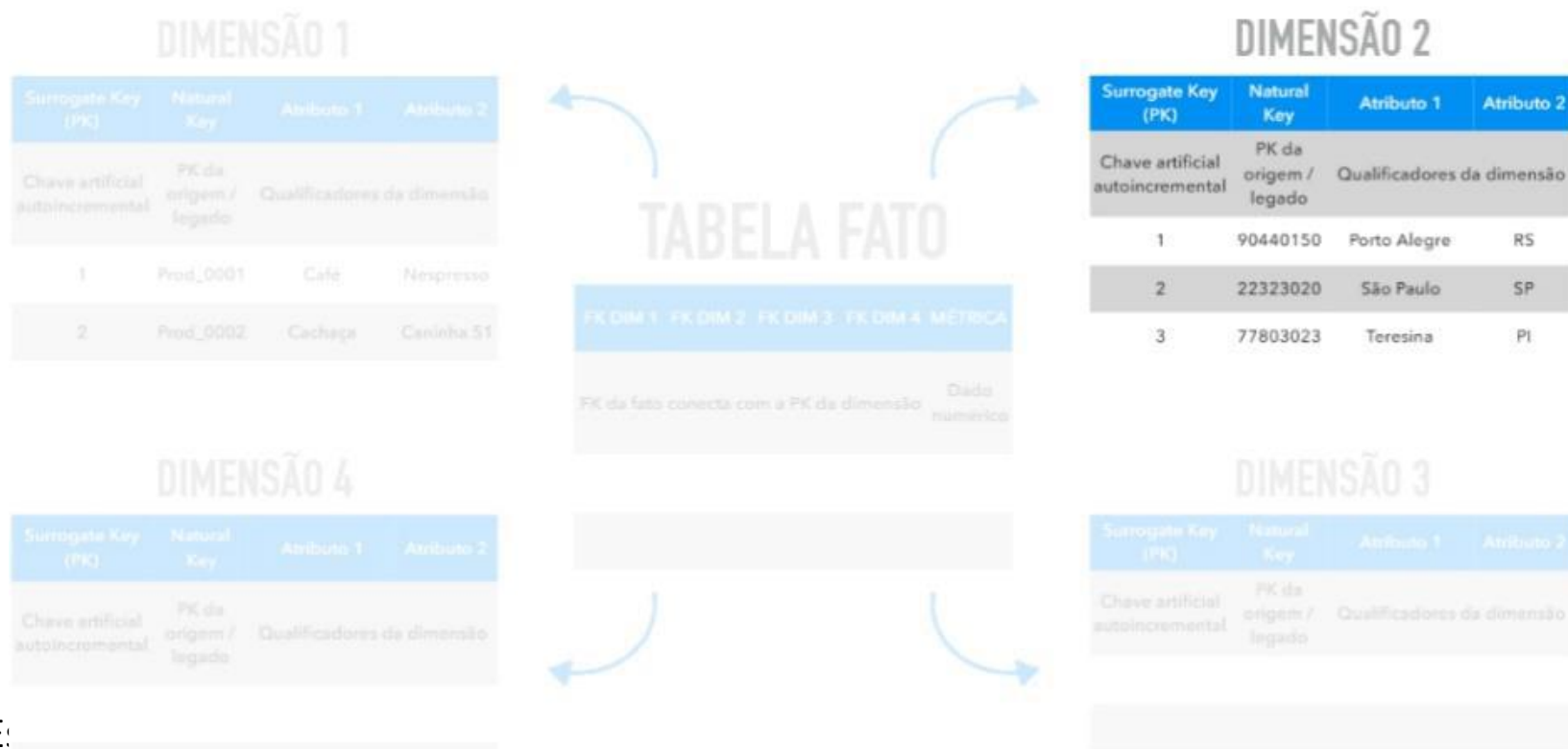
A

A

A natural key é o mesmo código que está na origem.

Os outros atributos nessa tabela são o nome e a marca do produto.

Construção de um Modelo Estrela



E:

A natural key, e o sur, que é o código único usado para identificar todos dados na origem.

E os atributos são o nome da cidade e do estado.

Construção de um Modelo Estrela

A dimensão 3 é a dimensão de tempo.

A surrogate key da dimensão de tempo normalmente é a mesma natural key, a própria data, embora não tenha problema se usar sequência como nas outras.

Aqui os atributos são o mês e o ano.

Fato



Na hora de inserir na fato, a gente vai lá na dimensão, pega a surrogate key e coloca na fato como foreign key.

Ou seja, esse 2 que está na dimensão, simplesmente vai ser inserido na linha da fato.

Fato

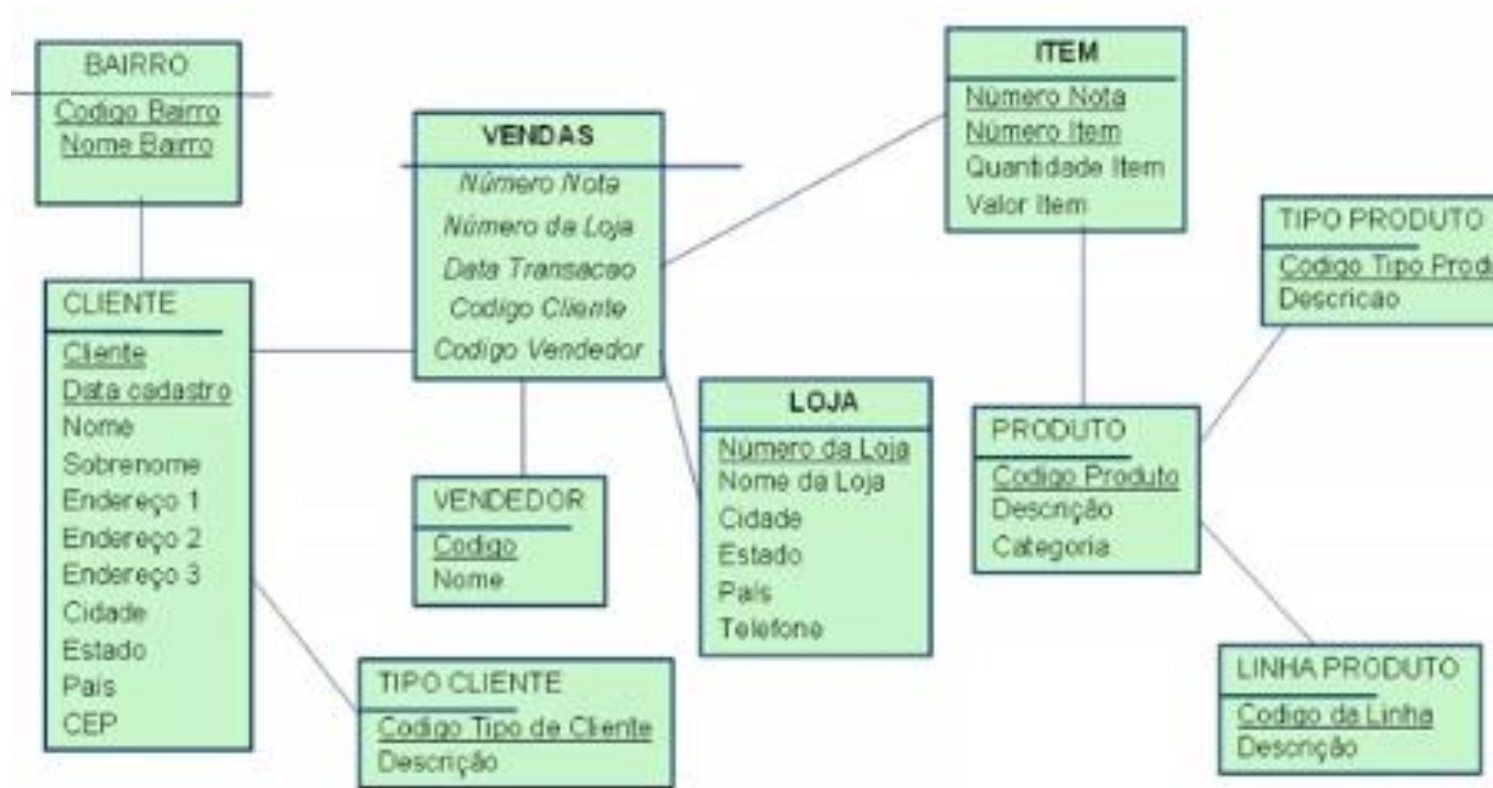


No fim, depois de fazer os cruzamentos, a gente vai na fato e insere a métrica.

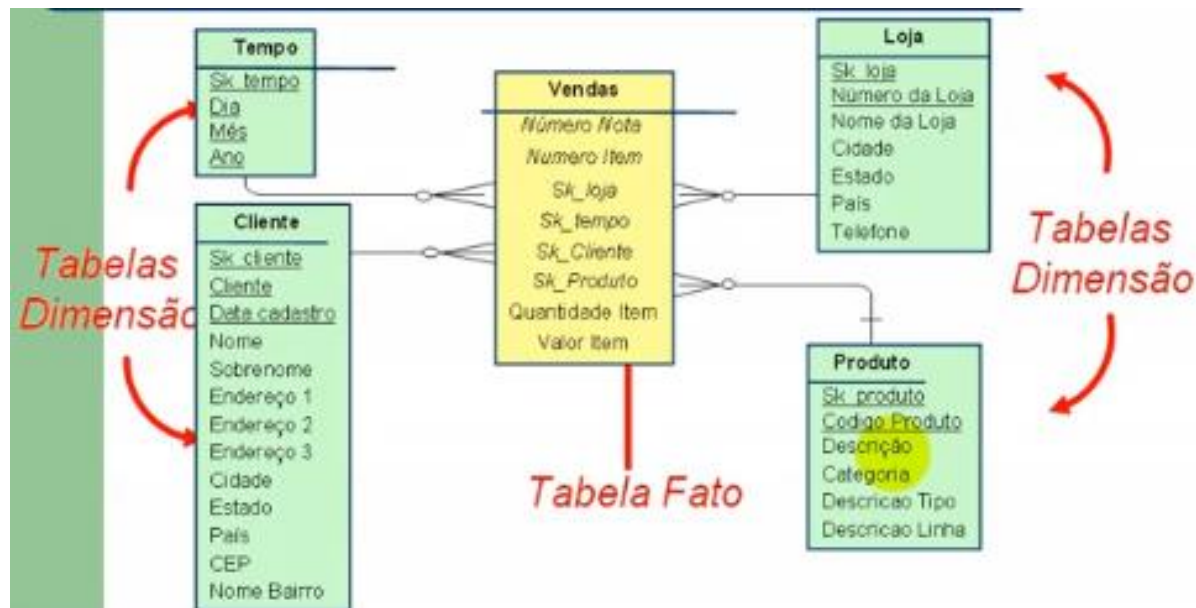
Agora vamos fazer uma leitura dessa linha que foi registrada na fato para você entender o que isso aqui quer dizer. **No dia 05/01/2017, o Rafael comprou uma cachaça da marca Caninha 51, em São Paulo, no valor de R\$200,00.**

Aplicar o DW (Dimensões e Fatos)

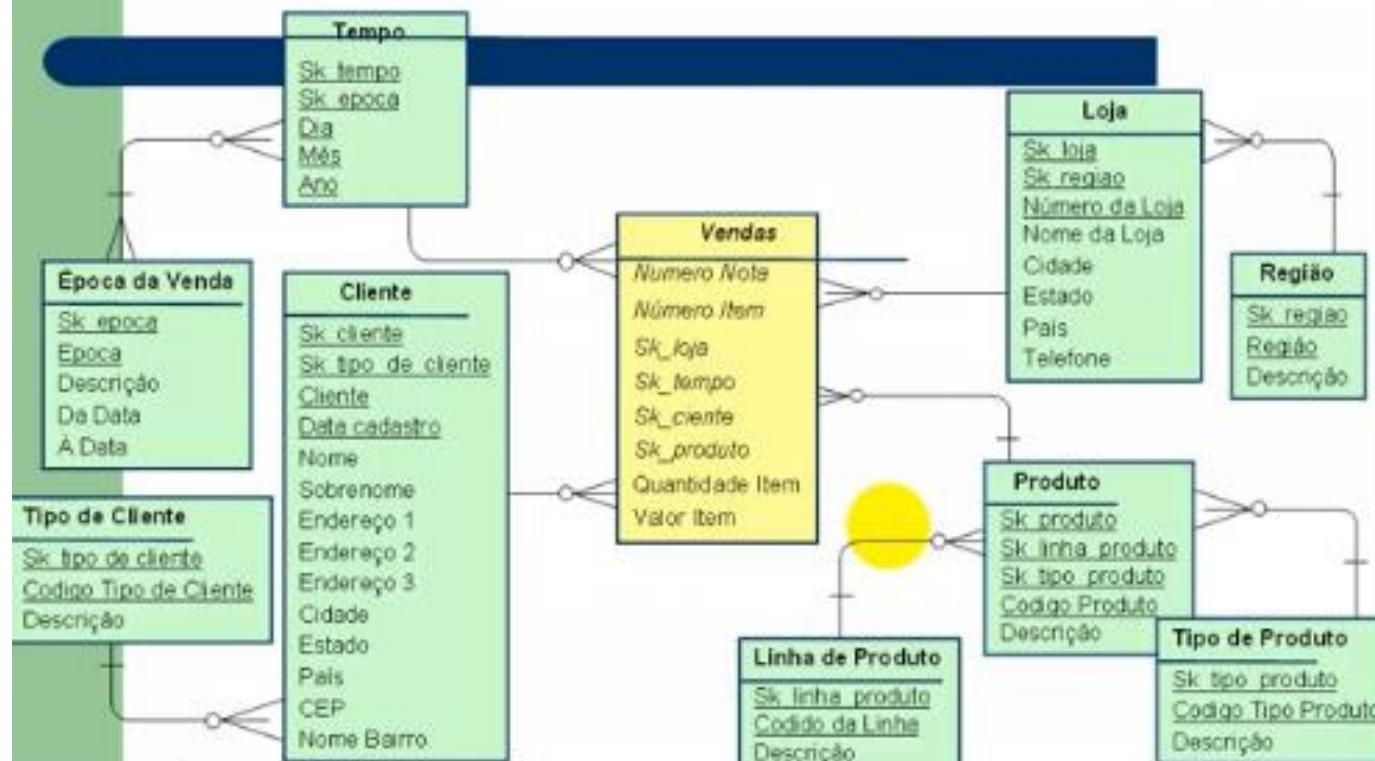
Star schema e Snow Flake



Resultado – Star Schema



Data Warehouse – Snow-Flake



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.

Fontes de Dados

Informações dos Descritores nas colunas

Fonte de Dados

| Descritor | Base Original | | | |
|-------------------|-----------------|---------------|---------|-----------------------|
| | Tabela do Banco | Atributo | Relac. | Domínio/Regra |
| Hospede | | | | |
| | Faturamento | Cod_hospede | CHAVE-1 | |
| | Hospede | Cod_hospede | CHAVE-1 | |
| Nome do Hóspede | Hospede | Nom_hospede | | |
| | | | | |
| Região | | | | |
| | Faturamento | Cod_hospede | CHAVE-1 | |
| | Hospede | Cod_hospede | CHAVE-1 | |
| | Hospede | Cod_Cidade | CHAVE-2 | |
| | Cidade_Origem | Cod_Cidade | CHAVE-2 | |
| Cidade | Nom_cidade | | | |
| | Cidade_Origem | Cod_pais | CHAVE-3 | |
| | | Cod_pais | CHAVE-3 | BRASIL,FRANÇA,ESPANHA |
| | Pais_Origem | | | |
| Pais | Pais_Origem | Nom_Pais | | |
| | | | | |
| | | | | |
| Aeroporto | | | | |
| | Faturamento | Cod_hospede | CHAVE-1 | |
| | Hospede | Cod_hospede | CHAVE-1 | |
| | Hospede | Cod_aeroporto | CHAVE-2 | |
| | Aeroporto | Cod_aeroporto | CHAVE-2 | cod_aeroporto<>99 |
| Nome do Aeroporto | Aeroporto | Des_aeroporto | | |
| | | | | |

Fonte de Dados

Vai ser utilizada através da Matriz de Necessidades (junto aos gestores) e depois criado esse documento que nos dá a visão das informações e onde se encontra no banco de dados (tabelas e atributos).

MODELO DE DADOS – HOTEL DALLAS

Hospede

Cod_hospede
Nom_hospede
Des_endereco
Cod_cidade
Cod_aeroporto
Dtc_cadastro

Cidade Origem

Cod_cidade
Nom_cidade
Cod_pais

Aeroporto Saída

Cod_aeroporto
Des_aeroporto
Nom_Localidade

Tipo classe quarto

Cod_classe_quarto
Des_classe_quarto

Diaria

Cod_diaria
Cod_hospede
Cod_tipo_quarto
Cod_classe_quarto
Val_diaria
Qtd_diaria
Qtd_pessoas
Dtc_registro_primeira_diaria

Pais Origem

Cod_pais
Nom_pais

Tipo quarto

Cod_Tipo_quarto
Des_Tipo_quarto

Faturamento

Cod_faturamento
Cod_hospede
Dtc_entrada
Dtc_saida
Val_hospedagem
Val_turismo
Val_bar

| Descritor | Base Original | | | |
|--------------------------|--------------------|--------------------|---------|-------------------------|
| | Tabela do Banco | Atributo | Relac. | Domínio/Regra |
| Hospede | Faturamento | Cod_hospede | CHAVE-1 | |
| | Hospede | Cod_hospede | CHAVE-1 | |
| Nome do Hóspede | Hospede | Nom_hospede | | |
| Região | Faturamento | Cod_hospede | CHAVE-1 | |
| | Hospede | Cod_hospede | CHAVE-1 | |
| | Hospede | Cod_Cidade | CHAVE-2 | |
| | Cidade_Origem | Cod_Cidade | CHAVE-2 | |
| Cidade | Nom_cidade | | | |
| | Cidade_Origem | Cod_pais | CHAVE-3 | |
| | Pais_Origem | Cod_pais | CHAVE-3 | BRASIL, FRANÇA, ESPANHA |
| Pais | Pais_Origem | Nom_Pais | | |
| Aeroporto | Faturamento | Cod_hospede | CHAVE-1 | |
| | Hospede | Cod_hospede | CHAVE-1 | |
| | Hospede | Cod_aeroporto | CHAVE-2 | |
| | Aeroporto | Cod_aeroporto | CHAVE-2 | cod_aeroporto<>99 |
| Nome do Aeroporto | Aeroporto | Des_aeroporto | | |

Fonte de Dados (1ª Tabela)

A primeira tabela com o atributo -> parte da Métrica (onde está a métrica relacionada com

| C | Descritor | Base Original | | |
|---|-------------------|-----------------|---------------|-------------------|
| | | Tabela do Banco | Atributo | Relac. |
| | Hospede | | | |
| | | Faturamento | Cod_hospede | CHAVE-1 |
| | | Hospede | Cod_hospede | CHAVE-1 |
| | Nome do Hóspede | Hospede | Nom_hospede | |
| | | | | |
| | Região | | | |
| | | Faturamento | Cod_hospede | CHAVE-1 |
| | | Hospede | Cod_hospede | CHAVE-1 |
| | | Hospede | Cod_Cidade | CHAVE-2 |
| | | Cidade Origem | Cod_Cidade | CHAVE-2 |
| | Cidade | Nom_cidade | | |
| | | Cidade Origem | Cod_pais | CHAVE-3 |
| | | | Cod_pais | CHAVE-3 |
| | | Pais Origem | | |
| | Pais | Pais Origem | Nom_Pais | |
| | | | | |
| | | | | |
| | Aeroporto | | | |
| | | Faturamento | Cod_hospede | CHAVE-1 |
| | | Hospede | Cod_hospede | CHAVE-1 |
| | | Hospede | Cod_aeroporto | CHAVE-2 |
| | | Aeroporto | Cod_aeroporto | CHAVE-2 |
| | Nome do Aeroporto | Aeroporto | Des_aeroporto | cod_aeroporto<>99 |
| | | | | |

Tarefa

A partir da Modelagem abaixo, criar o seu

M

