FIAP GRADUAÇÃO

TECNOLOGIA EM DESENVOLVIMENTO DE SISTEMAS

Enterprise Analytics e Data Warehousing

Contextualizando a Inteligência de Negócios com foco em BI

PROF. FERNANDO LIMA proffernando.lima@fiap.com.br

PROF. SALVIO PADLIPSKAS salvio@fiap.com.br

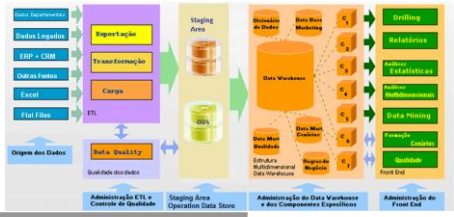
Copyright 2017

AGENDA

FUNDAMENTOS DA MODELAGEM DIMENSIONAL

- Aspectos Filosóficos (Inmon)
- Conceitos Básicos
- Introdução a Modelagem Dimensional (Kimball)
- Breve menção a Modelagem E-R
- Fatos e Dimensões
- Chaves

DATA WAREHOUSE (DW)



Data warehouse é o conjunto de dados de apoio às decisões gerenciais, integrado, não-volátil, variável em relação ao tempo e baseado em assuntos.

William Inmon

Ao processo de preparar os dados de um sistema de informação operacional de forma a se ter uma fonte de informações que possam dar suporte à tomada de decisões deu-se o nome de *data warehousing* .

Kimball

Data Warehousing não é um produto, e sim uma estratégia que reconhece a necessidade de se armazenar dados, separadamente, em sistemas de informação e consolidá-los, de forma a assistir diversos profissionais de uma empresa na tomada de decisões de modo rápido e eficaz.

CONCEITOS: DATA WAREHOUSE (DW)

- Criado pela IBM na década de 60 com o nome de *Information Warehouse.*
- Termo concebido por Bill Inmon, considerado o pai desta tecnologia.
- Foi projetado para organizar os dados que facilite e viabilize o acesso a grandes volumes de informações, o que não é possível no modelo tradicional de armazenamento de dados.
- Tem como objetivo armazenar dados de forma consolidada a partir dos sistemas transacionais (OLTP) e inúmeras outras origens de dados.

CONCEITOS: DATA WAREHOUSE (DW)

- Pode ser considerada a estrutura centralizadora das informações da empresa oriundas das inúmeras fontes de informações, com a vantagem dos dados de origens serem analisados, qualificados e disponibilizado em todos os níveis organizacionais.

- No DW as informações são armazenados de acordo com linha do tempo, permitindo uma análise histórica e comparativa dos fatos.

- Como conceito básico, no DW há apenas a carga de dados e posterior consulta.

CONCEITOS: DATA WAREHOUSE (DW)

- Devido a sua complexidade, um dos maiores desafios na implementação de um *Data Warehouse* é o seu tempo de implementação.

ASPECTOS FILOSÓFICOS GERAIS

Para Bill Inmon, um data warehouse é “um conjunto de dados baseados em assuntos, integrado, não-volátil e variável em relação ao tempo, de apoio às decisões gerenciais”.

- Orientado a assuntos

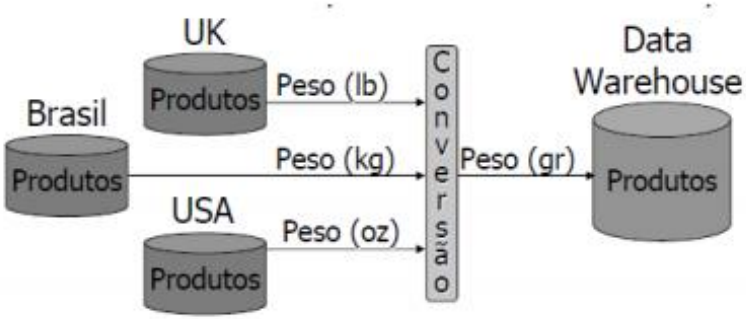
  Um data warehouse está organizado de forma a descrever o desempenho dos negócios. Os bancos de dados operacionais são voltados aos processos dos negócios.

ASPECTOS FILOSÓFICOS GERAIS

Para Bill Inmon, um data warehouse é “um conjunto de dados baseados em assuntos, integrado, não-volátil e variável em relação ao tempo, de apoio às decisões gerenciais”.

- Integrado

Diferentes nomenclaturas, formatos e estrutura das fontes de dados precisam ser acomodadas em um único esquema para prover uma visão unificada e consistente da informação



ASPECTOS FILOSÓFICOS GERAIS

Para Bill Inmon, um data warehouse é “um conjunto de dados baseados em assuntos, integrado, não-volátil e variável em relação ao tempo, de apoio às decisões gerenciais”.

- Variável em relação ao tempo

  Fica reconhecido explicitamente que o desempenho do negócio somente pode ser avaliado em relação a pontos cronológicos determinados e comparados com relação ao tempo.

Produto	Preço Vigente
Borracha	R$ 3,50
Caderno	R$ 12,00
Caneta	R$ 2,50
Lápis	R$ 1,70

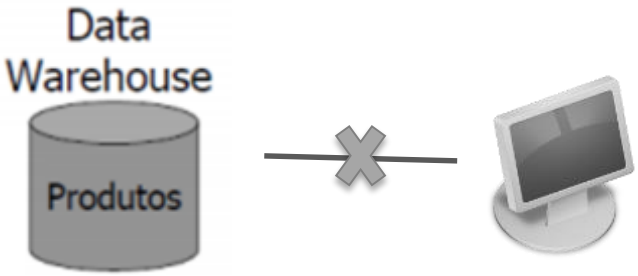
Produto	Janeiro 2013	Março 2014	Janeiro 2015	Fevereiro 2016
Borracha	R$ 2,50	R$ 2,80	R$ 3,10	R$ 3,50
Caderno	R$ 8,50	R$ 9,40	R$ 11,20	R$ 12,00
Caneta	R$ 1,50	R$ 1,90	R$ 2,10	R$ 2,50
Lápis	R$ 0,70	R$ 0,95	R$ 1,50	R$ 1,70

ASPECTOS FILOSÓFICOS GERAIS

Para Bill Inmon, um data warehouse é “um conjunto de dados baseados em assuntos, integrado, não-volátil e variável em relação ao tempo, de apoio às decisões gerenciais”.

- Não volátil

  Uma vez que o dado foi incorporado no warehouse esse é fixo e imutável, contrariamente ao que ocorre com os dados tipicamente transacionais



DATA WAREHOUSE: CONCEITOS BÁSICOS

A visão de Kimball

Ralph Kimball, talvez o autor mais referenciado em trabalhos práticos na modelagem e construção de data warehouse (a partir de Data Marts), distingue os projetos de data warehouse em **Entidade-relacionamento e Dimensiona**l, que foi um marco na consolidação da Modelagem Dimensional, hoje amplamente aceita e dominante.

DATA WAREHOUSE: CONCEITOS BÁSICOS

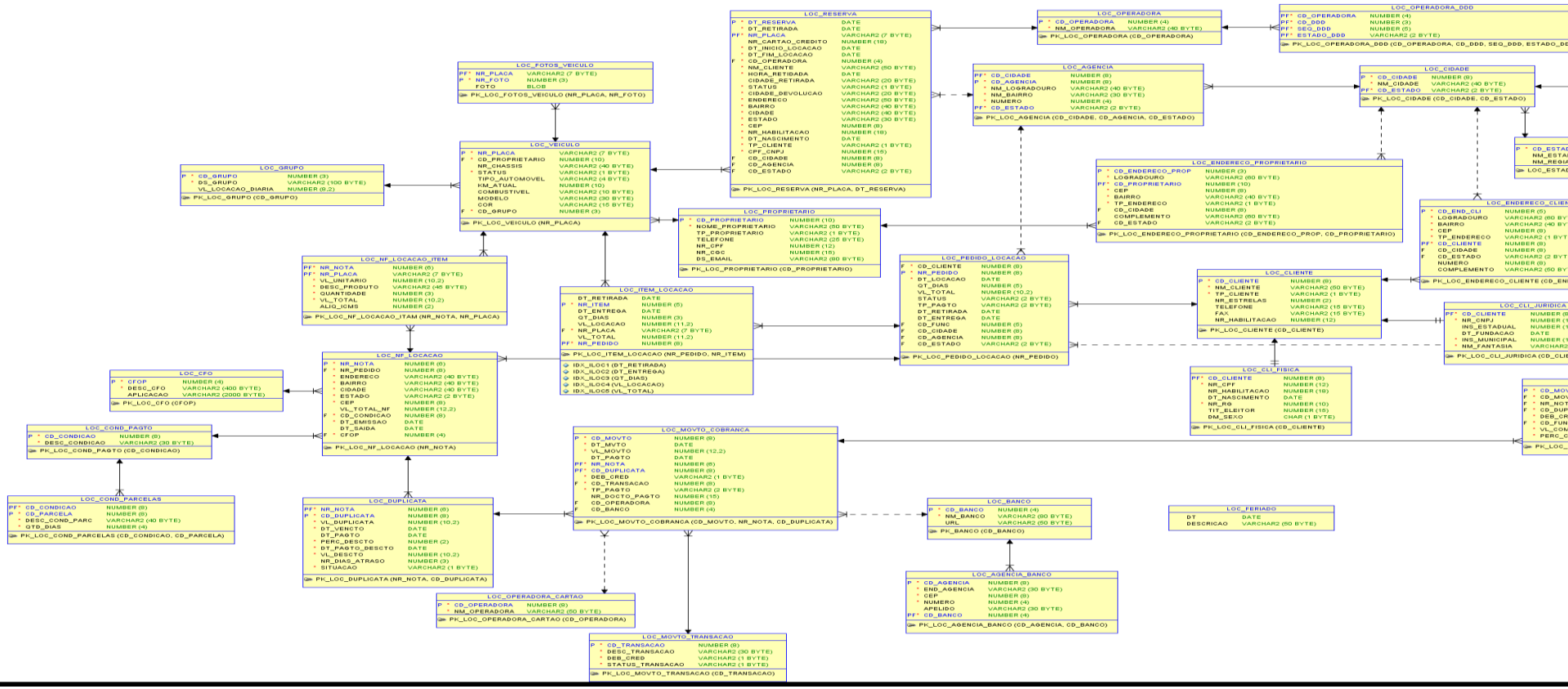
Vejamos o porquê disso...

- Enquanto o modelo **Entidade-relacionamento** é uma técnica de modelo de Banco de Dados relacional normalizado que enfatiza a eliminação das redundâncias de dados no projeto.

- É uma técnica que tem como seu principal objetivo o armazenamento estruturado, otimizado, consistente e por final, com menor nível de redundância possível.

- Devido ao alto número de estruturas, o entendimento desses modelos normalmente é facilmente interpretado somente por especialistas.

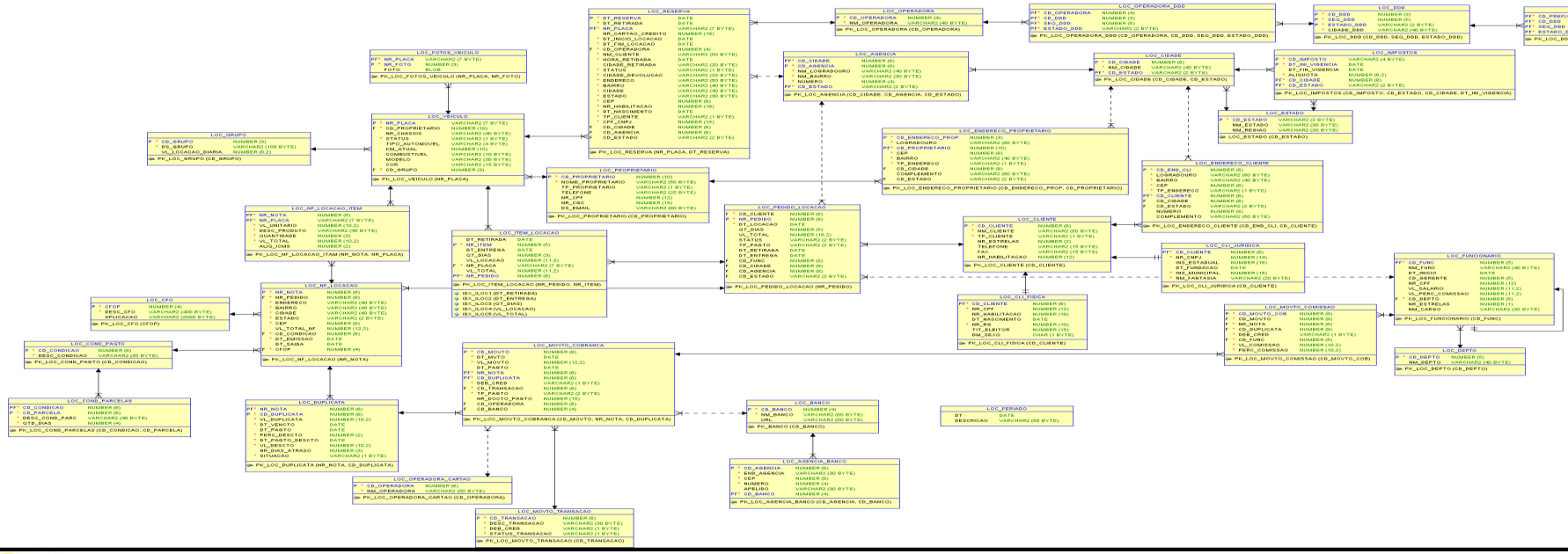
MODELO ENTIDADE-RELACIONAMENTO



- No exemplo, nota-se que através da busca das entidades e de seus relacionamentos, é possível modelar um DW voltado à compreensão do negócio.
- Contudo, essa abordagem é mais tecnicista do que voltada ao negócio, o que compromete seus resultados.

CARACTERÍSTICAS DO MODELO E-R

No modelo E-R, se observa que há uma fortíssima semelhança entre o projeto de um banco de dados voltados ao controle de transações diárias, excetuado o fato da união desses dados a outros mais gerenciais, como os



Verifique um modelo de dados Entidade Relacionamento completo no arquivo **Aula 10 OLTP_Carloca.jpg**

PRÓS E CONTRAS DO MODELO E-R

- Nesse tipo de modelo o acesso é realizado campo a campo. Não há uma adequação maior a forma como o usuário navega, pois há uma separação lógica entre os movimentos e os cadastros.

- Este tipo de situação, frequentemente está em desacordo as exigências de flexibilidade desejadas pelos usuários, portanto exige um projeto mais consistente, pois é grande a possibilidade dos usuários “se perderem” na navegação.

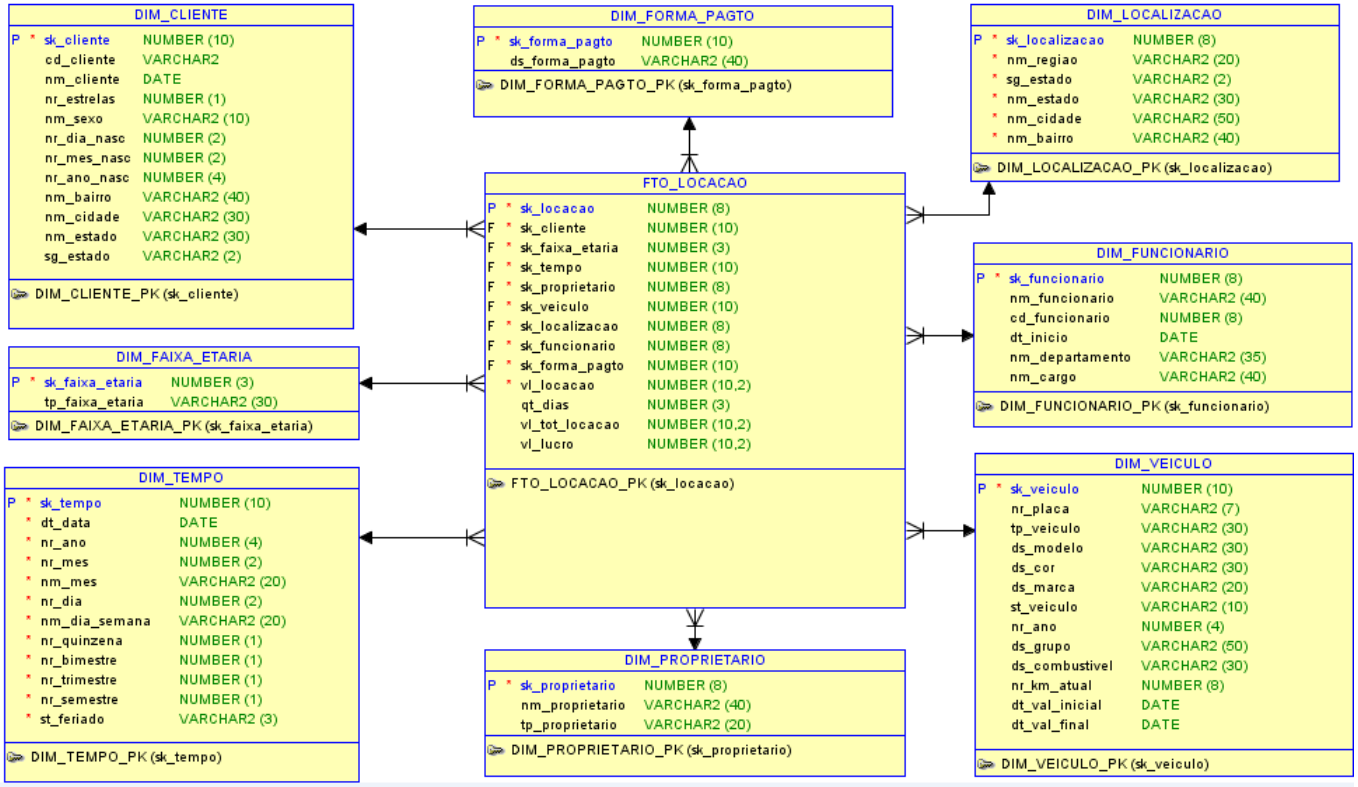
- Por outro lado, na medida em que a replicação dos dados é controlada, ocorre uma (pequena) economia de espaço.

ABORDAGEM DIMENSIONAL

Já a **abordagem dimensional** resulta em um projeto de banco de dados voltado a otimização da performance nos caminhos previstos para o acesso aos dados pelos usuários.

Assim vários cálculos intermediários são armazenados, criando úteis redundâncias que tornam possível evitar-se cálculos repetitivos.

MODELO DIMENSIONAL

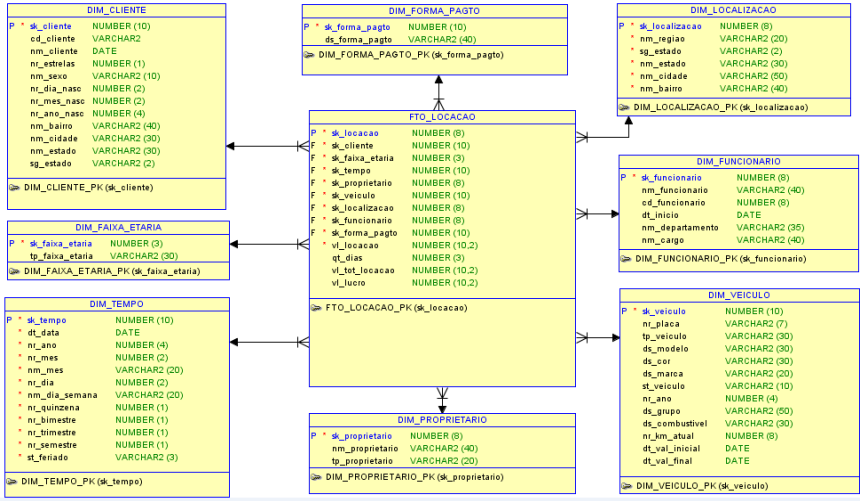


Já no exemplo acima é fácil observar que a separação não é feita por modelos técnicos e sim a partir do negócio em si.

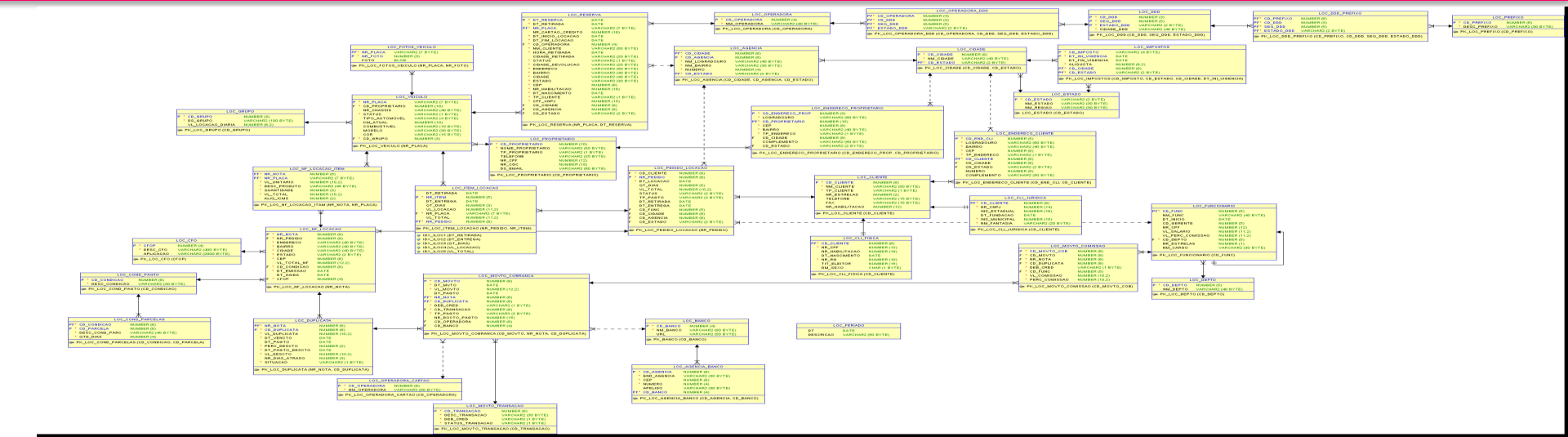
Essa será a abordagem que norteará nossos estudos.

MODELAGEM DIMENSIONAL: COMPARATIVO

Modelagem Dimensional



Modelagem Relacional



MODELAGEM DIMENSIONAL: COMPARATIVO

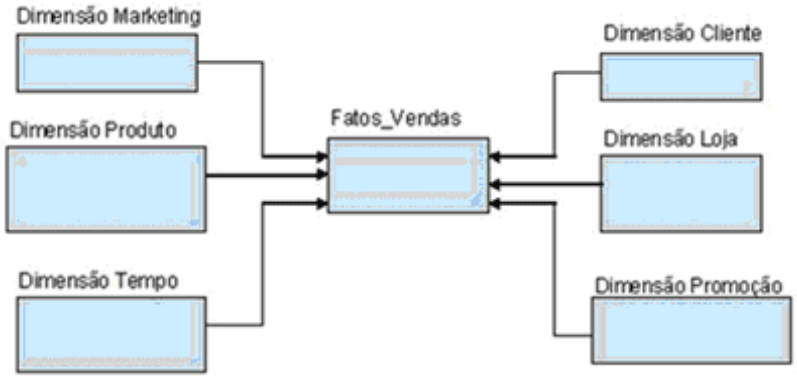
Modelo Dimensional	Modelo Relacional
Estrutura Intuitiva	Estrutura Complexa
Concebida nos Anos 60	Concebida nos Anos 70
Tabelas de Fato (Núcleo) Normalizadas	Todas Tabelas Normalizadas
Dimensões, funcionam como filtros e são usualmente desnormalizadas. São ponto de entrada nas pesquisas.	Não se aplica esse conceito
Joined simplificado	Joined complexo
Somente Leitura (Otimizado para Queries)	Atualizável(Otimizado para Atualizações)

MODELAGEM DIMENSIONAL: CONCEITOS BÁSICOS

Bases da Modelagem Dimensional: Tabelas Dimensionais e Fato

- **Tabelas Dimensionais**
  - São tabelas que fornecem a base para analisar dados das tabelas fato. Fragmentam os dados em pedaços menores, permitindo a visualização em faces diferentes.
  - As dimensões normalmente responde as questões do tipo “Quem”, “O que”, “Quando”, “Onde” e “ Por que”.

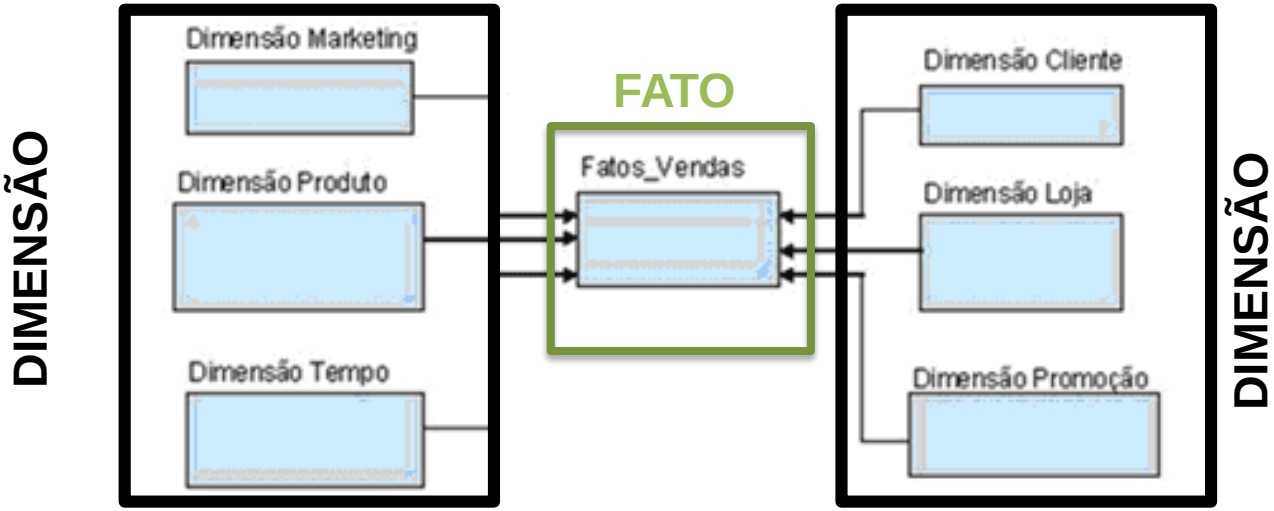
- **Tabelas Fatos**
  - São normalmente tabelas que representam processos de negócio, por exemplo, PEDIDO, ENTREGA, PAGAMENTO. Possuem valores numéricos que permitem análises, por exemplo, “TOTAL EM REAIS”, “TOTAL EM DÓLARES”.



MODELAGEM DIMENSIONAL: CONCEITOS BÁSICOS

Modelagem dimensional é uma técnica de organização de dados para um ambiente de Data Warehouse. Seu design é criado para obter as melhores performances em consultas a grandes volumes de dados. Segundo seu criador Prof. Kimball, a Modelagem dimensional é a única técnica viável Banco de dados de consulta de alta performance.

É composta por dois tipos de Entidades:

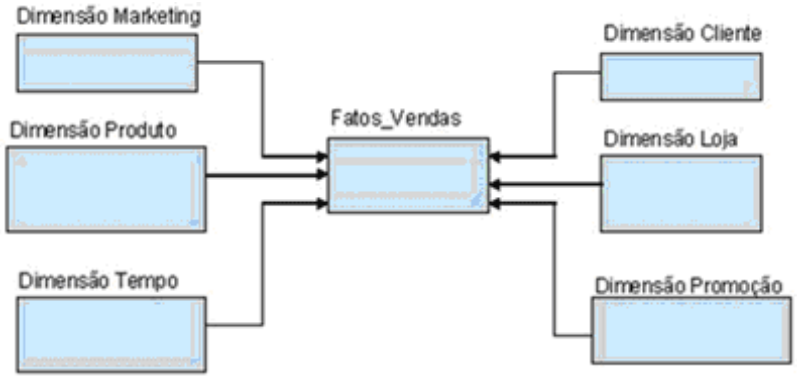


MODELAGEM DIMENSIONAL: CONCEITOS BÁSICOS

Bases da Modelagem Dimensional: Tabelas Dimensionais e Fato

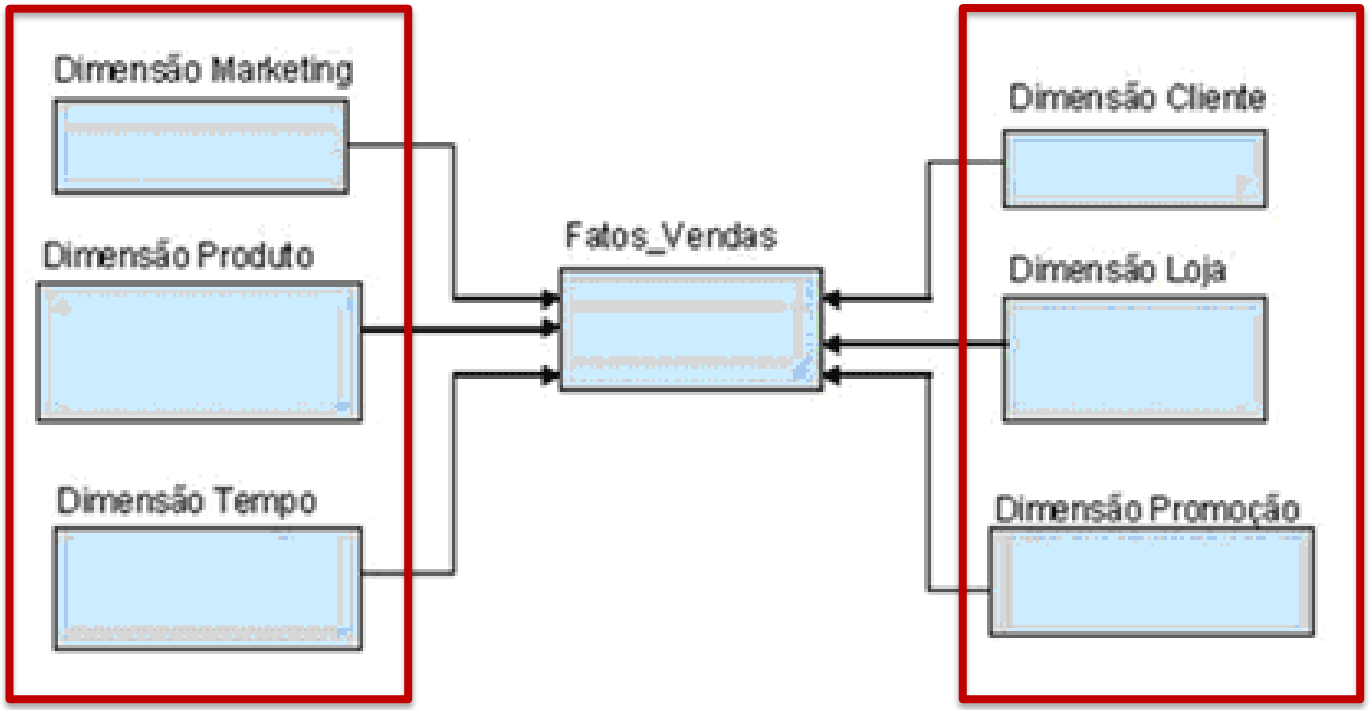
- **Tabelas Dimensionais**
  - São tabelas que fornecem a base para analisar dados das tabelas fato. Fragmentam os dados em pedaços menores, permitindo a visualização em faces diferentes.
  - As dimensões normalmente responde as questões do tipo “Quem”, “O que”, “Quando”, “Onde” e “ Por que”.

- **Tabelas Fatos**
  - São normalmente tabelas que representam processos de negócio, por exemplo, PEDIDO, ENTREGA, PAGAMENTO. Possuem valores numéricos que permitem análises, por exemplo, “TOTAL EM REAIS”, “TOTAL EM DÓLARES”.



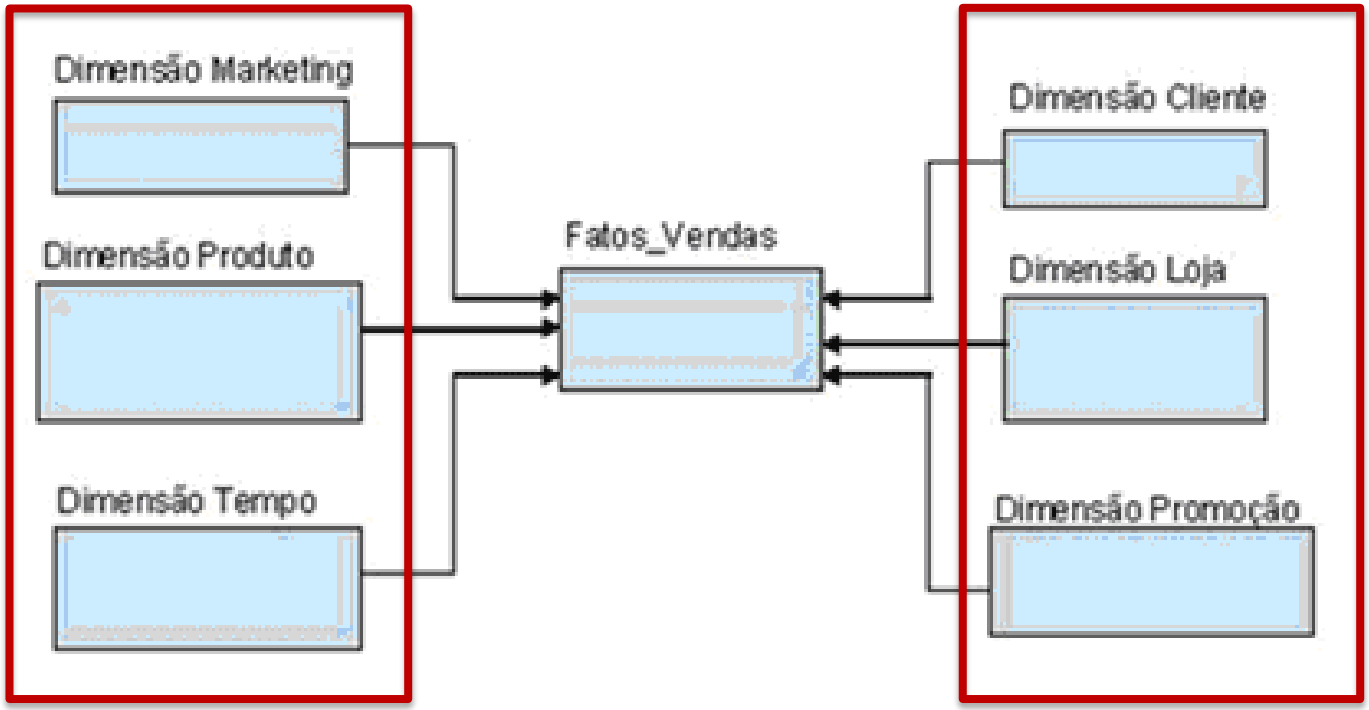
BASES DA MODELAGEM DIMENSIONAL

- As **tabelas dimensionais** contêm informações sobre as dimensões dos dados, como períodos de tempo, produtos, mercados, organizações, contas, fornecedores e clientes. Essas informações incluem descrições e atributos de dimensão, como grupos de produtos em marcas e categorias, em cidades e estados e em distritos e regiões, e demais atributos derivados das demais tabelas.



TABELAS DIMENSIONAIS

São tabelas que representam as entidades as quais se quer analisar o comportamento ao longo do tempo. Devemos notar que diferentemente do que ocorre nos modelos transacionais, uma série de replicações entre entidades dependentes é normalmente verificada.



TABELAS DIMENSIONAIS

- As tabelas dimensionais devem conter informações sobre as dimensões dos dados (períodos de tempo, produtos, mercados, etc.). Estas devem ser projetadas em função de pesquisas centrada no usuário, aonde os campos devem conter descrições significativas aos usuários e ideais para descrições de relatórios.

- As colunas de atributos devem servir para filtrar o conteúdo de uma dimensão, suportando questões do tipo **“e se?”** usualmente utilizadas para tomada de decisão.

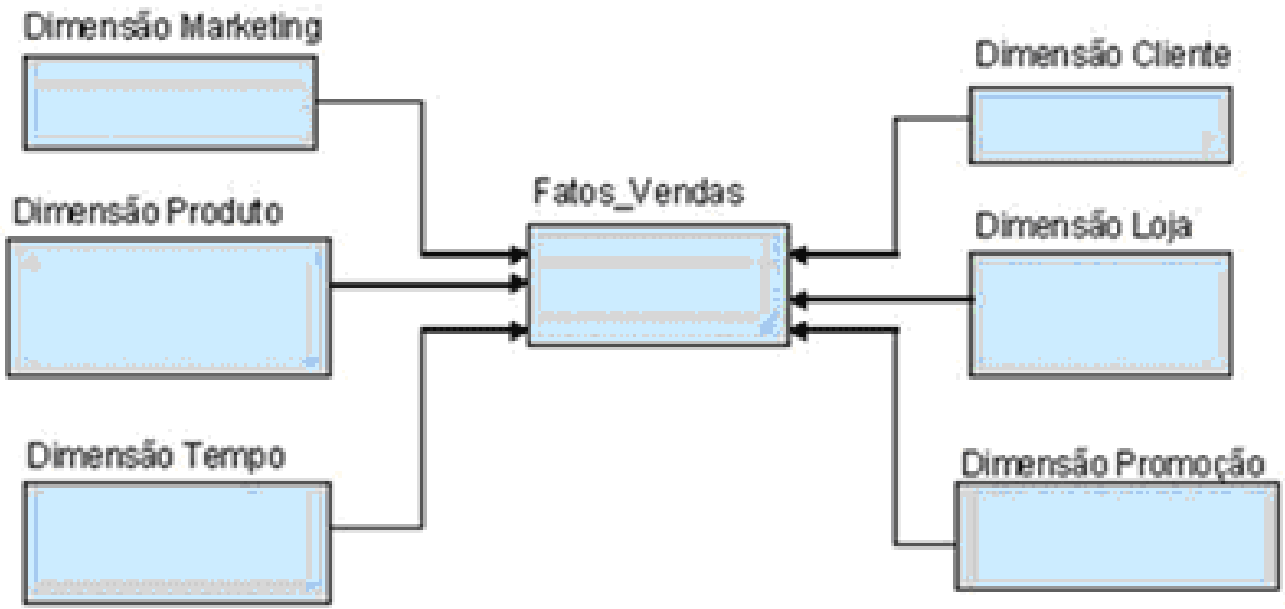


TABELA DIMENSIONAL - EXEMPLO

Chave de Região	Descrição da Região	Cidade	Estado	Área	Nível
101	Belo Horizonte	101	201	301	Cidade
102	Porto Alegre	102	202	302	Cidade
201	Minas Gerais	NA	201	301	Estado
202	Rio Grande do Sul	NA	202	302	Estado
301	Região Sudeste	NA	NA	301	Área
302	Região Sul	NA	NA	302	Área

BASES DA MODELAGEM DIMENSIONAL

- As **tabelas de fatos** que se **originam das tabelas de movimentos e de operações dos sistemas transacionais**, incluem dados temporais nos níveis de estruturas indicadas. Assim, se a tabela de dimensão contém uma estrutura que reflete a cidade, estado, região e o total de uma dimensão, então as tabelas de fato devem conter informações sobre todas as estruturas existentes na tabela de dimensão. O nível mais baixo dos dados é resumido em períodos de tempo, em vez de ser organizado de acordo com transação ou ocorrência.

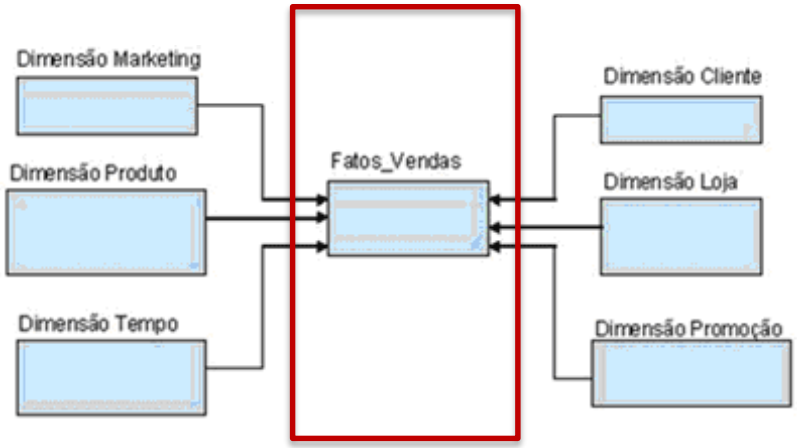


TABELA DE FATO

- Os fatos, nada mais que as medidas de negócio de uma organização, são mantidos em linhas. Cada uma dessas linhas tem uma chave indexada composta logicamente de uma dimensão relevante ao negócio.

Chave de Período	Chave de Produto	Chave de Região	Venda em Valores	Venda em Unidades	Vendas s/ Imposto
20020207	101	101	300	42	160
20020208	101	101	310	53	180
20020209	101	101	290	40	130
20020210	101	101	330	66	180

CHAVE SIGNIFICATIVA: UMA DISCUSSÃO...

Uma chave significativa geralmente é obtida por composição entre uma série de chaves estrangeiras, o que termina por determinar uma chave primária segmentada ou composta, que prejudica o desempenho do sistema de controle do banco de dados, quando esses estão montados em modelos relacionais.

MODELAGEM DIMENSIONAL: CONCEITOS BÁSICOS

Chave significativa: **uma discussão...**

Chave de Período	Chave de Produto	Chave de Região	Venda em Valores	Venda em Unidades	Vendas s/ Imposto
20020207	101	101	300	42	160
20020208	101	101	310	53	180
20020209	101	101	290	40	130
20020210	101	101	330	66	180

MODELAGEM DIMENSIONAL: CONCEITOS BÁSICOS

***Chave significativa:* uma discussão...**

Chave de Mercado	Venda em Valores	Venda em Unidades	Vendas s/ Imposto
10110120020207	300	42	160
10110120020208	310	53	180
10110120020209	290	40	130
10110120020210	330	66	180

Chave Artificial a partir da junção das chaves estrangeiras

Chave de Mercado	Chave de Período	Chave de Produto	Chave de Região	Venda em Valores	Venda em Unidades	Vendas s/ Imposto
1	20020207	101	101	300	42	160
2	20020208	101	101	310	53	180
3	20020209	101	101	290	40	130
4	20020210	101	101	330	66	180

Chave artificial gerada pelo sistema

MODELAGEM DIMENSIONAL: CONCEITOS BÁSICOS

Chave significativa para o Modelo Dimensional

- **Flexibilidade ideal:** uma definição única de metadados pode ser usada para qualquer nível de detalhes ou sumários que compartilhem tal chave.

- **Estrutura consistente:** com a utilização do *Data Warehouse*, habitualmente novos níveis de sumários são acrescentados ou removidos, a estrutura física das tabelas que compartilhem a chave gerada não mudam.

- **Tamanho:** Reduz o tamanho da chave primária indexada (por usar menos colunas como índice), acarretando aumento de desempenho.

MODELAGEM DIMENSIONAL: CONCEITOS BÁSICOS

Surrogate Key

- Surrogate key é a chave substituta para a chave primária natural dos dados de origem.
- Ela deva ser um identificador único (numérico) para cada linha utilizada em qualquer estrutura dimensional.
- O único requerimento para uma surrogate key é ser única para cada linha.
- Não devem ser dependente em nenhuma hipótese de qualquer chave de origem e com isso geralmente acaba sendo um valor artificial criado para associar informações.

MODELAGEM DIMENSIONAL: CONCEITOS BÁSICOS

Surrogate Key

Sales person dimension

Surrogate key	Business key	Name	Territory	Is Current
1	JONL	James O'Neill	N	Y
2	SVLM	Steve Lamb	E	N
3	SVLM	Steve Lamb	SW	Y

Sales facts

Surrogate key	Date key	Sales Amount	Sales Quantity
1	20070818	1000	3
2	20070818	500	4
2	20070818	700	14
3	20070821	900	18
3	20070821	1200	36
3	20070821	1500	75

- O conteúdo é gerado automática na fase de carga pelo sistema.
- O conteúdo não é manipulável em nenhuma hipótese.
- O conteúdo não contém nenhum significado semântico.
- O conteúdo não é visível para o usuário ou aplicação (no caso de modelagem relacional).
- O conteúdo não é composto de vários valores a partir de diferentes domínios.

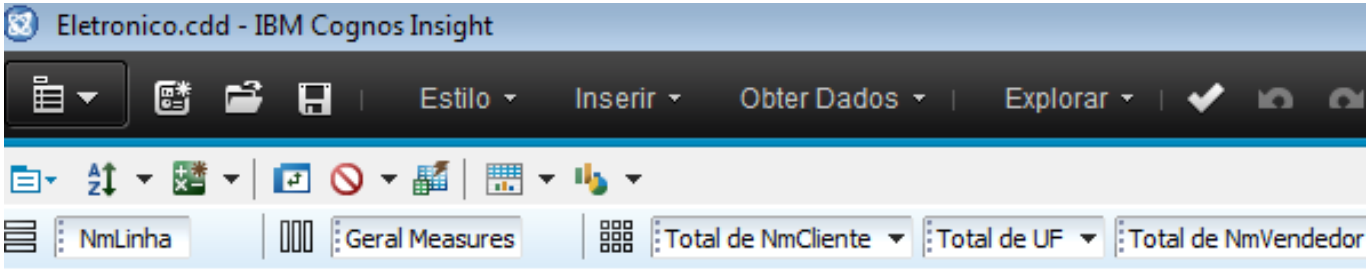
Exercício

Detentora de um marketshare de mais 20% do mercado nacional a Patatinhas é uma empresa nacional que produz bens de consumo realizando entrega em todo o território nacional, nas categorias de Linha Branca (Eletrodoméstico e produtos para a residência em geral). Hoje, o mercado nacional é atendido por mais de 50 produtos agrupados em 4 Linhas.

Para aumentar sua participação no mercado, a empresa vem realizando reestruturações na área de vendas e recentemente foi contratado um novo Diretor Comercial que deseja ter uma visão dos números de vendas por Linha de Produto, Produto, Vendedor, Cliente e Região.

Nesse sentido, analise os 2 slides a seguir e com a orientação do professor crie os relatórios que alcancem esse objetivo.

DRILLING: CONCEITO DE AGRUPAMENTO



	Qtde	Preco	PrecoICMS
Total de NmLinha	8.168	R$ 7.187.324	R$ 8.866.237
ELETRODOMESTICOS	2.329	R$ 2.311.118	R$ 2.819.590
TI	3.166	R$ 2.605.301	R$ 3.246.926
SOM & IMAGEM	1.378	R$ 1.145.022	R$ 1.416.896
TELEFONIA	1.295	R$ 1.125.882	R$ 1.382.825

ELETRODOMESTICOS	2.329	R$ 2.311.118	R$ 2.819.590
ELETRODOMESTICOS, LINHA BRANCA	509	R$ 367.588	R$ 456.072
ELETRODOMESTICOS, ELETRODOMESTICOS DE COZINHA	1.076	R$ 1.353.346	R$ 1.639.292
ELETRODOMESTICOS, APARELHOS DE AR	744	R$ 590.184	R$ 724.227

DRILLING: CONCEITO DE AGRUPAMENTO

	Qtde	Preco	PrecoICMS
⊟ Total de NmLinha	8.168	R$ 7.187.324	R$ 8.866.237
⊟ ELETRODOMESTICOS	2.329	R$ 2.311.118	R$ 2.819.590
⊟ ELETRODOMESTICOS, LINHA BRANCA	509	R$ 367.588	R$ 456.072
⊞ ELETRODOMESTICOS, LINHA BRANCA, FOGAO	220	R$ 157.458	R$ 196.835
⊞ ELETRODOMESTICOS, LINHA BRANCA, GELADEIRA	251	R$ 181.139	R$ 222.800
⊞ ELETRODOMESTICOS, LINHA BRANCA, LAVADORA	38	R$ 28.990	R$ 36.437
⊞ ELETRODOMESTICOS, ELETRODOMESTICOS DE COZINHA	1.076	R$ 1.353.346	R$ 1.639.292
⊞ ELETRODOMESTICOS, APARELHOS DE AR	744	R$ 590.184	R$ 724.227
⊟ TI	3.166	R$ 2.605.301	R$ 3.246.926
⊞ TI, MICRO-COMPUTADOR	451	R$ 358.106	R$ 447.132
⊟ TI, NETBOOK	1.347	R$ 1.046.627	R$ 1.305.692
⊞ TI, NETBOOK, XING-LING	338	R$ 285.773	R$ 354.906
⊞ TI, NETBOOK, BARATO	829	R$ 600.584	R$ 753.825
⊞ TI, NETBOOK, MARCA	180	R$ 160.269	R$ 196.961
⊞ TI, IPAD	913	R$ 768.697	R$ 955.943
⊞ TI, NOTEBOOK	455	R$ 431.872	R$ 538.159
⊞ SOM & IMAGEM	1.378	R$ 1.145.022	R$ 1.416.896
⊞ TELEFONIA	1.295	R$ 1.125.882	R$ 1.382.825

Exercício

1) Abra o arquivo **Aula 10 Eletro.cdd** na ferramenta Cognos Insight (*siga as orientações do professor)

2) A partir da planilha **Aula 10 Eletro.xls** sugira o modelo de dados Dimensional utilizando a técnica de engenharia Reversa.

Copyright © 2017 Prof. Fernando Lima e Prof. Salvio Padlipskas

Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proíbido sem o consentimento formal, por escrito, do Professor (autor).