



## Objetivos da Aula:

### - Modelagem de Dados – DW

Professora: Juliana

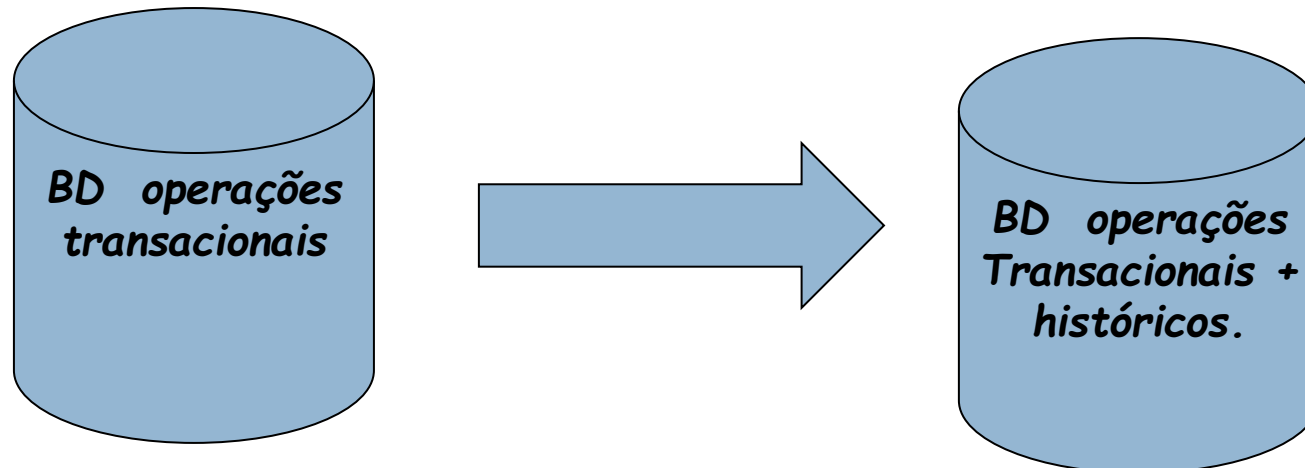
22-8-2025

# Modelagem de Dados para Data Warehouse

2

- A Modelagem de Dados para um Data Warehouse é diferente da utilizada para os sistemas de operações transacionais (OLTP);

## Exemplo do que Não é DW!!!



- Complexidade alta para o usuário realizar consulta;
- Modelos de dados transacionais estão na 3FN, requerem 5 ou mais joins de tabelas para realização de queries de apoio à decisão;

# **Qual a importância da Modelagem de Dados?**

# Modelagem Multidimensional

4

- A Modelagem multidimensional é uma técnica de concepção e visualização de um modelo de dados;
- Utilizada para projetar a sumarização e reestruturação de dados para fornecer a análise dos valores desses dados.

# Modelo Multidimensional

5

Um modelo multidimensional é formado por três elementos básicos:

- Fatos;
- Dimensões;
- Medidas (variáveis).

# Medidas e Dimensões

6

- **Uma Medida** é um valor numérico totalizado utilizado para analisar/observar um determinado negócio.  
Exemplo: Quant. De Vendas.
- **Dimensões** filtros que afetam a maneira de visualizar as medidas.

## Relatório Unidades Vendidas

7

113
-----

**Precisamos de dimensões para entender esse valor?**

## Análise Mensal

8

Elementos

January	February	March	April
14	41	33	25

- Unidade total de vendas (Medida) foi dividida em outros valores (meses).
- Quatro elementos da dimensão tempo.

**Podemos inserir novas dimensões?**



## Análise por mês e produto

	January	February	March	April
Road-650			6	17
Mountain-100	6	16	6	8
Cable Lock	8	25	21	

- Duas Dimensões: - Tempo(Mês) e Produto.
- Membros da dimensão tempo (Mês) = 4.
- Membros da dimensão produto = 3
- Quantidade de dados =  $4 \times 3 = 12$ .

## Análise por mês e produto (Nova Visualização)

10

Road-650	January	
Road-650	February	
Road-650	March	6
Road-650	April	17
Mountain-100	January	6
Mountain-100	February	16
Mountain-100	March	6
Mountain-100	April	8
Cable Lock	January	8
Cable Lock	February	25
Cable Lock	March	21
Cable Lock	April	

**Podemos inserir novas dimensões?**

## Análise por mês, produto e estado

11

		January	February	March	April
WA	Road-650			3	10
	Mountain-100	3	16	6	
	Cable Lock	4	16	6	
OR	Road-650			3	7
	Mountain-100	3			8
	Cable Lock	4	9	15	

- Quantidade de dados =  $2 \times 3 \times 4 = 24$ .
- Itens da dimensões são chamados de Membros.
- Dimensões? Mês, Estado e Produto.
- Membros? Itens das Dimensões. Exemplo: Estado: WA (*Washington*) e OR (*Oregon*)
- Uma Medida? Quantidade de Unidades Vendidas

## Análise por mês, produto e estado

12

***Inserindo mais uma medida!***  
***Medidas são independentes!***

		January		February		March		April	
		U	\$	U	\$	U	\$	U	\$
WA	Road-650					3	7.44	10	24.80
	Mountain-100	3	7.95	16	42.40	6	15.90		
	Cable Lock	4	7.32	16	29.28	6	10.98		
OR	Road-650					3	7.44	7	17.36
	Mountain-100	3	7.95					8	21.20
	Cable Lock	4	7.32	9	16.47	15	27.45		

U = Units; \$ = Dollars

## Tabela Dimensão, Fato e Medidas

		Membros da dimensão Mês							
		January		February		March		April	
		U	\$	U	\$	U	\$	U	\$
WA	Road-650					3	7.44	10	24.80
	Mountain-100	3	7.95	16	42.40	6	15.90		
	Cable Lock	4	7.32	16	29.28	6	10.98		
OR	Road-650					3	7.44	7	17.36
	Mountain-100	3	7.95					8	21.20
	Cable Lock	4	7.32	9	16.47	15	27.45		

U = Units; \$ = Dollars

**Medidas**

Como ficaria o Modelo Estrela?

Quando?

O quê?

Onde?

Quem?

## Fato, Dimensão e Medidas

<b>FATO</b>	<ul style="list-style-type: none"><li>▪ Reflete a evolução dos negócios.</li><li>▪ Representado por valores numéricos (medida).</li><li>▪ Implementado em tabelas denominadas tabelas de fatos (fact tables).</li></ul>
<b>MEDIDA</b>	<ul style="list-style-type: none"><li>▪ Medida é um valor numérico totalizado utilizado para analisar/observar um determinado negócio. Exemplo: Quant. De Vendas.</li><li>▪ Medidas são independentes.</li><li>▪ É um atributo de um fato.</li><li>▪ Determinada pela composição das dimensões que participam de um fato.</li></ul>
<b>DIMENSÃO</b>	<ul style="list-style-type: none"><li>▪ Elementos que participam de algum fato, o “por” dos dados.</li><li>▪ Filtros que afetam a maneira de visualizar as medidas.</li></ul>
<b>MEMBROS</b>	<ul style="list-style-type: none"><li>▪ Itens de uma dimensão.</li></ul>

# Tabela Fato, Dimensão e Medidas

15

Um DW pode conter muitas tabelas fatos e muitos conjuntos de dimensões e medidas?

## Armazenamento dos dados

### TABELA FATO

16

O DW dimensional possui uma tabela que irá armazenar as medidas chamada tabela **FATO**.

STATE_ID	PROD_ID	Month	Sales_Units	Sales_Dollars
1	347	1	3	7.95
1	447	1	4	7.32
2	347	1	3	7.95
2	447	1	4	7.32
1	347	2	16	42.40

- Linhas no menor nível de detalhe.
- Chave primária e Chaves Estrangeira.



## Armazenamento dos dados

### TABELA DIMENSÃO PRODUTO

17

- As tabelas de dimensões possuem os membros.
- A dimensão contém atributos.
- PROD\_ID = MEMBRO CHAVE

PROD_ID	Product_Name
347	Mountain-100
339	Road-650
447	Cable Lock

## Armazenamento dos dados

### TABELA DIMENSÃO PRODUTO

18

- As tabelas de dimensões possuem os membros.
- A dimensão contém atributos.
- PROD\_ID = MEMBRO CHAVE
- COLUNAS (product\_name, category, color, size, price)=ATRIBUTOS.

PROD_ID	Product_Name	Category	Color	Size	Price
347	Mountain-100	Bikes	Black	44	782.99
339	Road-650	Bikes	Silver	48	3,399.99
447	Cable Lock	Accessories	NA	NA	25.00

## Agrupar

19

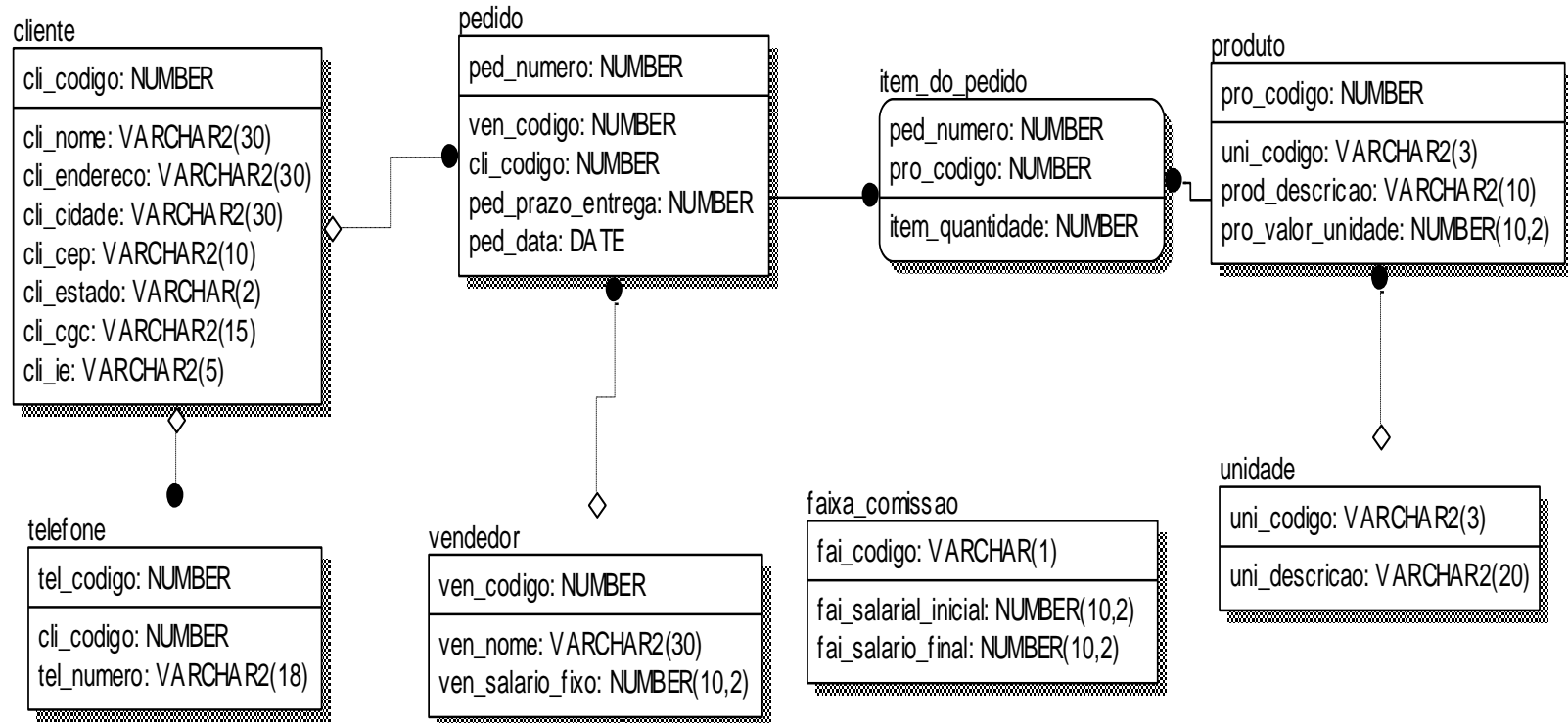
PROD_ID	Product_Name	Category	Color	Size	Price
347	Mountain-100	Bikes	Black	44	782.99
339	Road-650	Bikes	Silver	48	3,399.99
447	Cable Lock	Accessories	NA	NA	25.00

- Alguns atributos podem ser agrupados;
- Cria-se uma hierarquia quando agrupamos membros de uma dimensão.

# Estudo de Caso: Venda

20

O Atacadista São José utiliza um sistema para controlar as transações diárias das vendas. Segue o modelo entidade relacionando deste sistema:



# Necessidade do Cliente

21

- Os gerentes do atacadista São José desejam realizar consultas gerenciais para acompanhar a evolução das vendas e tomar decisões estratégicas.
- Crie um **Modelo estrela** que possibilite consultas gerenciais como as listadas a seguir:
  - Determinar o valor e a quantidade de produtos vendidos em uma determinada região (estado, cidade) em um determinado período de tempo (ano/mês/trimestre).
  - Determinar o valor e quantidade de produtos vendidos em um determinado (ano/mês/trimestre) por um determinado vendedor.
  - Determinar o valor e quantidade de produtos comprados em um determinado (ano/mês/trimestre) por um determinado cliente.

# Estudo de Caso: Venda

## Modelo estrela (STAR SCHEMA) -ERWIN

22

**Quem?**

**Dimensão Cliente**

Cliente

IDCliente: INTEGER

Nome\_Cliente: INTEGER

**Onde?**

**Dimensão Região**

Regiao

IDRegiao: INTEGER

Nome\_Cidade: VARCHAR(60)

Nome\_Estado: VARCHAR(60)

Nome\_Pais: VARCHAR(60)

**O quê?**

**Dimensão Produto**

Produto

IDProduto: INTEGER

Nome\_Produto: INTEGER

Nome\_Categoria: VARCHAR(60)

**TABELA FATO**

Vendas

IDTempo: INTEGER

IDProduto: INTEGER

IDCliente: INTEGER

IDRegiao: INTEGER

IDVendedor: INTEGER

Valor\_Venda: NUMERIC(11,2)

Qtde\_Venda: INTEGER

**Quem?**

**Dimensão Vendedor**

Vendedor

IDVendedor: INTEGER

Nome\_Vendedor: VARCHAR(60)

**Quando?**

**Dimensão Tempo**

Tempo

IDTempo: INTEGER

Ano: INTEGER

Semestre: INTEGER

Trimestre: INTEGER

Mes: INTEGER

# Estudo de Caso: Venda

## VENDAS TOTAIS CLASSIFICADAS POR REGIÃO DE VENDA

23

Região	Venda
Sul	R\$ 1.500,00
Sudeste	R\$ 5.000,00
Nordeste	R\$ 2.350,00
Norte	R\$ 1.890,00
Centro-Oeste	R\$ 1.732,00
<b>Total de Vendas</b>	<b>R\$ 12.472,00</b>

# Estudo de Caso: Venda

## VENDAS TOTAIS CLASSIFICADAS POR REGIÃO DE VENDA, DIVIDIDOS POR TRIMESTRE DE NEGÓCIO

24

Região	Trimestre	Venda
Sul		
	1	R\$250,00
	2	R\$700,00
	3	R\$250,00
	4	R\$300,00
Sudeste		
	1	R\$1.000,00
	2	R\$2.500,00
	3	R\$500,00
	4	R\$1.000,00



# Estudo de Caso: Venda

VENDAS TOTAIS CLASSIFICADAS POR REGIÃO DE VENDA, POR TRIMESTRE DE NEGÓCIO e  
POR CATEGORIA

25

Região	Trimestre	CATEGORIA	Venda
Sul	1		
		Enlatados	R\$50,00
		Cereais	R\$200,00
	2		
		Enlatados	R\$300,00
		Cereais	R\$500,00
	3		
		Enlatados	R\$25,00
		Cereais	R\$125,00
	4		
		Enlatados	R\$200,00
		Cereais	R\$100,00

## Os passos para modelagem dimensional ou multidimensional

26



Passo	Perguntas a serem feitas para o usuário	Elementos a serem definidos no modelo
<b>1</b>	O que estamos avaliando?	Fatos ou métricas (sempre um valor numérico).
<b>2</b>	Como serão avaliados ou analisados?	Dimensões de negócios relacionadas às métricas.
<b>3</b>	Qual o nível mais baixo de detalhe das informações?	Granularidade das informações em cada dimensão.
<b>4</b>	Como se espera agrupar ou sumarizar as informações?	Hierarquia de agrupamento das informações em cada dimensão.

## 1. Como identificar os fatos de um modelo dimensional?

- Identificar quais informações serão analisadas e acompanhadas pelo cliente com o passar do tempo.

Exemplo: - *Valor da Venda;*  
- *Quant. De Itens.*

A seguinte pergunta pode ser utilizada:

- “O que estamos avaliando?”.

## 2. Como identificar as dimensões?

28

Identificar como as **métricas** (medidas) serão analisadas com o usuário. Podemos realizar a seguinte pergunta:

**“Como os fatos serão analisados?”**, ou seja, sob quais dimensões de negócio avaliaremos os fatos?

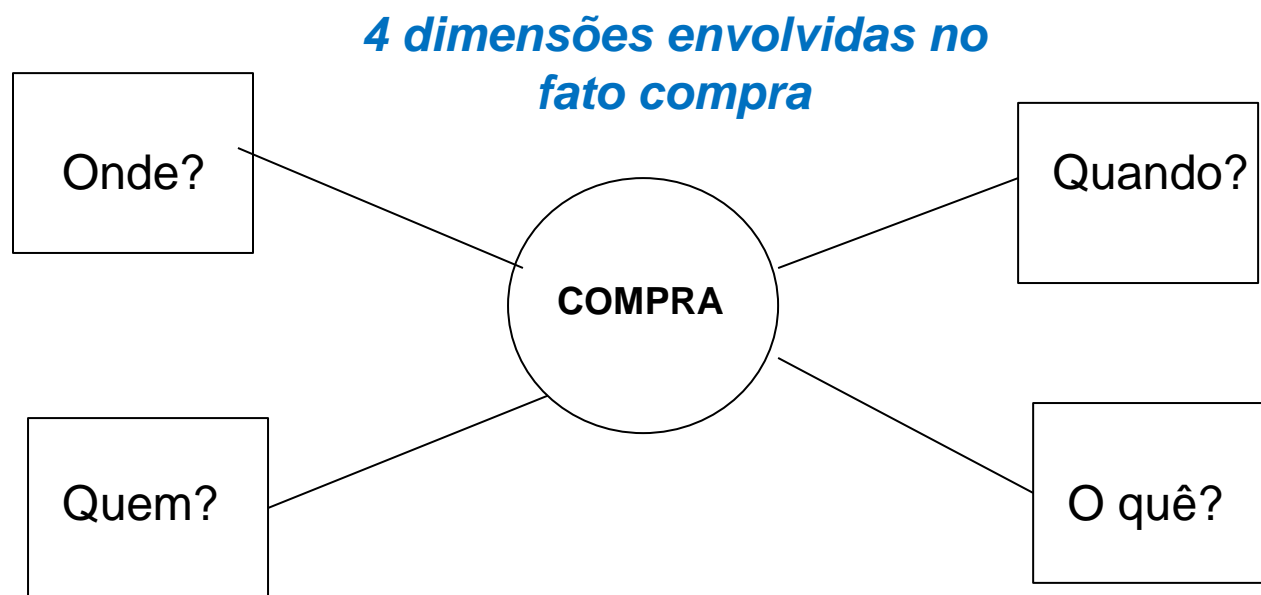
(O que? / Quando?/Onde?/Quem?)

### DICA – ARTIGO Modelagem de Data Warehouses e Data Marts – Parte 1:

*Temos que verificar se cada métrica(medida) se relaciona com todas as dimensões definidas, já que cada conjunto de métricas (medidas) deve ser analisado através do mesmo conjunto de dimensões. Para isto, podemos perguntar se cada métrica pode ser analisada ao longo de cada dimensão, por exemplo: “Faz sentido analisar o valor das vendas por produto? E por loja? E ao longo do tempo?”.*

## 1º. Exemplo – Identificação de dimensões

29



- Elementos participantes:
- Quando foi realizada a compra;
  - Onde foi realizada a compra;
  - Quem realizou a compra;
  - O que foi comprado.

## 1º. Exemplo – Identificação de dimensões

30

- Por meio de combinações entre os elementos participantes de um fato podemos analisar as informações de compras.

Quantos calçados o comprador Paulo comprou em outubro de 2000 nas Lojas Renner?

<b>Onde:</b>	Renner
<b>Quando:</b>	Outubro de 2000
<b>Quem:</b>	Paulo
<b>O quê:</b>	Calçados

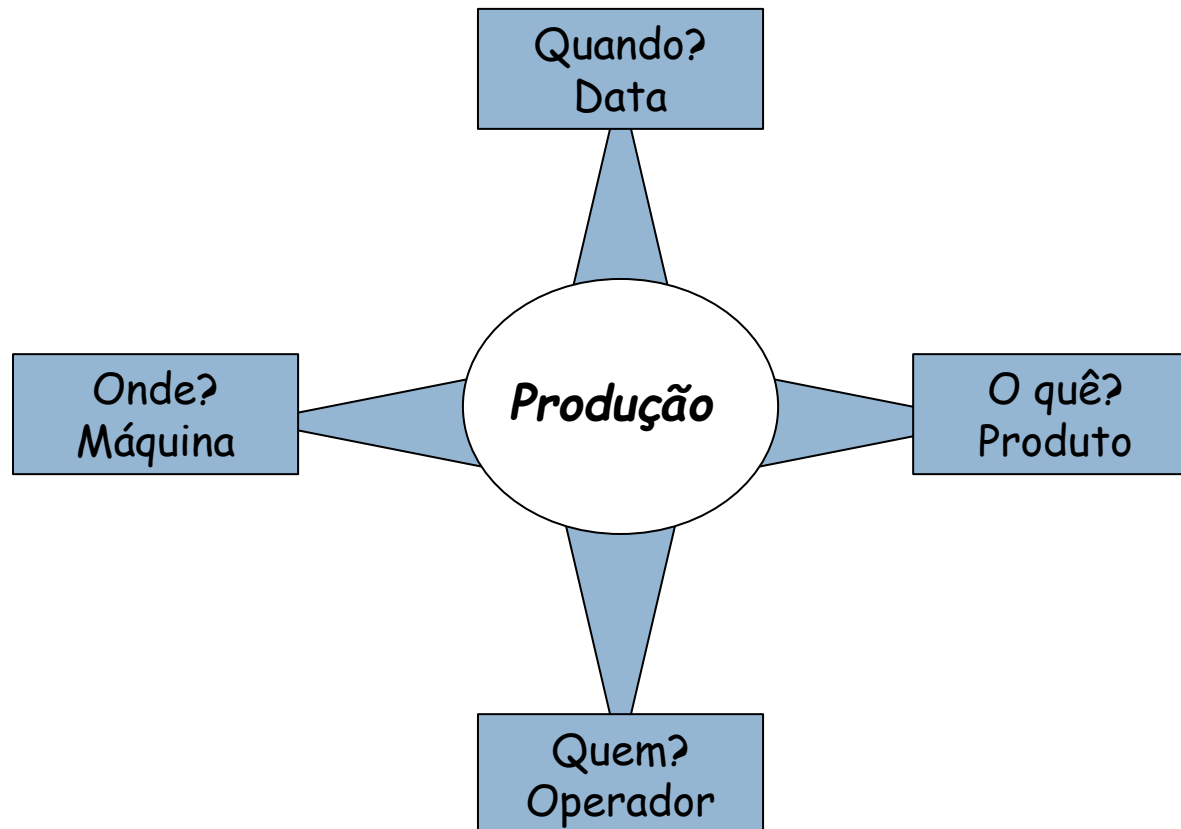
Quantos calçados foram comprados nas lojas Renner em outubro de 2000?

<b>Onde:</b>	Renner
<b>Quando:</b>	Outubro de 2000
<b>O quê:</b>	Calçados

## 2º. Exemplo – Identificação de dimensões

31

- Quais são os principais elementos (pontos cardeais) que participam de um fato produção?



### 3. Como identificar a granularidade das dimensões?

32

Qual o nível de detalhe, ou granularidade, mais baixo que o fato será avaliado?

#### Perguntas que podem ser utilizadas:

“Qual o nível de detalhe desejado?”

“Faz sentido avaliar a métrica(medida) quantidade vendida por dia?”

#### DICA – ARTIGO **Modelagem de Data Warehouses e Data Marts – Parte 1:**

*Assim, para cada uma das métricas(medida) definidas, vamos identificar qual o nível mais baixo de detalhe que será armazenado no data warehouse. Se, para a dimensão Tempo o nível mais baixo de detalhe for dia, então todas as métricas deverão ser obtidas com valores por dia.*



•:

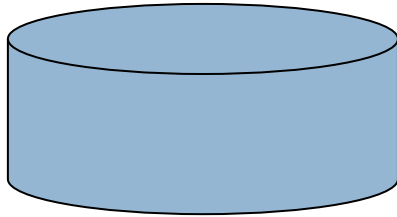
- Nível de detalhamento ou de resumo dos dados existentes no DW.
- Quanto maior o nível de detalhes, menor a granularidade do dado.

*A escolha do nível afeta de maneira direta o volume de dados armazenados no DW?*

## Vendas de 5 anos -> Qual o nível da granularidade necessária?

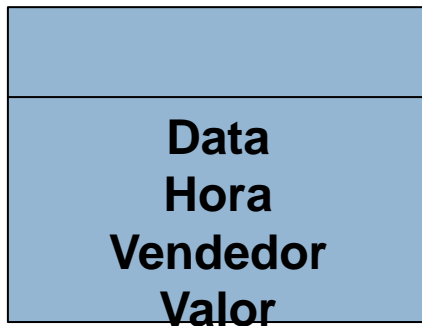
### Granularidade de Dados no Data Warehouse (1)

**Alto** nível de detalhes  
**Baixo** nível de granularidade



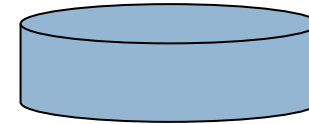
**Exemplo:**

- Detalhe de cada transação de venda de um vendedor realizada durante um mês



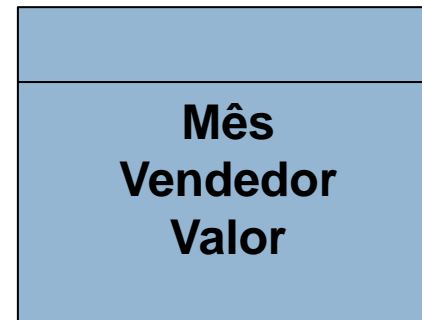
50 registros por mês

**Baixo** nível de detalhes  
**Alto** nível de granularidade



**Exemplo:**

- Sumarização das transações de venda de um vendedor realizada durante um mês.



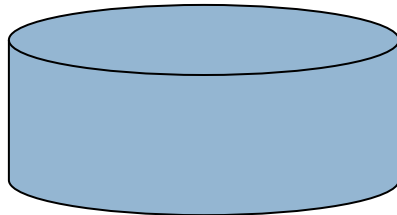
1 registro por mês

# Granularidade

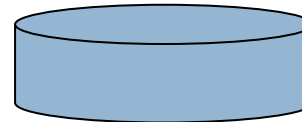
35

- Considerado o mais **importante aspecto do projeto de Data Warehouse**;
- Refere-se ao nível de sumarização dos elementos e de detalhe disponível nos dados;
- Quanto mais detalhe nos dados (**Mais dados**) = **Mais baixo** o nível de granularidade;
- Quanto Menos detalhe nos dados (**Menos dados**) = **Mais alto** o nível de granularidade.

**Baixa  
Granularidade**



**Alta Granularidade**



## 4. Como identificar a hierarquia de agrupamento?

36

- Identificar como o usuário deseja realizar o agrupamento da informações em relação a cada uma das dimensões.

Exemplo: - “***Qual o total de vendas no último semestre?***”.

Esta pergunta já indica que deveremos nos preocupar com o agrupamento, ou sumarização das informações armazenadas no data warehouse.

### Exemplos de agrupamento:

- ❑ **Dimensão Tempo:** mês/semestre/ano;
- ❑ **Dimensão Produto:** descrição de produto/categoria;
- ❑ **Dimensão Geografia:** loja/cidade/estado.

## Chave primaria das tabelas de dimensões

37

- A chave primária de uma tabela dimensão deve ser sempre um atributo único e definido pelo sistema com um valor genérico, inteiros atribuídos sequencialmente. (Chaves Substitutas).
- **Justificativa:**

Por questões de desempenho, não se utilizam chaves compostas. Também não são utilizados as chaves dos sistemas de origem, como código do cliente ou código do produto, pois tendo um valor genérico temos como guardar o histórico no DW dos valores atualizados no sistema transacional(origem).

# Exemplo Chave Substituta e Chave Natural

38

## Sistema Transacional (Origem)

### Vendedor

<u>Código_Vendedor</u>	<u>Nome_Vendedor</u>	<u>Região</u>
101	Paulo Silva	<u>1</u>
204	Maria Cristina Costa	<u>2</u>
305	Pedro Lima	<u>1</u>

### Região

<u>Código_Região</u>	<u>Nome_Região</u>
<u>1</u>	Sul
<u>2</u>	Sudeste
<u>3</u>	Nordeste
<u>4</u>	Norte
<u>5</u>	Centro-Oeste

## DW

### Vendedor

<u>Código_Vendedor</u> (Chave Substituta)	<u>Nome_Vendedor</u>	<u>Região</u>	<u>Código_Origem</u> (Chave Natural)
<u>1</u>	Paulo Silva	<u>1</u>	101
<u>2</u>	Maria Cristina Costa	<u>2</u>	204
<u>3</u>	Pedro Lima	<u>2</u>	305
<u>4</u>	Pedro Lima	<u>1</u>	305

## **Vantagem da redundância nas tabelas que representam as dimensões em um modelo Estrela (Star).**

39

### **1 – Melhor de desempenho nas consultas.**

Justificativa: Como as dimensões não são normalizadas existe um número menor de tabelas para retornar as informações necessárias.

### **2 – Guardar histórico dos dados.**

Justificativa: - A redundância proporciona armazenar a história de um determinado valor.

# Referência Bibliográfica

40

N.	Título	Autor
01	TECNOLOGIA E PROJETO DE DATA WAREHOUSE	FELIPE NERY RODRIGUES MACHADO
02	ARTIGO Modelagem de Data Warehouses e Data Marts – Parte 1	Isabel Cristina Italiano e Luiz Antonio Esteves