

PÓS-GRADUAÇÃO

Modelagem e arquitetura do DW (data warehouse)



Modelagem de Dados para um Data Warehouse

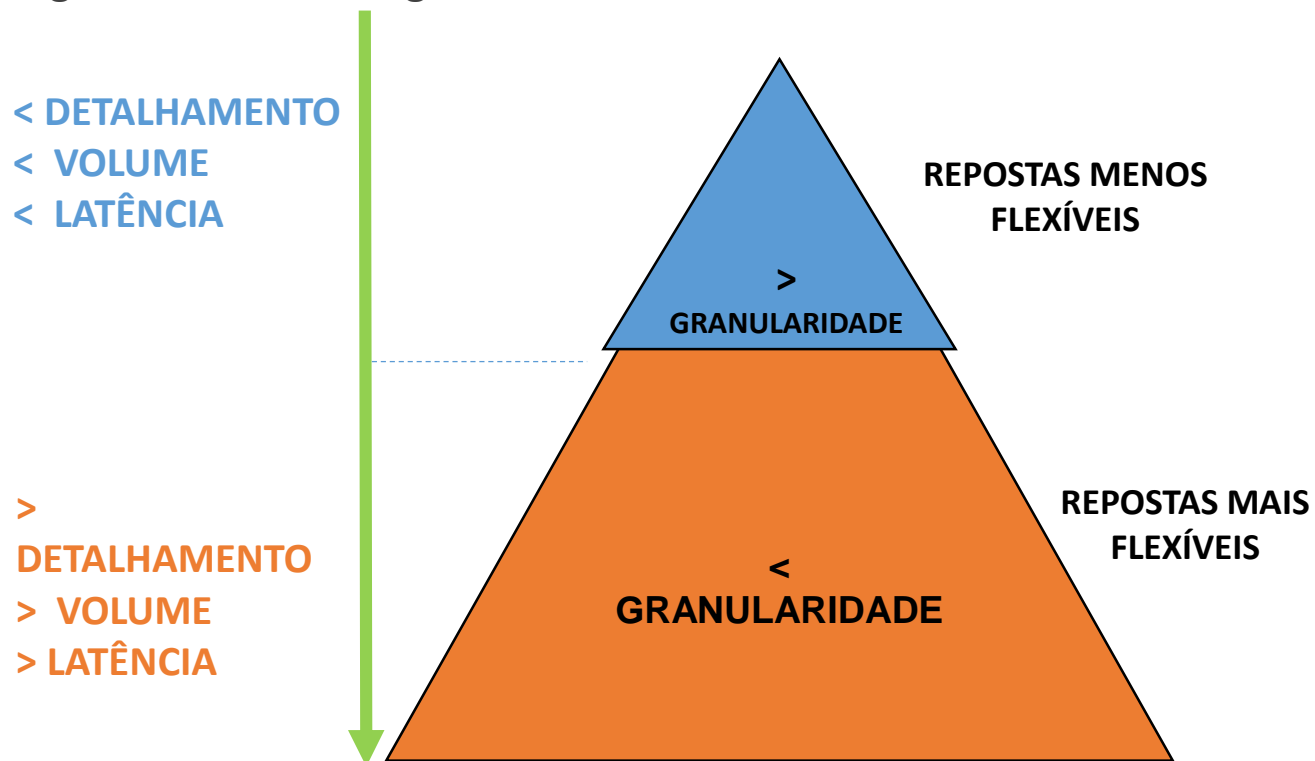
Bloco 1

Marise Miranda

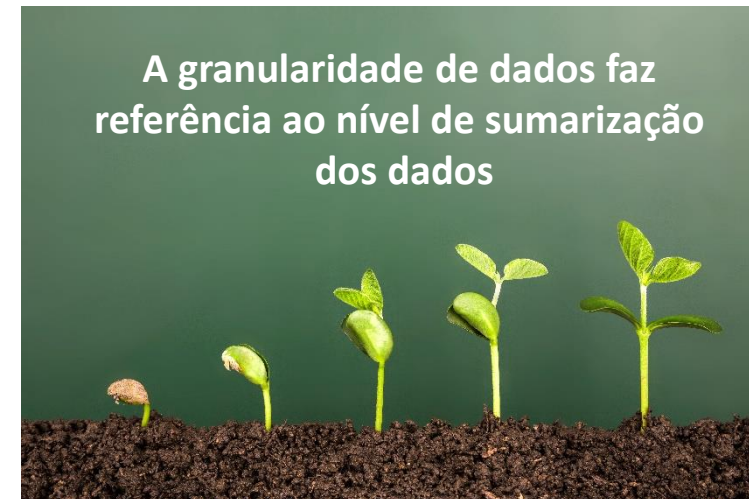


► Granularidade de dados

Figura 1 – Níveis de granularidade dos dados



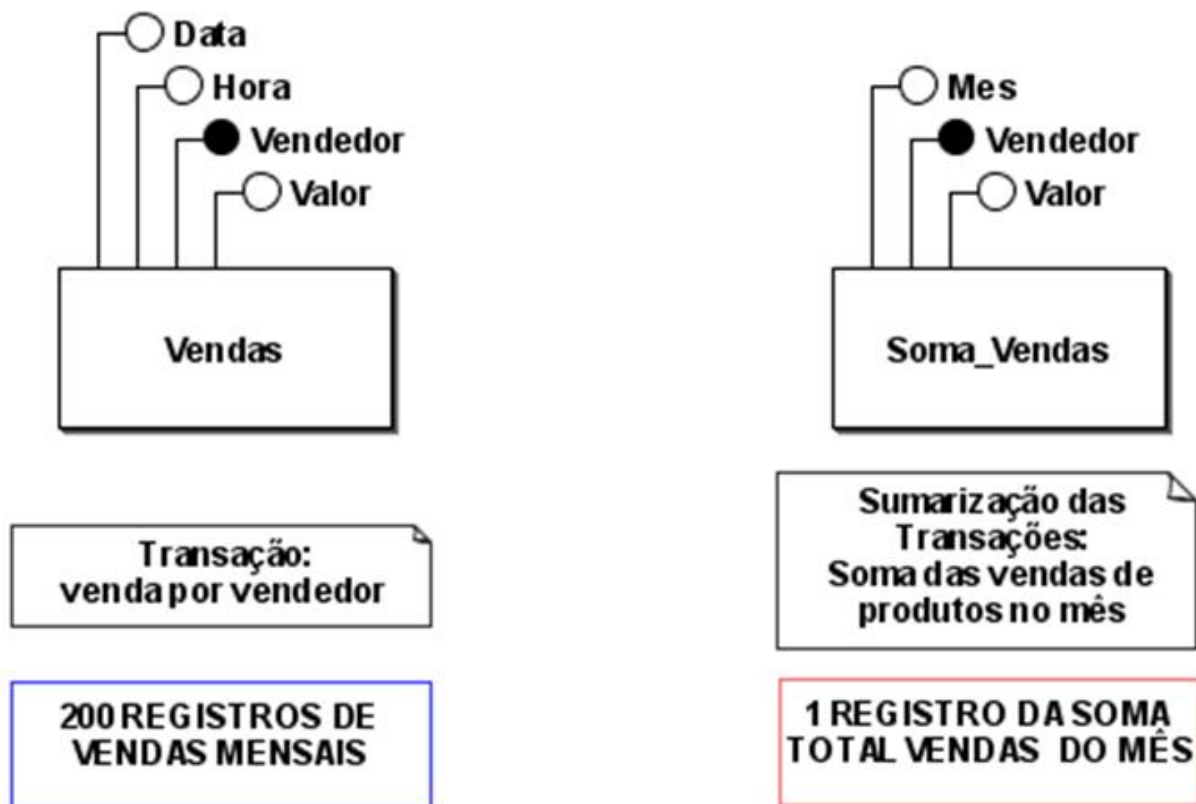
Fonte: adaptada de Kimball (2002, p. 133-134).



Fonte: adaptado de feellife/iStock.com

► Exemplo de granularidade

Figura 2 – Exemplo de granularidade e nível de detalhamento



Fonte: izusek /iStock.com

Fonte: elaborada pela autora.

► Exemplo de baixa granularidade e alto nível de detalhamento

Tabelas do banco de dados transacional

Id_Vendas	Data	Hora	Vendedor	Valor
1	14/06/2012	16:03:12	Paulo M	R\$ 2.500,00
2	15/06/2012	16:33:01	Julio W	R\$ 1.850,00
3	16/06/2012	18:50:10	Paulo M	R\$ 3.300,00
4	17/06/2012	16:33:01	Paulo M	R\$ 4.500,00
5	18/06/2012	17:25:06	Maria T	R\$ 200,00
6	13/07/2012	12:03:12	Julio W	R\$ 2.135,00
7	14/07/2012	14:40:59	Paulo M	R\$ 100,00
8	15/07/2012	13:22:08	Paulo M	R\$ 130,00
9	16/07/2012	14:15:16	Maria T	R\$ 4.750,00
10	17/07/2012	15:17:30	Maria T	R\$ 1.889,00

Baixa granularidade

Alto detalhamento

Fonte: elaborada pela autora.



Relatório Vendedor x Valor	
Vendedor	Valor
Paulo M	R\$2.500,00
Julio W	R\$1.850,00
Paulo M	R\$ 3.300,00
Paulo M	R\$ 4.500,00
Maria T	R\$ 200,00
Julio W	R\$ 2.135,00
Paulo M	R\$ 100,00
Paulo M	R\$ 130,00
Maria T	R\$ 4.750,00
Maria T	R\$ 1.889,00

► Exemplo de alta granularidade e baixo nível de detalhamento

Dados sumarizados para análises

Mês	Soma_Vvalor
jun/12	R\$ 12.350,00
jul/12	R\$ 9.004,00

Fonte: elaborada pela autora.



Fonte: adaptada de Deagreez/ iStock.com



Alta granularidade



Baixo detalhamento

Estudo de caso: simulação a partir de quatro lojas - parte 1

Loja 1

Id_Vendas	Data	Hora	Vendedor	Valor
1	14/01/2012	16:03:12	Paulo M	R\$2.200,00
2	15/01/2012	16:33:01	Julio W	R\$1.550,00
3	16/01/2012	18:50:10	Paulo M	R\$2.500,00
4	17/02/2012	16:33:01	Paulo M	R\$1.850,00
5	18/02/2012	17:25:06	Maria T	R\$3.300,00
6	13/02/2012	12:03:12	Julio W	R\$4.500,00
7	14/03/2012	14:40:59	Paulo M	R\$200,00
8	15/03/2012	13:22:08	Paulo M	R\$430,00
9	16/03/2012	14:15:16	Maria T	R\$2.110,00
10	17/03/2012	15:17:30	Maria T	R\$1.123,00

Loja 2

Id_Vendas	Data	Hora	Vendedor	Valor
1	14/08/2012	16:03:12	Paulo M	R\$2.600,00
2	15/08/2012	16:33:01	Julio W	R\$1.550,00
3	16/08/2012	18:50:10	Paulo M	R\$1.300,00
4	17/09/2012	16:33:01	Paulo M	R\$4.200,00
5	18/09/2012	17:25:06	Maria T	R\$800,00
6	13/09/2012	12:03:12	Julio W	R\$1.136,00
7	14/10/2012	14:40:59	Paulo M	R\$200,00
8	15/10/2012	13:22:08	Paulo M	R\$430,00
9	16/10/2012	14:15:16	Maria T	R\$4.110,00
10	17/10/2012	15:17:30	Maria T	R\$123,00

Loja 3

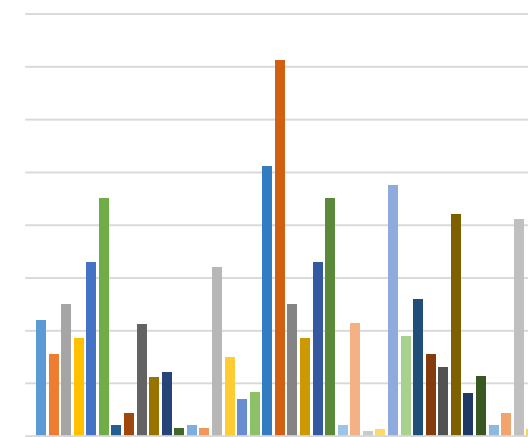
Id_Vendas	Data	Hora	Vendedor	Valor
1	14/04/2012	16:03:12	Paulo M	R\$1.200,00
2	15/04/2012	16:33:01	Julio W	R\$150,00
3	16/04/2012	18:50:10	Paulo M	R\$200,00
4	17/04/2012	16:33:01	Paulo M	R\$150,00
5	18/04/2012	17:25:06	Maria T	R\$3.200,00
6	13/04/2012	12:03:12	Julio W	R\$1.500,00
7	14/05/2012	14:40:59	Paulo M	R\$700,00
8	15/05/2012	13:22:08	Paulo M	R\$830,00
9	16/05/2012	14:15:16	Maria T	R\$5.110,00
10	17/05/2012	15:17:30	Maria T	R\$7.123,00

Loja 4

Id_Vendas	Data	Hora	Vendedor	Valor
1	14/06/2012	16:03:12	Paulo M	R\$2.500,00
2	15/06/2012	16:33:01	Julio W	R\$1.850,00
3	16/06/2012	18:50:10	Paulo M	R\$3.300,00
4	17/06/2012	16:33:01	Paulo M	R\$4.500,00
5	18/06/2012	17:25:06	Maria T	R\$200,00
6	13/07/2012	12:03:12	Julio W	R\$2.135,00
7	14/07/2012	14:40:59	Paulo M	R\$100,00
8	15/07/2012	13:22:08	Paulo M	R\$130,00
9	16/07/2012	14:15:16	Maria T	R\$4.750,00
10	17/07/2012	15:17:30	Maria T	R\$1.889,00

Valores vendas/dia

R\$8.000,00
R\$7.000,00
R\$6.000,00
R\$5.000,00
R\$4.000,00
R\$3.000,00
R\$2.000,00
R\$1.000,00
R\$0,00



Valor

Estudo de caso: simulação a partir de quatro lojas – parte 2

BG

Id_Vendas	Data	Hora	Vendedor	Valor
1	14/01/2012	16:03:12	Paulo M	R\$2.200,00
2	15/01/2012	16:33:01	Julio W	R\$1.550,00
3	16/01/2012	18:50:10	Paulo M	R\$2.500,00
4	17/02/2012	16:33:01	Paulo M	R\$1.850,00
5	18/02/2012	17:25:06	Maria T	R\$3.300,00
6	13/02/2012	12:03:12	Julio W	R\$4.500,00
7	14/03/2012	14:40:59	Paulo M	R\$200,00
8	15/03/2012	13:22:08	Paulo M	R\$430,00
9	16/03/2012	14:15:16	Maria T	R\$2.110,00
10	17/03/2012	15:17:30	Maria T	R\$1.123,00

BG

Id_Vendas	Data	Hora	Vendedor	Valor
1	14/08/2012	16:03:12	Paulo M	R\$2.600,00
2	15/08/2012	16:33:01	Julio W	R\$1.550,00
3	16/08/2012	18:50:10	Paulo M	R\$1.300,00
4	17/09/2012	16:33:01	Paulo M	R\$4.200,00
5	18/09/2012	17:25:06	Maria T	R\$800,00
6	13/09/2012	12:03:12	Julio W	R\$1.136,00
7	14/10/2012	14:40:59	Paulo M	R\$200,00
8	15/10/2012	13:22:08	Paulo M	R\$430,00
9	16/10/2012	14:15:16	Maria T	R\$4.110,00
10	17/10/2012	15:17:30	Maria T	R\$123,00

AG

Mês	Soma Valor
jan/12	R\$ 6.250,00
fev/12	R\$ 9.650,00
mar/12	R\$ 3.863,00
abr/12	R\$ 6.400,00
mai/12	R\$ 13.763,00
jun/12	R\$ 12.350,00
jul/12	R\$ 9.004,00
ago/12	R\$ 5.450,00
set/12	R\$ 6.136,00
out/12	R\$ 4.863,00

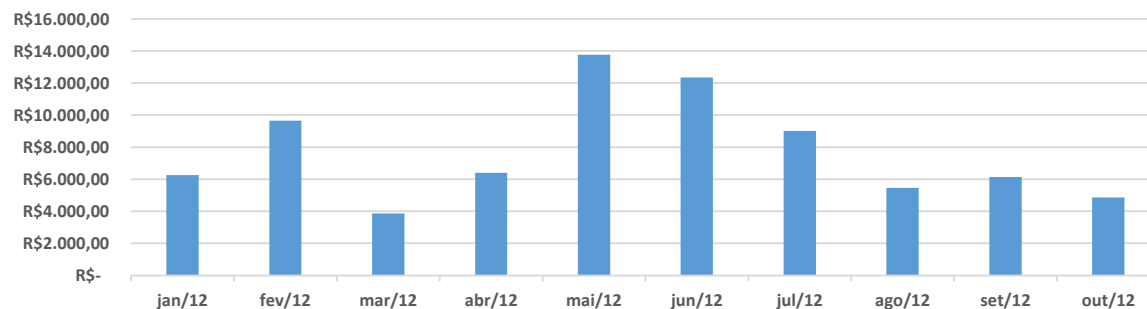
BG

Id_Vendas	Data	Hora	Vendedor	Valor
1	14/06/2012	16:03:12	Paulo M	R\$2.500,00
2	15/06/2012	16:33:01	Julio W	R\$1.850,00
3	16/06/2012	18:50:10	Paulo M	R\$3.300,00
4	17/06/2012	16:33:01	Paulo M	R\$4.500,00
5	18/06/2012	17:25:06	Maria T	R\$200,00
6	13/07/2012	12:03:12	Julio W	R\$2.135,00
7	14/07/2012	14:40:59	Paulo M	R\$100,00
8	15/07/2012	13:22:08	Paulo M	R\$130,00
9	16/07/2012	14:15:16	Maria T	R\$4.750,00
10	17/07/2012	15:17:30	Maria T	R\$1.889,00

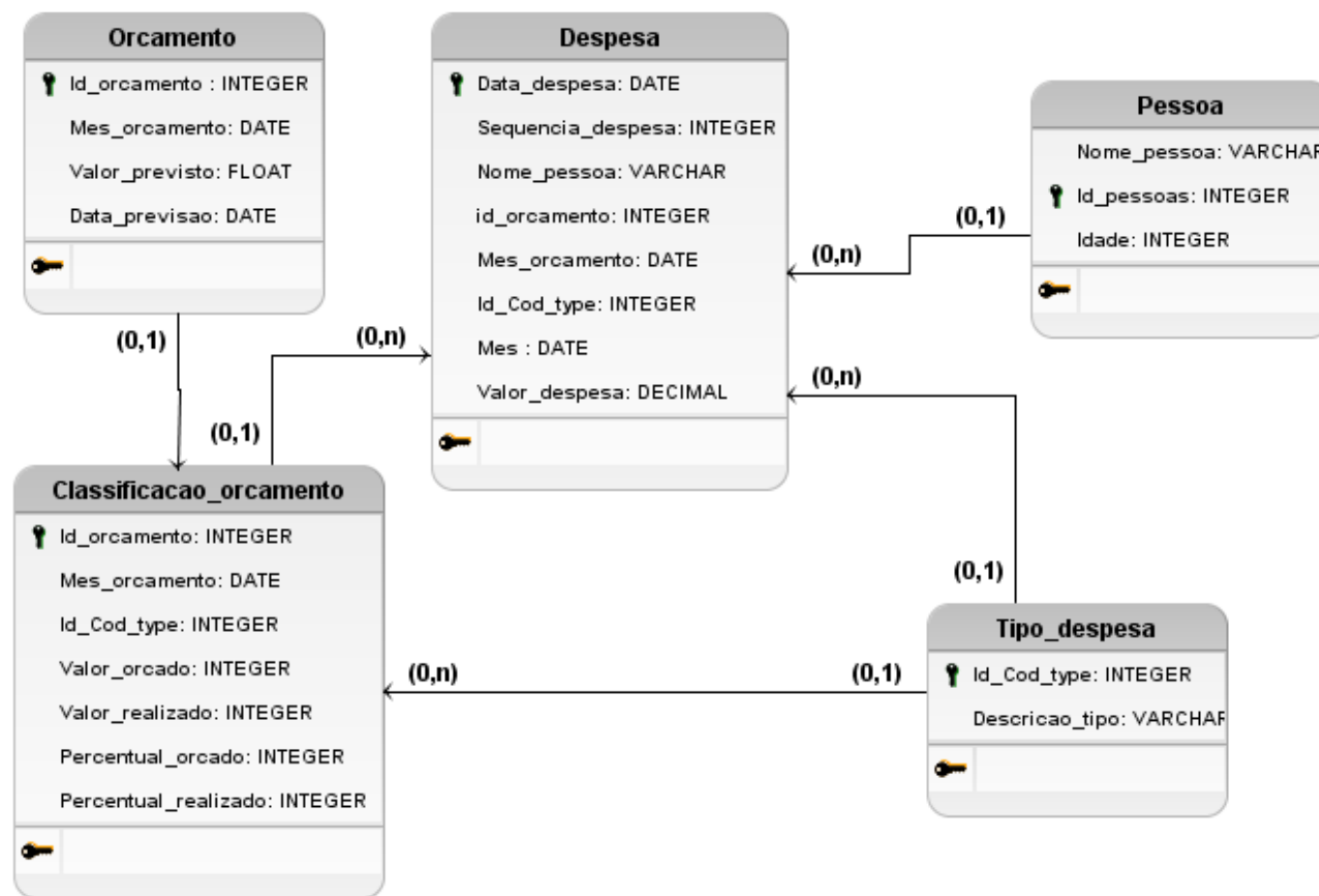
BG

Id_Vendas	Data	Hora	Vendedor	Valor
1	14/04/2012	16:03:12	Paulo M	R\$1.200,00
2	15/04/2012	16:33:01	Julio W	R\$150,00
3	16/04/2012	18:50:10	Paulo M	R\$200,00
4	17/04/2012	16:33:01	Paulo M	R\$150,00
5	18/04/2012	17:25:06	Maria T	R\$3.200,00
6	13/04/2012	12:03:12	Julio W	R\$1.500,00
7	14/05/2012	14:40:59	Paulo M	R\$700,00
8	15/05/2012	13:22:08	Paulo M	R\$830,00
9	16/05/2012	14:15:16	Maria T	R\$5.110,00
10	17/05/2012	15:17:30	Maria T	R\$7.123,00

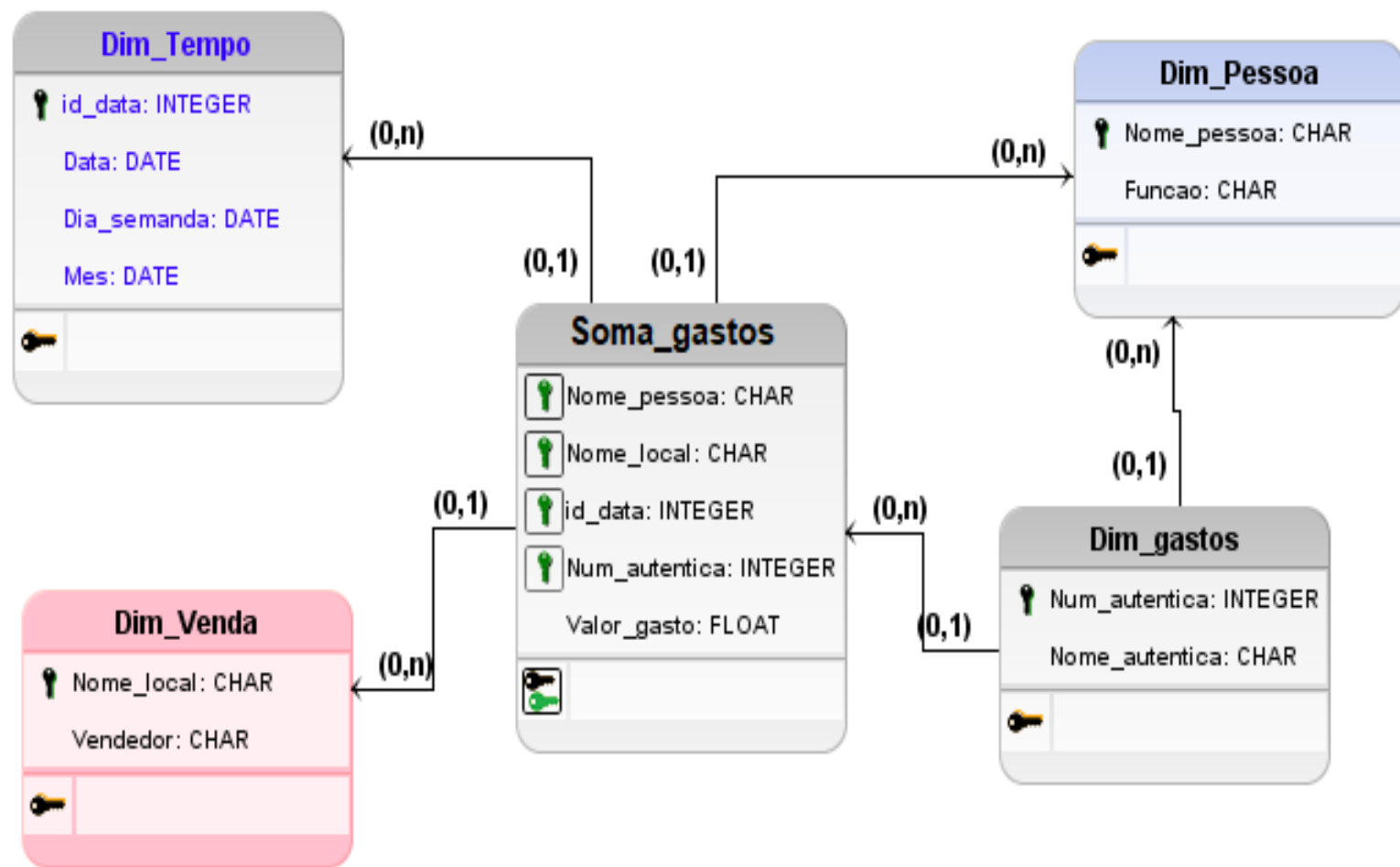
Sumarização: Soma_Vvalor



► Modelagem operacional e multidimensional – parte 1



► Modelagem operacional e multidimensional – parte 2



MODELO MULTIDIMENSIONAL

► Exemplo de modelo multidimensional

Tabela dimensão 1

Vendedor	
Id_vendedor	nome
1	Mario
2	Luis
3	Sergio
4	Augusto
5	Juvenal

Tabela fato

Desempenho vendedores		
Id_vendedor	Nome	total_vendedor
1	Mario	R\$ 4.139,00
2	Luis	R\$ 1.440,00
3	Sergio	R\$ 2.789,00
4	Augusto	R\$ 3.230,00
5	Juvenal	R\$ -

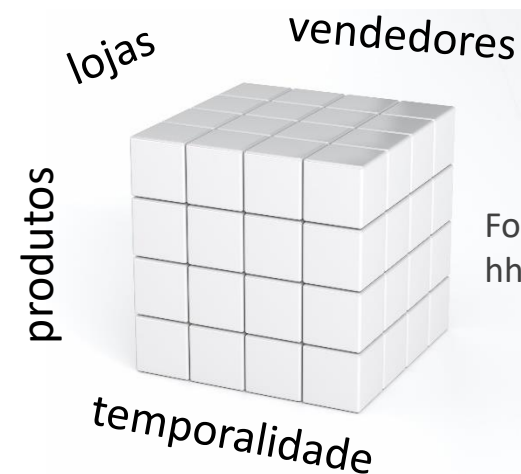


Tabela dimensão 2

Produto Venda			
Id_produto	tipo_produto	valor_produto	id_vendedor
1	geladeira soft	R\$ 2.500,00	4
2	fogão galaxi	R\$ 899,00	3
3	geladeira alfa	R\$ 1.890,00	3
4	microondas wave	R\$ 450,00	1
5	ferro pass	R\$ 120,00	2
6	aquecedor ice5	R\$ 3.689,00	1
7	microndas wavelet	R\$ 730,00	4
8	geladeira first	R\$ 1.320,00	2



► Conclusão

A modelagem de dados é o início de um projeto de DW. Deve levar em consideração os níveis de granularidade e detalhamento dos dados.

A abordagem deve começar pelo desenho do modelo “entidade relacionamento” a fim de sistematizar os conceitos de granularidade, correlacionando tabelas dimensões e tabelas fatos

A modelagem multidimensional irá se tornar mais eficiente na medida em que a base do modelo ER estiver compreendida, pondo em prática a construção de cubos multidimensionais.

Processo de desenho de modelo dimensional

Bloco 2


Marise Miranda





► Passos para desenho do processo de modelagem

Passo 1 – Desenhe os processos de negócio da organização

- Entender as operações diárias.
 - Desenhar os processos.
 - Verificar as ligações do processo.
 - Cada processo tem entradas e saídas.
 - Verificar o fluxo de funcionamento entre processos.
- 



► Passos para desenho do processo de modelagem

Passo 2 – Represente o grão

- O grão transmite o nível de detalhe associado à tabela de fatos.
- Uma grão pode ser o registro na linha da tabela de um produto adquirido por um cliente.
- As transações de uma conta bancária.



► Passos para desenho do processo de modelagem

Passo 3 – Identifique as dimensões

- As dimensões descrevem os dados resultantes dos eventos de medição do processo de negócio.
- Incluir nas tabelas de fatos um conjunto robusto de dimensões, representando todas as descrições possíveis que assumem valores únicos no contexto de cada medição.
- As dimensões normalmente podem ser facilmente identificadas, pois representam “Quem, o quê, onde, quando, por que e como”, associado ao evento.



► Passos para desenho do processo de modelagem

Passo 4 – Identifique as fatos

- Os fatos devem dar respostas sobre o processo de mediação.
- Analisar as métricas de desempenho das tabelas fato.
- Tomar decisões com base nos conjuntos de dados de origem.



► Conclusão

As tabelas fato são usadas para expor as métricas em gráficos ou tabelas sumarizadas

As dimensões devem ser desenhadas de modo a obedecer a uma organização hierárquica. Dimensões e dimensões filhas.

Métricas são geralmente observadas nas tabelas fato. Existem casos de realizar sumarizações, transformações, agregações ou cálculos em tabelas dimensões.

PÓS-GRADUAÇÃO

Teoria em prática

Bloco 3

Marise Miranda





► Estudo de Caso – Vendas no varejo

Uma empresa produtora de grãos para exportação e abastecimento no mercado brasileiro deseja iniciar um projeto de *data warehouse*. A empresa compra os grãos in natura de três fazendas, faz a coleta nos silos, leva para a produção de lotes para comercialização interna e internacional. Faça uma simulação de uma modelagem multidimensional, considerando, cinco tipos de grãos: feijão fradinho, feijão carioca, feijão preto, feijão branco e feijão fava.

A produção é trimestral, com variações de 10% em cada trimestre, a melhor safra é nos três últimos meses do ano, quando não é afetada pelas variações climáticas e chuvas. As fazendas estão localizadas no Norte, Nordeste e Centro-Oeste, a fábrica de separação e ensacamento fica no Sul do país. A logística para transporte dos silos das fazendas é quinzenal. A produção dos feijões é afetada pela variação climática, nos seis primeiros meses do ano, clima muito quente e chuvas.

Com base nesse cenário, projete um banco de DW por meio das técnicas de modelagem multidimensional. Simule o ambiente populando as tabelas dimensões com dados fictícios de produção agrícola, temperatura, tempo e localidade.



► Utilizando os quatro passos

Verificar as regras de negócios e os processos da organização.

Represente a granularidade.

Construa as tabelas dimensões.

Construa as tabelas fato.



PÓS-GRADUAÇÃO

Dica do professor

Bloco 4

Marise Miranda





► Características de capacidade do banco de dados de um *data warehouse* na prática

A empresa Microsoft disponibilizou um conjunto de informações sobre dimensionamento da carga de trabalho de um *data warehouse*.

O objetivo é dividir as consultas ao banco de dados balanceando a carga de trabalho.

Supondo que a base de dados de um DW seja de 20 milhões de registros e com base nas informações técnicas do SQL Data Warehouse (Microsoft), há a necessidade de dimensionar a memória acelerando as consultas.

Faça uma busca na internet com informações de Data Warehouse da Microsoft.



► Referências

MACHADO, F. N. R. **Tecnologia e Projeto de Data Warehouse**. 5. ed. São Paulo, Editora Érica, 2010.

INMON, W. H. **Building the Data Warehouse**. 4. ed. New York: Wiley Computer Publishing, 2005.

KIMBALL, R. ***The Data Warehouse Toolkit***: guia completo para modelagem dimensional. Rio de Janeiro: Campus, 2002.



