



Modelagem, Extração e Manipulação de Dados

BLOCO: B.I. E ANÁLISE DE DADOS

PROF. RODRIGO EIRAS, M.SC.

[ETAPA 2] AULA 3 – TIPOS DE DADOS E MODELAGEM DIMENSIONAL



No último trimestre...

- Introdução à disciplina de Modelagem, Extração e Manipulação de Dados
- Modelagem de dados
- Modelagem Relacional x Dimensional
- Modelagem Dimensional
- Quiz



Agenda

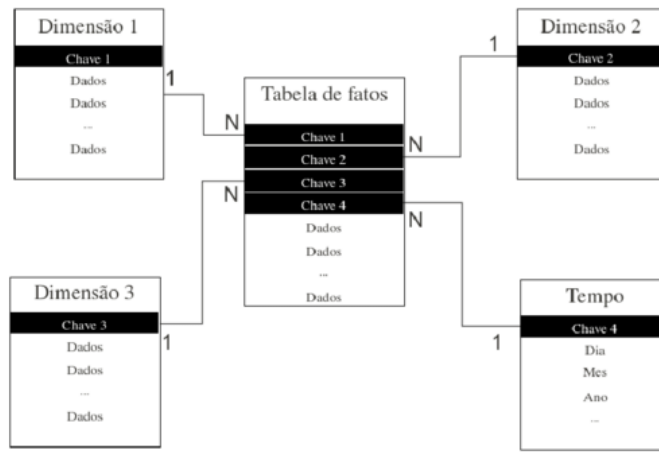
- Medidas
- Hierarquias
- Atributos
- Leitura do TP 1
- Modelagem dimensional (Exercício)



Medidas

- Tudo que a empresa for mensurar é uma medida.
- Na maioria das vezes, a medida vai ser o que o usuário quer medir.
- Exemplos de medidas são: quantidade vendida, valor faturado, quantidade devolvida, margem bruta e custo da venda.
- As medidas são armazenadas na Fato (tabela que armazena os dados quantitativos).

Medidas



- Existem 4 tipos de medidas: Aditivas, Semi-aditivas, Não-aditivas e Derivadas.
- Essa classificação permite: "você saber que tipo de cálculos e cruzamentos pode fazer com cada uma delas.
- Por mais que todas as métricas sejam numéricas, elas têm significados e interpretações diferentes, e é para identificar isso durante a modelagem que servem os tipos".

Medidas aditivas

Fato Venda		
FK	FK	Métrica ativa
DIM_DATA	DIM_PRODUTO	Quantidade da Venda
10122016	1	200
11122016	1	300
12122016	1	200
...

- As medidas aditivas são aquelas que podem ser sumarizadas independente das dimensões utilizadas.
- Estes tipos de medidas são flexíveis o suficiente para gerar informações em qualquer perspectivas.
- Por exemplo, métricas como quantidade e valores de determinados itens podem ser, em geral, sumarizados por data (dia, mês ou ano), local, clientes, entre outras dimensões, sem perder a consistência da informação.
- É a métrica mais simples que conhecemos.

Dimensão Produto			
SK	NK	Nome	Preço unitário
1	2323	Coca-Cola	3,55
2	1111	Pizza	23,60
3	4545	café	7,80
...

Dimensão Data			
SK	NK	Ano	Mês
10122016	10/12/2016	2016	Dezembro
11122016	11/12/2016	2016	Dezembro
12122016	12/12/2016	2016	Dezembro
...

Métricas aditivas

Alguns exemplos de métricas aditivas:

- Quantidade de vendas
- Valor da venda (se não for calculado)
- Quantidade de colaboradores
- Quantidade de demissões
- Quantidade de admissões

Medidas semi-aditivas

- As medidas semi-aditivas são aquelas que podem ser sumarizadas em alguns casos.
- Isso porque a depender da situação empregada à medida, ela pode perder sentido para a análise caso seja agregada.
- Neste caso a sumarização só fará sentido com algumas dimensões específicas.
- Por exemplo, a métrica saldo bancário.
- O saldo é um valor que reflete a situação atual da conta, que pode ter o saldo credor ou devedor.

FATO SALDO CONTA CORRENTE					
FK	FK		Métrica aditiva	Métrica aditiva	Métrica Semi-Aditiva
DATA	CLIENTE	AGENCIA	ENTRADA	SAIDA	SALDO
01/03/2017	PITON	ITAU-SP01	1.500,00	300,00	1.200,00
01/03/2017	PITON	ITAU-SP01		400,00	800,00
01/03/2017	PITON	ITAU-SP01		200,00	600,00
01/03/2017	ROBSON	ITAU-SP01	2.500,00	700,00	1.800,00
01/03/2017	MARCOS	ITAU-SP01	3.120,00	300,00	2.820,00
01/03/2017	DORIVAL	ITAU-SP01	7.500,00	300,00	7.200,00
			14.620,00	2.200,00	14.420,00

Medidas semi-aditivas

- Faria sentido, por exemplo, somar os saldos de todos os dias de um mês para uma determinada conta bancária? Não.
- Pois se um dia o saldo for de R\$1.000,00 e no dia seguinte ter os mesmos R\$1.000,00, a soma irá me devolver um saldo negativo de R\$2.000,00, o que não é verdade.
- Mas há casos onde a métrica semi-aditiva adquire característica de aditiva.
- Se por acaso somar os saldos de várias contas bancária em um determinado dia, poderemos ver o saldo geral, o que tem total sentido e utilidade para uma instituição bancária.
- Exemplos de medidas semi-aditivas são saldo de estoque e saldo bancário, quando representado de forma monetária, porque são aditivas em todas as dimensões, exceto na tempo.

FATO SALDO CONTA CORRENTE					
FK	FK		Métrica aditiva	Métrica aditiva	Métrica Semi-Aditiva
DATA	CLIENTE	AGENCIA	ENTRADA	SAIDA	SALDO
01/03/2017	PITON	ITAU-SP01	1.500,00	300,00	1.200,00
01/03/2017	PITON	ITAU-SP01		400,00	800,00
01/03/2017	PITON	ITAU-SP01		200,00	600,00
01/03/2017	ROBSON	ITAU-SP01	2.500,00	700,00	1.800,00
01/03/2017	MARCOS	ITAU-SP01	3.120,00	300,00	2.820,00
01/03/2017	DORIVAL	ITAU-SP01	7.500,00	300,00	7.200,00
			14.620,00	2.200,00	14.420,00

Medidas não-aditivas

- As medidas não-aditivas são aquelas que não podem ser sumarizadas ao longo das dimensões. Essas medidas não podem ter agregações pois perdem a veracidade do valor.
- Percentuais são exemplos de valores armazenados nas medidas que não permitem sumarização. Por exemplo, não faz sentido algum somar o percentual de vendas de um item “A” que teve 50% de saída com um item “B” que teve 60%.
- A soma resultaria em um valor agregado de 110%.
- A sugestão de modelagem para essas medidas é salvar as medidas que levam àquela não-aditiva e deixar para que ela seja calculada na ferramenta de BI

Medidas não-aditivas

- E é por causa dessas somas e cruzamentos que podem ou não ser realizadas numa tomada de decisão que é importante entender cada um dos tipos de medida e principalmente saber como usar.
- Pegando o exemplo do saldo, é uma métrica muito fácil de ser erroneamente tratada como uma aditiva, o que vai alterar os resultados e afetar a tomada de decisão.
- Muito importante conhecer o negócio para utilizar os dados de forma correta!

Medida derivada

- É uma medida calculada, uma medida que você calcula para ter um segundo número.
- Esse cálculo é sempre em cima de medidas que já estão na Fato.
- Exemplo de métrica derivada é o valor total de um pedido, a medida derivada é a multiplicação da quantidade da venda pelo o preço unitário da dimensão de produto.
- É importante ficar atento a essas diferenças de medidas, para que no desenvolvimento do DW possamos efetuar o tratamento adequado em cada um desses casos.
- Ticket médio também é um bom exemplo de medida derivada

R\$ 8.750,00

50 clientes ativos

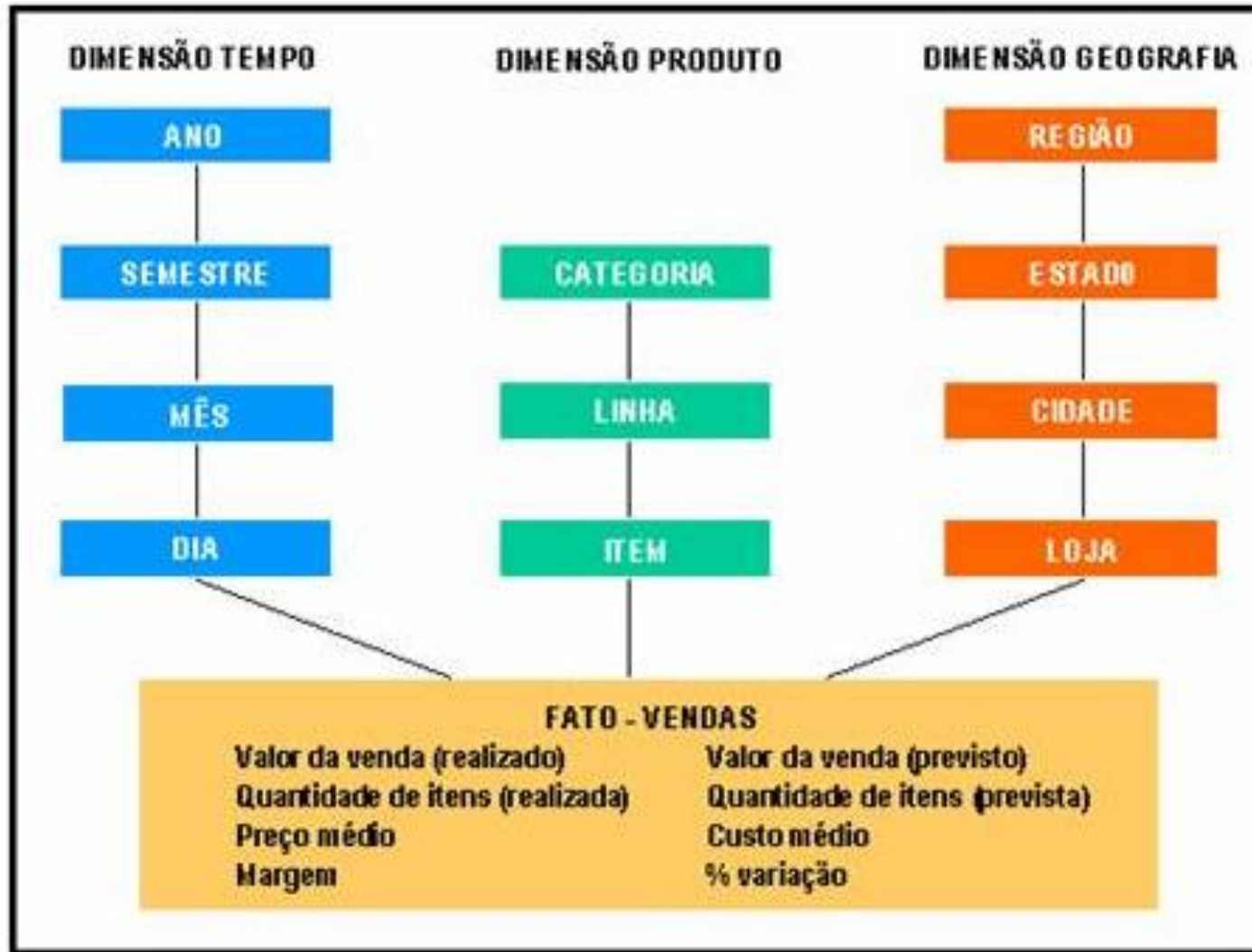
Ticket médio = R\$ 175,00.

Medida derivada

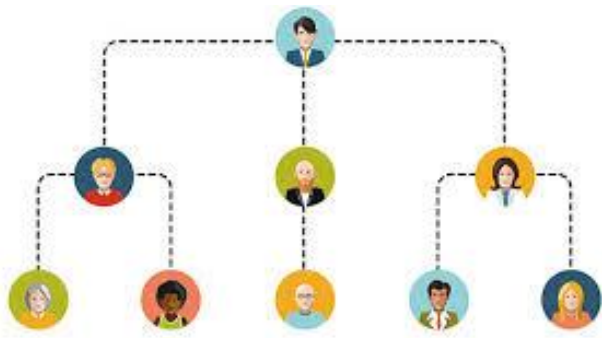
- Lembrando que quanto menos flexível for a utilização das medidas, mais complexa será a utilização pelos usuários, o que pode ser um fator crítico de sucesso para a solução de BI ou pior, levar a uma tomada de decisão errônea.
- Sempre que possível devemos gerar as agregações em tempo de execução para as métricas semi-aditivas e não-aditivas, facilitando a utilização e deixando transparente aos usuários os cálculos efetuados.
- Medidas derivadas podem ser facilmente reutilizáveis por diferentes dimensões

Hierarquias

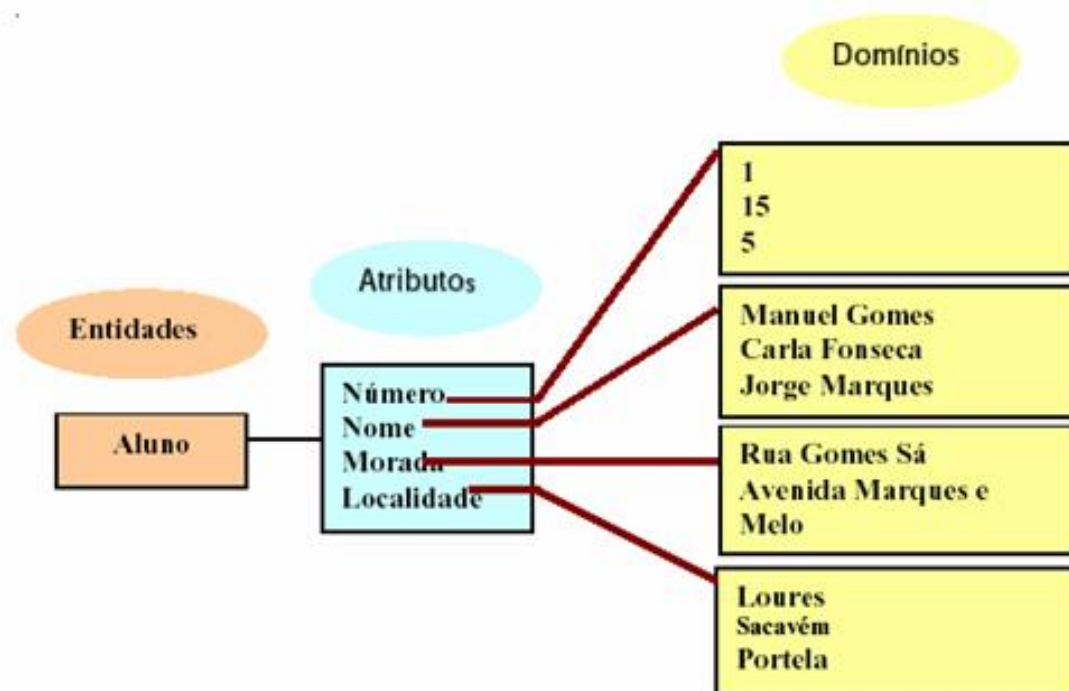
- Hierarquias são basicamente, uma forma hierárquica de organizar os dados nas dimensões.
- Elas permitem que atributos de uma dimensão relacionem-se com outros atributos da mesma dimensão.



Hierarquias



- É comum existir apenas uma hierarquia por dimensão, mas não é erro ter duas na mesma dimensão. Por exemplo, análise do Tempo por trimestre ou quadrimestre.
- O menor nível da hierarquia é o grão, onde é armazenado o dado.
- Então olhando a dimensão tempo o dia é o grão ou nível zero.
- A contagem de níveis começa no zero e vai subindo.

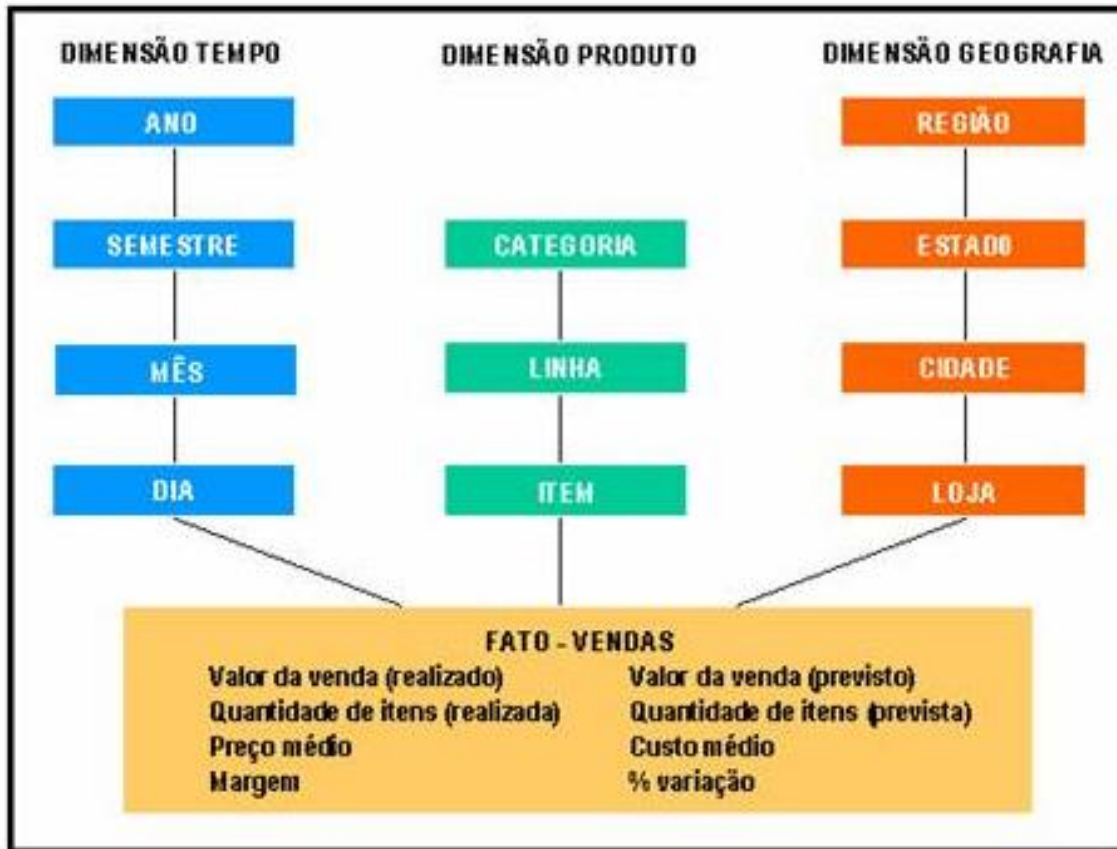


Atributos

- A principal função de uma tabela de dimensão é reunir os atributos que serão utilizados para qualificar as consultas e cujos valores serão utilizados para agrupar e sumariar as medidas (ou fatos).
- Ou seja, as tabelas Dimensão contêm atributos textuais que funcionam como filtros para as análises do usuário.
- Os atributos de uma dimensão podem compor uma hierarquia ou serem apenas descritivos.

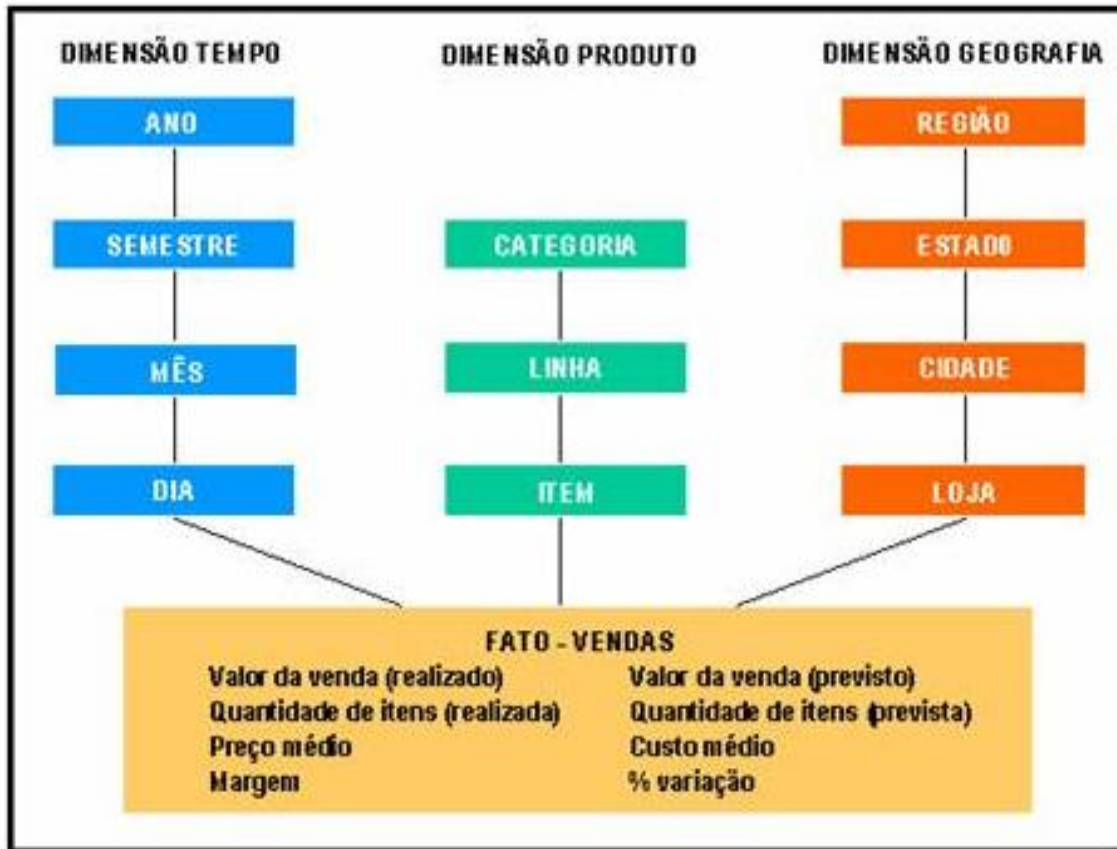
Atributos

- Em nosso exemplo, a dimensão Produto contém uma hierarquia composta pelos atributos produto, linha e categoria
- Indica que os itens de produto estão agrupados em linhas de produto, que por sua vez estão agrupadas em categorias de produto

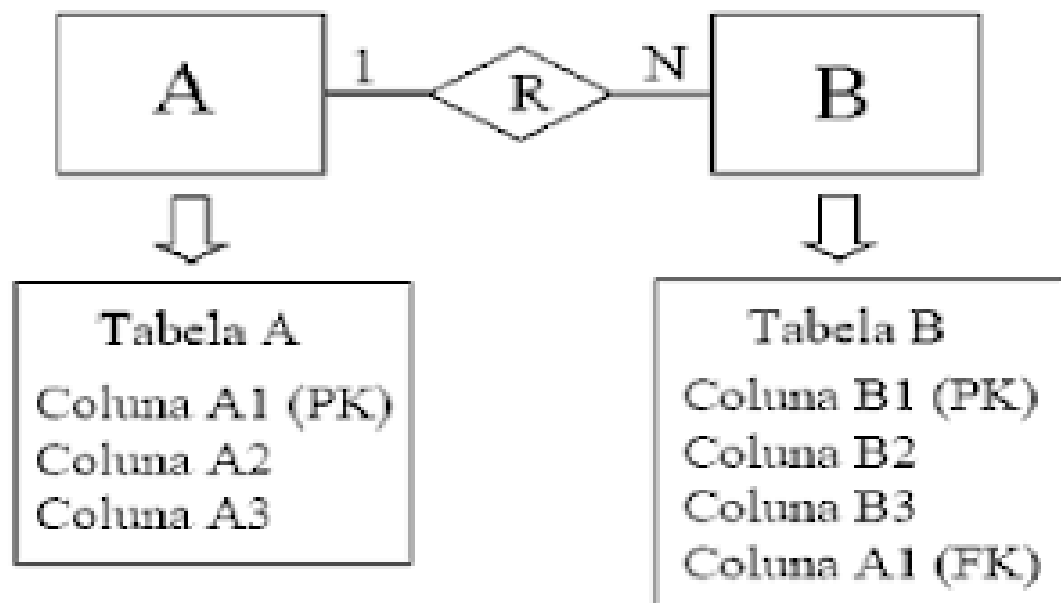


Atributos

- Podemos também incluir uma série de outros atributos descritivos que não fazem parte desta hierarquia, como o tipo de empacotamento (caixas com 12 unidades etc.), peso, valor unitário e outros que sejam relevantes para o processo de análise.
- Ralph Kimball sugere que as tabelas de dimensões possuam o maior número possível de atributos textuais para “enriquecer” o modelo de dados e, por consequência, as possibilidades de análise de informações.



Atributo chave



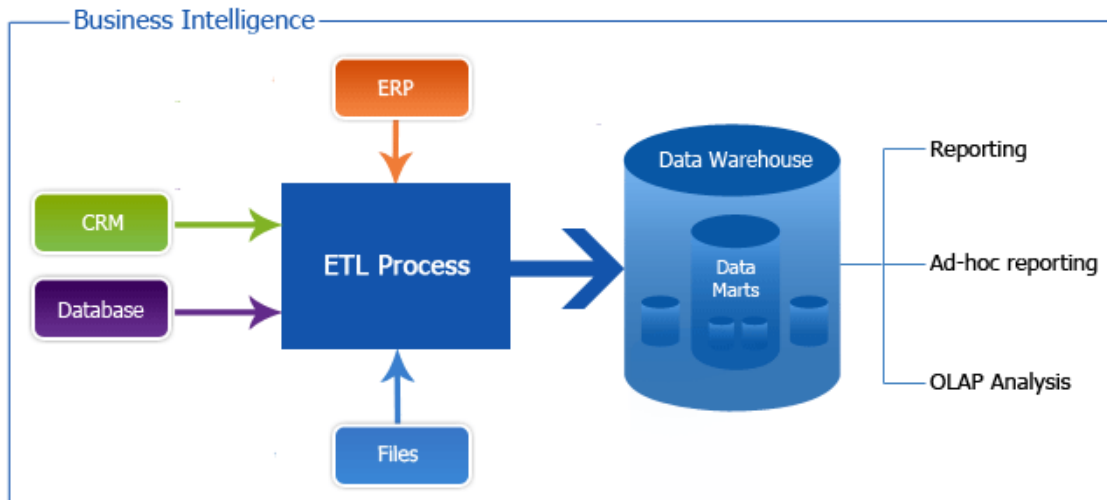
Mapeamento para relacionamento 1:N sem atributos

- Um atributo muito importante da tabela Dimensão é a sua chave.
- A chave primária de uma tabela Dimensão deve ser sempre um atributo único e definido pelo sistema com um valor genérico, inteiros atribuídos sequencialmente.
- Por questões de desempenho, não se utilizam chaves compostas por várias partes, nem tampouco chaves concatenadas.
- A ordem aqui é a da simplicidade, para facilitar o acesso aos grandes volumes de dados armazenados. As chaves serão utilizadas apenas para possibilitar a junção de tabelas entre uma tabela Dimensão e a tabela Fato.
- Também não são utilizados as chaves ou identificadores provenientes de outros sistemas, como código do cliente ou código do produto.

Atributo chave <https://youtu.be/sbIT5UXTEg8>

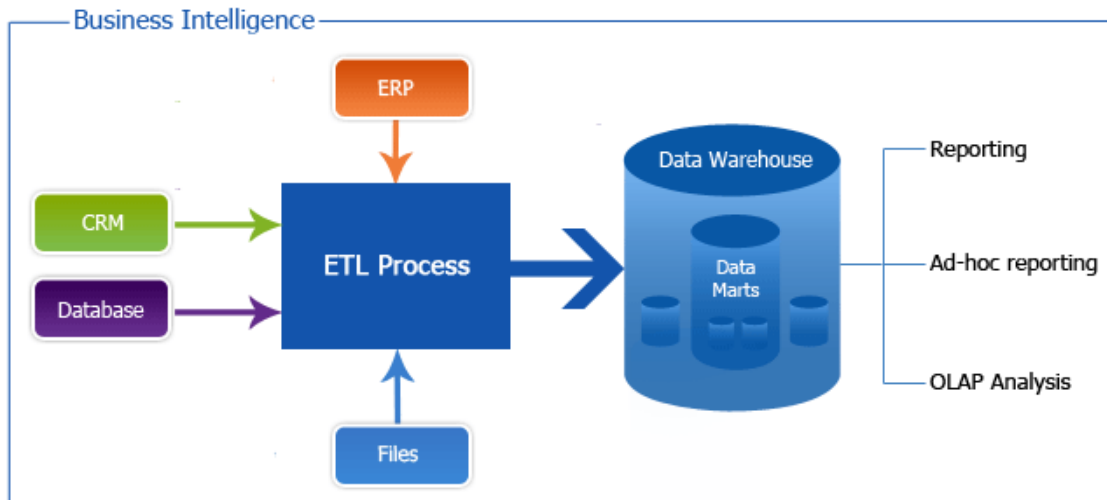


Atributos chave



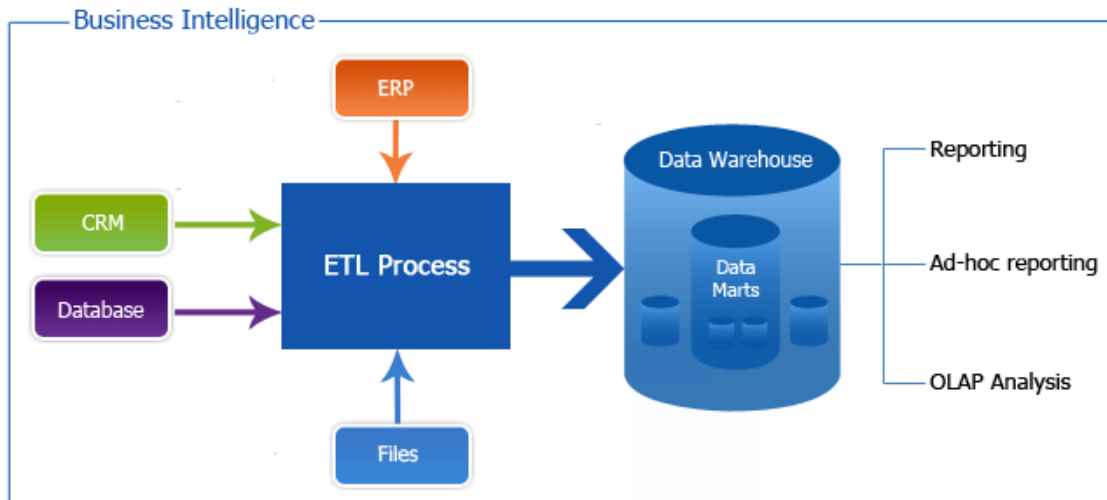
- Existem várias razões para se utilizar chaves genéricas, também chamadas de artificiais, substitutas ou surrogate keys.
- De acordo com Ralph Kimball, uma das razões é que o Data Warehouse deve se manter isolado das regras operacionais para gerar, atualizar, excluir, reciclar e reutilizar os códigos utilizados nos sistemas transacionais.
- O Data Warehouse manterá as informações durante muito tempo (normalmente vários anos) e não pode ficar vulnerável a problemas de sobreposição de chaves, no caso de aquisição ou consolidação de dados.

Atributos chave



- Outra razão é o melhor desempenho no acesso às informações.
- Muitas vezes, o código utilizado em um sistema transacional é um *string* de caracteres alfanuméricos que apresentam desempenho pior nas operações de acesso à base de dados do que as chaves genéricas, que utilizam o menor inteiro possível.
- Isto não significa que não se pode incluir, por exemplo, o código do produto utilizado nos sistemas transacionais, normalmente bastante conhecido dos usuários.

Atributos chave



- Se for bastante conhecido pelos usuários, este código pode ser um dos atributos textuais da dimensão Produto, favorecendo a análise das informações, mas, não deve ser a chave primária da dimensão, nem fazer parte da chave.
- É muito importante que a modelagem dimensional seja simples, pois essa simplicidade facilitará seu uso pelo analista de negócio e tomadores de decisão.



TP 1

Test Drive de Modelagem 2

Na pasta da disciplina de Modelagem, no Github.

- Procurar o arquivo: Exercício 2 – Modelagem Dimensional.pdf



Na próxima aula...

Vamos conectar o Power BI a bancos de dados

