

Dado x Informação x Conhecimento

Para início de conversa, vamos focalizar no que realmente agrega valor no Business Intelligence, que está relacionado aos conceitos de Dado, informação e conhecimento. Qualquer coisa diferente disso é meramente ilustração ao entorno.

Dado

O dado é forma bruta, a transação, o grão, o nível folha, ou seja, o detalhe. NÃO apoia a tomada de decisão.

O dado representa o que tem de mais importante em qualquer empresa, seja individual, micro, média, grande ou enorme. TODAS as Empresas só continuam existindo por conta dos dados, das suas transações do dia a dia. Tanto que atualmente, o dado é considerado um Ativo Empresarial.

Por exemplo: Em uma loja que venda canetas, o dado será a caneta azul marca X, também teremos como dado, caneta azul marca Y. e ainda teremos, a caneta preta marca Y.

Notem que são, de fato, o produto em si. Isto é um dado. Sempre que você realiza uma compra nas tradicionais lojas de varejo de seu bairro, quando o operador de caixa registra a venda do seu produto, aquela operação gera um dado (como exemplificado acima) e, este dado fica armazenado em um banco de dados.

Figura 1: Tabela representativa de dados – Excel.

	A	B	C	D	E	F
1	Produto	Tipo	Marca	qtde	Vlr	Total
2	Caneta azul	Caneta	X	1	R\$ 1,00	R\$ 1,00
3	Caneta azul	Caneta	X	2	R\$ 1,00	R\$ 2,00
4	Caneta azul	Caneta	Y	4	R\$ 1,00	R\$ 4,00
5	Caneta preta	Caneta	Y	1	R\$ 2,00	R\$ 2,00
6	Caneta preta	Caneta	Z	2	R\$ 4,00	R\$ 8,00
7	Caneta azul	Caneta	X	10	R\$ 1,00	R\$ 10,00
8	Lápis ponta fina	Lápis	Z	1	R\$ 0,50	R\$ 0,50

Fonte: Autoria própria

Informação

A informação é a forma inteligente, consolidada e **precisa do dado para existir**.

Por exemplo: Utilizando a mesma loja de venda de canetas, observando os dados da tabela ilustrada na figura 1, observamos as seguintes informações:

- 4 canetas azul vendidas
- 2 canetas pretas vendidas
- 1 Lápis vendido
- 6 canetas vendidas
- 3 canetas da marca X vendidas
- 2 canetas da marca Y vendidas
- 1 caneta da marca Z vendida

Perceba que, através da consolidação dos dados (ilustrados na figura 1), foi possível gerar diversas informações e, acredite, ainda mais informações que podem ser geradas a partir do mesmo conjunto de dados.

Então, como você pode observar, só é possível gerar informação, quando temos um conjunto de dados, ou seja, quando há transações e a partir das consolidações dessas, obtêm-se a possibilidade de gerar informações.

Conhecimento

Já o conhecimento é tudo aquilo que a informação te gerará de ideias e sugestões.

Seguindo no mesmo exemplo, note que neste momento, você tem conhecimento, baseado em informações, que te diz que a

marca X vende mais que a Y e a Z e, ainda que as canetas azuis são as preferidas.

Logo, sabendo que as canetas de marca X vendem bem, você pode negociar com o fornecedor um preço melhor ou a extensão do prazo de pagamento, para comprar um volume maior e destacar a marca em um ponto da loja de maior movimentação.

Perceba que isso é apenas um exemplo do conhecimento que pode ser gerado, a partir do conjunto de dados e informações que estamos trabalhando.

Certamente, existem outros conhecimentos que podem ser gerados, porém esse conhecimento, depende necessariamente, do seu conhecimento acerca do negócio que está sendo analisado.

É primordial que você entenda do assunto que irá aplicar os conceitos de business intelligence. Precisa conhecer os principais indicadores, os desafios, as oportunidades, o processo. É um momento que precisamos sair do escritório e “mergulhar” no processo.

Se você trabalha em uma administradora de shopping centers, vá para o shopping aprender o trabalho daqueles que o administram. Se você trabalha em uma loja com 1.000 franquias espalhadas pelo Brasil, vá para dentro da loja aprender o trabalho dos que lá estão.

VIVENCIAR o processo é fundamental para que você seja capaz de obter conhecimento a partir das informações geradas.

Dessa forma, podemos resumir os conceitos até o momento, utilizando a figura abaixo

Tabela 1 – Dado x Informação x Conhecimento

Dado	Informação	Conhecimento
Detalhe	Consolidado	Consolidado
Não apoia a decisão	Apoia a decisão	Provê insights e ideias
Operação	Gestão	Gestão
É criado a partir de uma transação	É criada a partir de um conjunto de dados	É criado a partir de um conjunto de informações

Fonte: Autoria Própria

Portanto, já é possível perceber que o Business Intelligence (BI), tem como objetivo principal, nos fornecer informações,

que a partir do entendimento do negócio, gere conhecimento para suportar tomadas de decisão.

Mas, guentaí! Vamos melhorar essa definição.

Ciclo de vida da informação Inteligente

Acontece que para que possamos gerar conhecimento, a partir de informações obtidas através dos dados, faz-se necessário entender a caminhada que o dado realiza até gerar um resultado (ou não).

Essa caminhada é chamada de ciclo de vida da informação inteligente.

Ciclo porque não tem um fim, visto que o objetivo primordial de se analisar dados para tomar decisões é a busca por resultados melhores. E, cá para nós, sempre é possível melhorar um cadinho!

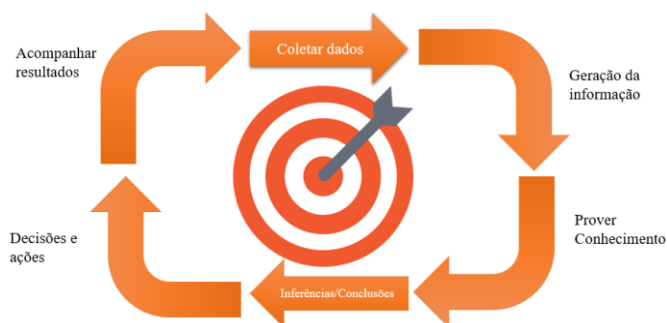
Informação inteligente é justamente pelo fato de que o conhecimento gerado é fundamentado em dados, de forma

íntegra e confiável, com toda a inteligência do seu negócio aplicada para se gerar essas informações.

Tá ficando mais claro o porquê devemos “mergulhar” nos processos do negócio para o qual vamos analisar dados?

A figura 2 nos ilustra a caminhada do dado, através do ciclo.

Figura 2 – Ciclo de vida da informação inteligente



Fonte: Autoria própria

Coletar Dados

Somente essa etapa daria um conteúdo dedicado, pois é o momento em que se precisa definir quais os dados serão coletados e, para se chegar a essa definição, você precisará atentar-se para 2 pontos cruciais:

1 – Conhecimento do negócio (já falei sobre isso mas, vou continuar repetindo)

2 – Requisitos¹ de negócio, que bem resumidamente, é descobrir quais as perguntas que precisam de respostas, onde essas respostas darão o suporte para tomada de decisão

Exemplos:

1. Qual o volume de vendas dos últimos 12 meses?
2. Quais os/as vendedores/as que atingiram as metas de vendas?
3. Qual o *turnover* da Empresa, por centro de custo?
4. Qual o gênero que mais compra?
5. Qual o horário com mais atendimentos?
6. Qual a faixa etária que mais compra, por canal de venda?
7. Qual a previsão de cancelamentos de compras, para os próximos 6 meses?
8. É possível vender o produto X na região A?

¹ Em meu livro “Metodologia e engenharia de requisitos para projetos de Business Intelligence”, abordo detalhadamente como trabalhar a etapa de “coleta de dados”, utilizando os fundamentos da engenharia de requisitos.

9. Qual a taxa de conversão por produto, através dos meses, dos últimos 2 anos?
10. O realizado está alinhado com o orçado?

Essa descoberta das perguntas que precisam de resposta deve ser identificada, no que chamamos de “a 4 mãos”, ou seja, entre você a pessoa que tem interesse em consumir as informações que você vai gerar.

Certamente, se você está realizando a coleta para si mesmo, pode “assumir” os 2 papéis. Agora, se vai gerar informações para outra pessoa tomar decisão, é de extrema importância que essa pessoa esteja envolvida no processo de identificação dos requisitos do negócio.

A fase de coleta de dados precisa, necessariamente, obter aqueles que responderão, na exata medida, todas as perguntas identificadas.

Geração da Informação

Uma vez que os dados foram coletados, vamos consolidá-los e organizá-los em grupos de informação.

A tabela do campeonato brasileiro é uma informação gerada a partir dos dados dos jogos, ilustrando assim a classificação (ordenação) dos clubes.

O gráfico do IBGE, em tempos de eleição, ilustra em uma linha do tempo o percentual de intenções de voto por candidato, ilustrando assim a tendência de um candidato ser eleito ou não. Essa informação é gerada a partir da coleta de dados de intenção de voto de cidadãos e cidadãs brasileiras, em periodicidades específicas.

A tabela com dados relacionados aos voos de um determinado Aeroporto, ilustram informações acerca do embarque, através de dados que são coletados através das Companhias Aéreas.

Uma conta de um restaurante, ilustra o valor que você precisa pagar, a partir de dados relacionados ao seu consumo.

Poderia escrever dezenas, talvez centenas de páginas dando exemplos de como a geração da informação está presente em nossa vida e, conseqüentemente, nos apoia para tomar decisões, desde as mais simples, como a de uma conta de restaurante, até as mais importantes, como a escolha de um/a candidato/a para governar nosso país.

Mas, como chego a uma decisão?

Prover Conhecimento

A partir das informações a que sou exposto e que tenho entendimento do negócio, sou capaz de obter conhecimento.

Ora, a partir do painel de informações dos voos de um determinado aeroporto, tendo eu o conhecimento do número do meu voo e da hora de embarque, identifico o portão, bem como observo se há (ou não) um tempinho para tomar um café.

A partir das informações acerca dos/as vendedores/as que atingiram a meta e do entendimento que possuo sobre o negócio, vou buscar conhecimento sobre o trabalho que eles/as realizaram e que obtiveram sucesso, para replicar àqueles/as vendedores/as que não atingiram suas metas.

Mais uma vez, percebe-se o quão importante é ter o entendimento do negócio e de seus requisitos. Pois é a partir deles que todo o ciclo será realizado.

Conclusões/Inferências

Inferência é um termo oriundo da estatística que, parte da premissa que a partir do conhecimento gerado acerca de

informações de um conjunto de dados específico, é possível assumir que esse conhecimento pode ser aplicado no todo.

Vamos ao exemplo para facilitar a compreensão.

Uma vez que a partir da informação acerca do atingimento das metas de vendas dos/as vendedores/as e, consequente, entendimento do trabalho que eles/as executaram e que gerou resultado positivo, é possível inferir (ou concluir) que se os/as vendedores/as que não atingiram suas metas executarem esse mesmo trabalho de sucesso, poderão atingir, em um futuro próximo, as metas de vendas.

Vale ressaltar que a inferência NÃO significa que o resultado será o mesmo, mas sim uma alta probabilidade do resultado se repetir.

Decisão e ação

Momento crucial da vida profissional de todo tomador de decisão, uma vez que é detentor do conhecimento e já obteve as conclusões necessárias, é necessário planejar e executar suas decisões.

Perceba que mencionei “planejar”. Sim! Decisão não se toma de qualquer jeito. É preciso planejamento e aqui eu vou te

apresentar a maneira mais simples e objetiva de planejar suas decisões, para colocá-las em ação que, inclusive é a forma como faço no meu dia a dia.

Você já deve ter ouvido falar do 5W2H. É uma ferramenta oriunda da administração que vai te auxiliar nesse processo de determinação de ações que precisam ser executadas.

Chama-se 5W2H pelo fato de você ter de identificar o seguinte:

- *WHAT* (O que?)
 - O que será feito?
 - A resposta é de fato o que você decidiu fazer.
 - Exemplo: Estruturar um treinamento, baseado nas ações dos/as 3 primeiros vendedores/as, segundo suas metas e capacitar àqueles/as que não atingiram as metas, visando dobrar a base de vendedores que atingiram suas metas.
- *WHY* (Por que?)
 - Por que preciso fazer isso?
 - Você deve elencar os motivos reais que ilustram a necessidade da sua decisão ser tomada. Procure ser bem claro, objetivo e direto ao ponto. Evite detalhamentos.

- Exemplo: É necessário ampliar a base de vendedores/as que atingem as metas, pois dessa forma, haverá um impacto positivo no resultado da Empresa.
- *WHERE* (Onde?)
 - Onde será feito?
 - Apesar de parecer óbvio (e é). É preciso ser exposto. Eu sempre falo que “o óbvio precisa ser dito!”. Então, aqui você vai descrever onde a sua decisão acontecerá, descreva exatamente o local, endereço, departamentos envolvidos etc.
 - Exemplo: Acontecerá no auditório do 7º andar, com a presença do comercial e do marketing.
- *WHEN* (Quando?)
 - Quando será feito?
 - Todo planejamento precisa de um prazo, precisa de datas. Então é o momento dessa definição.
 - Exemplo: Dias 20/02, 22/02 e 24/02, das 14:00 às 18:00
- *WHO* (Quem?)

- Quem vai fazer?
- É preciso pensar em quem são as pessoas que te auxiliarão a resolver o problema. É preciso dos nomes, afinal elas serão envolvidas no processo.
- Exemplo: Vendedora Maria (Zona Sul), Vendedores João e Manuela (Zona Norte), Carlos (Gerente da regional do Rio de Janeiro) e Claudia (Instrutura de vendas).
- *HOW* (Como?)
 - Como será feito?
 - Nesse ponto, você precisa detalhar a caminhada para atingir o que você descreveu no “*What*”. É preciso pensar em todo o processo que será necessário desenvolver para atingir o objetivo. Aqui é importante a especificidade.
 - Exemplo: Reunir-se com o 3 primeiros/as vendedores, segundo o atingimento de metas, para mapear as ações bem-sucedidas. Estruturar um plano de ensino. Desenvolver uma apostila eletrônica. Desenvolver os slides da apresentação. Alinhar o conteúdo com todos/as

os/as envolvidos/as (mapeados no *HOW*).
Ajustar material, conforme necessidades.
Reservar a sala nas datas e horários definidos (*WHEN*). Enviar agenda para todos/as os participantes. Executar o treinamento.

- *HOW MUCH* (Quanto?):
 - Quanto vai custar?
 - Sabemos que no mundo corporativo, tudo envolver dinheiro, seja ele físico, seja ele virtual ou seja ele representado pela sua hora de trabalho. Aqui, você precisa detalhar os gastos com equipamentos, serviços de terceiros, horas trabalhadas pelos profissionais envolvidos etc. Essa etapa é importante até para você observar a viabilidade da sua decisão, pois pode acontecer do custo estar fora do seu orçamento e você precisar replanejar suas decisões e ações. Logo, trata-se de uma etapa bem importante.
 - Exemplo:
 - Gráfica: R\$ 3.000,00
 - Aluguel Projetor: R\$ 250,00

- Valor das horas de preparação do material: R\$ 2.000,00
- Valor das horas dos/as 3 vendedores/as que apoiarão o treinamento: R\$ 6.000,00
- Valor das horas do gestor: R\$ 2.500,00
- Papelaria: R\$ 250,00
- Total: R\$ 14.000,00

Assim como eu, você pode utilizar o Excel, para detalhar suas decisões. Para cada decisão que você vai tomar, defina um 5W2H. Pode ter certeza que essa ferramenta vai te auxiliar e você nunca mais ficará preocupado com suas tomadas de decisão.

Acompanhar resultados

Decisão tomada e ação executada, chega-se o momento de acompanhar os resultados gerados. Surtiu efeito? Gerou o resultado Esperado?

Se o resultado foi positivo, comemore! Valorize o seu trabalho e o das pessoas envolvidas. Dê publicidade ao feito, pois o seu

resultado vai motivar outras áreas e outras pessoas. Semeie a positividade dentro da sua Empresa!

Agora, aqui é importante entender uma coisa. Você NÃO vai acertar sempre! E tudo bem! Aprenda a lidar e conviver com isso. Só erra quem trabalha! Só erra quem “bota a cara a tapa”.

O problema não estar em errar, mas sim, em se omitir, em se esquivar ao erro! Não se acovarde diante dos erros! Encare-os de frente!

O Erro é uma grande oportunidade de aprendizado. Então, se a sua decisão não surtiu o efeito desejado, busque o entendimento do/s porquê/s. Perguntas que você deve procurar responder nesse momento:

- Qual era o resultado esperado (veja no *What*) e qual foi o realizado?
- Quais os possíveis motivos para não alcançar o resultado esperado? (Converse com os/as envolvidos e identifique se as ações aconteceram como deveriam acontecer)

- Quais decisões e ações precisam de um novo planejamento? É importante ajustar decisões e ações que NÃO surtiram efeito, por novas.

E o ciclo nunca se acaba, independente da resposta que você terá.

Business Intelligence e Ciência de dados

Imagino que você já tenha tomado conhecimento que, nos tempos atuais, fale-se muito desses dois conceitos. Eu não vou replicar aqui, o que blogs, gurus, youtubers, marketeiros vem falando por aí, vou te apresentar os fundamentos, baseado na ciência, que por sinal, muitas pessoas utilizam o termo “ciência”, mas na realidade não sabem o que significa.

A etimologia da palavra “Ciência” origina-se no latim “Scientia”² e que significa “Conhecimento”.

² Recuso-me a colocar palavras em latim no itálico, visto que o latim é nossa língua mãe.

A partir desse entendimento, podemos conceituar “Ciência de Dados”, dizendo que é o conhecimento obtido através dos dados.

Quando partimos para a definição de *Business Intelligence*, no primeiro momento precisamos traduzir o termo que, em português, significa Inteligência de Negócios.

Agora, vamos a etimologia da palavra “Inteligência” que já é a derivação do latim de “Intelligentia” que, por sua vez, origina-se em “Intelligere” e, que significa “saber escolher a melhor alternativa, dentre as várias opções”, “saber ler as entrelinhas”.

Dessa forma, podemos dizer que inteligência de negócios é a capacidade de observar a fundo, para se escolher a melhor opção.

Agora, perceba que, uma vez que temos os significados de ciência de dados e de business intelligence, podemos pressupor que ambos estão conectados e que não faz sentido separarmos os conceitos, visto que o conhecimento obtido a partir dos dados (ciência de dados) é capaz de dar o suporte para que seja possível observar de forma mais profunda e,

consequentemente, escolher a melhor opção, tomar uma decisão mais assertiva (business intelligence).

Perceba que o objetivo aqui é te entregar a verdadeira fundamentação e o conceito real do que é business intelligence, sem falsas promessas, sem falácias, sem marketing, sem guruzisses. Aqui partiremos sempre do que a ciência pode nos apresentar.

Por falar em ciência, é legal destacar que existe um amplo campo de estudo e pesquisas científicas, que vai abordar de forma mais detalhada esses termos, chamado “Sistemas de Apoio à Decisão” (SAD).

Bom, até aqui creio que você já consegue discernir sobre o que é o business intelligence porém, conforme formos caminhando vamos melhorando ainda mais essa definição.

Tipos de variável

Importante entendermos que quando vamos analisar informações para obter conhecimento que suportará tomadas de decisão, vamos interagir com dois tipos de variável: Variáveis qualitativas e Variáveis quantitativas.

Variáveis Qualitativas

As variáveis qualitativas estão divididas em dois grupos:

- Nominais:
 - Por exemplo: Gênero, cor, grupo sanguíneo, região, departamento, lojas etc.
- Ordinais:
 - Por exemplo: Pressão arterial (baixa, média, alta), Empregado (Sim, Não) etc.

Perceba que as variáveis qualitativas, independentemente dos grupos, têm por finalidade a categorização acerca de uma determinada informação.

Variáveis Quantitativas

Já as quantitativas são divididas em 2 grupos:

- Discretas: Remetem a contagem.
 - Por exemplo: Quantidade de acidentes, número de filhos, quantidade de lojas etc.
- Contínuas: Remetem a mensuração
 - Por exemplo: Km, peso, altura, m³, valor vendido etc.

Assim, é possível perceber que as variáveis quantitativas, independentemente do grupo, terão a finalidade de quantificar dados, através de operações matemáticas.

Tenha cuidado com variáveis que remetem a quantificação, porém NÃO devem ser quantificadas, como por exemplo: Ano, telefone, número da moradia, códigos numéricos etc. Estes dados devem ser tratados como variáveis qualitativas.

Variáveis qualitativas e quantitativas juntas

Vamos analisar os dados da figura 4.

Figura 3 - Variáveis qualitativas e quantitativas

	A	B	C	D	E	F	G	H	I
1	Identificação da venda	Ano	Mês	Produto	Tipo	Marca	qtde	Vlr	Total
2	1253	2022	10	Caneta azul	Caneta	X	1	R\$ 1,00	R\$ 1,00
3	1254	2022	10	Caneta azul	Caneta	X	2	R\$ 1,00	R\$ 2,00
4	1255	2022	10	Caneta azul	Caneta	Y	4	R\$ 1,00	R\$ 4,00
5	1256	2022	10	Caneta preta	Caneta	Y	1	R\$ 2,00	R\$ 2,00
6	1257	2022	10	Caneta preta	Caneta	Z	2	R\$ 4,00	R\$ 8,00
7	1258	2022	10	Caneta azul	Caneta	X	10	R\$ 1,00	R\$ 10,00
8	1259	2022	10	Lápis ponta fina	Lápis	Z	2	R\$ 0,50	R\$ 1,00
9	1260	2022	11	Caneta azul	Caneta	X	5	R\$ 1,00	R\$ 5,00
10	1261	2022	11	Caneta azul	Caneta	X	3	R\$ 1,00	R\$ 3,00
11	1262	2022	11	Caneta azul	Caneta	Y	1	R\$ 1,00	R\$ 1,00
12	1263	2022	11	Caneta preta	Caneta	Y	2	R\$ 2,00	R\$ 4,00
13	1264	2022	11	Caneta azul	Caneta	X	20	R\$ 1,00	R\$ 20,00

Fonte: Autoria própria

Observando o conjunto de dados dispostos na figura 4, podemos identificar:

- Variáveis qualitativas: Identificação de venda, ano, mês, produto, tipo e marca
- Variáveis quantitativas: Qtde, vlr e total

É possível existirem outras variáveis quantitativas, nesse mesmo conjunto de dados, de acordo com o algum requisito.

Por exemplo: Quantas vendas foram realizadas entre os meses 10 e 11?

Figura 4 - Quantidade total mensal

K	L
Meses <input type="button" value="v"/>	Qtde total mensal
10	22
11	31
Total Geral	53

Fonte: Autoria própria

A figura 5 ilustra a resposta ao requisito, apresentando uma nova variável quantitativa, intitulada “Qtde total mensal”, que é o somatório da variável quantitativa “Qtde” (figura 4), através da variável qualitativa “Mês” (figura 4).

Ressalta-se também a possibilidade de novas variáveis qualitativas surgirem, dependendo do requisito.

Por exemplo: Qual o valor total vendido pelas cores?

Figura 5 - Valor total vendido por cores

	A	B	C	D	E	F	G	H	I	J
1	Identificação	Ano	Mês	Produto	Cor	Tipo	Marca	qtde	Vlr	Total
2	1253	2022	10	Caneta azul	azul	Caneta	X	1	R\$ 1,00	R\$ 1,00
3	1254	2022	10	Caneta azul	azul	Caneta	X	2	R\$ 1,00	R\$ 2,00
4	1255	2022	10	Caneta azul	azul	Caneta	Y	4	R\$ 1,00	R\$ 4,00
5	1256	2022	10	Caneta preta	preta	Caneta	Y	1	R\$ 2,00	R\$ 2,00
6	1257	2022	10	Caneta preta	preta	Caneta	Z	2	R\$ 4,00	R\$ 8,00
7	1258	2022	10	Caneta azul	azul	Caneta	X	10	R\$ 1,00	R\$ 10,00
8	1259	2022	10	Lápis ponta fina preta	preta	Lápis	Z	2	R\$ 0,50	R\$ 1,00
9	1260	2022	11	Caneta azul	azul	Caneta	X	5	R\$ 1,00	R\$ 5,00
10	1261	2022	11	Caneta azul	azul	Caneta	X	3	R\$ 1,00	R\$ 3,00
11	1262	2022	11	Caneta azul	azul	Caneta	Y	1	R\$ 1,00	R\$ 1,00
12	1263	2022	11	Caneta preta	preta	Caneta	Y	2	R\$ 2,00	R\$ 4,00
13	1264	2022	11	Caneta azul	azul	Caneta	X	20	R\$ 1,00	R\$ 20,00
14										
15	Cor									
16	azul									
17	preta									
18	Total Geral									

Fonte: Autoria própria

Observe na figura 6, que para entregar a resposta a esse requisito foi necessário a criação de uma variável qualitativa (Coluna “Cor”, em destaque na figura 6) e uma variável qualitativa, intitulada “Valor total” (na tabela dinâmica, abaixo do conjunto de dados).

Sendo assim, percebe-se que o conjunto de dados original vai te dispor variáveis qualitativas e quantitativas, porém para você obter as respostas necessárias para gerar o conhecimento necessário para suas tomadas de decisão, certamente novas variáveis serão criadas.

Um outro fator importante a observar é que as variáveis qualitativas interferem diretamente nas quantitativas.

Figura 6 - Qualitativa interfere na quantitativa

15	Cor	<input type="button" value="v"/>	Valor total
16	azul		46
17	preta		15
18	Total Geral		61

Fonte: Autoria própria

Observe a figura 7 e perceba que o tipo de cor (variável qualitativa) interfere diretamente no resultado do “Valor total” (variável quantitativa). Conforme a perspectiva de análise (qualitativa) varia, o valor (quantitativa) altera.

Em business intelligence, as variáveis qualitativas, independentemente do grupo, são chamadas de dimensões de negócio e as variáveis quantitativas, independentemente do grupo, são chamadas de métricas.

Dimensão de negócio

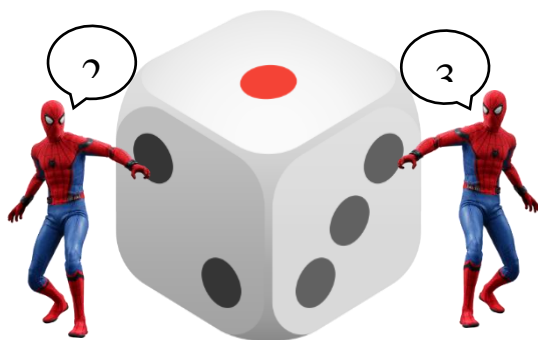
Como já aprendemos anteriormente, as variáveis qualitativas em business intelligence são chamadas de Dimensões de

negócio, porém é importante se atentar para os tipos de dados, visto que um conjunto de colunas pode representar uma mesma dimensão (veremos esse assunto de forma mais detalhada em “hierarquias”).

Chamam-se dimensões, por remete a visão tridimensional (mais a frente veremos o conceito de multidimensão).

Para facilitar o seu entendimento, imagine um dado tradicional de jogatina, ele é uma figura geométrica espacial, chamada de Cubo (mais a frente falaremos da metáfora do cubo), com seis faces, onde dependendo da posição do observador, da dimensão que se observa, o número apresentado é diferente.

Figura 7 - Perspectiva dimensional



Fonte: Autoria própria

Dependendo do ponto de vista do homem aranha, ou seja, de que lado da dimensão da figura espacial ele se encontra, a observação é diferente e, conseqüentemente, o valor. Por esse motivo, o nome de “Dimensão”.

E é de negócio, pois já sabemos que a dimensão se refere a uma variável qualitativa, porém lembre-se que *Business Intelligence*, em português significa inteligência de negócios, esse último está diretamente ligado ao negócio (com o perdão da redundância) que você pretende obter conhecimento para tomar decisões.

Ou seja, as dimensões de negócio são representadas por variáveis qualitativas inerentes ao negócio.

Métricas

As variáveis quantitativas, independentemente do seu grupo, em *Business Intelligence* são chamadas de métricas, pois remetem a mensuração.

Tudo que vimos sobre variáveis quantitativas valem para as métricas. Ou seja, é um conceito bem mais simples de se entender.

Fato

Em *Business Intelligence* existe um conceito que representa a união das dimensões com as métricas, que é intitulado de fato.

Chama-se fato, pois “contra fatos não há argumentos”. É isso mesmo! O fato é a verdade que o dado apresenta (partindo da premissa que estamos trabalhando com dados confiáveis e íntegros).

Figura 8 - Fato

	A	B	C	D	E	F	G	H	I
1	Identificação da venda	Ano	Mês	Produto	Tipo	Marca	qtde	Vlr	Total
2	1253	2022	10	Caneta azul	Caneta	X	1	R\$ 1,00	R\$ 1,00
3	1254	2022	10	Caneta azul	Caneta	X	2	R\$ 1,00	R\$ 2,00
4	1255	2022	10	Caneta azul	Caneta	Y	4	R\$ 1,00	R\$ 4,00
5	1256	2022	10	Caneta preta	Caneta	Y	1	R\$ 2,00	R\$ 2,00
6	1257	2022	10	Caneta preta	Caneta	Z	2	R\$ 4,00	R\$ 8,00
7	1258	2022	10	Caneta azul	Caneta	X	10	R\$ 1,00	R\$ 10,00
8	1259	2022	10	Lápis ponta fina	Lápis	Z	2	R\$ 0,50	R\$ 1,00
9	1260	2022	11	Caneta azul	Caneta	X	5	R\$ 1,00	R\$ 5,00
10	1261	2022	11	Caneta azul	Caneta	X	3	R\$ 1,00	R\$ 3,00
11	1262	2022	11	Caneta azul	Caneta	Y	1	R\$ 1,00	R\$ 1,00
12	1263	2022	11	Caneta preta	Caneta	Y	2	R\$ 2,00	R\$ 4,00
13	1264	2022	11	Caneta azul	Caneta	X	20	R\$ 1,00	R\$ 20,00

Fonte: Autoria própria

Cada linha, da 2 a 13, é a junção de dimensões e métricas e representam um fato, uma verdade ilustrada pelos dados que estamos garantindo que são confiáveis e íntegros.

Modelo de dados

Devido a importância da modelagem de dados para uma solução de *business intelligence*, dedico um capítulo especial, para que você aprenda de uma vez por todas, como modelar um dado de forma simples e objetiva.

Em um primeiro momento, vamos entender que modelagem de dados é um conceito, sendo assim, uma vez que você o assimila pode aplica-lo em qualquer tecnologia, aliás...até desenhar em guardanapo de bar é possível.

Modelar dados é simplesmente entender e projetar como que os diversos dados que você identificou nos requisitos de negócio, se correlacionam entre si, como se conectam, como conversam, como dependem um do outro. Então, perceba que a definição do modelo da dados independe que tecnologia.

Já a construção do seu modelo de dados, necessariamente, precisará de uma tecnologia, essa por sua vez, está diretamente ligada a complexidade da sua arquitetura de dados: Baixa, média ou alta.

Como já citado anteriormente, em uma arquitetura de dados, encontraremos o *Data Warehouse* e quando estamos lidando com arquiteturas de baixa complexidade, o modelo de dados pode ser construído dentro do Excel ou do Google planilhas, através das abas, onde cada aba representará um conjunto de dados específico e que se correlacionam. Outra possibilidade é a construção dentro de uma tecnologia de acesso ao dado, como o Power BI ou Google Looker Studio.

Quando estamos lidando com arquiteturas de média e alta complexidade, possivelmente o modelo de dados será constituído por tabelas de banco de dados.

Bancos de dados são tecnologias com a finalidade específica de armazenamento de dados, onde vai te prover uma performance maior e com mais agilidade de consulta aos dados, diferente de uma tecnologia de planilhas, como Excel, que possui limitações.

Tabelas são os objetos dentro do banco de dados, onde os dados ficam armazenados e são compostas por linhas e colunas. Se fizermos uma analogia (guardada as devidas proporções e limitações) com o Excel, cada aba seria uma tabela no banco de dados.

Para que estas tabelas se comuniquem umas com as outras, ou seja, para que seja possível relacionar dados de lojas, com produtos e vendas, faz-se necessário um modelo de dados, no qual será responsável por prover a relação entre os dados das diversas tabelas.

Vamos trazer um exemplo prático do dia a dia: Imagina que você vai a uma loja e efetua a compra de uma calça, uma camisa e um tênis. Vamos atentar para os processos que acontecem:

Processo de compra:

- a. Você escolhe os produtos
- b. Encaminha ao caixa
- c. Caixa registra os produtos
- d. Você efetua o pagamento
- e. Recebe a NF e seus produtos
- f. Finaliza a compra

Neste simples processo, quando a pessoa do caixa registra seus produtos, esses dados estão sendo inseridos em uma tabela de banco de dados, responsável por armazenar as vendas. Ficando da seguinte forma:

Tabela 2 – Exemplo da tabela de vendas

Data_Compra	CPF_Cliente	COD_Vendedor	Cod_Produto	Qtde	Vlr_Produto	Vlr_Desconto	Vlr_Impostos	Vlr_Total
25-01-21	000.111.222-00	99999	123	1	100	0	20	80
25-01-21	000.111.222-00	99999	345	1	50	0	10	40
25-01-21	000.111.222-00	99999	678	1	80	0	16	64

Fonte: Autoria própria

Agora, imagine que este processo de compra aconteça inúmeras vezes, durante os dias, meses e anos. Ou seja, constantemente dados são inseridos nesta tabela, registrando as sucessivas vendas.

Em paralelo a este processo, existe um outro, chamado processo gerencial, responsável pela gestão da loja, onde o gestor necessita tomar algumas decisões e para isso alguns requisitos de negócio, baseado nas perguntas que precisam de respostas, precisam ser mapeados, como por exemplo:

- Quais os três produtos que mais venderam na semana?
- Quais os três produtos que menos venderam neste mês?
- Qual o melhor vendedor da semana?
- Qual o cliente que mais comprou neste mês?
- Qual a evolução do meu volume de vendas no ano de 2020?
- Qual o valor vendido nos últimos quinze dias?

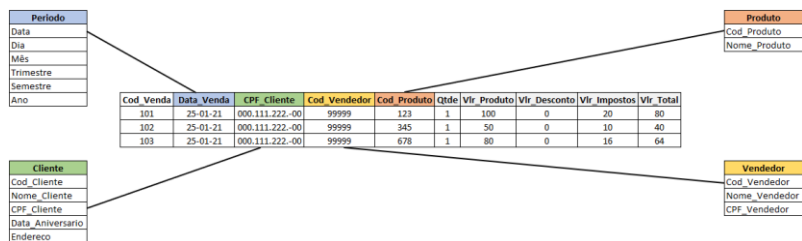
g) Quais os vendedores não bateram a meta mensal?

Enfim, citei algumas perguntas, mas certamente cada gestor necessitará de diversas respostas, para tomar decisões de forma a conduzir o negócio da melhor forma possível.

Para que estas perguntas (e outras) sejam respondidas, se faz necessário um modelo de dados capaz de sustentá-las, de forma que seja possível prover as informações de produtos, vendedores, períodos, clientes etc.

Estas informações acerca dos dados serão as tabelas do modelo de dados que, para responder as perguntas hipotéticas listadas acima, ficariam modeladas da seguinte forma:

Figura 9 - Modelo de dados do exemplo



Fonte: Autoria própria

Perceba que, orbitam ao redor da tabela com o registro das vendas, as demais tabelas com os dados inerentes as respostas

que serão fornecidas ao gestor da loja, em seu processo gerencial de tomada de decisão, de acordo com os requisitos que foram identificados.

Desta forma, entenda que a tabela que possui as MÉTRICAS de vendas é o FATO, afinal perceba que todas as dimensões e métricas (variáveis quantitativas) se encontram nela e as tabelas que a orbitam são as DIMENSÕES, àquelas que possuem os dados e metadados específicos de cada variável qualitativa.

O modelo de dados clássico, utilizado em um *Data Warehouse* e *Data Mart* é o Modelo Dimensional, também chamado de *star schema*, em português, modelo estrela (KIMBALL, 2002) ou *star join*, em português, junção estrela (INMON, 2005). Estrela, pois a tabela fato está no meio do modelo e as dimensões nas pontas, lembrando uma estrela.

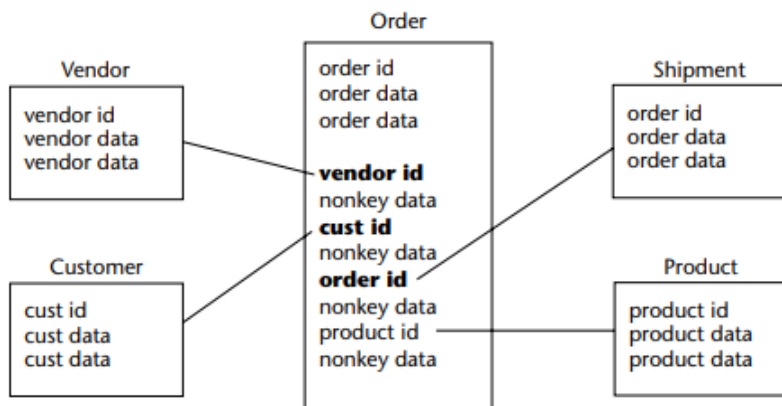
Modelo dimensional é uma metodologia para modelagem lógica de dados que visa desempenho de consulta e facilidade de uso começando a partir de um conjunto de eventos de aferição dos dados. No banco de dados, uma tabela de fatos é construída geralmente com um registro para cada medição. Esta tabela de fatos é rodeada por um conjunto de tabelas de dimensão que descrevem precisamente o que é conhecido no contexto de cada registro de

medição. Por causa da estrutura característica de um modelo dimensional, é freqüentemente chamado de esquema em estrela. (KIMBALL, 2002, p. 399)

Este tipo de modelagem tornou-se a principal busca em projetos de *business intelligence*, por ser de fácil compreensão, baixa manutenção, facilmente escalável e que possibilita consultas rápidas, ágeis e de fácil assimilação pelos/as usuários/as. Por este motivo que os requisitos de negócio é que serão o insumo para a criação do seu modelo de dados.

O benefício de criar modelos estrela é agilizar os dados para o processamento. Interligando os dados e criando redundância seletiva, o modelo simplifica e agiliza os dados para acesso e análise, que é exatamente o que é necessário para as consultas [...] A grande vantagem do modelo multidimensional é sua eficiência de acesso. Quando projetado corretamente, o modelo estrela é muito eficiente na entrega de dados para o usuário final. Para tornar a entrega de informações eficiente, os requisitos do usuário final devem ser reunidos e assimilados. São o cerne da definição de como será a estrutura multidimensional. Uma vez que os requisitos do usuário final são compreendidos, serão usados para o desenvolvimento da estrutura ideal do modelo estrela. (INMON, 2005, p. 129-130-362)

Figura 10 – Modelo Estrela (Star Join) proposto por Inmon



Fonte: Um simples modelo estrela (INMON, 2005, p. 129)

É importante analisar que ambos os autores trazem abordagem semelhantes com relação ao modelo estrela, principalmente no sentido de desempenho, agilidade e facilidade de consulta pelo/a usuário/a. E isto é um fator preponderante de sucesso da sua solução, a facilidade e agilidade com que o usuário conseguirá realizar consultas para obter respostas aos seus requisitos de negócio.

Metáfora do cubo

Neste ponto, eu ousou acrescentar ao tema, a metáfora do cubo, nome popular ao modelo estrela, que visa facilitar ainda mais

o acesso, fazendo com que seu modelo estrela seja também uma camada semântica, utilizando uma nomenclatura de fácil entendimento, por exemplo:

- Ao invés de DSC_Produto, utilize Descrição do Produto.
- Ao invés de NM_Cliente, utilize Nome do Cliente.
- Ao invés de Vlr_Total, utilize Valor Total.

E assim por diante, sempre que disponibilizar para acesso, faça com que seja de simples entendimento e alinhado com a terminologia que o tomador de decisão utiliza. No momento do mapeamento de requisitos, você precisa identificar a semântica dos dados.

O nome cubo, como já falado anteriormente, surge pelo fato de que um modelo dimensional ou, no nosso caso, multidimensional – pois apresenta duas ou mais dimensões – possibilita a análise de uma determinada informação sobre vários pontos de vista, várias perspectivas. No exemplo que citamos (vide figura 23), a venda poderá ser analisada por produto e esta mesma venda poderá ser analisada por vendedor.

subconjunto de dados altamente inter-relacionados organizado para permitir aos usuários combinar quaisquer atributos em um cubo (p.ex., lojas, produtos, clientes, fornecedores) com quaisquer métricas no cubo (p.ex., vendas, lucro, unidades, idade). (TURBAN *et al.*, 2009, p. 241-242)

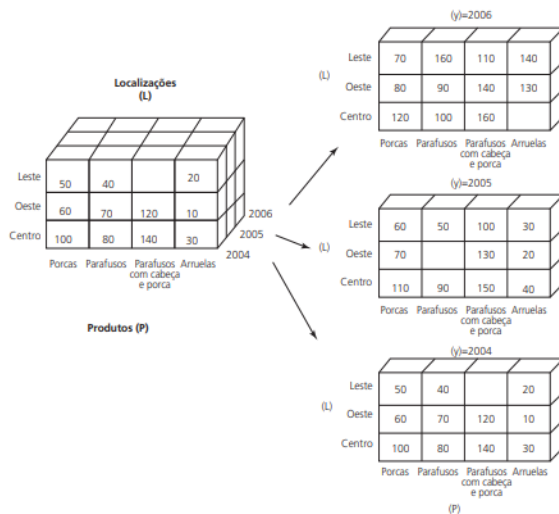
Perceba que Turban *et al.* (2009) ainda reforça a possibilidade de se analisar todas as métricas (variáveis quantitativas) existentes na tabela fato, com todas as dimensões (variáveis qualitativas) existentes em seu cubo.

Neste ponto, é preciso ter cuidado pois esta funcionalidade pode (e vai) abrir possibilidade de se analisar uma métrica ou um indicador com alguma dimensão que não faça sentido, por exemplo: Imagine o que na sua empresa existam vendedores com exclusividade em produtos – isso é comum quando se requer um vendedor técnico - Maria só vende parafusos e João só vende pregos. Observa-se então, uma dimensão de Vendedor (Maria e João), uma dimensão de Produto (Parafuso e Prego) e uma fato (Valor da Venda).

Como o cubo fornece a possibilidade de se analisar qualquer dimensão com qualquer métrica, será possível realizar uma análise sobre o valor das vendas de parafusos feitas por João e o valor das vendas de pregos feitas por Maria. O que resultará

em valores nulos. É importante discernir e ter essa consciência, pois você identificará isso no momento do mapeamento de requisitos e precisa saber como proceder nestes casos. Uma simples supressão ou filtro de valores em branco já resolveria.

Figura 11 – Exemplo de possibilidades de análises em um cubo.



Fonte: Visões e análise do cubo (TURBAN *et al.*, 2009, p. 119)

Lembre-se que o Cubo é um modelo de dados estrela que, por sua vez, é um conceito, logo pode ser aplicado em planilhas eletrônicas (baixa complexidade) ou, em arquiteturas de média e alta complexidade, pode ser desenvolvido em qualquer banco

de dados relacional, utilizar tecnologias próprias para isso, como por exemplo o *Oracle Essbase* ou *SQL Server Analysis Services*.

A estrutura de um cubo de dados, como já mencionada, pode utilizar tabelas de um banco de dados relacional, bastando implementar a arquitetura e modelagem corretas. Outra opção, normalmente utilizada em arquitetura de dados corporativa é utilizar tecnologias específicas (como já citado), porém deve-se atentar a estrutura desta, visto que existem tecnologias que trabalham utilizando armazenamento em disco e outras que trabalham utilizando memória.

O cubo em uma estrutura que envolve armazenamento, por exemplo, *Oracle Essbase* e *SQL Server Analysis Services Multidimensional*, necessitarão de uma infraestrutura que suporte a escala de volume de dados, de crescimento. Estas tecnologias armazenam em disco todos os dados referentes a todas correlações entre as dimensões e métricas, independentemente se faça ou não sentido.

Elas trabalham desta forma, pois visam a agilidade na consulta aos dados, visto que todos eles já estão disponíveis com os seus

resultados calculados, ou seja, a geração da informação será instantânea, muito rápida.

Já estruturas que utilizam tecnologias baseadas em memória (*in-memory*), não armazenam os dados, os cubos ficam conectados a sua fonte de origem e quando se realiza uma consulta, o cubo se utilizará de memória e processamento do computador para buscar a informação, calculá-la (se necessário) e apresentar na tela.

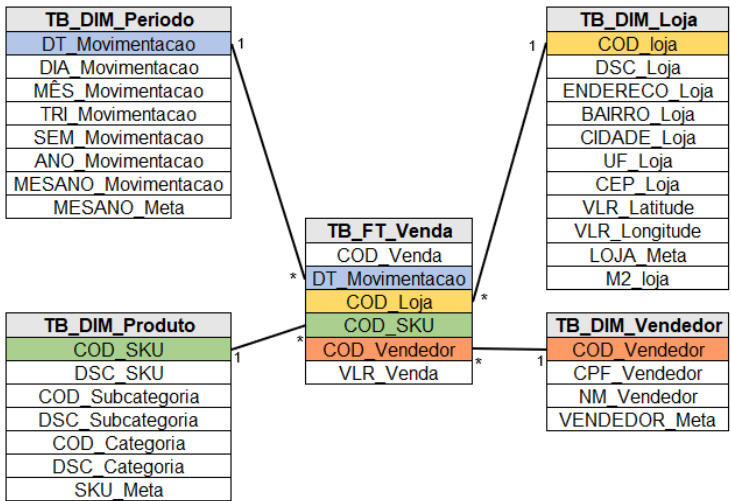
Todas as operações no cubo são realizadas com o suporte da memória. Neste caso, faz-se necessário que a sua estrutura garanta escalabilidade de memória que suporte o crescimento do volume de informações e transações. Atualmente, as grandes empresas optam por essa estrutura, utilizam serviços em nuvem, como o *Azure Analysis Services*.

Os cubos de dados oferecem a oportunidade de obter informações para suporte à decisão de maneira eficiente. A análise de cubo permite às pessoas realizar consultas através da busca de uma série de visualizações de relatórios, usando os recursos [...] como page-by, pivot, classificação, filtro e drill-up/drill-down, que permitem aos usuários fazer “visões parciais” de um cubo de dados ou analisar um cubo usando um simples clique do mouse. (TURBAN *et al.*, 2009, p. 118)

Como já mencionado anteriormente e reforçado neste momento, utilize o mapeamento de requisitos como seus principais insumos para o desenho do modelo de dados.

Vamos agora observar o nosso modelo de dados de exemplo (figura 26), utilizando somente os nomes das colunas e ilustrando as suas correlações.

Figura 12 – Modelo de dados estrela



Fonte: Autoria própria

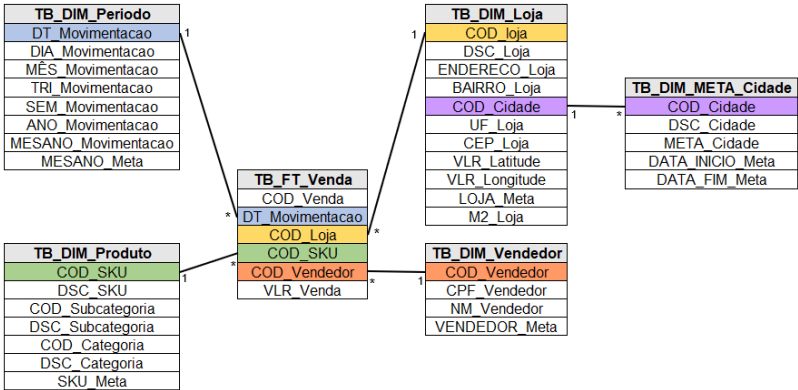
Temos alguns pontos a se observar com relação a este modelo de dados. O primeiro ponto a observar é que, como já explicado, um dimensão pode ser composta hierarquicamente

através de duas ou mais variáveis qualitativas, por exemplo na TB_Dim_Loja, que representa a dimensão Loja, é possível visualizar uma hierarquia de localidade (Endereço->Bairro->Cidade->UF), assim como na dimensão de Produto e na Período.

Importante lembrar que o modelo de dados é projetado, em concordância com os requisitos de negócio.

Imagine se houvesse um requisito que necessitasse da análise do atingimento das metas de venda por cidade, este modelo de dados (figura 26) sofreria uma variação, ficando da seguinte forma:

Figura 13 – Modelo de dados com meta por cidade



Fonte: Autoria própria

Este modelo é uma variação do estrela, chamado *snow flake*, em português floco de neve, que consiste em criar um relacionamento entre uma dimensão e uma tabela de atributo desta mesma dimensão.

Neste caso específico, como havia um requisito que demandava uma análise do atingimento das metas de vendas por cidade, que é diferente da meta da loja, foi necessária a criação de mais um tabela, como um atributo da dimensão loja, especificando a meta por cidade.

Isso já deixa exposto a importância do mapeamento dos Requisitos em sua solução de *business intelligence*, pois um pequeno detalhe pode alterar a forma como seu modelo de dados será projetado.

O **segundo ponto** a destacar está relacionado as métricas calculadas, que são as variáveis qualitativas que são criadas a partir dos requisitos, conforme estudamos no capítulo sobre métricas.

Imagine os requisitos:

- Atingimento de vendas: $(\text{Total de Venda}/\text{Meta}) \cdot 100$

- Variação ano sobre ano (*Year over Year*): $(\text{Vendas/m}^2 (\text{ano atual}) - \text{Vendas/m}^2 (\text{ano anterior})) / \text{Vendas/m}^2 (\text{ano anterior})$.

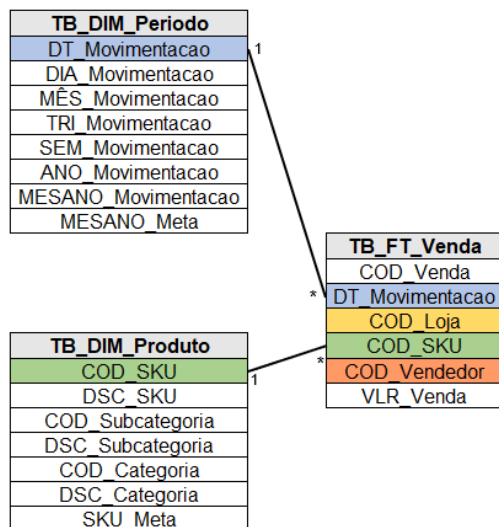
Como é possível observar no modelo de dados (figura 26), nós não as descrevemos na tabela fato, pois são métricas resultantes de cálculos utilizando métricas já existentes no modelo de dados.

Nesse caso, esses requisitos precisam estar mapeados, juntamente com a regra do cálculo, indicando quais as métricas que serão utilizadas a partir do modelo.

Se tivéssemos criado a medida $(\text{Total de Venda/Meta}) * 100$ dentro da tabela fato, teríamos muitas informações equivocadas, com análises que não fariam sentido, pois na tabela fato, encontra-se a correlação entre todas as dimensões e as metas são por dimensão, consequentemente este cálculo estaria errado.

Partindo para o **terceiro ponto**, falaremos sobre relacionamentos cardinais. Repare que em cada relação da dimensão com a tabela fato, existe uma linha conectando uma coluna da tabela de dimensão com uma coluna da tabela fato.

Figura 14 – Ilustração do Relacionamento entre tabelas



Fonte: Autoria própria

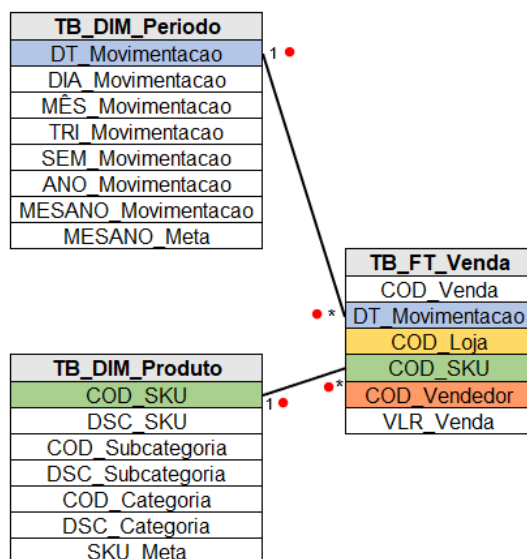
Perceba que existe uma correlação física entre a tabela fato e as tabelas de dimensão, através – do que chamamos – de chave de relacionamento. Na figura 28, as chaves de relacionamento estão conectadas e pintadas da mesma cor. Este relacionamento possibilita obter informações complementares, existentes somente na dimensão.

Por exemplo, para saber o total de vendas da categoria, que um produto pertence, utilizaremos a métrica VLR_Venda

(TB_FT_Vendas), e através da chave de relacionamento (COD_SKU), será possível obter a categoria, através da coluna DSC_Categoria (TB_Dim_Produto). Se o relacionamento não existisse, não seria possível analisar o total de vendas da categoria.

Ainda em referência aos relacionamentos, é possível observar que existe o número “1”, de um lado do relacionamento e o “*” do outro lado, conforme ilustra a imagem a seguir:

Figura 15 – Relacionamento cardinal



Fonte: Autoria própria

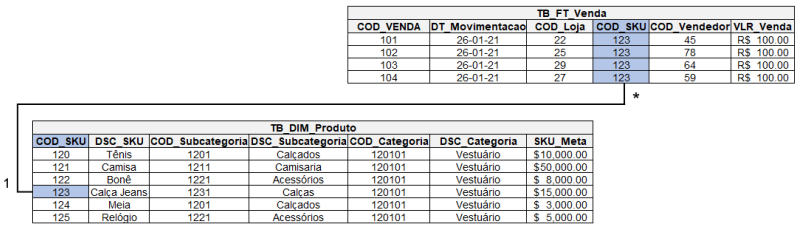
O nome que se dá a este procedimento é cardinalidade, criando assim um relacionamento cardinal, que pode ser de três tipos:

- 1:* - Leia “um para muitos”
- 1:1 – Leia “um para um”
- *.* - Leia “muitos para muitos”

O numeral “1” significa que na tabela não existe repetição da chave de relacionamento. Na tabela TB_DIM_Produto, o COD_SKU é único, nunca se repetirá.

O símbolo “*”, que significa “muitos”, informando que na tabela a chave do relacionamento se repete frequentemente. Na tabela TB_FT_Venda, o COD_SKU se repetirá.

Figura 16 – Relacionamento cardinal 1:* (um para muitos)



Fonte: Autoria própria

O tema acerca da modelagem de dados é mais amplo do que o que foi abordado aqui, porém o suficiente para fundamentar o conceito da modelagem de dados para *business intelligence*.

É possível observar, após toda a explicação sobre o modelo de dados, o que venho repetindo constantemente acerca da necessidade de se conhecer o negócio a fundo, bem como mapear os requisitos, identificar as perguntas que precisam de respostas. Somente assim, você será capaz de projetar um modelo de dados, por mais simplório que seja, identificando as correlações, as chaves de relacionamento e interpretar a cardinalidade.

Volto a repetir, sem o embasamento e profundidade acerca do negócio para o qual você pretende gerar informações e prover conhecimento para suportar tomadas de decisão, não será possível modelar o dado, por mais que você seja um especialista em modelagem de dados.

Referências

INMON, Willian H. *Building the Data Warehouse*. 4. ed. Indiana/EUA: Wiley Publishing, Inc., 2005.

KIMBALL, Ralph. *The data warehouse toolkit: the complete guide to dimensional modeling*. 2. ed. Nova Iorque/EUA: John Wiley and Sons, Inc., 2002.

TURBAN, Efraim *et al.* **Business Intelligence: um enfoque gerencial para a inteligência do negócio**. Porto Alegre: Bookman, 2009.