

TECNOLOGIA DA INFORMAÇÃO

Noções de Modelagem
Dimensional – Parte I



Livro Eletrônico



SUMÁRIO

Apresentação	4
Noções de Modelagem Dimensional – Parte I	5
1. Conceitos Básicos	5
1.1. <i>Data Warehouse (DW)</i>	5
2. Processo de <i>Data Warehousing</i>	12
2.1. <i>Data Warehouse (DW) x Visões</i>	13
3. <i>Data Mart (DM)</i>	14
4. <i>Data Mining (Mineração de Dados)</i>	15
5. Abordagens para o Projeto de DW	17
6. ETL (<i>Extract Transform Load – Extração, Transformação e Carga de Dados</i>) em DW21	
7. <i>Operational Data Storage (ODS) ou Staging Area (SA)</i>	26
7.1. Coleta, Tratamento, Armazenamento, Integração, Recuperação e Descarte de Dados	29
8. Projeto do <i>Data Warehouse</i>	31
8.1. Visão Geral	31
8.2. Tabela de Fatos	34
8.3. Classificação de Fatos	37
8.4. Cubo de Dados	39
9. Tabela de Dimensão	42
9.1. Medidas e Dimensões	43
10. Esquemas Multidimensionais	46
10.1. Esquema Estrela (Star Schema)	46
10.2. Esquema Floco de Neve (Snowflake)	50
11. Esquema Estrela x Esquema Floco de Neve	52
12. Constelação de Fatos	54
13. Sete Passos para Construir um DW	54
Resumo	64

Questões Comentadas em Aula	72
Questões de Concurso	79
Gabarito.....	116
Referências.....	117

APRESENTAÇÃO

Vamos lá, querido(a) amigo(a)!

Pra chegar até essa pontinha do iceberg chamada “sucesso” é preciso enfrentar uma longa jornada. Todo merecimento é fruto de várias tentativas e motivação pra não desistir dos nossos sonhos. Lembre-se que **você é capaz de chegar lá!**



Uma ótima semana. 

NOÇÕES DE MODELAGEM DIMENSIONAL – PARTE I

1. CONCEITOS BÁSICOS

1.1. DATA WAREHOUSE (DW)

Um **Data Warehouse** (**armazém de dados**, ou ainda **depósito de dados**), é um repositório de informações colhidas de várias origens, armazenadas sob um esquema unificado, em um único local, **que propõe sustentar a tomada de decisão com dados**.



Assim, uma das características fundamentais de um ambiente de data warehouse está em proporcionar um ambiente que permita realizar análise dos negócios de uma empresa com base nos dados por ela armazenados.

Para que serve?

- Para criar uma **visão única** e **centralizada** dos dados que estavam dispersos em diversos Bancos de Dados.
- Permite que usuários finais executem consultas, gerem relatórios e façam **análises**.

Características: Data Warehouse pode ser considerado como **uma coleção de dados orientada por assunto, integrada, não volátil, variante no tempo**, que dá apoio às decisões da administração.

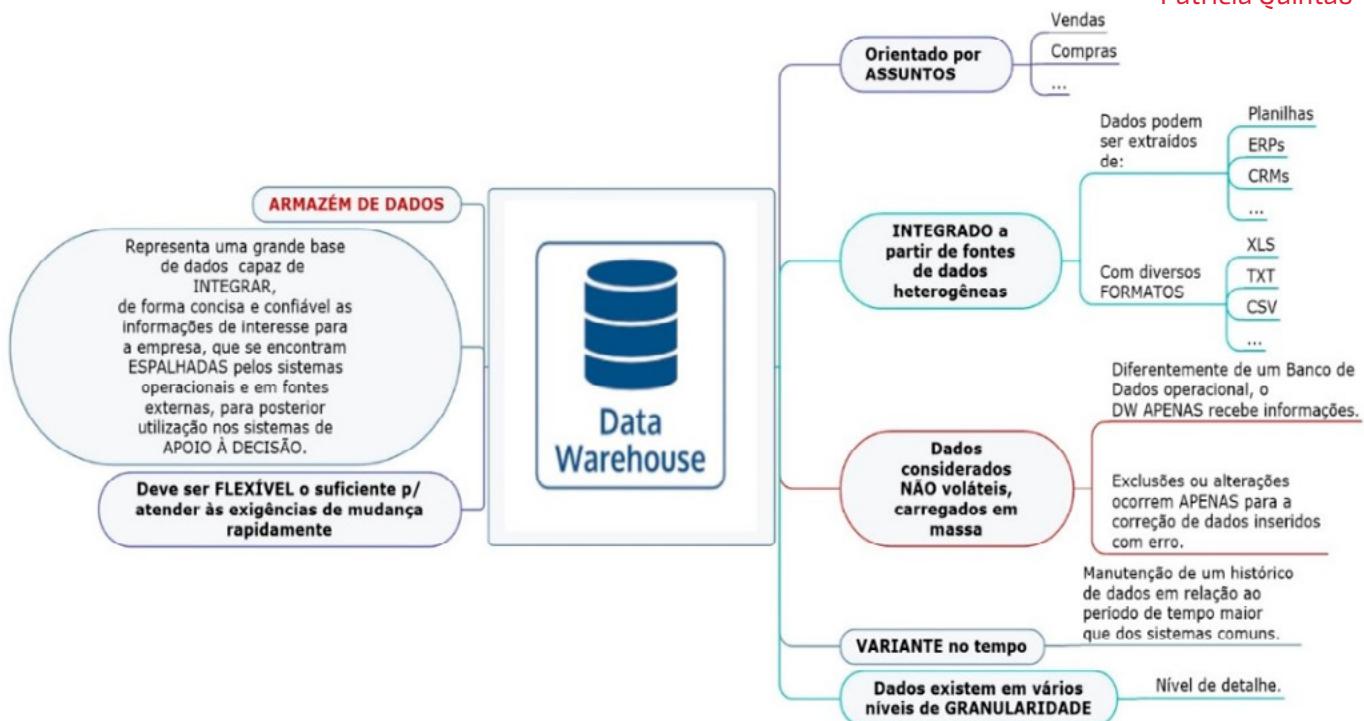


Figura. Características do Data Warehouse. Fonte: Quintão (2020)

Orientado a assunto: refere-se ao fato do Data Warehouse (DW) ser organizado conforme diferentes visões de negócio, ou seja, armazena informações sobre temas específicos importantes para o negócio da empresa. Exemplo: Vendas, Compras etc.



Figura. Visões de Negócio.

Integrado a partir de fontes de dados heterogêneas: refere-se à consistência de nomes, das unidades, das variáveis, etc., no sentido de que os dados foram transformados até um estado uniforme.

Por exemplo, considere sexo como um elemento de dado. Uma aplicação pode codificar sexo como M/F, outra como 1/0 e uma terceira como H/M.

Assim, **conforme os dados são inseridos no Data Warehouse, eles são convertidos para um mesmo padrão.**

Sexo é codificado apenas de uma forma.

Da mesma maneira, se um elemento de dado é medido em centímetros em uma aplicação, em polegadas em outra, ele será convertido para uma representação **única** ao ser colocado no Data Warehouse.

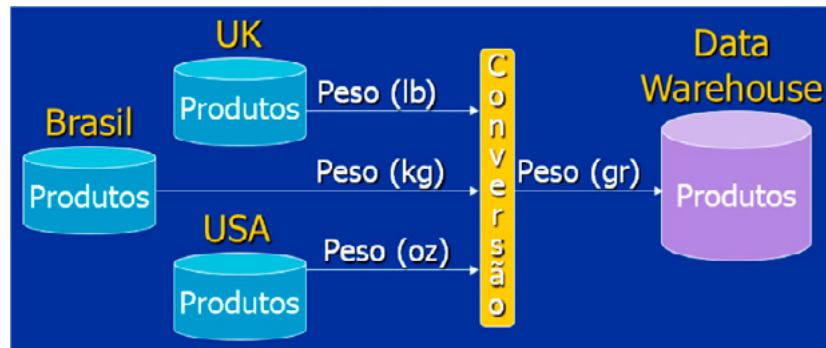


Figura. Exemplo de Integração de Dados de Fontes Heterogêneas.

Não volátil: os dados são sempre inseridos, nunca excluídos. Isso significa que **em um DW não existem alterações de dados, somente a carga inicial e as consultas posteriores**. No ambiente operacional, ao contrário, os dados são, em geral, **atualizados** registro a registro, em múltiplas transações.

Por definição, **os dados em um Data Warehouse não são voláteis**, ou seja, eles **não mudam, salvo quando é necessário fazer correções de dados previamente carregados**. Os dados estão disponíveis somente para leitura e não podem ser alterados.

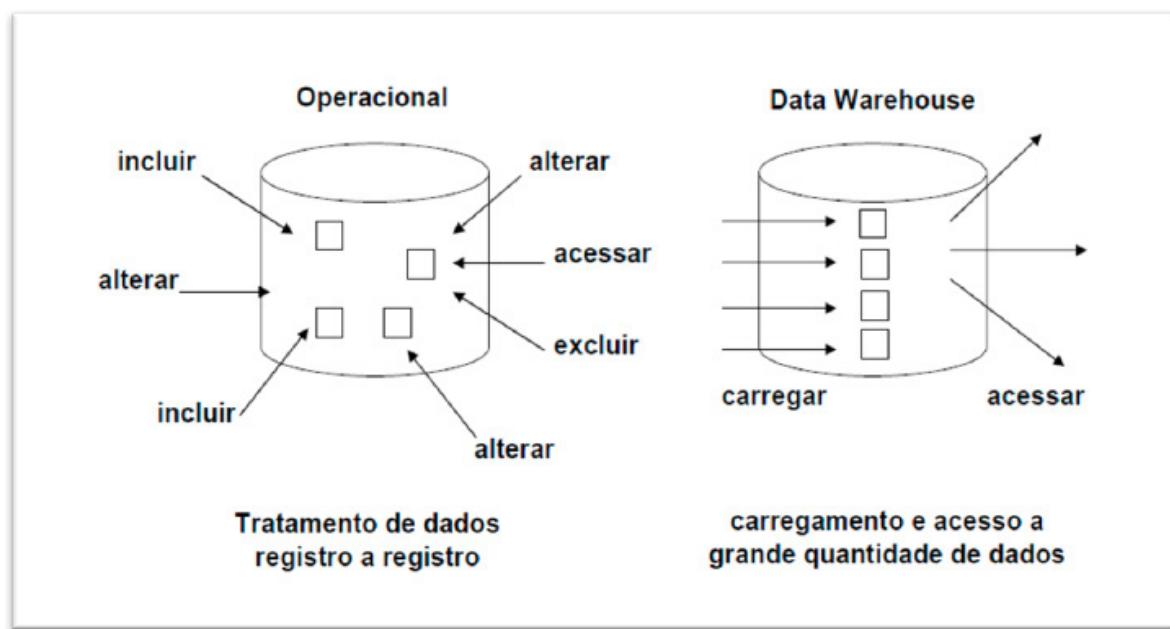


Figura. Carregamento e Acesso a Dados no DW.

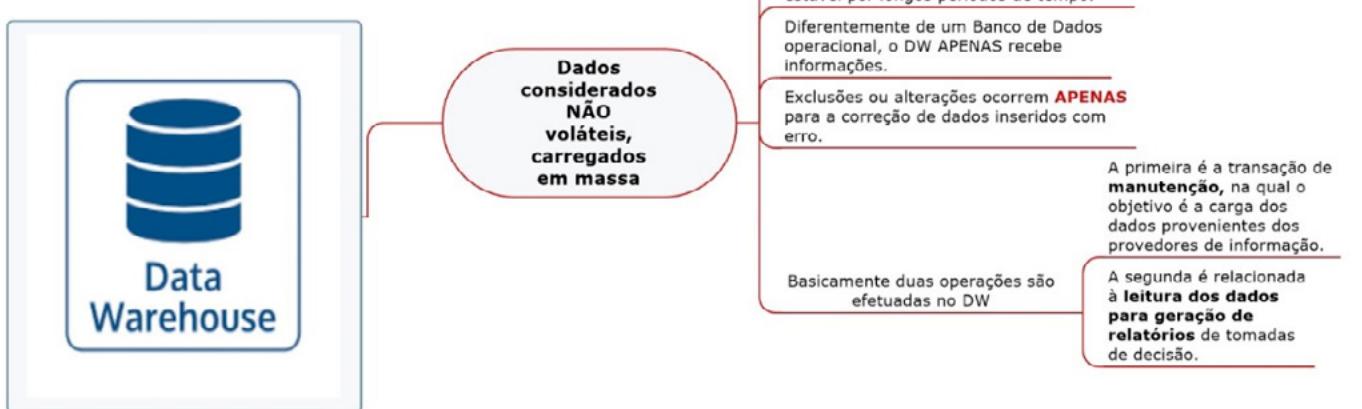


Figura. Dados Não Voláteis

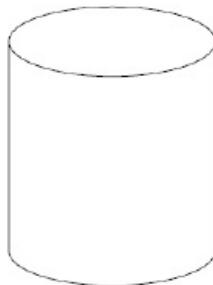
Variante no tempo: posições históricas das atividades no tempo.

Refere-se ao fato de o **dado em um Data Warehouse referir-se a algum momento específico**, significando que ele **NÃO É ATUALIZÁVEL**.

Enquanto que o **dado de produção é atualizado de acordo com mudanças de estado do objeto em questão**, refletindo, em geral, o estado do objeto no momento do acesso. Em um **Data Warehouse**, a cada ocorrência de uma mudança, uma nova entrada é criada, para marcar esta mudança.

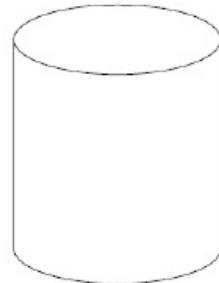
É Variante no Tempo

Operacional



- Horizonte de tempo – atual a 60-90 dias
- Atualização de registros
- A estrutura chave pode ou não conter um elemento de tempo

Data Warehouse



- Horizonte de tempo – 5-10 anos
- Instantâneos sofisticados dos dados
- A estrutura chave contém um elemento de tempo


Figura. Variação de Preços no Tempo

Granularidade de dados: refere-se ao **nível de sumarização** dos elementos e de detalhes disponíveis nos dados, considerado o **mais importante aspecto do projeto de um Data Warehouse**.

Um Data Warehouse é recomendado para armazenar **dados sumarizados** de toda a empresa para apoio à decisão e utilização de ferramentas **OLAP** (Online Analytical Processing).

DIRETO DO CONCURSO

001. (CESGRANRIO/BANCO DA AMAZÔNIA/TÉCNICO CIENTÍFICO/TECNOLOGIA DA INFORMAÇÃO/2018) Um Data Warehouse é recomendado para armazenar dados

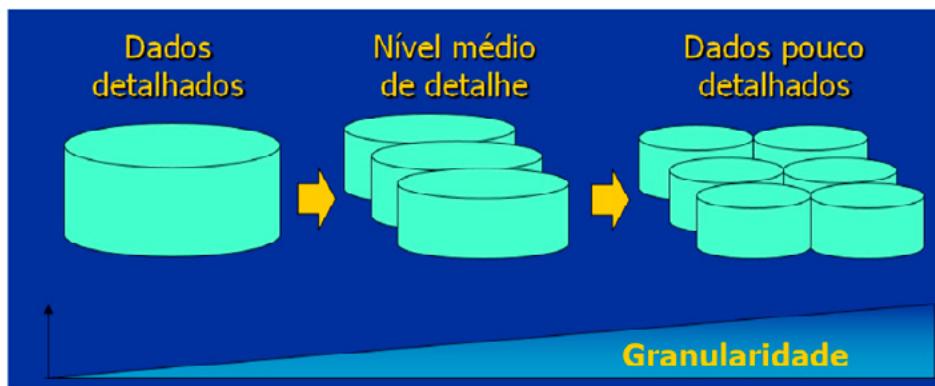
- sumarizados de um departamento.
- sumarizados de toda a empresa para apoio à decisão e utilização de ferramentas OLAP.
- detalhados de toda a empresa para apoio à decisão e utilização de ferramentas OLAP.
- detalhados gerados por sistemas de informação transacionais.
- históricos detalhados de todas as transações realizadas em um determinado período de tempo.



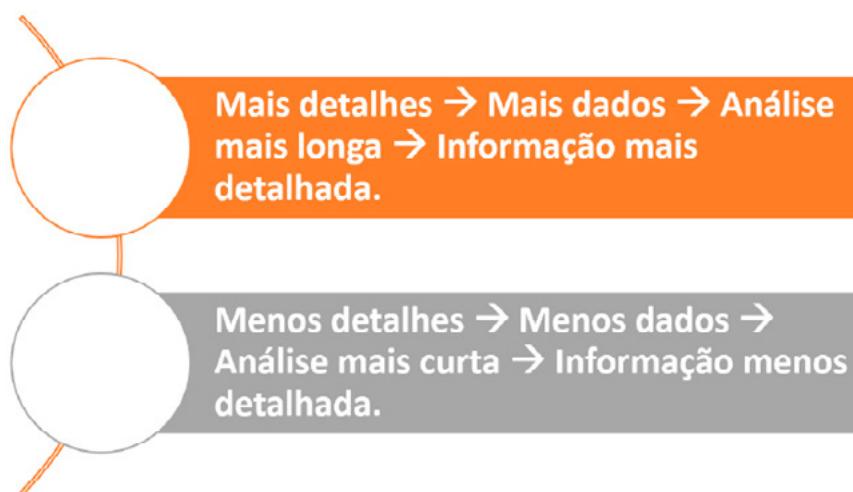
O DW, na definição de Barbieri (2001, p.49), é “um banco de dados, destinado a sistemas de **apoio à tomada de decisão** e cujos dados foram armazenados em estruturas lógicas dimensionais, possibilitando o seu processamento analítico por ferramentas especiais (OLAP e Mining)”. A partir desta interpretação, pode-se afirmar que o DW é a fonte para a exploração dos dados, enquanto o **OLAP (On-line Transaction Processing)** e o data mining são as técnicas utilizadas para explorar e investigar os dados. Um Data Warehouse é recomendado para armazenar dados **sumarizados** de toda a empresa para apoio à decisão e utilização de ferramentas OLAP.

Letra b.

Quanto menor a granularidade, mais detalhada é a informação disponível.



Definir a **granularidade** adequada é vital para que o DW atenda seus objetivos.



Para evitar que se perca informação são criados **vários níveis de granularidade**.



-
- | | |
|--|--|
| Em um nível de granularidade muito alto | o espaço em disco e o número de índices necessários se tornam bem menores; |
| | há, porém, uma diminuição da possibilidade de utilização dos dados para atender a consultas detalhadas. |
-

O Data Warehouse é um armazém centralizado de dados, ou seja, um banco de dados ou um agrupamento de bases de dados que contêm dados sobre os negócios organizados por assunto.

Por exemplo, uma indústria automotiva poderia ter um Data Warehouse com uma base de dados destinada a armazenar registros inerentes ao setor de Vendas. Poderia haver também uma outra base de dados que contivesse dados inerentes ao departamento de Produção de Automóveis.

A cada uma dessas bases de dados dá-se o nome de **Data Mart**, e ao agrupamento de todos esses Data Marts damos o nome de **Data Warehouse**.

Conforme visto, **Data Warehouse** é o processo de integração dos dados corporativos de uma empresa em um único repositório.

É um **ambiente de suporte à decisão** que alavanca dados armazenados em diferentes fontes e os organiza e entrega aos tomadores de decisões. Resumindo, é uma tecnologia de gestão e análise de dados.

DIRETO DO CONCURSO

002. (CESPE/DEPEN/AGENTE PENITENCIÁRIO FEDERAL/TECNOLOGIA DA INFORMAÇÃO/2015) Acerca de datawarehouse e datamining, julgue o item subsequente: [O datawarehouse possibilita a análise de grandes volumes de dados, que, por sua vez, permitem a realização de uma melhor análise de eventos futuros].



Um **Data Warehouse**, também chamado de **armazém de dados**, é um **repositório de informações colhidas de várias origens, armazenadas sob um esquema unificado, em um único local**. Quando reunidos, os dados são armazenados por muito tempo, permitindo o acesso a dados históricos. Ainda, o desenho da base de dados favorece os relatórios, a análise de grandes volumes de dados e a obtenção de informações estratégicas que podem **facilitar a tomada de decisão**.

O **Data Warehouse possibilita a análise de grandes volumes de dados**, coletados dos sistemas transacionais (OLTP - On-line Transaction Processing). São as chamadas **séries históricas que possibilitam uma melhor análise de eventos passados**, oferecendo **suporte às tomadas de decisões presentes e a previsão de eventos futuros**.

Errado.

2. PROCESSO DE DATA WAREHOUSING

O termo *Data Warehousing* pode ser descrito como uma **coleção de tecnologias de suporte à decisão, objetivando capacitar o trabalhador do conhecimento** (executivo, gerente, analista) **a tomar decisões melhores e mais rápidas**.

A figura seguinte apresenta uma **arquitetura típica de Data Warehousing**, mostrando todo o processo envolvido.

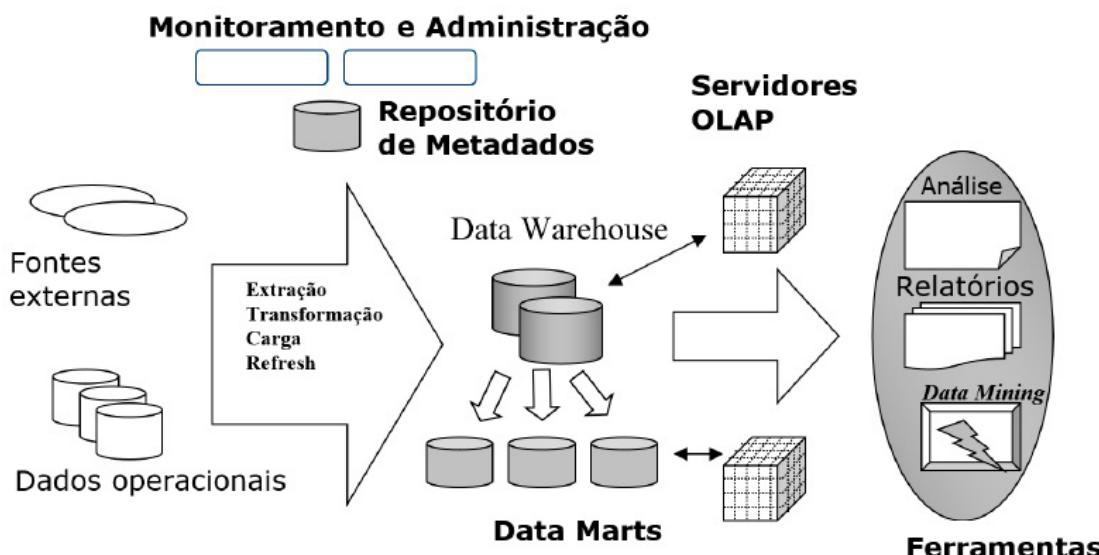


Figura. Arquitetura de Data Warehousing. Fonte: Chaudhuri e Dayal (1997)

Essa **arquitetura** inclui ferramentas para:

- **extração de dados de múltiplas bases de dados operacionais e de fontes externas** (por exemplo perfil dos clientes fornecida por consultores externos) usando **gateways** ou interfaces externas padrão suportadas pelo SGBD de suporte;
- **limpeza** dos dados para minimizar erros e preencher informações ausentes, quando possível;
- **transformação** para reconciliar erros semânticos;
- **integração** desses dados;
- **carga** dos dados no DW, que consiste na materialização das visões e no seu armazenamento no **Data Warehouse**; e para periodicamente executar o **refresh** do **Data Warehouse**, para que esse reflita as atualizações sofridas pela base operacional.

Além do Data Warehouse principal, podem existir diversos **Data Marts** (subconjuntos específicos focados em assuntos selecionados) **departamentais**.

Os dados contidos no Data Warehouse e nos Data Marts são armazenados e gerenciados por um ou mais **servidores de Data Warehouse**, que oferecem visões multidimensionais dos dados para uma variedade de ferramentas front-end: ferramentas de **consulta**, geradores de **relatórios**, ferramentas de **análises** específicas e ferramentas de **Data Mining**.

Finalmente, há um repositório para armazenamento e gerenciamento de dados de log e ferramentas para monitoramento e administração do sistema total.

2.1. DATA WAREHOUSE (DW) x VISÕES

Os *Data Warehouses* têm sido considerados por algumas pessoas como sendo uma extensão de funções e visões do banco de dados. No entanto, **visões fornecem somente um subconjunto das capacidades de Data Warehouses**.

Visões e Data Warehouses são similares nos seguintes aspectos:

- ambos têm dados extraídos de bancos de dados;
- são orientados ao assunto.

Data Warehouses são distintos de visões nos seguintes pontos:

- DW existem como **armazenamento persistente** ao invés de ser materializado sob demanda;
- DW não são usualmente relacionais, mas sim **multidimensionais**. Visões de um banco de dados relacional são relacionais.
- DW **podem ser indexados** para otimizar performance. Visões não podem ser indexadas independente dos bancos de dados utilizados.
- DW fornecem suporte específico de funcionalidade, visões não;
- DW fornecem **grande quantidade de dados integrados e frequentemente temporais**, geralmente mais do que está contido em um banco de dados, enquanto as visões são um extrato de um banco de dados.

DIRETO DO CONCURSO

003. (ESAF/MPOG/2008/ADAPTADA) Algumas pessoas têm considerado que os *Data Warehouses* são uma extensão de visões de banco de dados. Porém, as visões fornecem apenas um subconjunto das funções e das capacidades dos *Data Warehouses*. Com relação às diferenças e similaridades entre as visões e os *Data Warehouses*, é correto afirmar que tanto os *Data Warehouses* quanto as visões fornecem, frequentemente, grandes quantidades de dados integrados e temporais, geralmente mais do que é contido em um banco de dados.



Uma **View (Visão)** é uma tabela lógica, baseada em uma tabela ou em outra visão. Ela não possui dados próprios, é somente uma interface para a manipulação de um conjunto de dados. Ela pode ser utilizada para restringir o acesso a dados em uma tabela, facilitar consultas complexas e também otimizar o tempo dos desenvolvedores.

A View é uma maneira alternativa de observação de dados de uma ou mais tabelas, que compõem uma base de dados. Pode ser considerada como uma tabela virtual ou uma consulta

armazenada. Como exemplo de utilização de view, cita-se a restrição usuário x domínio controlando o acesso de um usuário específico a colunas de uma tabela.

Alguns benefícios da utilização das Views: economia de tempo com retrabalho; velocidade de acesso às informações; mascara a complexidade do banco de dados; organiza dados a serem exportados para outros aplicativos.

Um **Data warehouse** (ou armazém de dados) é um sistema de computação utilizado para armazenar informação relativa às atividades de uma organização em banco de dados, de forma consolidada. Ele possibilita a análise de grandes volumes de dados, coletados dos sistemas transacionais. Por definição, os dados em um Data Warehouse não são voláteis, ou seja, eles não mudam, são somente para leitura e não podem ser alterados.

Os *Data Warehouses* surgiram como conceito acadêmico na década de 80. Com o amadurecimento dos sistemas de informação empresariais, as necessidades de análise dos dados cresceram paralelamente. Como os sistemas transacionais não conseguiam cumprir a tarefa de análise com a simples geração de relatórios, os *Data Warehouses* são atualmente o núcleo dos sistemas de informações gerenciais e apoio a decisão das principais soluções de *Business intelligence* do mercado, devido a sua capacidade de sumarizar grandes volumes de dados e de possibilitar análises.

As ferramentas **OLAP (Online Analytical Processing)** têm como função a navegação nos dados de um Data Warehouse, possuindo uma estrutura adequada tanto para as pesquisas como para a apresentação das informações.

A assertiva, portanto, é falsa, tendo-se em vista que a **View** não tem como objetivo fornecer frequentemente grandes quantidades de dados integrados, e sim **fornecer um subconjunto dinâmico de dados (tabela virtual) a partir de uma ou mais tabelas**.

Observem ainda que o item afirma indevidamente que a quantidade de dados ofertada é maior do que em todo o banco de dados. Uma **View** não possui quantidade de dados maior do que o próprio banco de dados em que ela está inserida.

Errado.

3. DATA MART (DM)

O **Data Mart (DM)** nada mais é que um subconjunto de dados de um **Data Warehouse**, em que tipicamente desempenham o papel de um DW departamental, regional ou funcional.

- Segundo Ralph Kimball (2002)
 - Um Data Mart é um subconjunto lógico do Data Warehouse, normalmente visto como um Data Warehouse setorial.

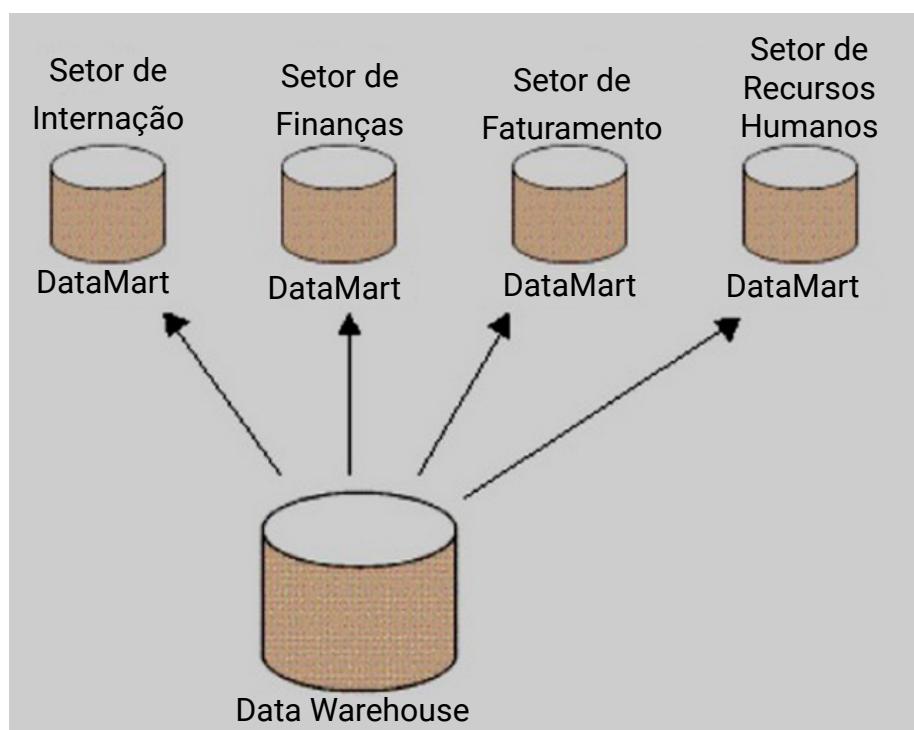


Figura. Data Warehouse e Data Marts

Alguns autores e especialistas dizem que o DW é uma evolução do DM que começou localizado e cresceu para atender um escopo maior.



4. DATA MINING (MINERAÇÃO DE DADOS)

É o processo de **análise de conjuntos de dados** que tem por objetivo a **descoberta de padrões interessantes** e que possam representar informações úteis.

A **mineração de dados** é um processo de negócio utilizado para explorar grandes quantidades de **dados** em busca de **padrões e regras significativas**.

O Data Mining apoia o **conhecimento indutivo**, que descobre novas regras e padrões nos dados fornecidos. (ELMASRI, NAVATHE, 2005).

Um **exemplo** clássico de utilização do Data Mining na área de *marketing* é o das **fraldas próximas das cervejas**. Uma das maiores redes de varejo dos Estados Unidos detectou, em seu gigantesco armazém de dados, que a venda de fraldas descartáveis estava associada à de cerveja. De uma maneira geral, os compradores eram homens, que saíam à noite para comprar fraldas e aproveitavam para levar algumas latinhas para casa. Os produtos foram postos lado a lado. **Resultado: a venda de fraldas e cervejas disparou.**

Veja mais: <http://exame.abril.com.br/revista-exame/edicoes/633/noticias/o-que-cerveja-tem-a-ver-com-fraldas-m0053931>.

Outro exemplo de Data Mining: imagine o histórico de vendas de uma empresa que está no mercado há muitos anos. Suponhamos que essa empresa desejasse saber qual a **tendência de vendas de um produto X para os próximos 10 anos**. O Data Mining possui uma série de etapas e técnicas apropriadas para responder essa questão.

Os processos de Data Mining são muito facilitados quando a empresa já possui seu Data Warehouse bem estruturado, por isso esses dois termos Data Warehouse e Data Mining caminham tão juntos.

As empresas comumente irão primeiramente amadurecer seus processos de organização dos dados sobre o negócio e agrupá-los por assunto, formando seus **Data Marts**. Em seguida, irão compor seu **Data Warehouse**, para após, iniciar os **processos de Data Mining** com a finalidade de encontrar algum conhecimento de valor em meio aos dados sobre o negócio.

5. ABORDAGENS PARA O PROJETO DE DW

No que se refere aos **métodos** de construção de **Data Warehouse (DW)**, existem formalmente **dois**:



Nesse caso é realizada a **modelagem integral do DW**, seguida pelas extrações de dados.

Considera que se deve desenvolver um DW completo e centralizado antes que partes dele, summarizadas, possam ser derivadas na forma de Data Marts.

A principal **vantagem** é a criação de um **modelo único**.

O **revés** fica por conta do **maior tempo de projeto**.



O **foco** é em uma área por vez, com o **crescimento gradual do DW**.

Considera que um DW possa ser composto a partir de Data Marts previamente desenvolvidos.

A **vantagem** é a obtenção de resultados a intervalos mais curtos, garantindo muitas vezes sustentação ao projeto.

A **desvantagem** é a **maior dificuldade de se consolidar informações entre as diversas áreas**.

Podem-se visualizar pela imagem seguinte as **duas formas de construção**, uma na qual o DW gera os DM (Data Mart) e outra em que os DM geram o DW.

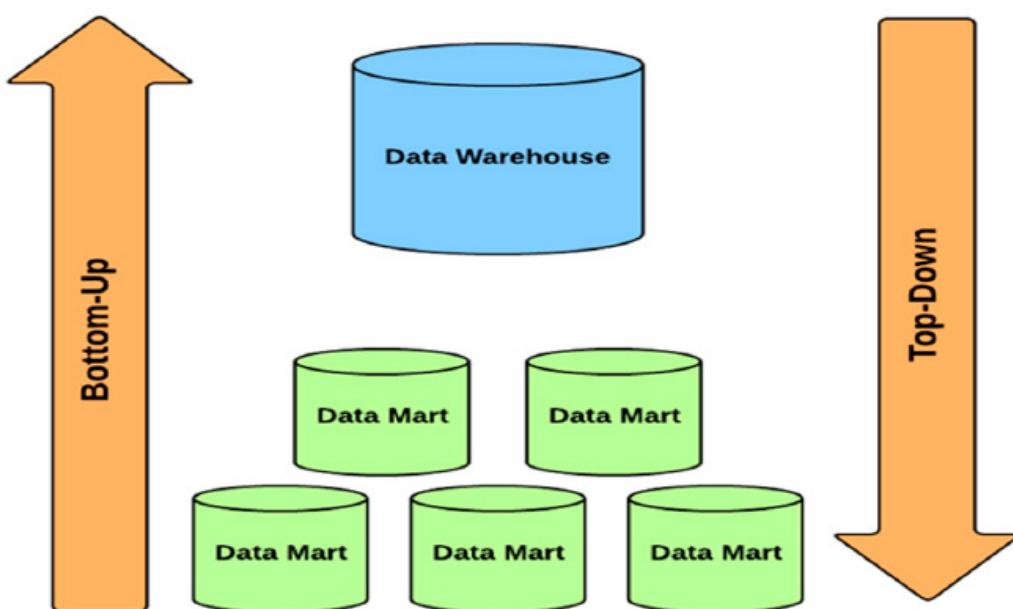


Figura. Abordagens Bottom-up e Top-Down.

Fonte: <https://canaltech.com.br/infra/a-abordagem-top-down-e-bottom-up-no-data-warehouse-21108/>

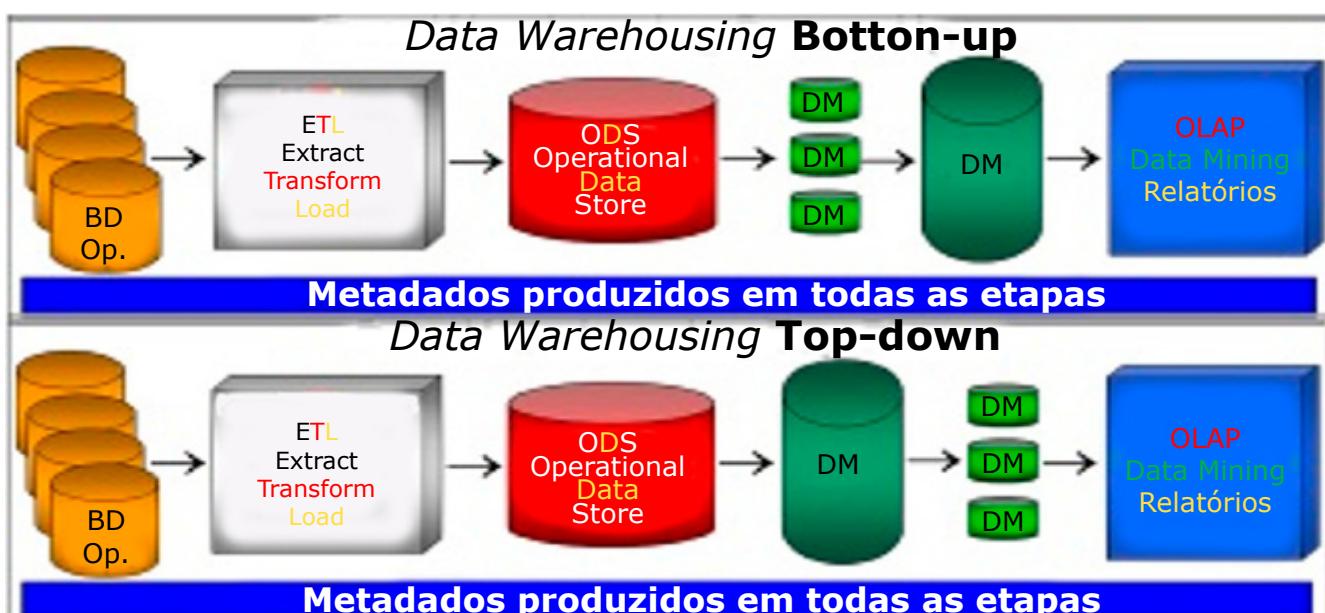


Figura. Exemplos das Duas Formas de Construção de um Data Warehouse

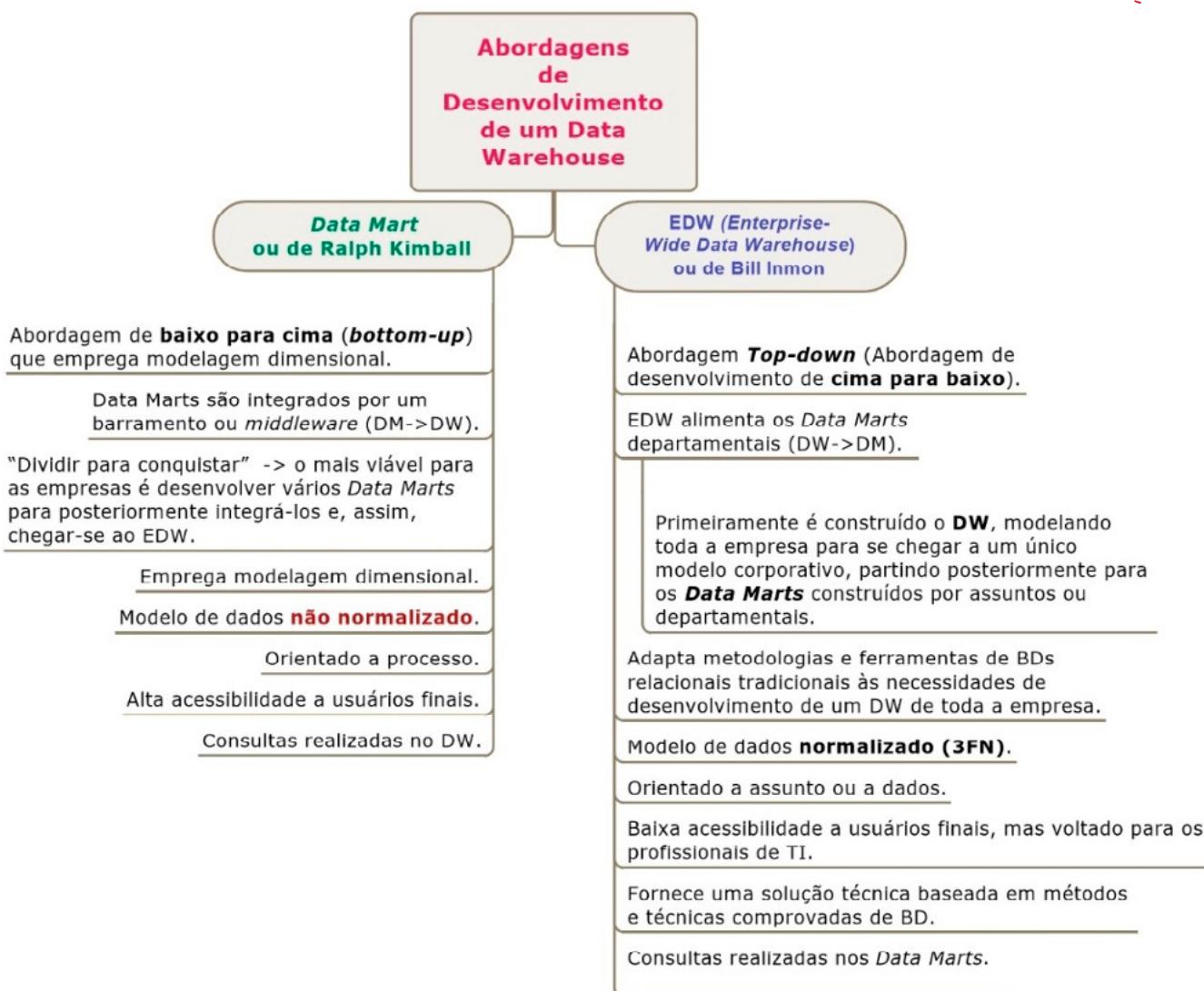


Figura. Abordagens de Desenvolvimento de um DW. Fonte: Quintão (2020)

Kimball concebeu uma arquitetura chamada **matriz de barramento** para auxiliar no desenvolvimento de um **Data Warehouse**.

De acordo com o autor, a matriz de barramento é uma **matriz de relação cruzada entre os processos de negócio e suas dimensões**, possibilitando visualizar aquelas dimensões que são compartilhadas entre os vários processos de negócio.

Observe pelo diagrama seguinte que, nas **linhas** são destacados os **processos de negócio** da organização. Já nas **colunas**, são dispostas as **dimensões** comuns a esses processos.

BUSINESS PROCESSES	COMMON DIMENSIONS						
	Date	Product	Warehouse	Store	Promotion	Customer	Employee
Issue Purchase Orders	X	X	X				
Receive Warehouse Deliveries	X	X	X				X
Warehouse Inventory	X	X	X				
Receive Store Deliveries	X	X	X	X			X
Store Inventory	X	X		X			
Retail Sales	X	X		X	X	X	X
Retail Sales Forecast	X	X		X			
Retail Promotion Tracking	X	X		X	X		
Customer Returns	X	X		X	X	X	X
Returns to Vendor	X	X		X			X
Frequent Shopper Sign-Ups	X			X		X	X

Fonte: Kimball (Group)

DIRETO DO CONCURSO

004. (VUNESP/TJ-SP/2012) Uma das técnicas utilizadas no projeto de um data warehouse corporativo consiste no uso da chamada matriz de barramento, na qual as linhas e colunas representam, respectivamente,

- a) cubos e medições.
- b) data staging e cubos.
- c) cardinalidades e hierarquias.
- d) dimensões e cardinalidades.
- e) processos de negócio e dimensões.



A matriz de barramento, proposta por Kimball, é uma matriz de relação cruzada entre **os processos de negócio e suas dimensões**, possibilitando visualizar aquelas dimensões que são compartilhadas entre os vários processos de negócio.

Essa matriz de barramento dispõe, nas suas **linhas**, os **processos de negócio** da organização e em suas **colunas** as **dimensões comuns**.

Letra e.

Para os **processos de construção de um Data Warehouse (DW)** é altamente importante a compreensão do negócio que envolve a empresa ou instituição em que se está desenvolvendo o trabalho.

6. ETL (Extract Transform Load – Extração, Transformação e Carga de Dados) em DW

O processo de **ETL** (do inglês Extract Transform and Load) é usado na criação do **DW (Data Warehouse)**.

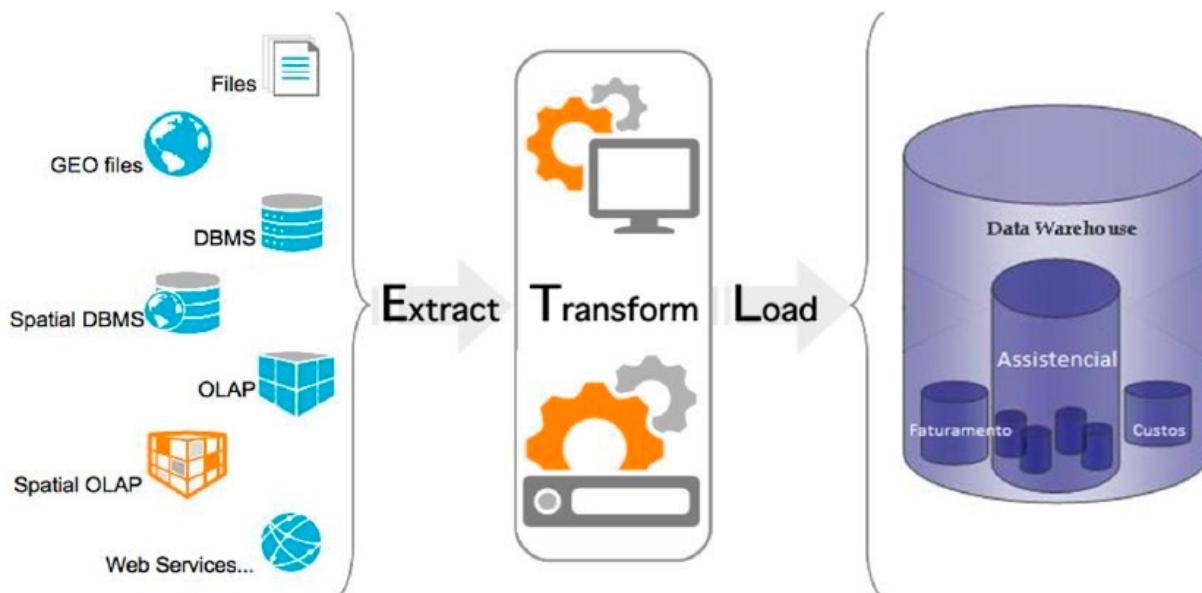


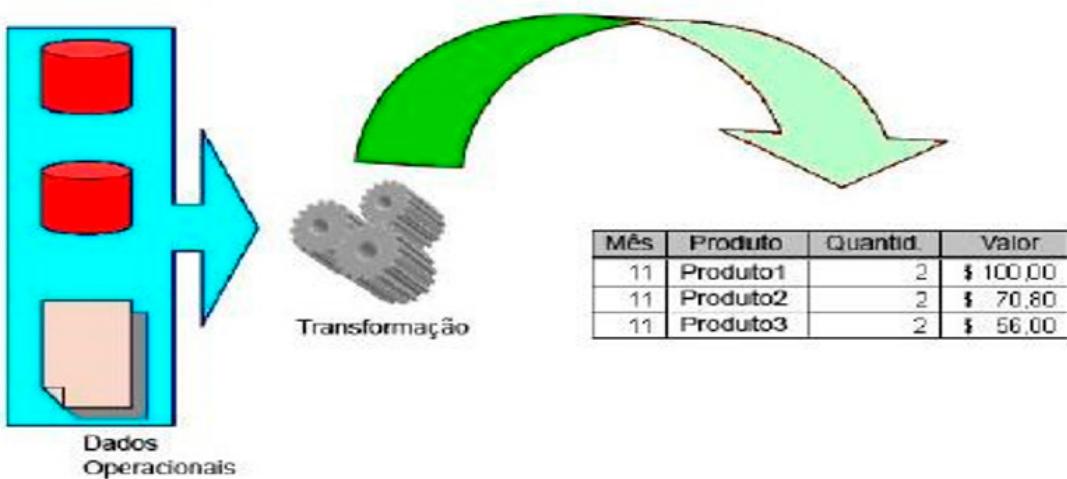
Figura. ETL. Fonte: http://laudosweb.hospitalorione.com.br/manual_html/etl_v4.1.0_painel.png



ETL
Os diferentes processos concentrados no conceito de extração, transformação e carga de dados em um Data Warehouse denomina-se ETL, em inglês Extract – Transform – Load.

O processo **inicial** faz a **extração (extract)** de dados das bases de dados transacionais, dados de sistemas ERP, dados locais, externos ou dados da Web. Conforme visto, busca obter os dados das diferentes fontes de dados para posterior transformação. Geralmente essa etapa envolve a conexão e acessos aos dados a partir de fontes como arquivos texto, XML, bancos de dados relacionais, logs e outros formatos.

Transformação (Transform): tem como objetivo transformar os dados modificando-os de modo que eles sejam adequados para o armazenamento no destino. O tipo de operação de transformação pode ser a validação, complementação, remoção, mudança de formato, consolidação ou outro tipo qualquer de manipulação de dados;



Load (Carga): basicamente armazena os dados transformados em uma fonte de dados de destino de acordo com um modelo que vai indicar como eles serão disponibilizados para as aplicações de um projeto de BI (*Business intelligence*).

Na etapa de extração, os dados relevantes ou de interesse são capturados de uma base de dados estruturada ou fontes diversas (exemplo redes sociais).

Em seguida, na etapa de transformação os dados são processados ou moldados para compor a informação na base do DW. Este processamento inclui atividades como normalização e geração de chave que identifica o registro na base.

Por fim, a etapa de carga consiste na inserção dos dados na base do Data Warehouse ou Data Mart.

Algumas das **ferramentas** conhecidas de ETL (Extract – Transform – Load) são:

- IBM InfoSphere DataStage;
- Informática Power Center;
- Business Objects Data Integrator;
- Data Transformation Services;
- Pentaho Data Integration;
- Oracle Data Integrator etc.

DIRETO DO CONCURSO

005. (FUNCAB/MDA/MINISTÉRIO DO DESENVOLVIMENTO AGRÁRIO/ANALISTA DE *Business intelligence*/2014)

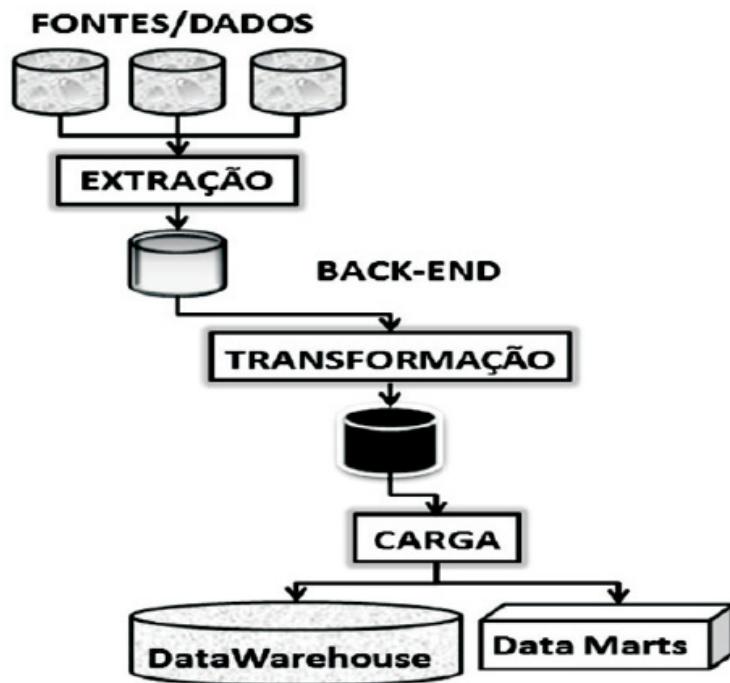


Figura. Modelo de organização para as próximas 3 questões.

A principal atividade de EXTRAÇÃO é:

- a) considerar logs de eventos e arquivos de controle.
- b) coletar dados das fontes externas transferindo-os para o ambiente de DW.
- c) carregar as dimensões considerando os tipos de hierarquias estáticas.
- d) tratar as inconsistências de dados resultantes da transcrição de dados.
- e) executar conversões de formatos para códigos geográficos dos países.



a) Errada. Considerar *logs* de eventos e arquivos de controle faz parte da etapa de **Extração**, mas não podemos considerá-lo como a principal atividade desta etapa.

b) Certa. A figura apresentada na questão destaca um modelo geral de como os dados são processados e armazenados em um Data Warehouse e em um Data Mart. Lembre-se de que este processo é conhecido como **ETL** (do inglês **Extract Transform and Load**), cujas etapas são as seguintes:

- O processo inicial é a **extração de dados das bases de dados transacionais**, dados de sistemas ERP, dados locais, de fontes externas ou web;
- **Após a extração, os dados devem ser transformados** para que seja possível a **carga dos dados** em um Data Warehouse ou Data Mart, dependendo do método de construção adotado.

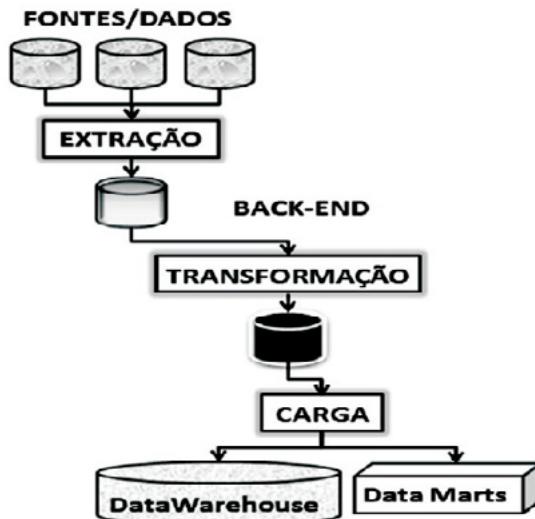
Baseado nas etapas de **Extração, Transformação e Carga**, aqui apresentadas, tem-se que a alternativa B é a resposta, pois é a que melhor descreve a principal atividade realizada durante a etapa da **Extração**.

c) Errada. Retrata uma atividade realizada na etapa de **Carga**.

d/e) Erradas. Estão relacionados a atividades da etapa de **Transformação**.

Letra b.

006. (FUNCAB/MDA/MINISTÉRIO DO DESENVOLVIMENTO AGRÁRIO /ANALISTA DE Business intelligence/2014)



A principal atividade de TRANSFORMAÇÃO é:

- a) aplicar regras aos dados extraídos para ajustá-los antes de serem carregados.
- b) analisar impactos das alternativas do código de aplicação no sistema fonte.
- c) avaliar aspectos de performance por meio do uso de paralelismo.
- d) obter dados de natureza estruturada e não estruturada.
- e) carregar tabelas Fato e fazer mapeamento das chaves.



O processo **ETL** (do inglês **Extract Transform and Load**) consiste na **extração** (leitura de dados de uma ou mais bases de dados), **transformação** (conversão dos dados extraídos de sua forma anterior para a forma em que precisa estar para que possa ser colocado em um Data Warehouse ou simplesmente em outro banco de dados), e **carga** (colocar os dados no Data Warehouse).

Vamos aos comentários das assertivas:

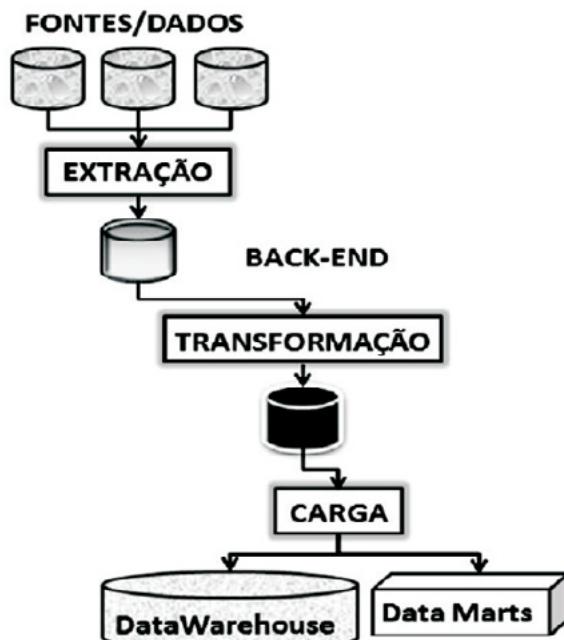
- a) Certa. No processo **ETL**, a etapa de **Transformação** vai manipular os dados para que eles fiquem prontos para serem carregados na fonte de dados de destino. Para isso, faz uso de regras predefinidas de formatação, validação ou alteração de dados para ajustar os dados antes de serem carregados.
- b) Errada. A análise de impacto e alternativas de código de aplicação, como apresentado nessa assertiva, não está definida em nenhuma das etapas do processo ETL.
- c) Errada. Também, a avaliação de aspectos de performance com paralelismo, indicada nessa assertiva, não faz parte das etapas do ETL.

d) Errada. A obtenção de dados de natureza estruturada e não estruturada é realizada na etapa de **Extração**.

e) Errada. A atividade de carregar **tabelas Fato** e fazer mapeamento das chaves é conduzida durante a etapa de **Carga**.

Letra a.

007. (FUNCAB/MDA/MINISTÉRIO DO DESENVOLVIMENTO AGRÁRIO /ANALISTA DE Business intelligence/2014)



A principal atividade de CARGA é:

- a) realizar a limpeza dos dados extraídos para melhorar a qualidade dos dados.
- b) definir a periodicidade da transcrição dos dados a serem coletados.
- c) otimizar a qualidade dos dados por meio da conversão de formatos.
- d) separar e concatenar dados visando eliminar inconsistências nos dados.
- e) estruturar e carregar os dados para o DW seguindo o modelo dimensional.



O **processo ETL** (do inglês **Extract Transform and Load**): processo de extração, tratamento e limpeza dos dados para inserção no DW.

- a/c/d) Erradas. As atividades aqui elencadas são realizadas na etapa de **Transformação**.
- b) Errada. A definição da periodicidade de dados geralmente é associada com a etapa de **Extração**.
- e) Certa. A etapa de **Carga** é a última do processo e sua responsabilidade é garantir que os dados já transformados serão armazenados corretamente no modelo indicado na fonte de dados de destino. Diferentes tipos de modelos podem ser utilizados para armazenar os dados

durante a etapa de Carga, porém o modelo multidimensional é o mais comum em sistemas de (*Business intelligence*).

Letra e.

7. OPERATIONAL DATA STORAGE (ODS) ou STAGING AREA (SA)

Staging Area representa a área de armazenamento intermediário criada a partir do processo de ETL. Auxilia na transição dos dados das fontes OLTP para o destino final no *Data Warehouse*.

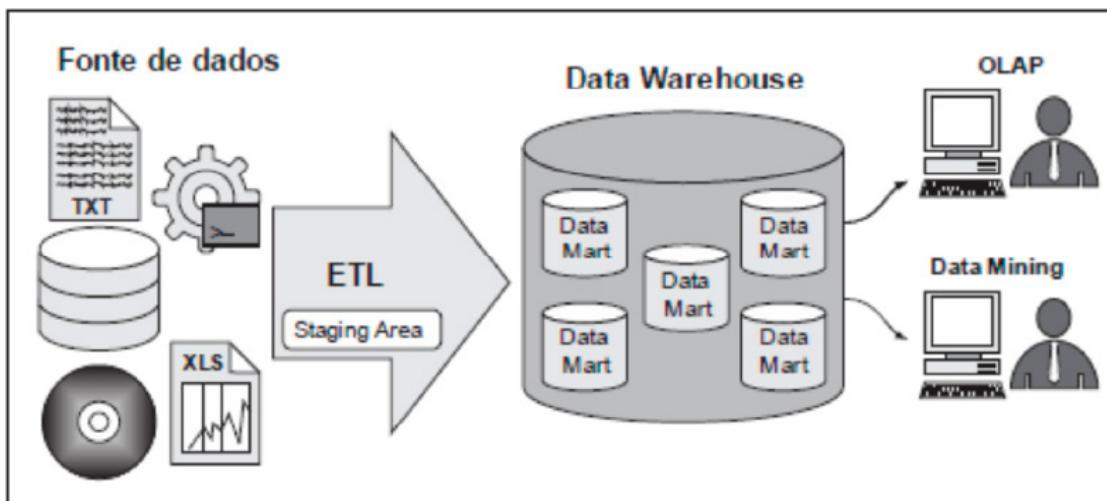


Figura. Arquitetura Geral de um Sistema de BI, com Destaque para Staging Area

No âmbito dos DWs, **Staging Area** representa um **armazenamento intermediário dos dados**, que facilita a integração dos dados do ambiente operacional antes de sua atualização no DW. Atualmente, alguns autores passaram a denominá-lo **Dynamic Data Storage (DDS)**.

Esse banco de dados intermediário é considerado uma **área de transição** em que se pode inserir os dados oriundos da extração e realizar as transformações necessárias, tudo isso antes de carregá-los no DW, que será o destino dos dados.

O uso dessa área de transição é útil para **não sobrecarregar** o ambiente de produção (os sistemas operacionais da empresa) e nem o próprio DW com as operações de transformação.

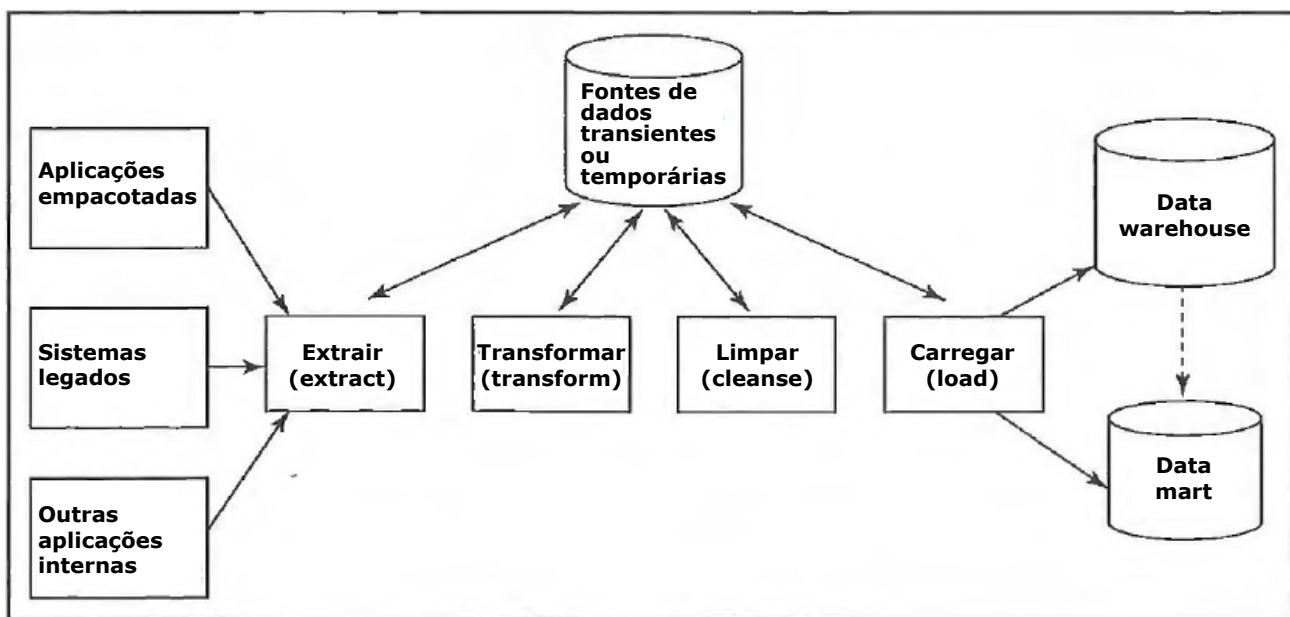
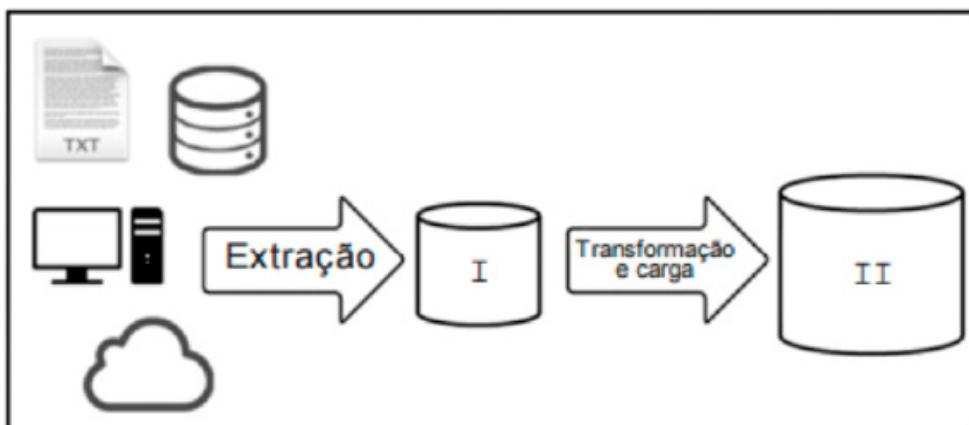


Figura. Staging Area

DIRETO DO CONCURSO

008. (FCC/AL-MS/TÉCNICO DE INFORMÁTICA/2016) Considere a figura abaixo.



No processo de ETL mostrado na figura, I e II correspondem, respectivamente, a

- OLTP e Data Warehouse.
- OLTP e Staging Area.
- Data Mart e Staging Area.
- Staging Area e OLTP.
- Staging Area e Data Warehouse.



O processo **ETL** (do inglês **Extract Transform and Load**) consiste na **extração** (leitura de dados de uma ou mais bases de dados), **transformação** (conversão dos dados extraídos de sua

forma anterior para a forma em que precisa estar para que possa ser colocado em um DW ou simplesmente outro banco de dados), e **carga** (colocar os dados no DW).

Os dados são carregados para uma área de preparação (conhecida como Staging Area (ou Operational Data Storage (ODS)) antes de serem transformados e carregados para o DW.

Assim, no processo de ETL mostrado na figura, I e II correspondem, respectivamente, a Staging Area (ou Operational Data Storage (ODS)) e Data Warehouse.

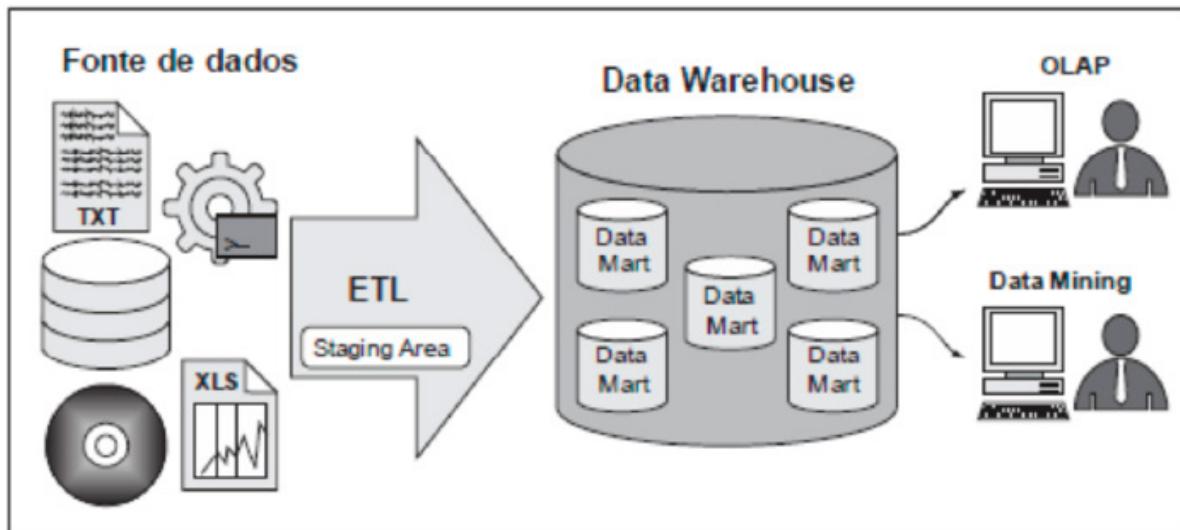


Figura. Ambiente de Data Warehouse

Letra e.

7.1. COLETA, TRATAMENTO, ARMAZENAMENTO, INTEGRAÇÃO, RECUPERAÇÃO E DESCARTE DE DADOS

Conforme vimos, o processo de **ETL** (do inglês **Extract Transform and Load**) envolve **coleta** (ou **extração**), **transformação** (nessa etapa, os dados são **tratados** para que se tornem mais corretos ou relevantes para os propósitos do BI) e **armazenamento** dos dados.

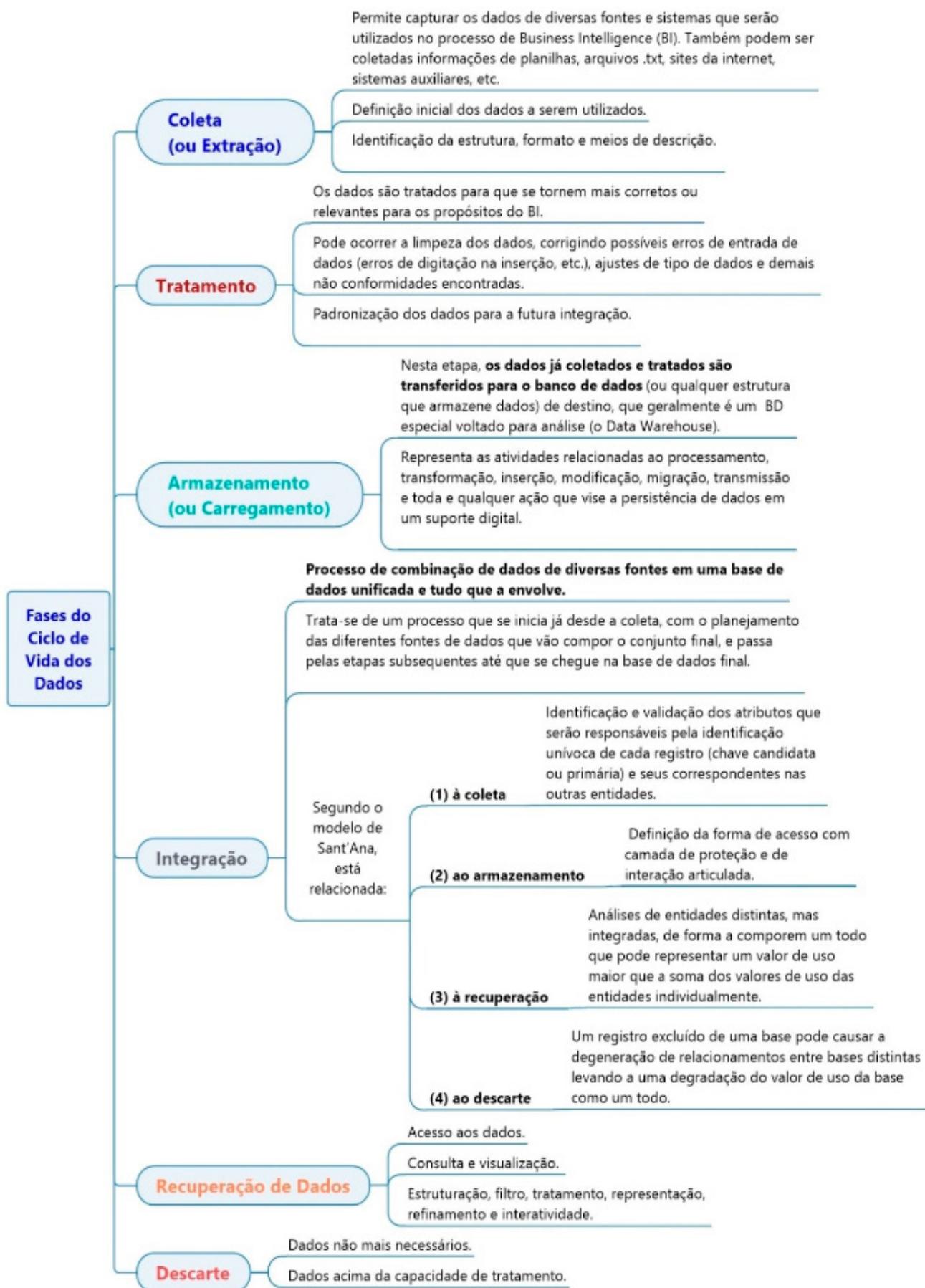
A **integração** é um processo de **combinação de dados de diversas fontes em uma base de dados unificada**. Esse processo já se inicia durante a coleta, com o planejamento das diferentes fontes de dados que vão compor o conjunto final, passando pelas etapas subsequentes até que se chegue na base de dados final.

Os dados já coletados, tratados, armazenados e integrados devem ser tornar acessíveis de modo a torná-los úteis à tomada de decisão. Nesse ponto, a literatura destaca dois sentidos para a etapa de **recuperação ou extração de dados**.

Ao tratar o **ciclo de vida dos dados**, a recuperação refere-se ao **acesso aos dados, por meio da consulta e visualização**, para subsidiar o processo de tomada de decisão.

Outra abordagem encontra na literatura envolve o **uso de um conjunto de técnicas e procedimentos para a extração de informações em dispositivos de armazenamento digital** (HD, Storages, dentre outros), **que não podem ser acessados de modo convencional pelo usuário ou pelo sistema** (Por exemplo, quando um dispositivo de armazenamento apresenta falhas. Diante desse fato, para recuperar os arquivos do dispositivo é preciso utilizar métodos não convencionais. Esses métodos não convencionais são conhecidos pelo termo **recuperação de dados** (ou **data recovery**, em inglês)).

A fase de **descarte** trata da reflexão sobre descarte de dados **que não são mais necessários ou que estejam acima da capacidade de tratá-los com eficiência para o sistema como um todo**.



DIRETO DO CONCURSO

009. (IADES/METRÔ-DF/TÉCNICO EM ELETRÔNICA/2014) É o conjunto de técnicas e procedimentos para a extração de informações em dispositivos de armazenamento digital, que não podem ser acessados de modo convencional pelo usuário ou pelo sistema. Com base no exposto, é correto afirmar que essas informações apresentam o conceito de

- a) recuperação de dados.
- b) backup corrompido.
- c) mineração de dados.
- d) backup interrompido.
- e) recuperação de dispositivos.



A **mineração de dados (Data Mining)** corresponde a processos usados para **explorar e analisar grandes volumes de dados, visando identificar padrões, previsões, erros e associações** etc. Esse conceito, geralmente, está associado ao **aprendizado de máquina**, que é uma área da inteligência artificial que desenvolve algoritmos capazes de fazer com que o computador **aprenda** a partir do passado, usando dados de eventos que já aconteceram. Este aprendizado é capaz de identificar padrões que dificilmente seriam identificados por meio de técnicas tradicionais de análise de dados, como filtros, junções, pivôs ou agrupamentos.

O conceito de **recuperação de dados** (Ou **Data Recovery**, em inglês) envolve o **uso de um conjunto de técnicas e procedimentos para a extração de informações em dispositivos de armazenamento digital** (HD, Storages, dentre outros), **que não podem ser acessados de modo convencional pelo usuário ou pelo sistema** (Por exemplo, quando um dispositivo de armazenamento apresenta falhas. Diante desse fato, para recuperar os arquivos do dispositivo é preciso utilizar métodos não convencionais. Esses métodos não convencionais são conhecidos pelo termo **recuperação de dados**).

Letra a.

8. PROJETO DO *DATA WAREHOUSE*

8.1. VISÃO GERAL

A **modelagem multidimensional** (alguns autores a chamam de **modelagem dimensional**) é a técnica de projeto mais frequentemente utilizada para a construção de um Data Warehouse.

O objetivo é buscar um **padrão de apresentação de dados** que seja facilmente visualizado pelo usuário final e que possua um bom desempenho para consultas.

Quando falamos em modelagem multidimensional, estamos nos referindo à definição de um **modelo que se destina à análise de dados**.

No que diz respeito à **análise de dados**, o que se espera do modelo de dados encontra-se listado a seguir:

- seja uma **representação simples** do modelo de negócios estudado;
- seja um **modelo físico de fácil interpretação**, de modo que usuários sem treinamento formal em Tecnologia da Informação (TI) possam entendê-lo;
- **facilite a implementação física do modelo** de modo a maximizar performance das consultas aos dados.

Portanto, **no modelo multidimensional**, deixamos de focar na **coleta de dados** para nos ocuparmos com a **consulta aos dados**. **E esta é uma mudança radical de foco!**

DIRETO DO CONCURSO

010. (COPEVE UFAL/MPE-AL/ANALISTA DO MINISTÉRIO PÚBLICO/2012) Um modelo de banco de dados multidimensional está mais fortemente relacionado com:

- a) data warehouse.
- b) modelo relacional.
- c) bancos hierárquicos.
- d) modelo em 3 camadas.
- e) banco de dados distribuídos.



O modelo de dados denominado “multidimensional” está mais fortemente relacionado com o **Data Warehouse, uma coleção de dados orientados por assuntos, integrados, variáveis com o tempo e não voláteis, para dar suporte ao processo de tomada de decisão com dados**.

Letra a.

011. (FUNCAB/MDA/ADMINISTRADOR DE BANCO DE DADOS/2014) O modelo de dados denominado “multidimensional” se aplica para banco de dados com a tecnologia:

- a) relacional.
- b) hierárquica.
- c) datamining.
- d) distribuída
- e) data warehouse.



O modelo de banco de dados multidimensional se aplica para banco de dados com a tecnologia: **Data Warehouse (DW)**.

A modelagem dimensional oferece diversas maneiras de realizar consultas de dados em um DW.
Letra e.

012. (CESPE/SE-DF/ANALISTA DE GESTÃO EDUCACIONAL/TECNOLOGIA DA INFORMAÇÃO/2017) Com relação aos conceitos de modelagem multidimensional de dados para inteligência computacional, julgue o seguinte item. [Diferentemente da estrutura relacional, a estrutura multidimensional oferece baixa redundância de dados e suporte a normalização até a segunda forma normal].



A **estrutura multidimensional** oferece alta redundância.

Modelagem Multidimensional

- Desnormalizado -> Alta redundância.

Modelagem Relacional

- Normalizado -> Baixa redundância.

Errado.

A finalidade de **bases de dados multidimensionais**, conforme destaca Nardi (2007), é **fornecer subsídio para realização de análises**. Para tanto, sua arquitetura e até mesmo a terminologia empregada são distintas das utilizadas para bancos de dados transacionais.

Assim, o fato de existirem diversas **informações a serem cruzadas (dimensões)** permite a realização de pesquisas tais como a ilustrada na figura seguinte:

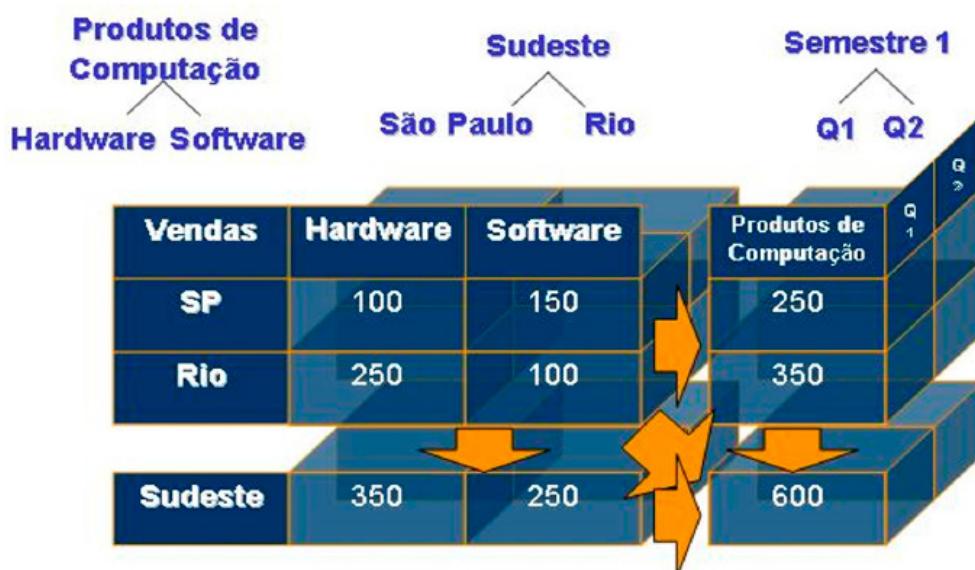


Figura. Exemplo de Pesquisa Multidimensional. Fonte: Nardi (2007)

Alguns **conceitos importantes** dentro da modelagem multidimensional estão listados a seguir:

8.2. TABELA DE FATOS

A **tabela de fatos** é a tabela primária do modelo dimensional e contém os valores do negócio que deseja analisar. **Cada tabela de fatos contém as chaves externas que se relacionam com suas respectivas tabelas de dimensões** e as colunas com os valores que serão analisados.

	Exemplos de Fatos <ul style="list-style-type: none"> ✓ Em um hospital: admissão de pacientes ✓ Em uma operadora telefônica: Tráfego telefônico
	Um fato é um conceito de interesse primário para o processo de tomada de decisões e corresponde aos eventos que ocorrem de forma dinâmica no negócio da empresa.

Fonte: Microsoft (2011)

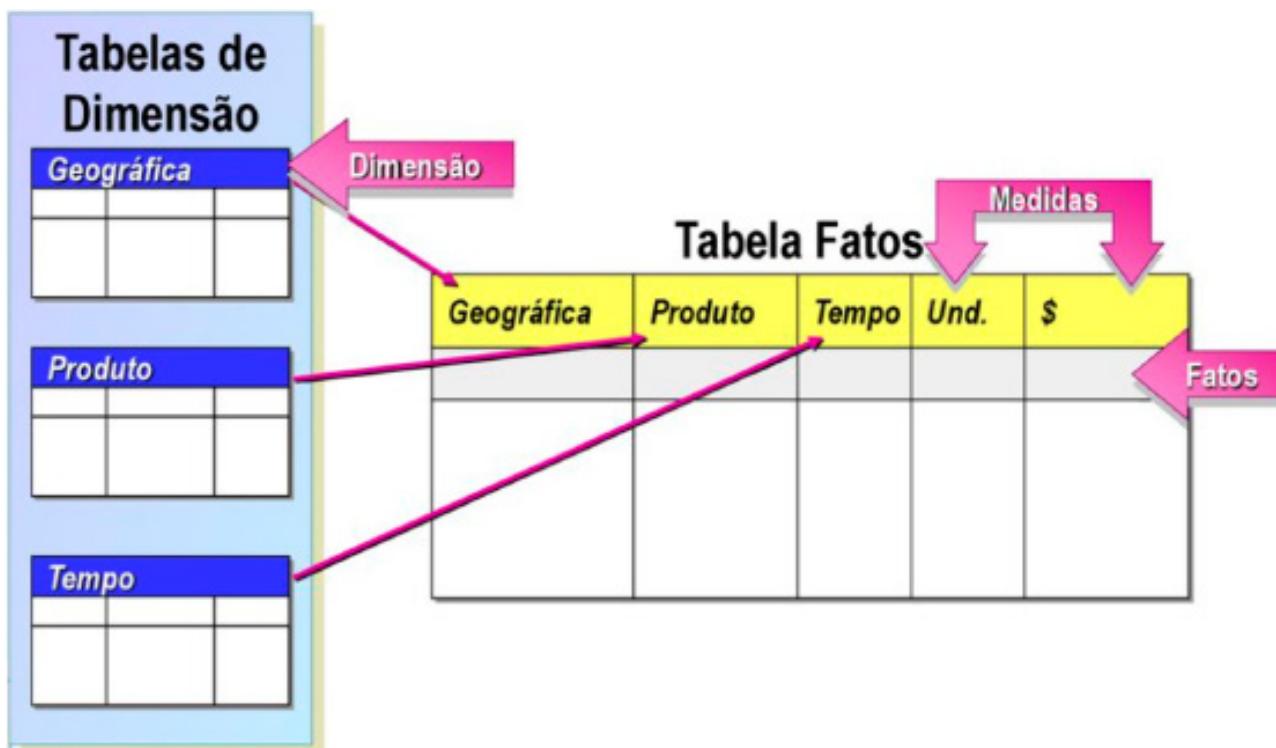


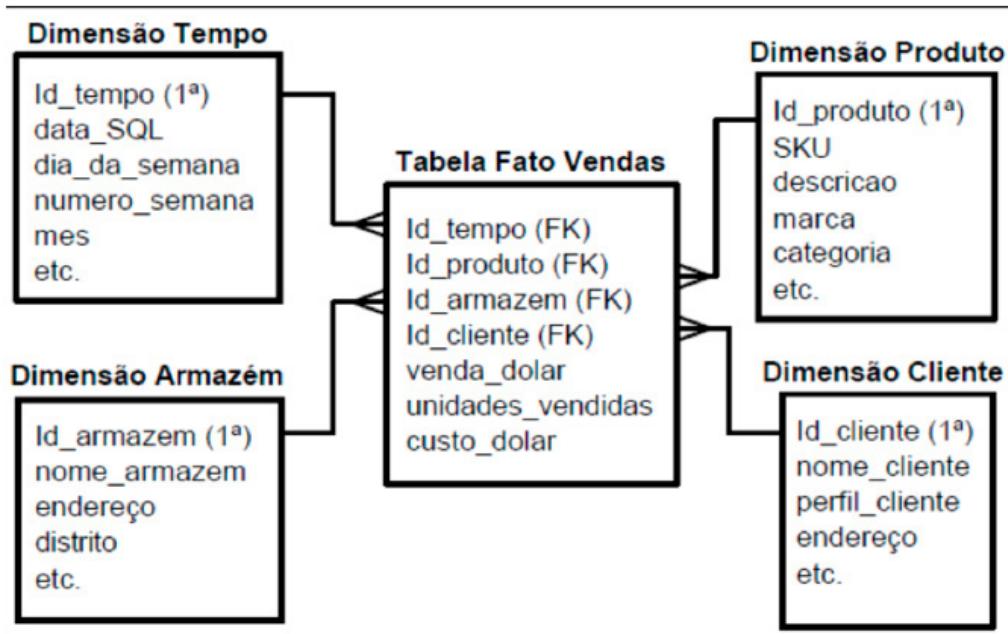
Figura. Componentes de Modelagem Dimensional. Fonte: <https://slideplayer.com.br/slide/14585808/>

As **tabelas fatos** indicam o **nível de granularidade das tabelas dimensões**. Elas orientam a criação das **tabelas dimensão**. Contudo, observe que os dados devem ser inseridos primeiramente nas **dimensões** para, em seguida, serem inseridos na **tabela fato**. Isso vai garantir a integridade referencial dos dados na **tabela fato**.

Cada atributo que compõe a chave primária da **tabela fato** também é uma chave estrangeira para uma das **tabelas dimensão**.

A **tabela fato** contém uma chave primária composta. Essa é formada pela composição de atributos que são chaves estrangeiras, cada um apontando para sua respectiva **tabela dimensão**.

Veja essa figura:



Fonte: http://2.bp.blogspot.com/-6sMFt0xK3Cw/Ty6UH15Zi9I/AAAAAAAABJU/71TfY-TiDlo/s1600/Modelo_estrela.png

Em uma **tabela de fatos**, uma linha corresponde a uma **medição**. **Todas as medições em uma tabela de fatos devem estar alinhadas na mesma granularidade** (KIMBALL; ROSS, 2002, p. 21).

DIRETO DO CONCURSO

013. (FAURGS/HCPA/ANALISTA DE TI- BANCO DE DADOS/2016) Em um modelo dimensional, a tabela fatos armazena:

- estatísticas sobre os metadados.
- as restrições de domínio do negócio.
- descrições textuais das dimensões.
- medições numéricas do negócio.
- o tempo de processamento das transações.



Em um modelo dimensional, a **tabela fatos** armazena medições numéricas de negócio.

Letra d.

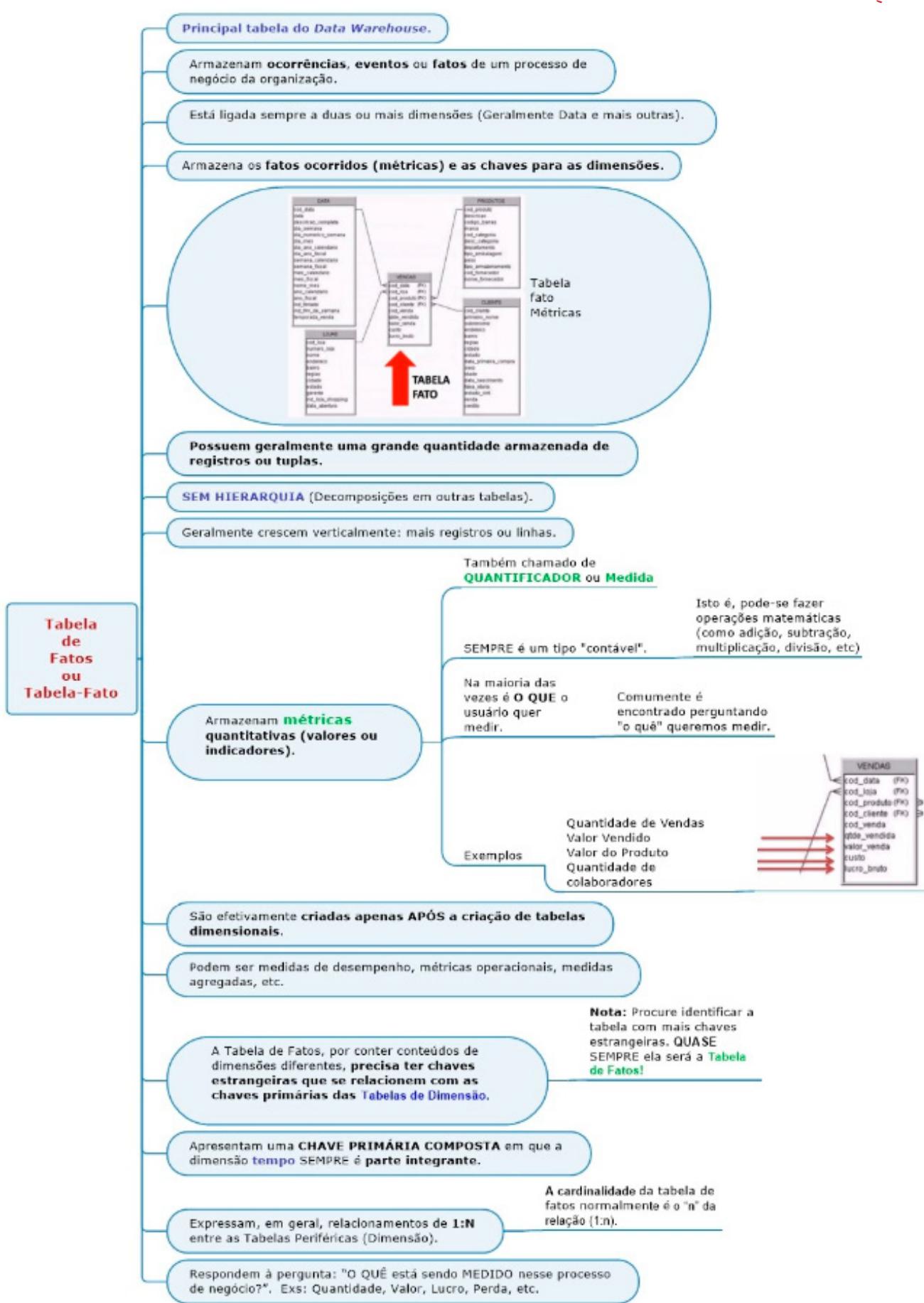


Figura. Tabela de Fatos. Fonte: Quintão (2020)

O conteúdo deste livro eletrônico é licenciado para MARIO LUIS DE SOUZA - 41250799864, vedada, por quaisquer meios e a qualquer título, a sua reprodução, cópia, divulgação ou distribuição, sujeitando-se aos infratores à responsabilização civil e criminal.

Um aspecto importante é que a tabela de fatos deve **representar uma unidade do processo do negócio, não devendo misturar assuntos diferentes numa mesma tabela de fatos.**

Os atributos mais comuns em uma **tabela de fatos** são valores numéricos. Estes valores são, em sua maioria, aditivos.

As **métricas aditivas** são as que permitem operações como adição, subtração e média de valores por todas as dimensões, em quaisquer combinações de registros, como “total de itens vendidos” por combinação de data, produto e loja. Métricas aditivas são importantes porque normalmente as aplicações de Data Warehouse não retornam uma linha da tabela de fatos, mas sim centenas, milhares e até milhões (HOKAMA et al, 2004, p. 36).

As **métricas não aditivas** são valores que **não podem ser manipulados livremente, como valores percentuais ou relativos**. Exemplos de métricas não aditivas são preço de custo e preço de venda de um produto em uma venda.

Já as **métricas semiaditivas** são valores que **não podem ser somados em todas as dimensões**. Por exemplo: numa tabela com o registro diário do saldo bancário dos clientes de uma agência, não faz sentido somar os saldos bancários diários de um cliente durante um mês, mas pode-se somar os saldos de todos os clientes de uma agência em determinada data.

Fonte: <http://meusite.mackenzie.com.br/rogerio/tgi/2004ModelagemDW.pdf>

8.3. CLASSIFICAÇÃO DE FATOS

Os **fatos** podem ser classificados em **transações individuais**, em **snapshots** e em **linhas de itens** (KIMBALL et al, 1998).

As **transações individuais** representam uma transação completa, normalmente, apresentam uma estrutura muito simples, possuem um campo acumulado que contém o valor total de uma transação.

Os **fatos snapshots** representam medidas de atividades extraídas em tempo determinado, como, por exemplo, final do dia ou final do mês, com medições acumuladas durante esse período.

Os **fatos do tipo linhas de itens** são aqueles que **representam exatamente uma linha de item**, são armazenados no nível mais detalhado dos fatos, como, por exemplo, itens de pedido, itens de entrega e itens de apólice de seguro.

Vamos agora falar dos **tipos de tabelas fatos**. Existem na literatura **seis tipos de fatos**:

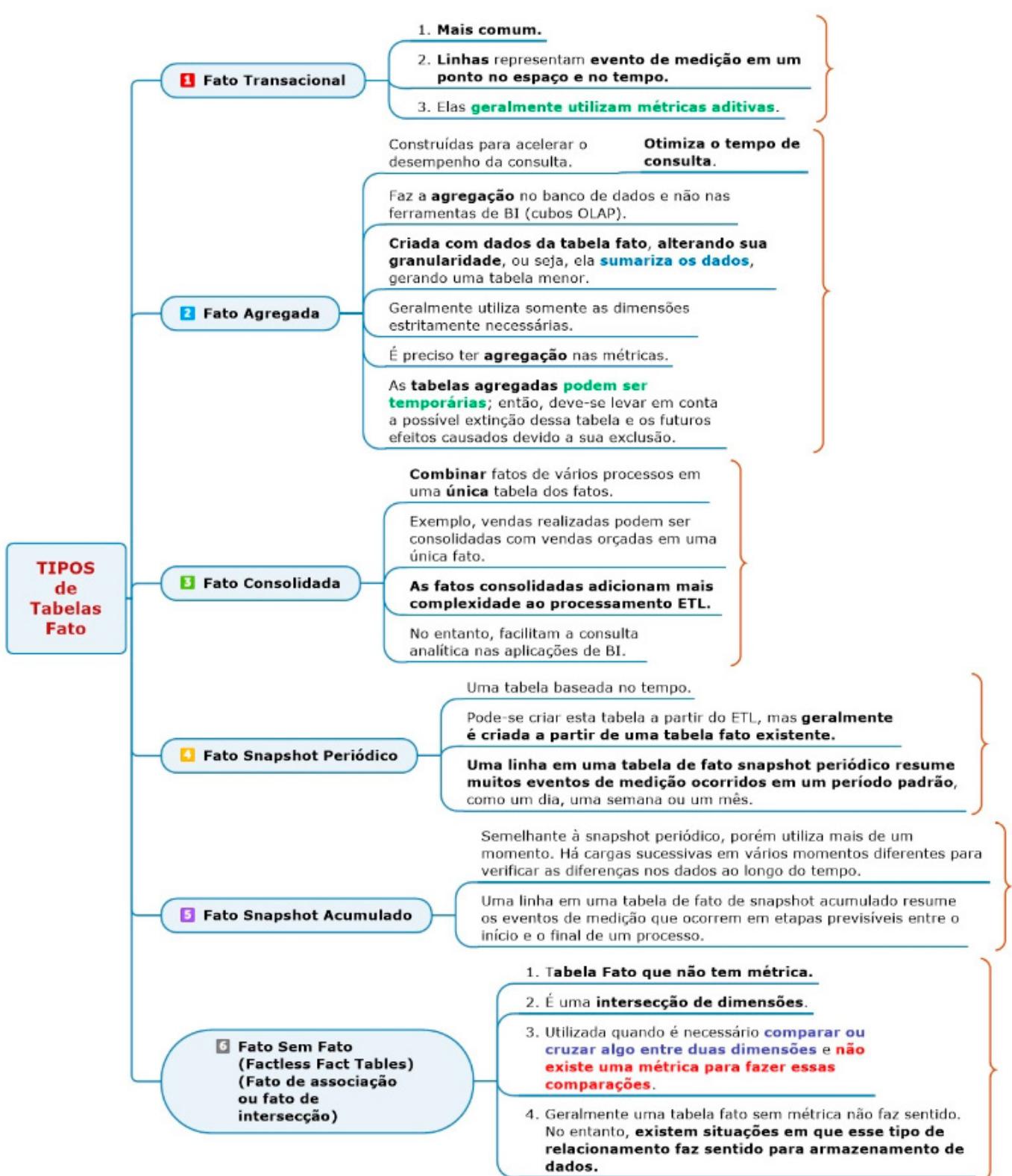


Figura. Tipos de Tabelas Fato. Fonte: Quintão (2020)

8.4. CUBO DE DADOS

O **modelo dimensional** é formado por uma **tabela central** (**tabela de fatos**) e várias outras a ela interligadas (**tabelas de dimensões**), sempre por meio de **chaves especiais, que associam o fato a uma dimensão do cubo.**

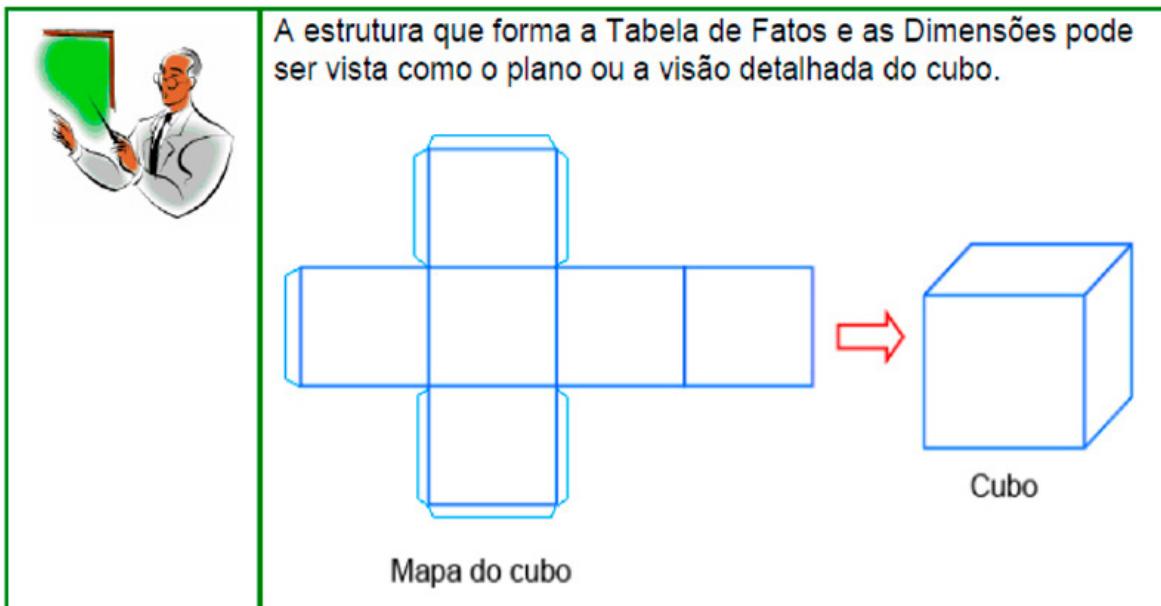


Figura. Cubo. Fonte: Microsoft (2011)

Cubo é uma **estrutura multidimensional** de dados que expressa a forma na qual os tipos de informações se relacionam entre si.

O cubo de uma forma genérica **armazena todas as informações relacionadas a um determinado assunto**, de maneira a permitir que sejam montadas várias combinações entre elas, resultando na extração de várias visões sobre o mesmo tema (Hokama et al., 2004).

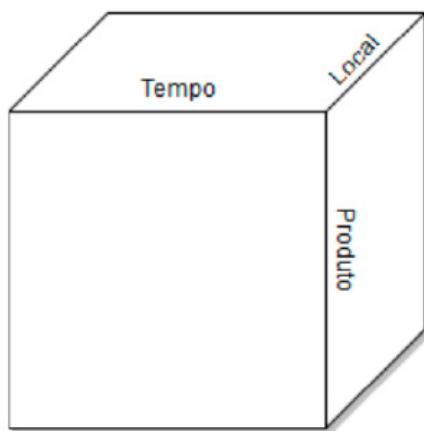


Figura. Representação Gráfica de um Cubo de Dados

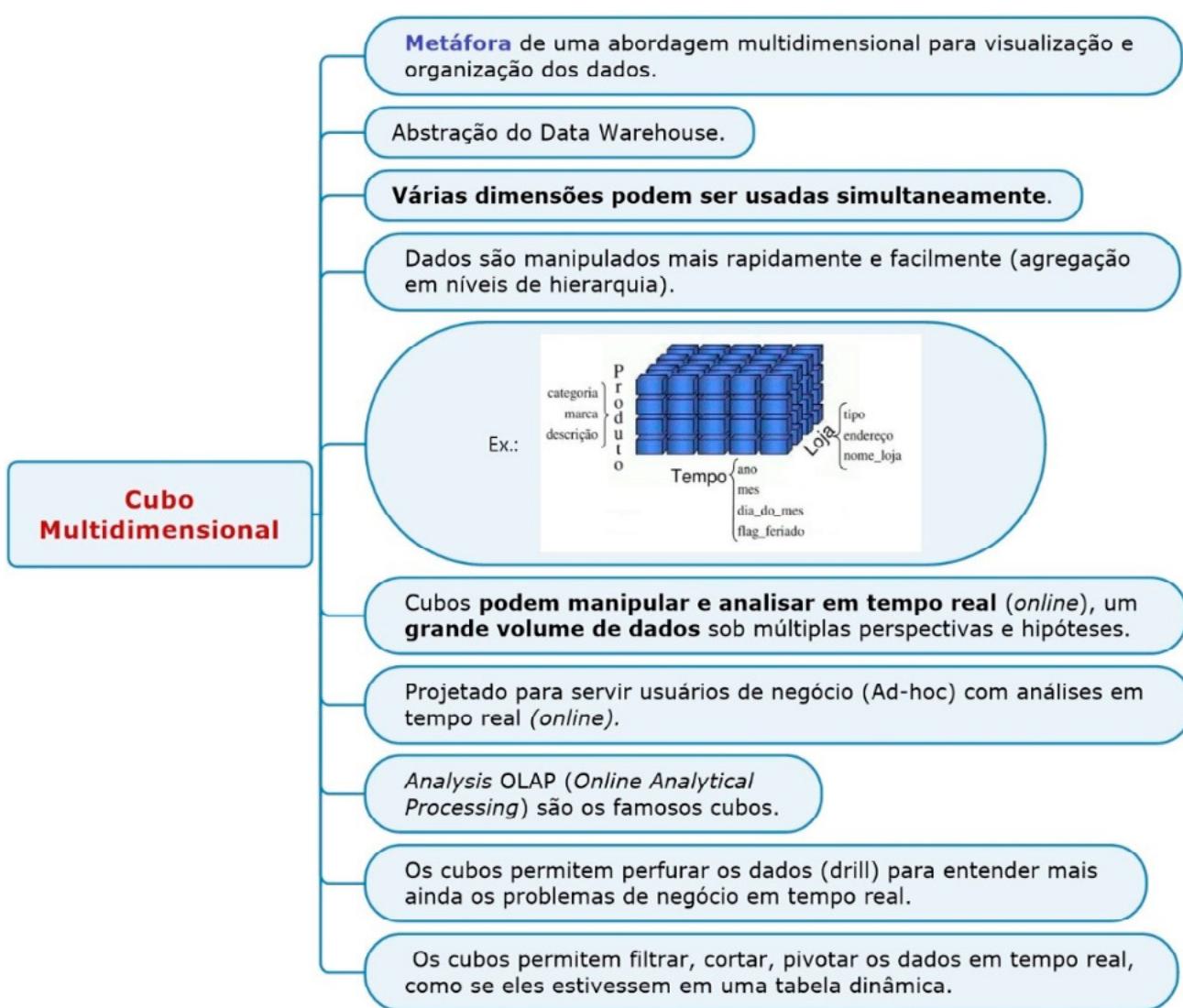


Figura. Cubo Multidimensional. Fonte: Quintão (2020)

DIRETO DO CONCURSO

014. (COSEAC/DATAPREV/ANALISTA DE TECNOLOGIA DA INFORMAÇÃO/ANO/ADAPTA-DA) Sobre modelagem multidimensional, o cubo:

I – é uma representação intuitiva, pois todas as dimensões coexistem para todo ponto no cubo e são independentes umas das outras;

II – é, de fato, apenas uma metáfora visual;

III – serve para descrever requisitos funcionais.

Acerca dos itens acima mencionados, apenas:

a) I e III estão corretos;

b) Somente I está correto;

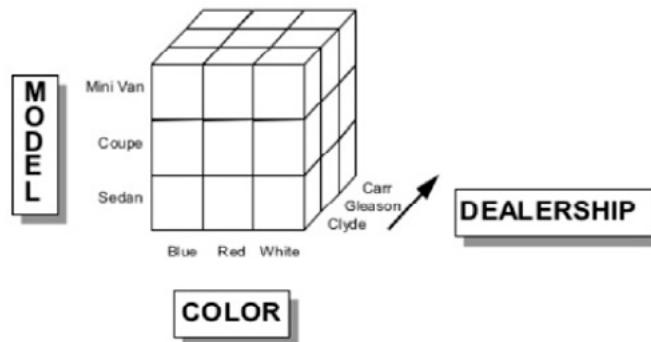
c) I e II estão corretos;

d) II e III estão corretos;

e) III está correto.



I – Certo. Os **cubos** são massas de dados que retornam das consultas feitas ao banco de dados e podem ser manipulados e visualizados por inúmeros ângulos e diferentes níveis de agregação. É uma **representação intuitiva do fato** porque **todas as dimensões coexistem para todo ponto no cubo e são independentes uma das outras**.



- O cubo é, de fato, apenas uma metáfora visual.
- É uma representação intuitiva do fato porque todas as dimensões coexistem para todo ponto no cubo e são independentes umas das outras.

Figura. Visão Multidimensional. Volume de Vendas. Fonte: <https://pt.slideshare.net/wbillacox/dicas-de-modelagem-dimensional>

II – Certo. O **cubo** é, de fato, apenas uma representação conceitual ou uma metáfora visual.

III – Errado. A análise multidimensional é uma das grandes utilidades da tecnologia OLAP (**Online Analytical Processing**), consistindo em ver determinados **cubos de informações** de diferentes ângulos (perspectivas) para análise e de vários níveis de agregação.

Portanto, I e II estão corretos.

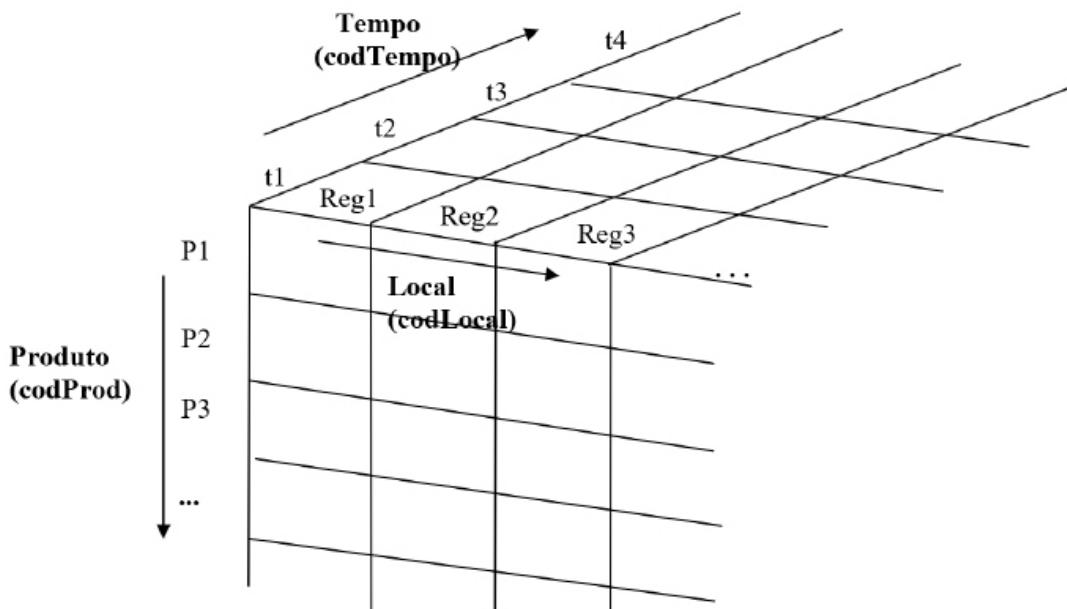
Letra c.

015. (CESGRANRIO/PETROBRAS/ANALISTA DE SISTEMAS JÚNIOR) Em um banco de dados multidimensional, os dados estão conceitualmente armazenados e organizados em:

- a) classes e objetos.
- b) cubos e hipercubos.
- c) partições e índices.
- d) consultas materializadas e sumários.
- e) estrelas e constelações.



Um **cubo de dados** é a **representação multidimensional de dados agregados**. A figura seguinte ilustra um **cubo** que organiza dados de venda de produto por data e regiões de venda. Cada célula representa a venda de um produto específico, em um período de tempo específico (ano, trimestre, mês,...) em um local específico.



Adicionando outras dimensões, teremos um hipercubo. Os dados podem ser consultados diretamente em qualquer combinação de dimensões, permitindo que consultas complexas no banco de dados original sejam realizadas de forma mais direta e com maior desempenho. Existem ferramentas que permitem a visualização dos dados de acordo com a escolha de dimensões do usuário.

Assim, em um banco de dados multidimensional, os dados estão conceitualmente armazenados e organizados em **cubos** e **hipercubos**.

Letra b.

016. (CESPE/MPE-PI/ANALISTA MINISTERIAL/2012) Um cubo de dados é a representação multidimensional dos dados não agregados na qual é necessário que as dimensões tenham o mesmo tamanho.



Um **cubo de dados** é a **representação multidimensional de dados agregados**, mas **não** é necessário que as dimensões tenham o mesmo tamanho!

Errado.

9. TABELA DE DIMENSÃO

As **tabelas de dimensões** são as companheiras das **tabelas de fatos**. Cada dimensão é definida pela sua **chave primária**, que serve para manter a integridade referencial na tabela de fatos à qual está relacionada.

Um **cubo** requer a definição de **pelo menos uma dimensão** no seu esquema.

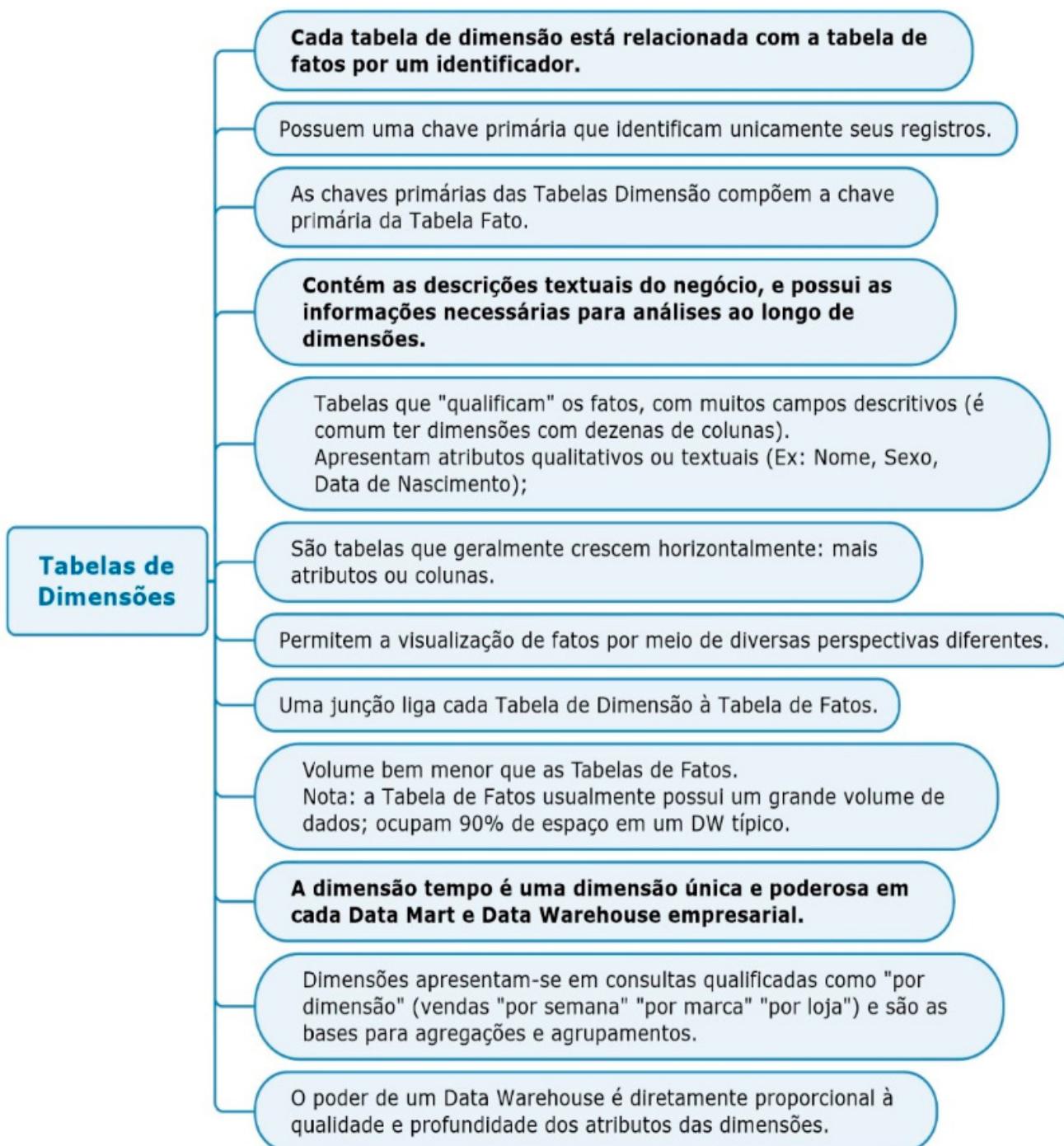


Figura. Tabela de Dimensão. Fonte: Quintão (2020)

9.1. MEDIDAS E DIMENSÕES

Cabe destacar que o modelo dimensional divide o mundo dos dados em dois grandes tipos: as **medidas** e as **dimensões** destas medidas.

As **medidas** são sempre numéricas e são armazenadas nas **tabelas de fatos** e as **dimensões** contextuais são armazenadas nas **tabelas de dimensões**.

As dimensões organizam os dados em função de uma área de interesse para os usuários.
Cada dimensão descreve um aspecto do negócio e proporciona o acesso intuitivo e simples aos dados. Uma dimensão oferece ao usuário um grande número de combinações e intersecções para analisar dados.

DIRETO DO CONCURSO

017. (IDECAN/INMETRO/ANÁLISE E DESENVOLVIMENTO DE SISTEMAS/2015) A modelagem multidimensional é uma técnica de concepção e visualização de um modelo de dados de um conjunto de medidas que descrevem aspectos comuns de negócios. Um modelo multidimensional é formado por três elementos básicos. Assinale-os:

- a) Esquema, fatos e itens.
- b) Fatos, dimensões e itens.
- c) Medidas, esquema e fatos.
- d) Fatos, dimensões e medidas
- e) Dimensões, medidas e esquema.

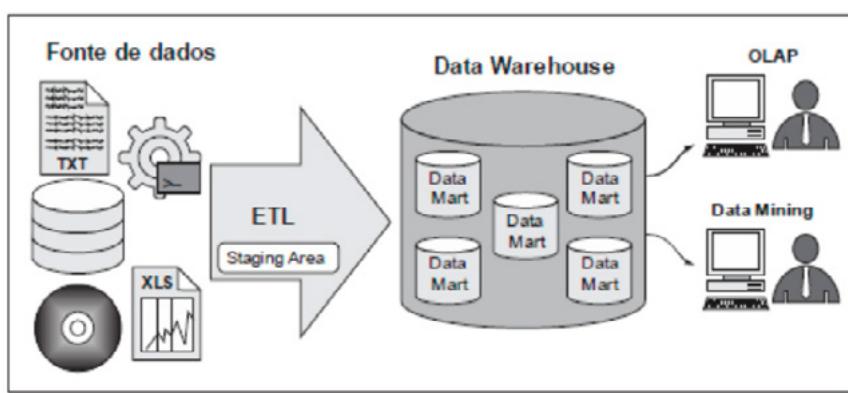


Conforme visto, os **três elementos básicos** são:

Fatos	São os dados a serem agrupados, contendo os valores de cada medida para cada combinação das dimensões existentes . O tamanho da tabela que contém os fatos merece atenção especial do analista.
Dimensões	Estabelecem a organização dos dados, determinando possíveis consultas/cruzamentos. Por exemplo: região, tempo, canal de venda,... Cada dimensão pode ainda ter seus elementos, chamados membros, organizados em diferentes níveis hierárquicos . A dimensão tempo, por exemplo, pode possuir duas hierarquias: calendário gregoriano (com os níveis ano, mês e dia) e calendário fiscal (com os níveis ano, semana e dia).
Medidas	São os valores a serem analisados, como médias, totais e quantidades.

Letra d.

018. (FCC/TRT-4ª REGIÃO/ANALISTA JUDICIÁRIO/TECNOLOGIA DA INFORMAÇÃO/2015)
 Considere a arquitetura geral de um sistema de BI- *Business intelligence* mostrada na figura abaixo.



Nesta arquitetura

- a) Data Marts representam áreas de armazenamento intermediário criadas a partir do processo de ETL. Auxiliam na transição dos dados das fontes OLTP para o destino final no Data Warehouse.
- b) OLAP é um subconjunto de informações extraído do Data Warehouse que pode ser identificado por assuntos ou departamentos específicos. Utiliza uma modelagem multidimensional conhecida como modelo estrela.
- c) os dados armazenados no Data Warehouse são integrados na base única mantendo as convenções de nomes, valores de variáveis e outros atributos físicos de dados como foram obtidos das bases de dados originais.
- d) o Data Warehouse não é volátil, permite apenas a carga inicial dos dados e consultas a estes dados. Além disso, os dados nele armazenados são precisos em relação ao tempo, não podendo ser atualizados.
- e) Data Mining se refere ao processo que, na construção do Data Warehouse, é utilizado para composição de análises e relatórios, armazenando dados descritivos e qualificando a respectiva métrica associada.



- a) Errada. **Staging Area** representa a área de armazenamento intermediário criada a partir do processo de ETL. Auxilia na transição dos dados das fontes OLTP para o destino final no Data Warehouse. **Data Mart** é um subconjunto de informações do DW que podem ser identificados por assuntos ou departamentos específicos.
- b) Errada. **Data Mart** é um subconjunto de informações extraído do Data Warehouse que pode ser identificado por assuntos ou departamentos específicos. Utiliza uma modelagem multidimensional conhecida como modelo estrela. **OLAP (Online Analytical Processing)** são ferramentas com capacidade de análise em múltiplas perspectivas das informações armazenadas.
- c) Errada. Os dados armazenados no Data Warehouse são integrados na base única mantendo as convenções de nomes, valores de variáveis e outros atributos físicos de dados. O **processo ETL** (do inglês **Extract Transform and Load**) faz a extração, tratamento e limpeza dos dados para inserção no DW.

- d) Certa. O **Data Warehouse** não é volátil, permite apenas a carga inicial dos dados e consultas a estes dados. Além disso, os dados nele armazenados são precisos em relação ao tempo, não podendo ser atualizados.

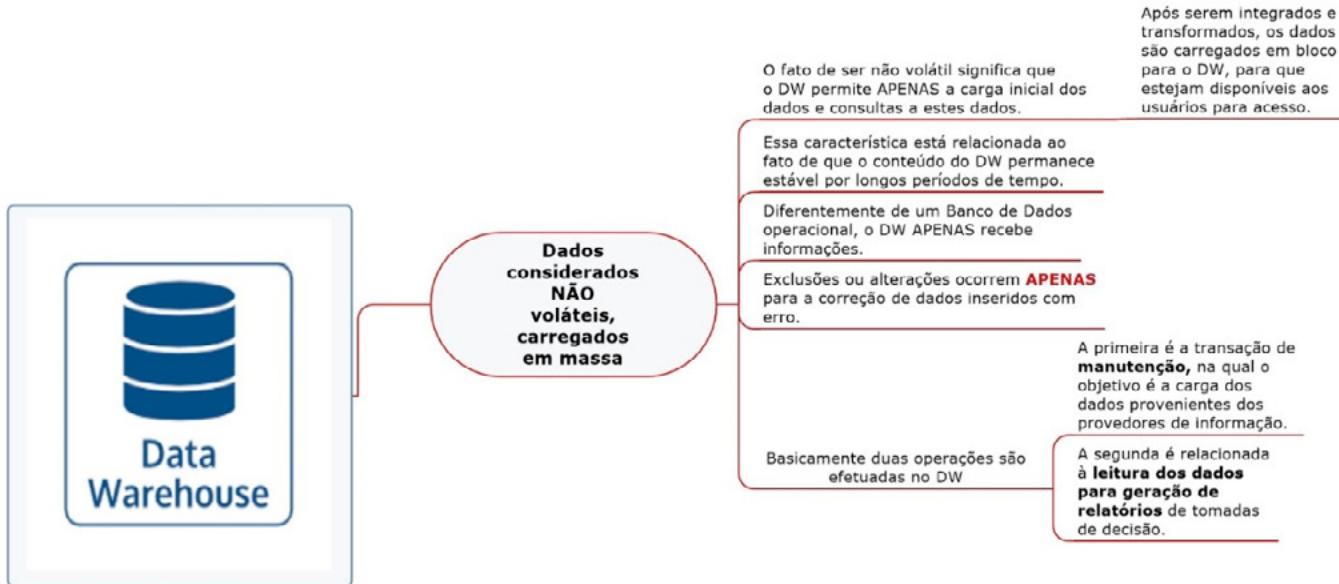


Figura. Dados não voláteis

- e) Errada. **Dimensão** se refere à **tabela** que, na construção do Data Warehouse, é **utilizada** para composição de análises e relatórios, **armazenando dados descritivos e qualificando a respectiva métrica associada**.

Letra d.

10. ESQUEMAS MULTIDIMENSIONAIS

Dois esquemas multidimensionais comuns são o de **Estrela (Star Schema)** e o **Snow Flake (Flocos de Neve)**.

10.1. ESQUEMA ESTRELA (STAR SCHEMA)

Esta estrutura está formada por uma **tabela central - tabela de fatos** - e um conjunto de tabelas organizadas ao redor dela, **as tabelas de dimensões**.

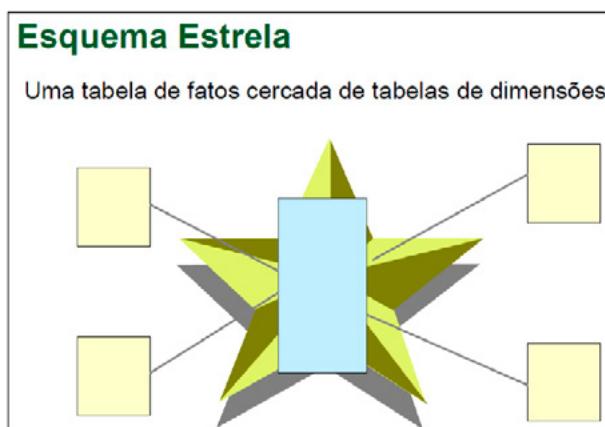


Figura. Esquema Estrela (Star Schema)

Segundo Elmasri e Navathe (2005, p. 651), uma **tabela de dimensão** consiste em *tuplas* de atributos da dimensão. Uma **tabela de fato** pode ser pensada como possuindo *tuplas*, uma por fato registrado.

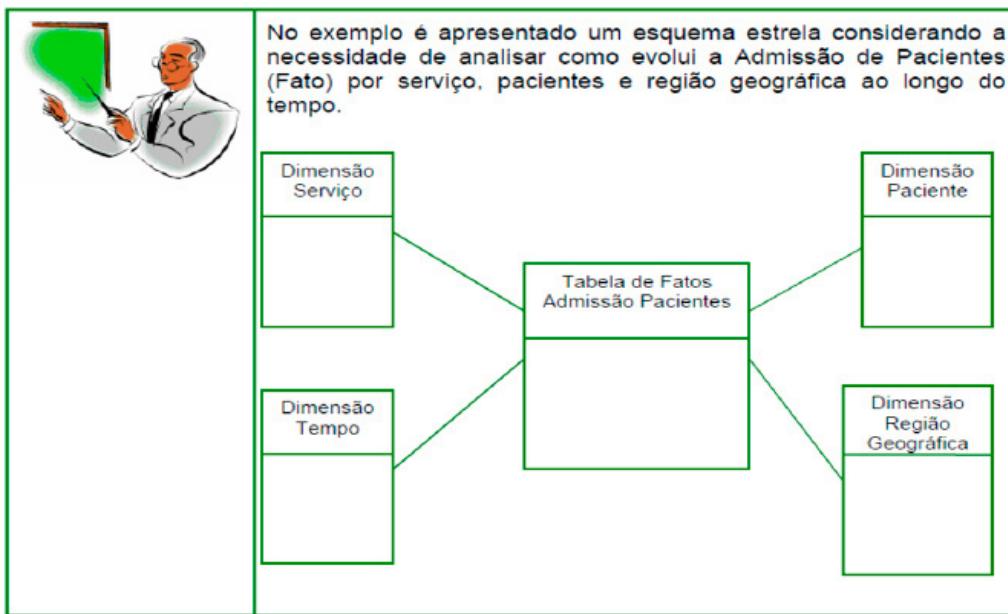


Figura. Exemplo de Esquema Estrela (Star Schema). Fonte: Microsoft (2011)

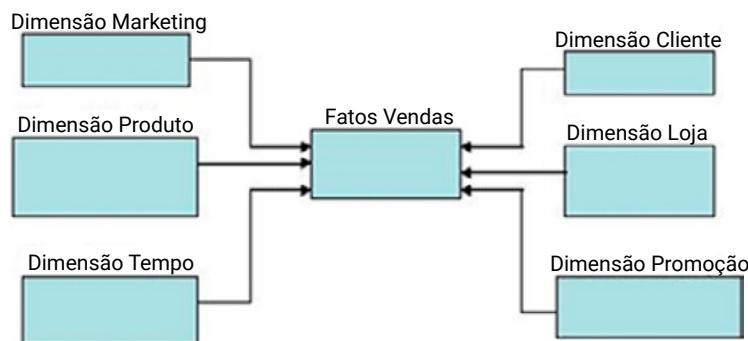


Figura. Exemplo de Esquema Estrela (Star Schema)

Nas pontas da estrela estão as **tabelas de dimensões** que contêm os **atributos** das aberturas que interessam ao negócio e que podem ser utilizadas como critérios de filtro e são relativamente pequenas.

Cada **tabela de dimensão** está relacionada com a **tabela de fatos** por um **identificador**.

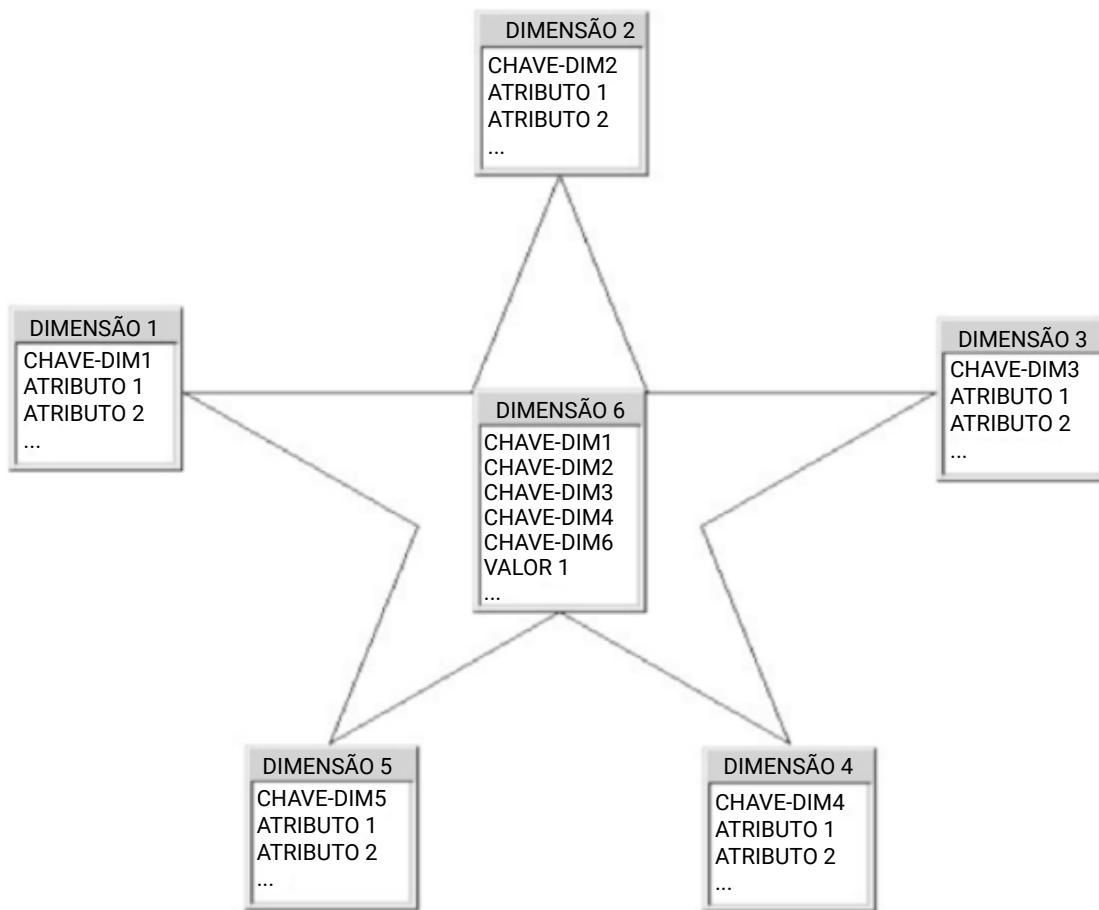


Figura. Representação da estrutura do Esquema Estrela. Fonte: <http://meusite.mackenzie.com.br/rogerio/tgi/2004ModelagemDW.pdf>

As **características** de um **esquema de estrela** são:

- O centro da estrela é a **tabela de fatos**.
- As pontas da estrela são as **tabelas de dimensões**.
- Cada esquema está formado por apenas **UMA tabela de fatos**.

DIRETO DO CONCURSO

019. (FCC/TRT-MS/ANALISTA JUDICIÁRIO-TI/2017) Uma das formas de apresentação de um banco de dados multidimensional é através do modelo estrela. No centro de um modelo estrela encontra-se a tabela de:

- a) dimensão e, ao seu redor, as tabelas de fatos.
- b) dimensão, cuja chave primária deve ser composta.
- c) núcleo e, ao seu redor, as tabelas de nível.
- d) fatos, cuja chave primária deve ser simples.
- e) fatos e, ao seu redor, as tabelas de dimensões.



O modelo Estrela (Star Schema) é formado por uma tabela central - **tabela de fatos** - e um conjunto de tabelas organizadas ao redor dela, as **tabelas de dimensões**.

Letra e.

020. (FCC/SABESP/ANALISTA DE GESTÃO/SISTEMAS/2018) Um Analista está trabalhando em um Data Warehouse – DW que utiliza no centro do modelo uma única tabela que armazena as métricas e as chaves para as tabelas ao seu redor (que descrevem os dados que estão na tabela central) às quais está ligada. O esquema de modelagem utilizado pelo DW, a denominação da tabela central e a denominação das tabelas periféricas são, respectivamente,

- a) floco de neve, base, granulares.
- b) estrela, fato, dimensões.
- c) constelação, fato, granulares.
- d) atomic, base, branches.
- e) anel, base, dimensões.



O **modelo Estrela (Star)** é formado por uma tabela central - **tabela de fatos** - e um conjunto de tabelas organizadas ao redor dela, as **tabelas de dimensões**.

Nas pontas da estrela estão as **tabelas de dimensões** que contêm os **atributos** das aberturas que interessam ao negócio e que podem ser utilizadas como critérios de filtro e são relativamente pequenas. **Cada tabela de dimensão está relacionada com a tabela de fatos por um identificador.**

Letra b.

021. (FCC/TRT-AL/ANALISTA JUDICIÁRIO/TI/2011) O modelo estrela, como estrutura básica de um modelo de dados multidimensional, possui uma configuração típica composta de uma entidade central:

- a) mining e um conjunto de entidades fatos.
- b) mining e um conjunto de entidades dimensões.
- c) mining e um conjunto de entidades roll-up.
- d) dimensão e um conjunto de entidades fatos.
- e) fato e um conjunto de entidades dimensões.



O Esquema ou Modelo Estrela é basicamente uma Tabela de Fatos central conectada a várias Tabelas de Dimensão em um relacionamento 1:N, sendo uma única tabela para cada dimensão.

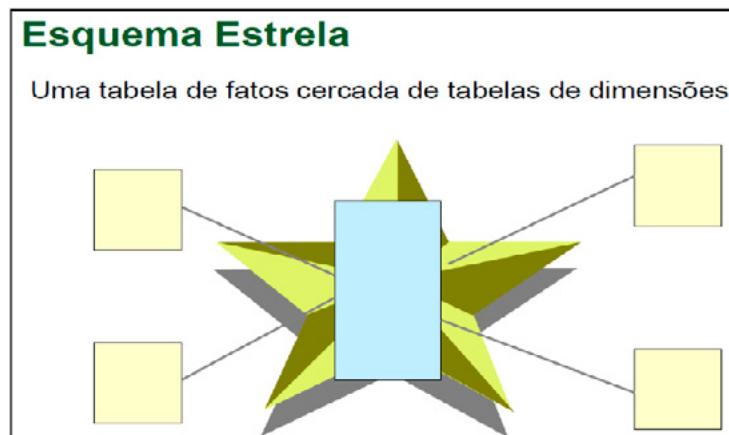


Figura. O Modelo Estrela (Star Schema)

Assim, o modelo estrela, como estrutura básica de um modelo de dados multidimensional, possui uma configuração típica composta de uma entidade central fato e um conjunto de entidades dimensões.

Letra e.

10.2. ESQUEMA FLOCO DE NEVE (SNOWFLAKE)

O **esquema floco de neve (Snowflake)** é uma variação do esquema estrela em que **alguma ponta da estrela explode em mais tabelas**.

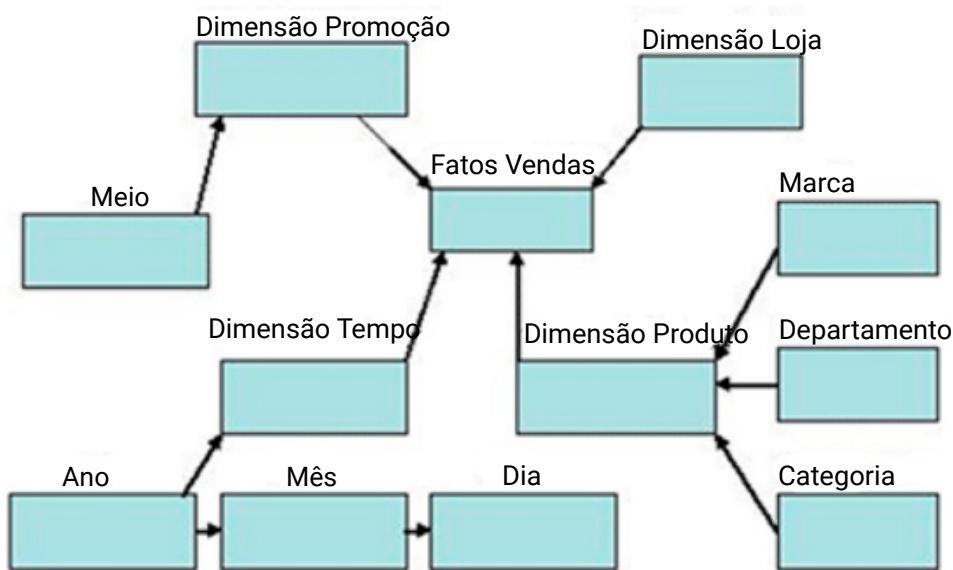
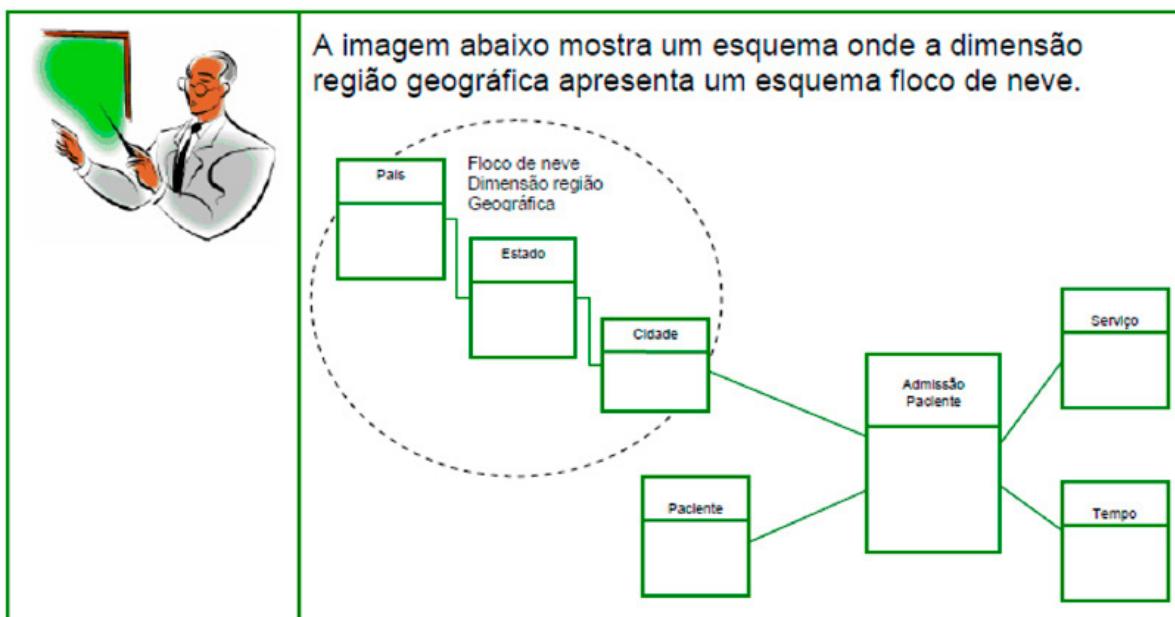


Figura. Esquema floco de neve (Snowflake)

Neste esquema, as **tabelas de dimensão floco de neve** estão **padronizadas para eliminar redundância de dados**. Diferente do esquema estrela, neste esquema os dados das dimensões são distribuídos em **múltiplas tabelas**.



Como **vantagem** do esquema destaca-se a **economia de espaço no armazenamento em disco, porém com um aumento na quantidade de tabelas**.

As **características** a seguir são parte de um **esquema floco de neve**:

- a **dimensão** está **padronizada**;
- os diferentes níveis estão armazenados em tabelas separadas;
- verifica-se economia de espaço.

DIRETO DO CONCURSO

022. (UPENET/UPE/ANALISTA DE SISTEMAS/ENGENHARIA DE SOFTWARE/2017) O modelo dimensional de um data warehouse, no qual todas as tabelas relacionam-se diretamente com a tabela de fatos, de forma que as tabelas dimensionais devem conter todas as descrições que são necessárias para se definir uma classe, é denominado de:

- a) Floco de neve.
- b) Estrela.
- c) Barramento.
- d) Árvore.
- e) Anel.



No **modelo Estrela (Star Schema)** todas as tabelas relacionam-se diretamente com a **tabela de fatos**. Sendo assim, as **tabelas dimensionais** devem conter todas as descrições que são necessárias para definir uma classe.

No **modelo Floco de Neve (Snow Flake)**, as tabelas dimensionais relacionam-se com a tabela de fatos, mas algumas dimensões relacionam-se apenas entre elas.

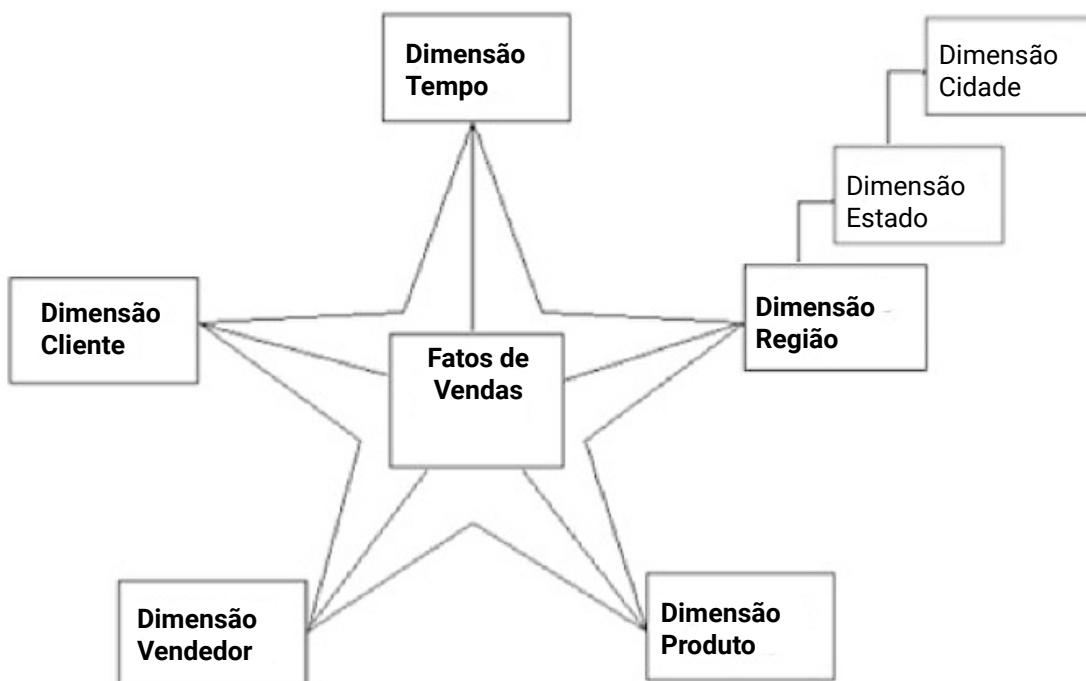


Figura. Modelo Floco de Neve. Fonte: Machado (2004)

Letra b.

11. ESQUEMA ESTRELA X ESQUEMA FLOCO DE NEVE

Vamos a um **quadro resumido** sobre a **diferença** entre o Esquema Estrela (Star Schema) e o Esquema Floco de Neve (Snow Flake).

	Star Schema (Estrela)	Snow Flake (Flocos de Neve)
Clareza	Mais fácil	Mais difícil
Quantidade de Tabelas	Menor (<) Nesse esquema existe uma Tabela Fato central conectada a várias Tabelas Dimensão.	Maior (>) Aqui tem-se também em uma Tabela Fato central (geralmente apenas uma) conectada a múltiplas dimensões , no entanto, estas Tabelas Dimensões são divididas em tabelas relacionadas . Essa divisão reduz a redundância dos dados, mas aumenta a complexidade do modelo.

Complexidade de Consultas	Consultas mais simples e fáceis de entender , pois há menor necessidade de junção de tabelas.	Consultas mais complexas e difíceis de entender , pois há necessidade de realizar muitas junções.
Desempenho de Consulta	Menos chaves estrangeiras, e consequentemente, menor tempo para as consultas (mais rápido). Melhora o desempenho.	Mais chaves estrangeiras , e consequentemente, mais tempo para as consultas (mais lento). Aumenta a quantidade de vínculos entre tabelas provocando redução do desempenho.
Tabelas Dimensão	Somente uma para cada dimensão.	Mais que uma para cada dimensão.
Normalização	Recomenda a não normalização das tabelas dimensão . Tabelas de Dimensões são desnortinalizadas.	Recomenda a normalização das Tabelas Dimensão.
	Star Schema (Estrela)	Snow Flake (Flocos de Neve)
	Geralmente é um esquema totalmente não padronizado e pode estar parcialmente padronizado nas tabelas de dimensões.	(A normalização das dimensões também tem efeito negativo na velocidade de saída dos resultados). Todas as tabelas de dimensão são normalizadas, geralmente, até a 3FN.
Armazenamento	Aumenta o espaço.	Economiza espaço.
Joins	Poucos	Muitos. (Devido à Hierarquia das Tabelas Dimensão). Este é outro fator que também torna mais lento .

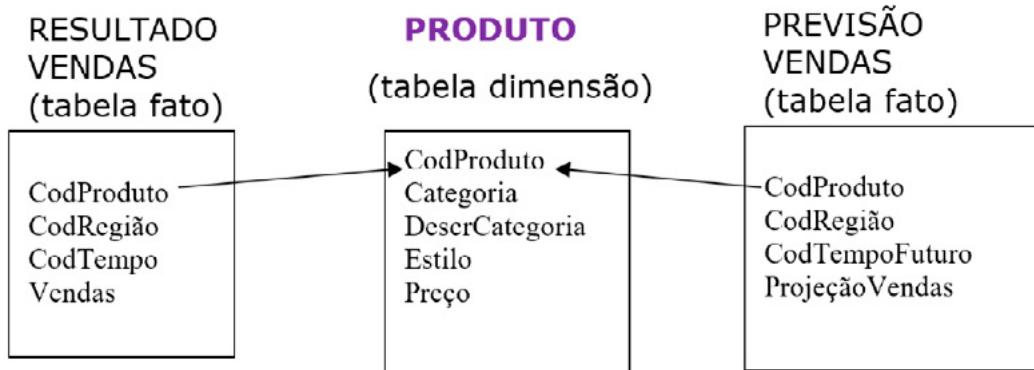
Facilidade de Manutenção/ Alteração	Como neste esquema o foco está nas consultas rápidas, pode existir redundância dos dados. As Tabelas Dimensão deste esquema não são normalizadas (mais difícil de manter e alterar).	Sem redundância (mais fácil de manter e alterar). O fato de não haver redundância faz com que haja otimização dos dados .
Quando usar	Quando Tabelas Dimensão são menores (têm menos linhas).	Quando Tabelas Dimensão são maiores.

Tabela. Modelo Estrela x Modelo Floco de Neve

12. CONSTELAÇÃO DE FATOS

Conforme destaca Elmasri e Navathe (2005, p. 651), uma **Constelação de Fatos** é um **conjunto de tabelas de fatos que compartilham algumas tabelas de dimensão**.

A figura seguinte mostra uma constelação de fatos com duas tabelas de fatos, intituladas “resultado vendas” e “previsão vendas”. Elas compartilham a tabela de dimensão chamada “produto”. As constelações de fatos limitam as possíveis consultas no *Data Warehouse*.


Figura. Uma constelação de fatos

13. SETE PASSOS PARA CONSTRUIR UM DW

A seguir, destacamos os 7 passos para construir um Data Warehouse (DW), conforme destacado em Piton (2020).

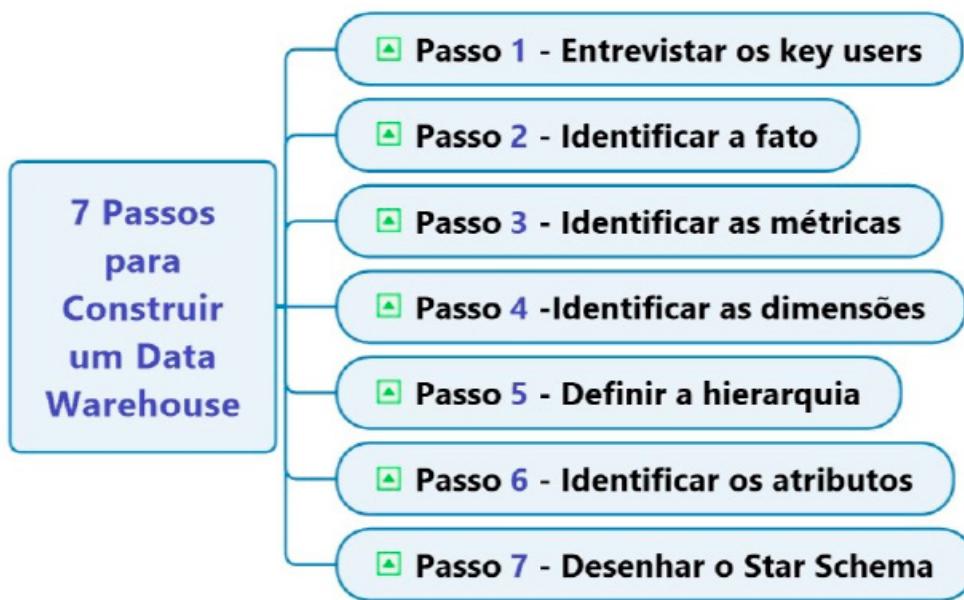


Figura. 7 Passos para Construir um DW. Fonte: Quintão (2020)

Passo 1 – Entrevistar os Key Users (Usuários Principais):

Embora o cliente deva saber o que ele quer, é preciso saber extrair as melhores respostas, então:

- Use perguntas que os usuários principais (Key Users) estão acostumados a ver no dia a dia;
- **Não** use termos técnicos **se não tiver necessidade**;
- O usuário de negócio não sabe o que são **fatos** ou **dimensões**, então é preciso induzir ele a dar as informações necessárias para a criação do DW. Use, por exemplo, a técnica 5W3H, que tem 9 perguntas simples para:
 - levantar as informações necessárias para montar o Star Schema;
 - falar a mesma língua dos usuários de negócio.

Qual fato aconteceu?

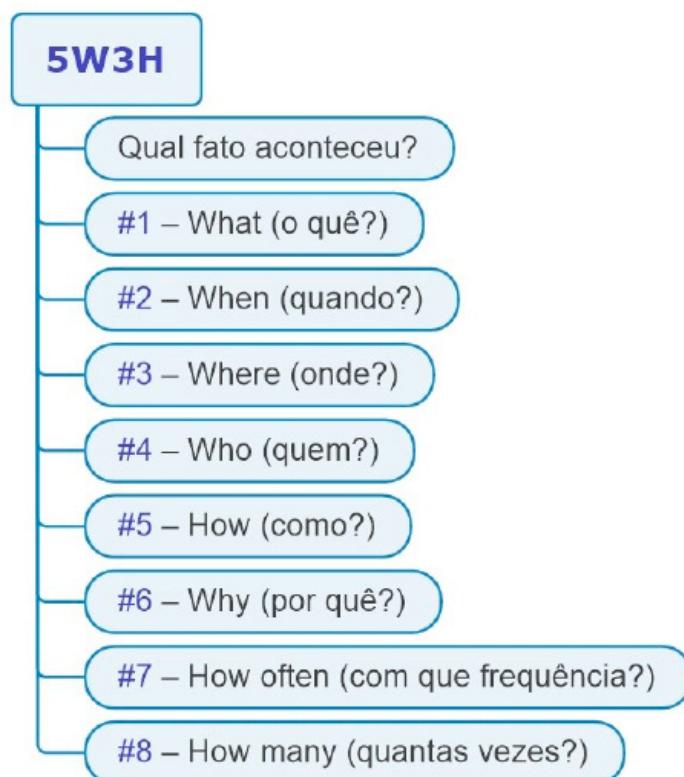


Figura. Técnica 5W3H

Como os usuários principais (*key users*) podem não estar habituados com as terminologias de BI, explique o que se espera dessas perguntas. Dê exemplo de fatos, como:

- venda;
- transação;
- pedido;
- faturamento.

Piton (2017) cita como exemplo a venda, que é o mais comum.

Qual o fato que aconteceu? **Uma venda.**

Juntando todas as informações, é montada a história de como aquele **fato** aconteceu. As próximas perguntas fazem referência à resposta desta. Nelas tem-se também um **template** para deixar a pergunta mais fácil de entender.

#1 – What – o quê / do quê?

Template:

Aconteceu _____. Do quê?

Aconteceu uma venda. Do quê?
De Coca-Cola

Uma Coca-Cola foi vendida

#2 When – quando?

Template:

Aconteceu _____. Quando?

Aconteceu uma venda.
Quando? 01/02/2017Uma Coca-Cola foi vendida no dia
01/02/2017

#3 Where – onde?

Template:

Aconteceu _____. Onde?

Aconteceu uma venda. Onde?
Na loja de São PauloUma Coca-Cola foi vendida, no dia
01/02/2017, em São Paulo

#4 Who – quem?

“Quem” são as pessoas envolvidas com o fato, e o papel delas varia de acordo com o negócio.

Template:

Aconteceu _____. Para quem?

Aconteceu _____. Quem entregou?

Para quem foi a venda? Para o Pedro

Quem fez a venda? Joana

Quem entregou? Transportadora
Mercúrio

Joana vendeu uma Coca-Cola para o Pedro, no dia 01/02/2017, em São Paulo e a entrega foi realizada pela Transportadora Mercúrio

#5 How – como?

Template:

Aconteceu _____. Como?

Aconteceu uma venda. Como?
Com pagamento em cartão

Joana vendeu uma Coca-Cola para o Pedro, que pagou com cartão, no dia 01/02/2017, em São Paulo e a entrega foi realizada pela Transportadora Mercúrio

#6 Why – por quê?

Template:

Aconteceu _____. Por quê?

Aconteceu uma venda. Por quê? Por causa de uma promoção de natal

Joana vendeu uma Coca-Cola para o Pedro, que pagou com cartão, em uma promoção de natal, no dia 01/02/2017 e a entrega foi realizada pela Transportadora Mercúrio em São Paulo

#7 How often – com que frequência?

Template:

Com que frequência acontece _____?

Com que frequência acontece uma venda? A cada 3h

#8 How many – quanto / quantas?

Template:

Quantas _____ aconteceram?

Quanto foi _____?

Quantas vendas aconteceram? 2
Quanto foi vendido? R\$200,00

Passo 2 – Identificar a Fato:

Com a entrevista feita, é hora de organizar essas informações.

A tabela **fato** possui 2 elementos:

- **foreign keys**, que conectam a fato nas dimensões;
- **métricas**, que são sempre dados numéricos.

Essa parte é fácil. É só fazer a primeira pergunta corretamente para se obter a fato definida.



Passo 3 – Identificar as Métricas:

Depois de ter a fato definida, daremos prosseguimento à **identificação das métricas** dela.

E para isso, pode-se usar as respostas da pergunta #8 (How many – quanto / quantas?), em que você tem a quantidade de vendas e o valor delas.



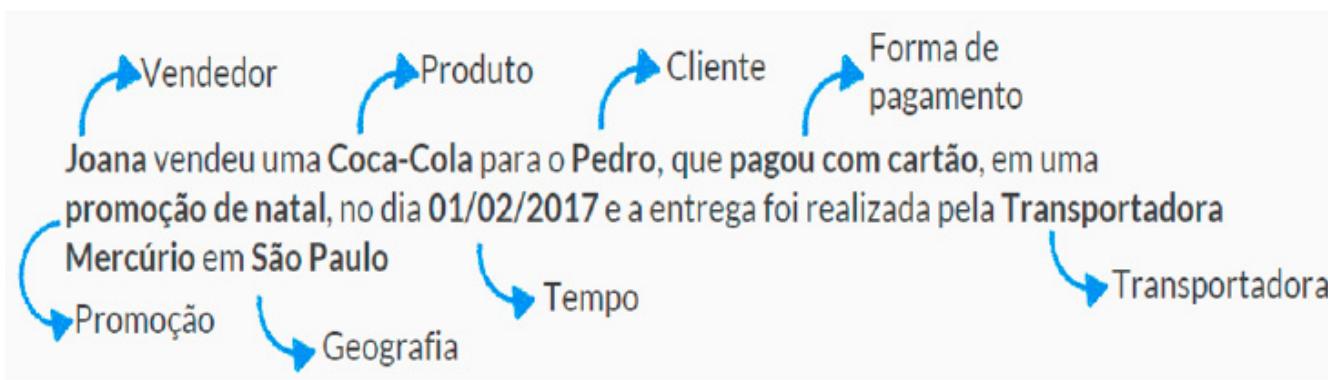
Passo 4 – Identificar as Dimensões:

As dimensões possuem 3 elementos:

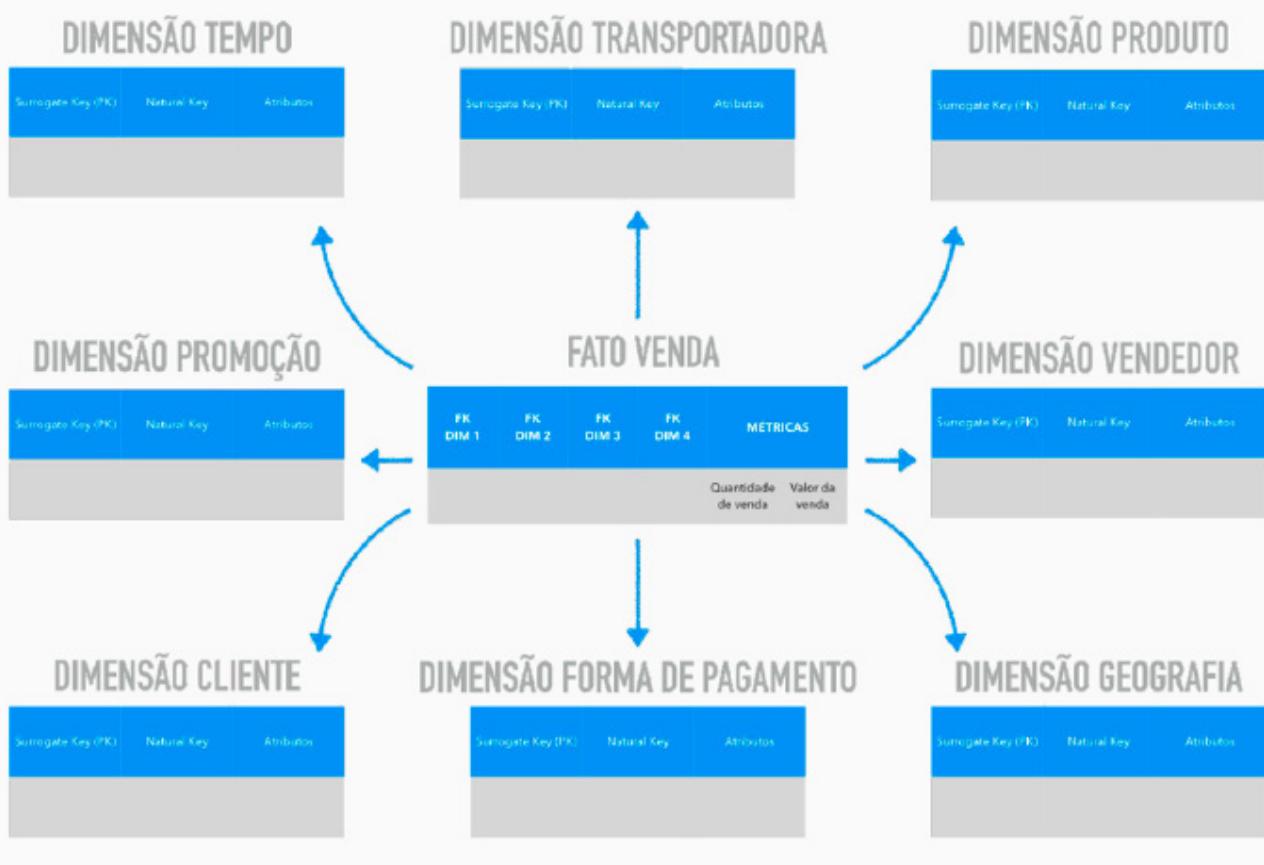
- **surrogate key**, que é a primary key da dimensão;
- **natural key**, que é a primary key da origem;
- **atributos**, que vão qualificar as **métricas** da **fato**.

As outras perguntas da entrevista vão definir suas dimensões.

Na frase final, já é possível identificar as **dimensões**:



E com esse levantamento foi possível **identificar 8 possíveis dimensões do Star Schema.**



Passo 5 – Definir a Hierarquia:

Com as dimensões identificadas, iremos **definir a hierarquia e o grão de cada uma delas.**

Neste momento devemos entender exatamente o que o usuário principal (key user) quer ver com aquela informação e se certificar de qual é o menor nível que ele vai precisar ver.

Para essa parte, é fundamental que você entenda como funciona a **hierarquia de dimensões.**

Na dimensão produto, você teria a seguinte **hierarquia:**

- categoria
 - subcategoria
 - produto → Esse é o grão, onde os dados vão de fato ser inseridos

DIMENSÃO PRODUTO

Surrogate Key (PK)	Natural Key	Categoria	Subcategoria	Produto	Código	Peso	Unidade de medida
		Bebidas	Refrigerante	Coca-Cola	P0001	600	ml

Passo 6 – Identificar os Atributos:

Com as dimensões definidas, é hora de identificar os **atributos** de cada uma delas.

Nesse momento, podemos fazer as perguntas para saber o que os usuários principais (key users) gostariam de analisar.

Exemplo:

Nessa primeira pergunta você respondeu Coca-Cola, que é um produto. O que você gostaria de analisar do produto?

E entrevistando as pessoas envolvidas com o fato, iremos identificar os **atributos** necessários. Iremos obter informações como:

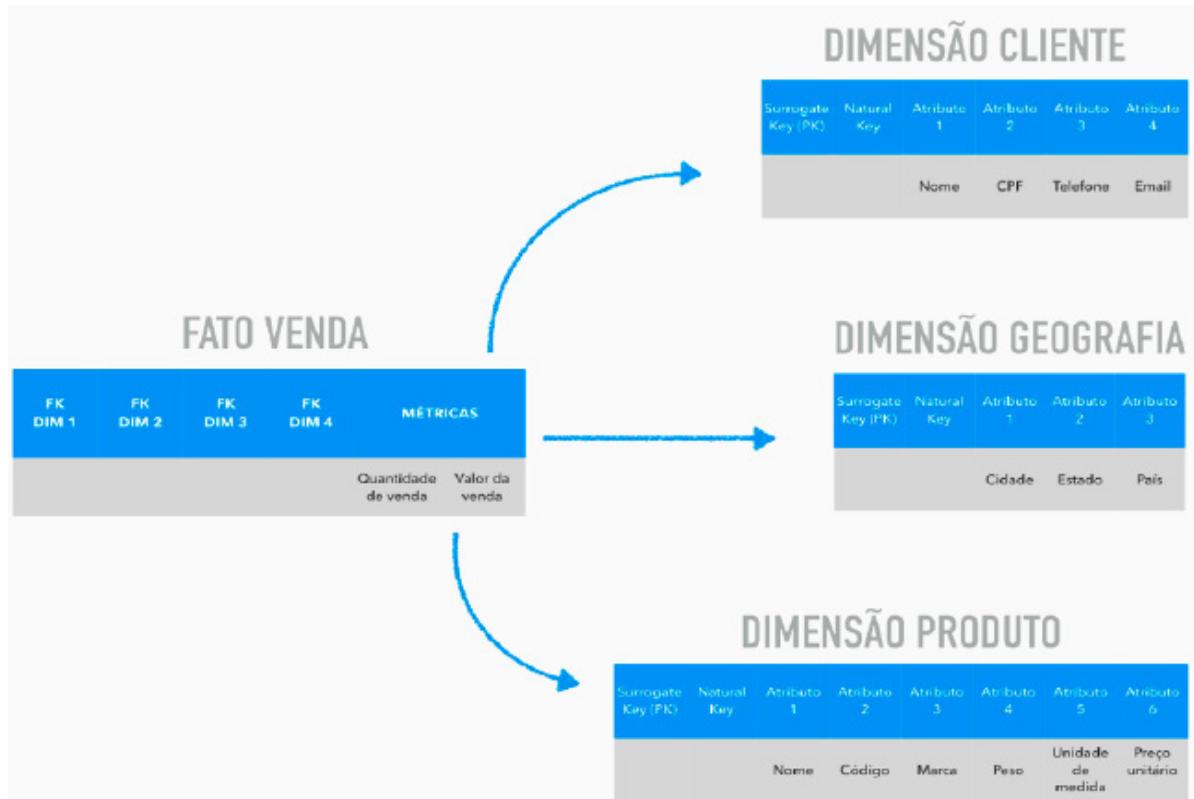
- Coca-Cola é o nome do produto;
- todo produto tem um código;
- o time de marketing vai dizer que precisa saber a marca do produto;
- o cara da logística vai precisar saber o peso do produto e a unidade de medida para planejar o caminhão;
- o pessoal do financeiro vai precisar saber o preço unitário do produto.

DIMENSÃO PRODUTO

Surrogate Key (PK)	Natural Key	Atributo 1	Atributo 2	Atributo 3	Atributo 4	Atributo 5	Atributo 6
		Nome	Código	Marca	Peso	Unidade de medida	Preço unitário

Passo 7 – Desenhar o Star Schema:

Nesse ponto já devemos ter todas as informações que devem constar no Star Schema levantadas e organizadas. Após analisá-las, **planeje as tabelas, defina seus atributos e hierarquia**, e teremos então a modelagem completa da primeira parte do Data Warehouse.

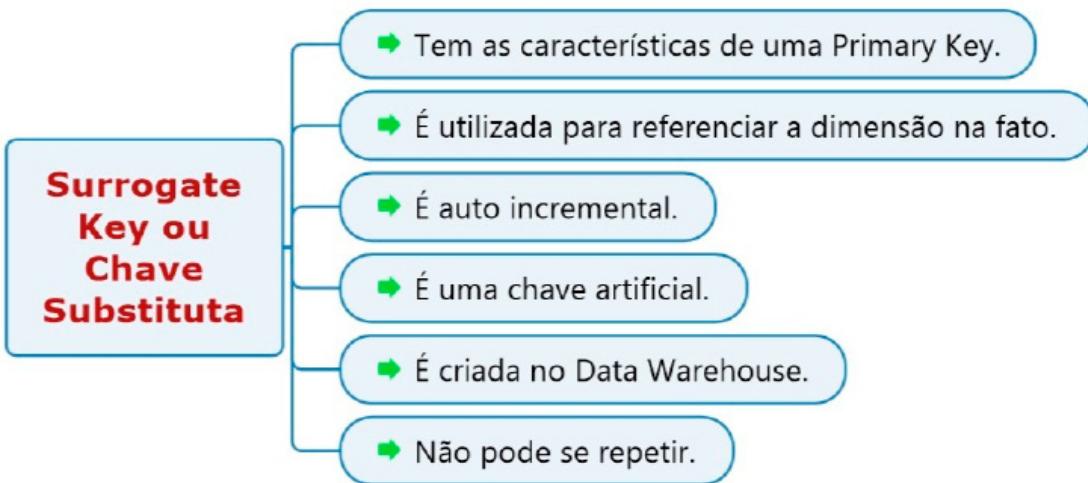


Com o desenho pronto, é só partir para a ferramenta de modelagem de banco de dados que preferir e o reproduzir, **fazer as conexões da fato com as dimensões** e colocar para funcionar.

Com o Star Schema criado, tem-se a primeira versão do Data Warehouse pronta para receber os dados do ETL.

NOTA:

- Star Schema já te permite fazer análises.
- A **Surrogate Key** nada mais é que o **campo de Primary Key da dimensão**.



- Ao invés de levar meses montando um DW gigante, entregue-o por partes e comece o quanto antes a apresentar os dados que os key users (usuários principais) precisam ver (PITON, 2020).

DIRETO DO CONCURSO

023. (CESPE/AUDITOR MUNICIPAL DE CONTROLE INTERNO/TECNOLOGIA DA INFORMAÇÃO/DESENVOLVIMENTO DE SISTEMAS/2018) Com relação à modelagem dimensional e à otimização de bases de dados para *business intelligence*, julgue o item subsequente.

Na modelagem multidimensional utilizada em *Data Warehouses* para se prover melhor desempenho, a tabela fato central deve relacionar-se às suas dimensões por meio da chave primária oriunda da fonte de dados original. O valor dessa chave deve ser idêntico ao da fonte, para que tenha valor semântico e garanta que o histórico das transações seja mantido.



Na modelagem multidimensional utilizada em *Data Warehouses* para se prover melhor desempenho, a tabela fato central deve relacionar-se às suas dimensões por meio da **chave primária oriunda da fonte de dados original chave substituta (ou surrogate key)**. O valor dessa chave deve ser **idêntico ao da fonte, para que tenha valor semântico numérico e auto incremental, não havendo valor semântico** e garanta que o histórico das transações seja mantido.

Errado.

RESUMO

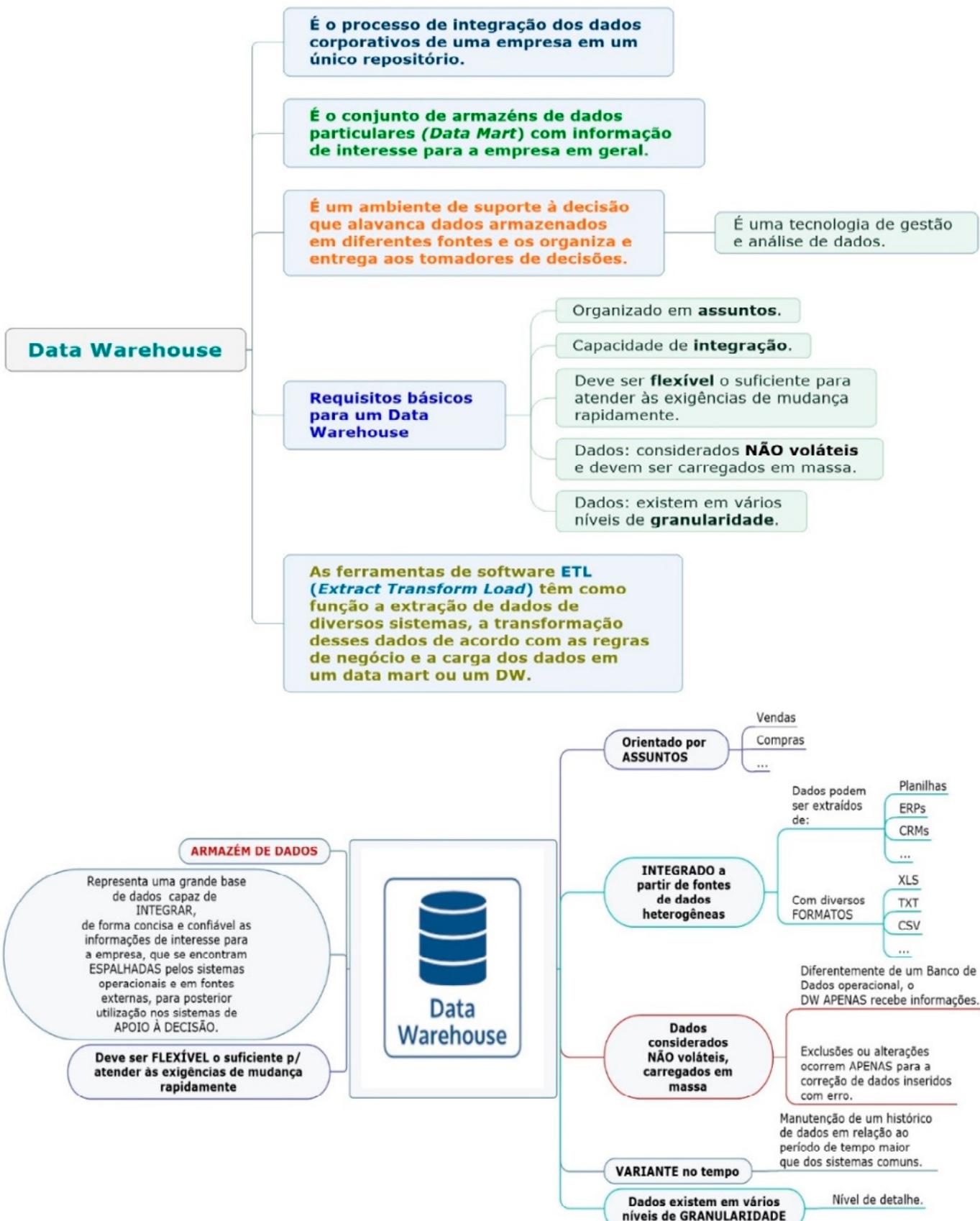


Figura. Data Warehouse. Fonte: Elaboração Própria (2020)

O conteúdo deste livro eletrônico é licenciado para MARIO LUIS DE SOUZA - 41250799864, vedada, por quaisquer meios e a qualquer título, a sua reprodução, cópia, divulgação ou distribuição, sujeitando-se aos infratores à responsabilização civil e criminal.

Patrícia Quintão

Após serem integrados e transformados, os dados são carregados em bloco para o DW, para que estejam disponíveis aos usuários para acesso.

O fato de ser não volátil significa que o DW permite APENAS a carga inicial dos dados e consultas a estes dados.

Essa característica está relacionada ao fato de que o conteúdo do DW permanece estável por longos períodos de tempo.

Diferentemente de um Banco de Dados operacional, o DW APENAS recebe informações.

Exclusões ou alterações ocorrem APENAS para a correção de dados inseridos com erro.

Basicamente duas operações são efetuadas no DW

A primeira é a transação de manutenção, na qual o objetivo é a carga dos dados provenientes dos provedores de informação.

A segunda é relacionada à leitura dos dados para geração de relatórios de tomadas de decisão.

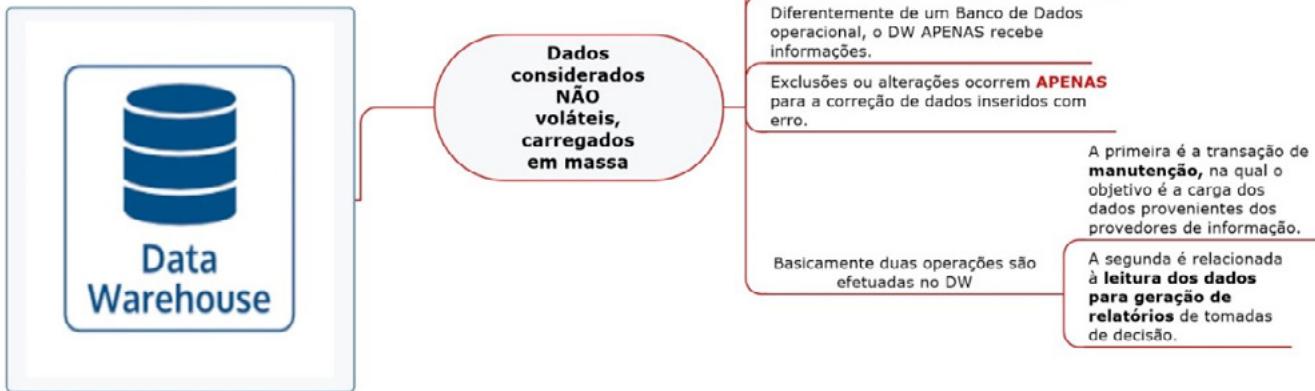


Figura. Dados Não Voláteis. Fonte: Quintão (2020)

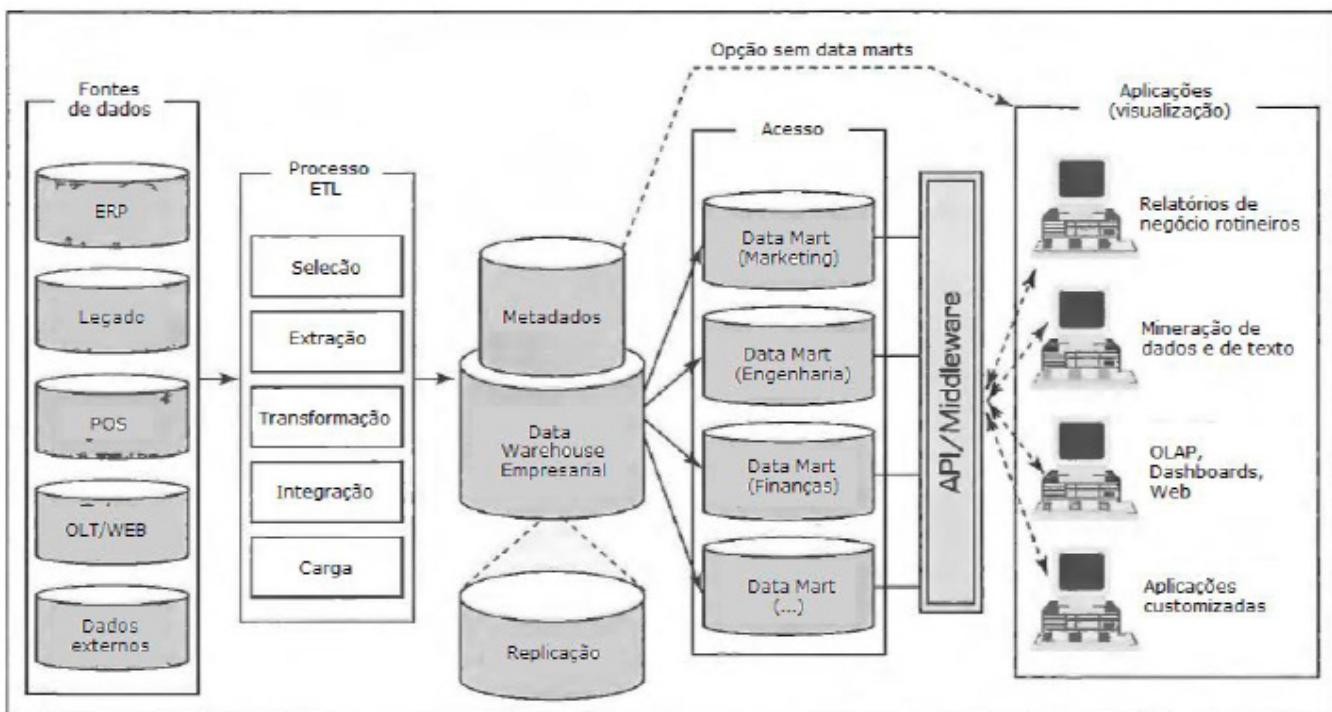


Figura. Processo de Data Warehousing

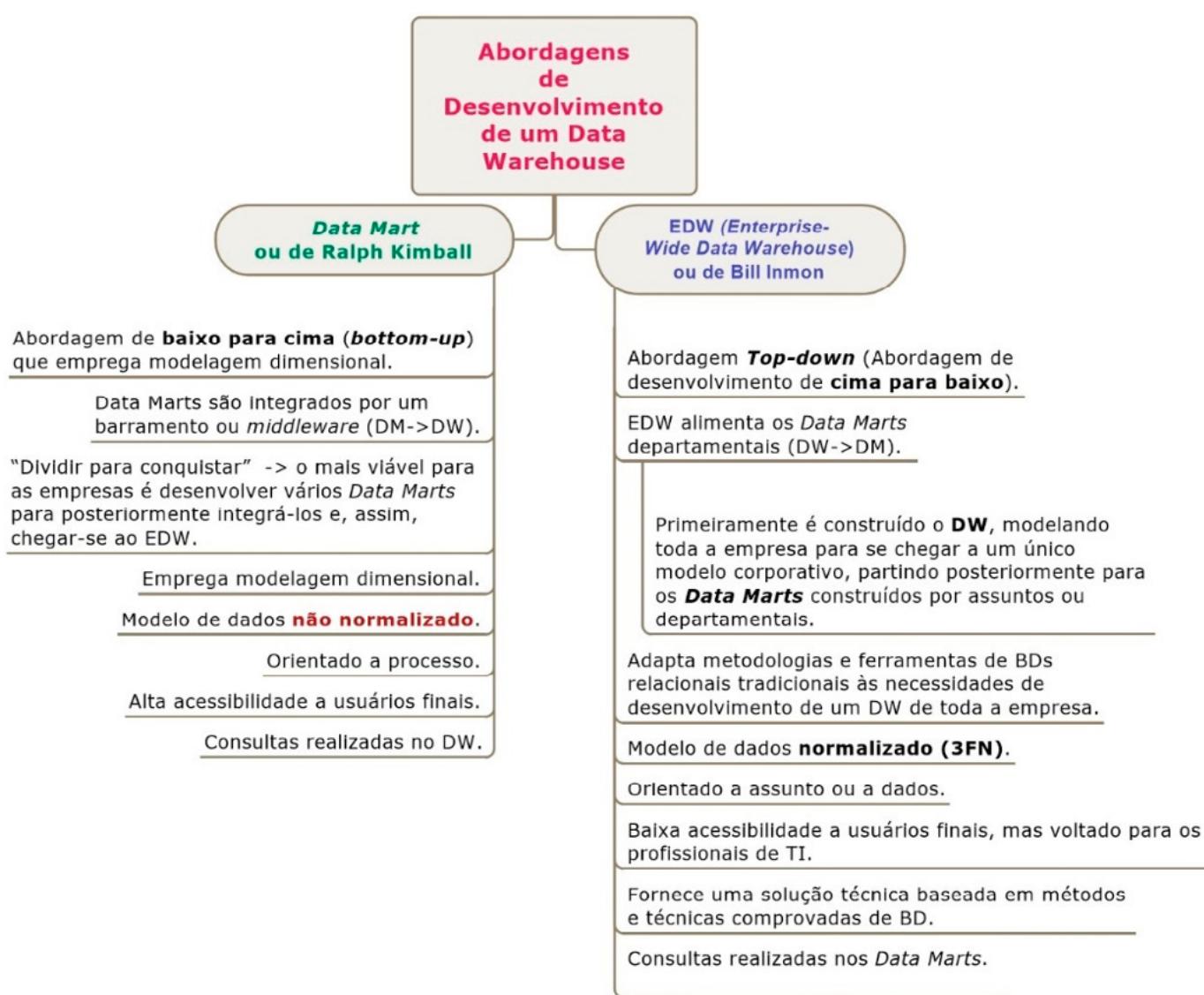


Figura. Abordagens de Desenvolvimento de um DW.

Fonte: Quintão (2020)

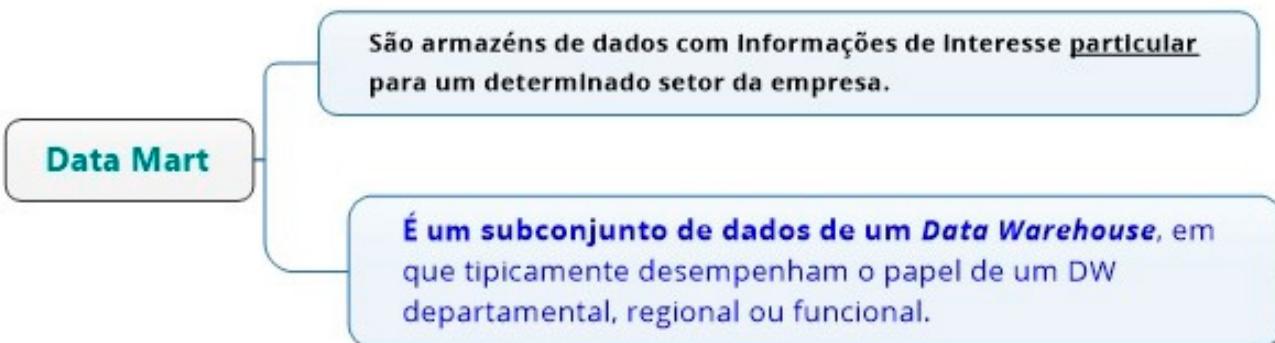


Figura. Data Mart. Fonte: Elaboração Própria (2020)

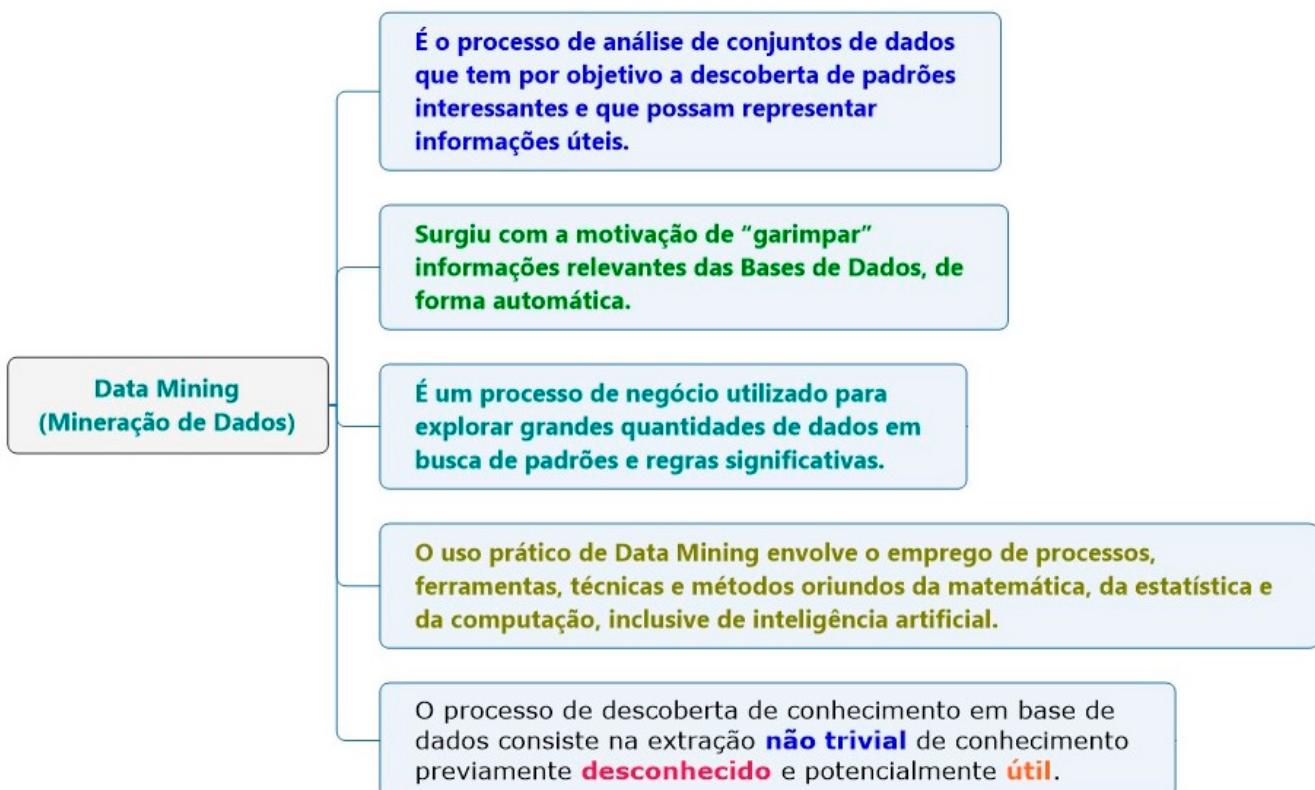


Figura. Data Mining. Fonte: Quintão (2020)

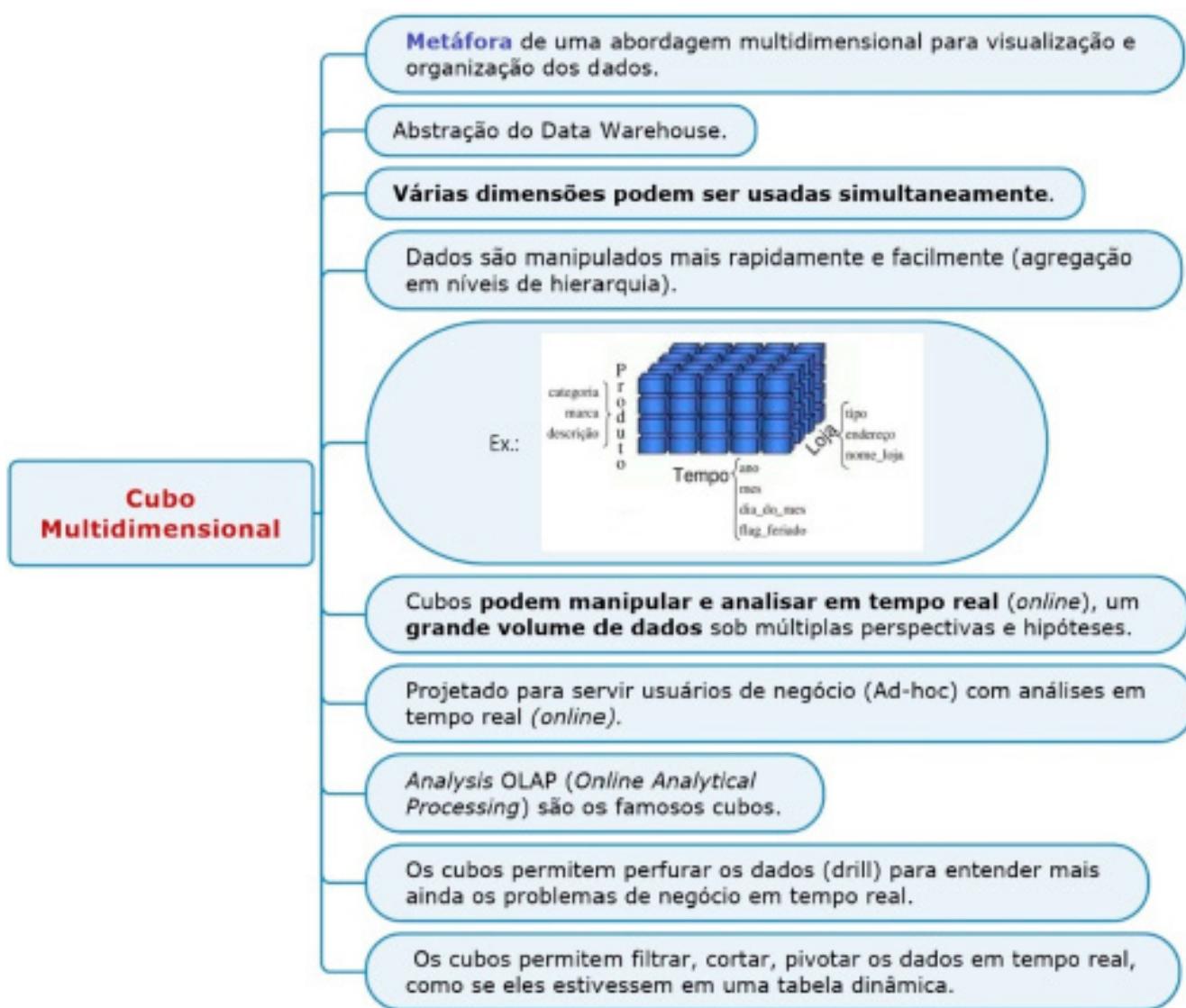


Figura. Cubo Multidimensional. Fonte: Quintão (2020)



Figura. Tipos de Tabelas de Fatos. Fonte: Quintão (2020)

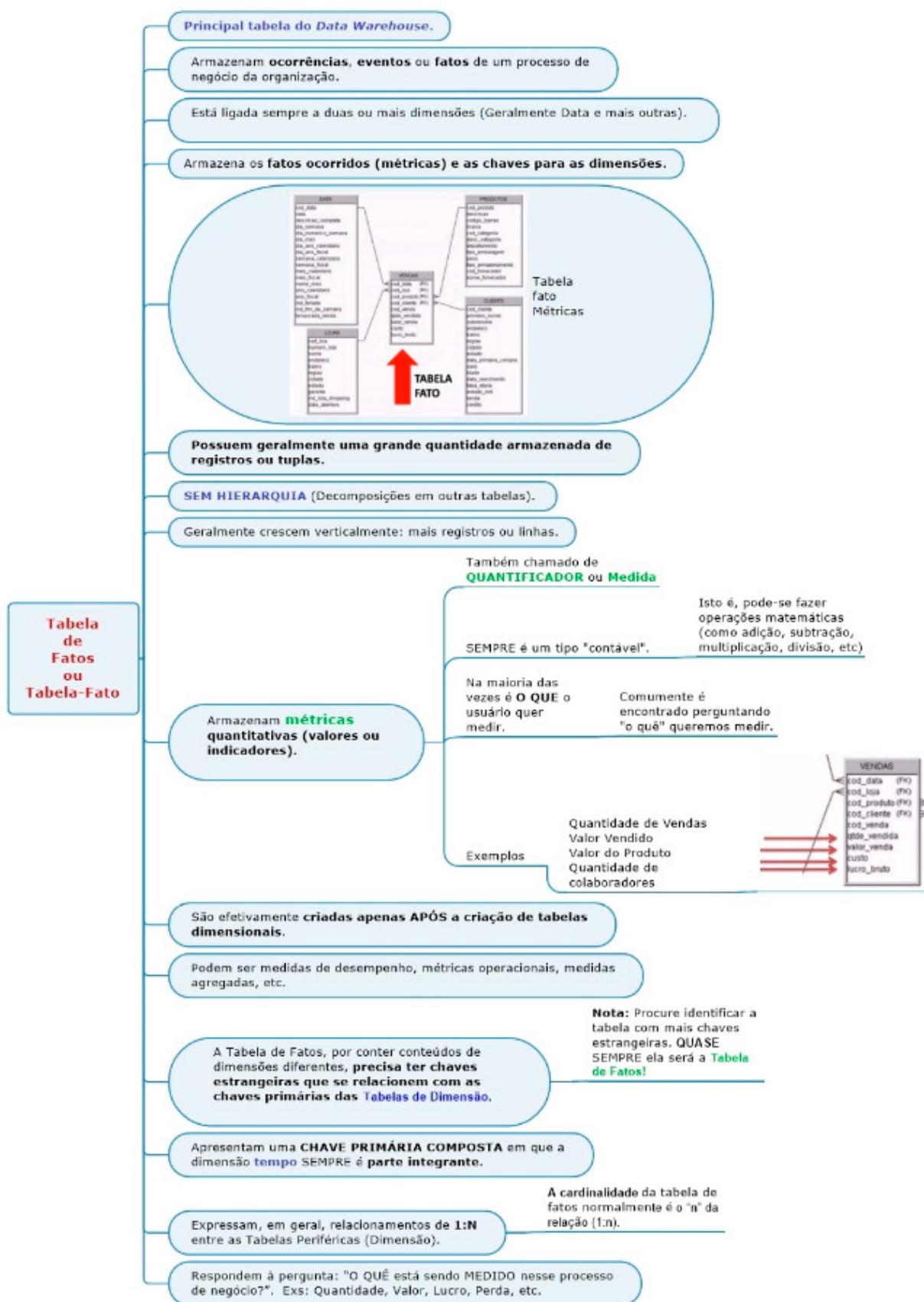


Figura. Tabela de Fatos. Fonte: Quintão (2020)

O conteúdo deste livro eletrônico é licenciado para MARIO LUIS DE SOUZA - 41250799864, vedada, por quaisquer meios e a qualquer título, a sua reprodução, cópia, divulgação ou distribuição, sujeitando-se aos infratores à responsabilização civil e criminal.

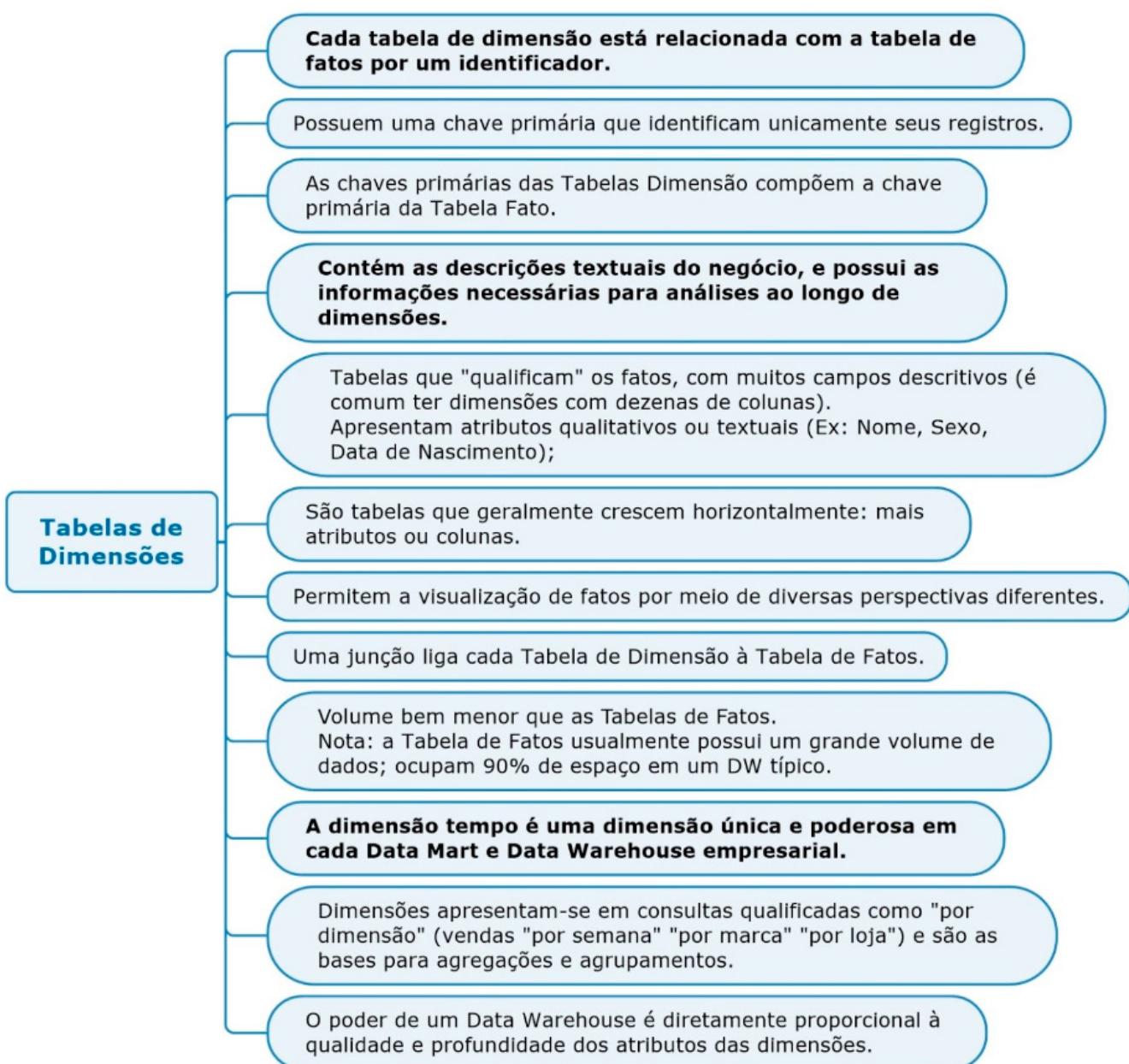


Figura. Tabelas de Dimensões. Fonte: Quintão (2020)

Granularidade refere-se ao nível de sumarização (**resumo**) dos elementos e de detalhes disponíveis. É considerado o aspecto mais importante de um DW.

Granularidade e **detalhamento** são grandezas **inversamente** proporcionais:

- Quanto mais detalhe, menor é a granularidade (Baixa granularidade = mais detalhado).
- Quanto menos detalhe, maior é a granularidade (Alta granularidade = menos detalhado).

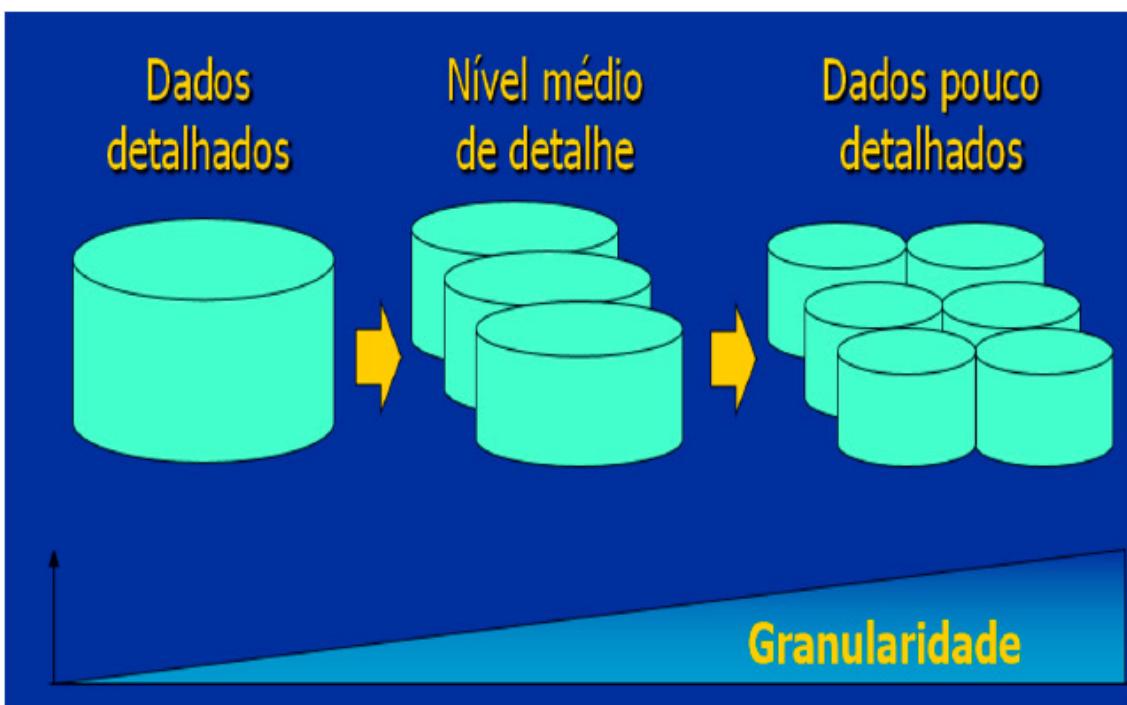


Figura. Granularidade e Detalhamento

Granularidade e **sumarização** são grandezas **diretamente** proporcionais:

- Quanto maior a sumarização, maior a granularidade.
- Quanto menor a sumarização, menor a granularidade.

QUESTÕES COMENTADAS EM AULA

001. (CESGRANRIO/BANCO DA AMAZÔNIA/TÉCNICO CIENTÍFICO/TECNOLOGIA DA INFORMAÇÃO/2018) Um Data Warehouse é recomendado para armazenar dados

- a) summarizados de um departamento.
- b) summarizados de toda a empresa para apoio à decisão e utilização de ferramentas OLAP.
- c) detalhados de toda a empresa para apoio à decisão e utilização de ferramentas OLAP.
- d) detalhados gerados por sistemas de informação transacionais.
- e) históricos detalhados de todas as transações realizadas em um determinado período de tempo.

002. (CESPE/DEPEN/AGENTE PENITENCIÁRIO FEDERAL/TECNOLOGIA DA INFORMAÇÃO/2015) Acerca de datawarehouse e datamining, julgue o item subsequente: [O datawarehouse possibilita a análise de grandes volumes de dados, que, por sua vez, permitem a realização de uma melhor análise de eventos futuros].

003. (ESAF/MPOG/2008/ADAPTADA) Algumas pessoas têm considerado que os *Data Warehouses* são uma extensão de visões de banco de dados. Porém, as visões fornecem apenas um subconjunto das funções e das capacidades dos *Data Warehouses*. Com relação às diferenças e similaridades entre as visões e os *Data Warehouses*, é correto afirmar que tanto os *Data Warehouses* quanto as visões fornecem, frequentemente, grandes quantidades de dados integrados e temporais, geralmente mais do que é contido em um banco de dados.

004. (VUNESP/TJ-SP/2012) Uma das técnicas utilizadas no projeto de um data warehouse corporativo consiste no uso da chamada matriz de barramento, na qual as linhas e colunas representam, respectivamente,

- a) cubos e medições.
- b) data staging e cubos.
- c) cardinalidades e hierarquias.
- d) dimensões e cardinalidades.
- e) processos de negócio e dimensões.

005. (FUNCAB/MDA/MINISTÉRIO DO DESENVOLVIMENTO AGRÁRIO /ANALISTA DE Business intelligence/2014)

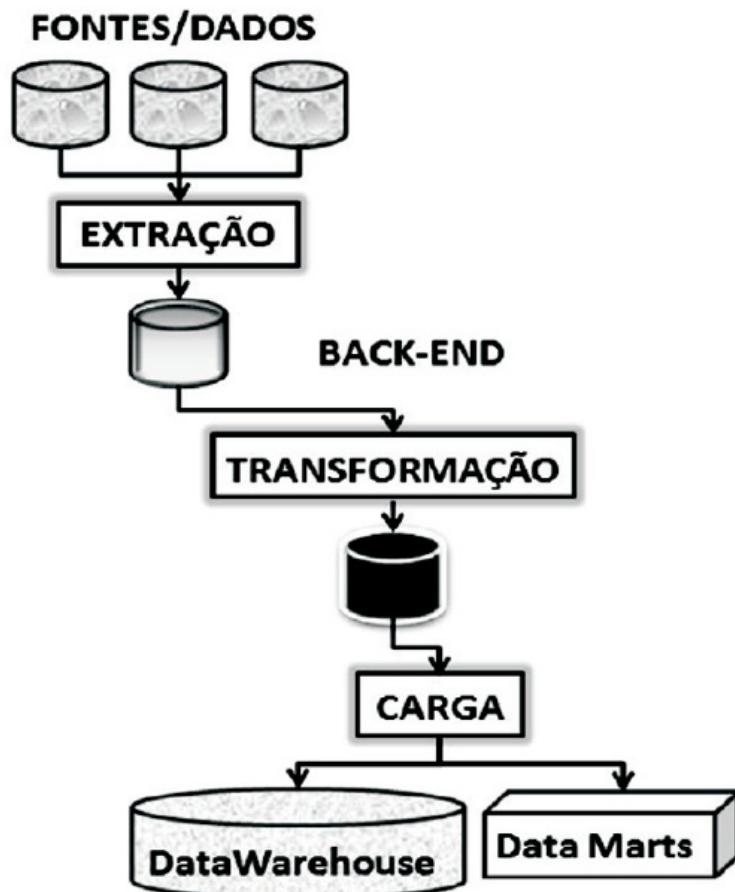


Figura. Modelo de organização para as próximas 3 questões.

A principal atividade de EXTRAÇÃO é:

- a) considerar logs de eventos e arquivos de controle.
- b) coletar dados das fontes externas transferindo-os para o ambiente de DW.
- c) carregar as dimensões considerando os tipos de hierarquias estáticas.
- d) tratar as inconsistências de dados resultantes da transcrição de dados.
- e) executar conversões de formatos para códigos geográficos dos países.

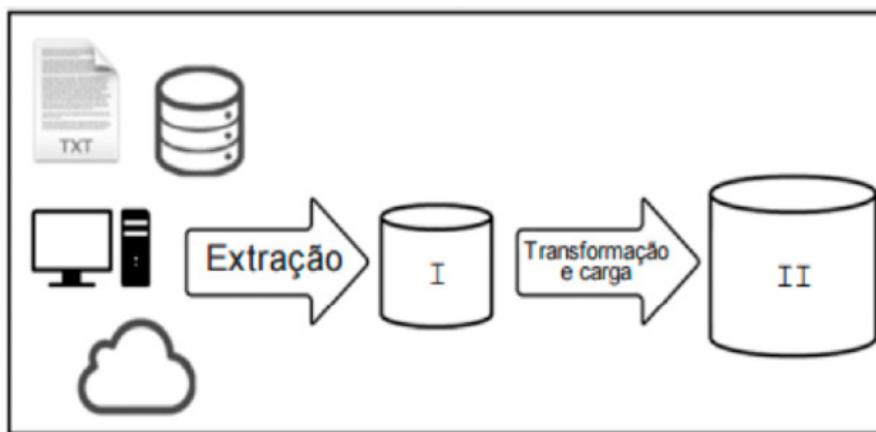
006. (FUNCAB/MDA/MINISTÉRIO DO DESENVOLVIMENTO AGRÁRIO /ANALISTA DE BUSINESS INTELLIGENCE/2014) A principal atividade de TRANSFORMAÇÃO é:

- a) aplicar regras aos dados extraídos para ajustá-los antes de serem carregados.
- b) analisar impactos das alternativas do código de aplicação no sistema fonte.
- c) avaliar aspectos de performance por meio do uso de paralelismo.
- d) obter dados de natureza estruturada e não estruturada.
- e) carregar tabelas Fato e fazer mapeamento das chaves.

007. (FUNCAB/MDA/MINISTÉRIO DO DESENVOLVIMENTO AGRÁRIO /ANALISTA DE BUSINESS INTELLIGENCE/2014) A principal atividade de CARGA é:

- a) realizar a limpeza dos dados extraídos para melhorar a qualidade dos dados.
- b) definir a periodicidade da transcrição dos dados a serem coletados.
- c) otimizar a qualidade dos dados por meio da conversão de formatos.
- d) separar e concatenar dados visando eliminar inconsistências nos dados.
- e) estruturar e carregar os dados para o DW seguindo o modelo dimensional.

008. (FCC/AL-MS/TÉCNICO DE INFORMÁTICA/2016) Considere a figura abaixo.



No processo de ETL mostrado na figura, I e II correspondem, respectivamente, a:

- a) OLTP e Data Warehouse.
- b) OLTP e Staging Area.
- c) Data Mart e Staging Area.
- d) Staging Area e OLTP.
- e) Staging Area e Data Warehouse.

009. (IADES/METRÔ-DF/TÉCNICO EM ELETRÔNICA/2014) É o conjunto de técnicas e procedimentos para a extração de informações em dispositivos de armazenamento digital, que não podem ser acessados de modo convencional pelo usuário ou pelo sistema. Com base no exposto, é correto afirmar que essas informações apresentam o conceito de

- a) recuperação de dados.
- b) backup corrompido.
- c) mineração de dados.
- d) backup interrompido.
- e) recuperação de dispositivos.

010. (COPEVE UFAL/MPE-AL/ANALISTA DO MINISTÉRIO PÚBLICO/2012) Um modelo de banco de dados multidimensional está mais fortemente relacionado com:

- a) data warehouse.
- b) modelo relacional.
- c) bancos hierárquicos.
- d) modelo em 3 camadas.
- e) banco de dados distribuídos.

011. (FUNCAB/MDA/ADMINISTRADOR DE BANCO DE DADOS/2014) O modelo de dados denominado “multidimensional” se aplica para banco de dados com a tecnologia:

- a) relacional.
- b) hierárquica.
- c) datamining.
- d) distribuída
- e) data warehouse.

012. (CESPE/SE-DF/ANALISTA DE GESTÃO EDUCACIONAL/TECNOLOGIA DA INFORMAÇÃO/2017) Com relação aos conceitos de modelagem multidimensional de dados para inteligência computacional, julgue o seguinte item. [Diferentemente da estrutura relacional, a estrutura multidimensional oferece baixa redundância de dados e suporte a normalização até a segunda forma normal].

013. (FAURGS/HCPA/ANALISTA DE TI- BANCO DE DADOS/2016) Em um modelo dimensional, a tabela fatos armazena:

- a) estatísticas sobre os metadados.
- b) as restrições de domínio do negócio.
- c) descrições textuais das dimensões.
- d) medições numéricas do negócio.
- e) o tempo de processamento das transações.

014. (COSEAC/DATAPREV/ANALISTA DE TECNOLOGIA DA INFORMAÇÃO/ANO/ADAPTA-DA) Sobre modelagem multidimensional, o cubo:

I – é uma representação intuitiva, pois todas as dimensões coexistem para todo ponto no cubo e são independentes umas das outras;

II – é, de fato, apenas uma metáfora visual;

III – serve para descrever requisitos funcionais.

Acerca dos itens acima mencionados, apenas:

- a) I e III estão corretos;
- b) Somente I está correto;
- c) I e II estão corretos;
- d) II e III estão corretos;
- e) III está correto.

015. (CESGRANRIO/PETROBRAS/ANALISTA DE SISTEMAS JÚNIOR) Em um banco de dados multidimensional, os dados estão conceitualmente armazenados e organizados em:

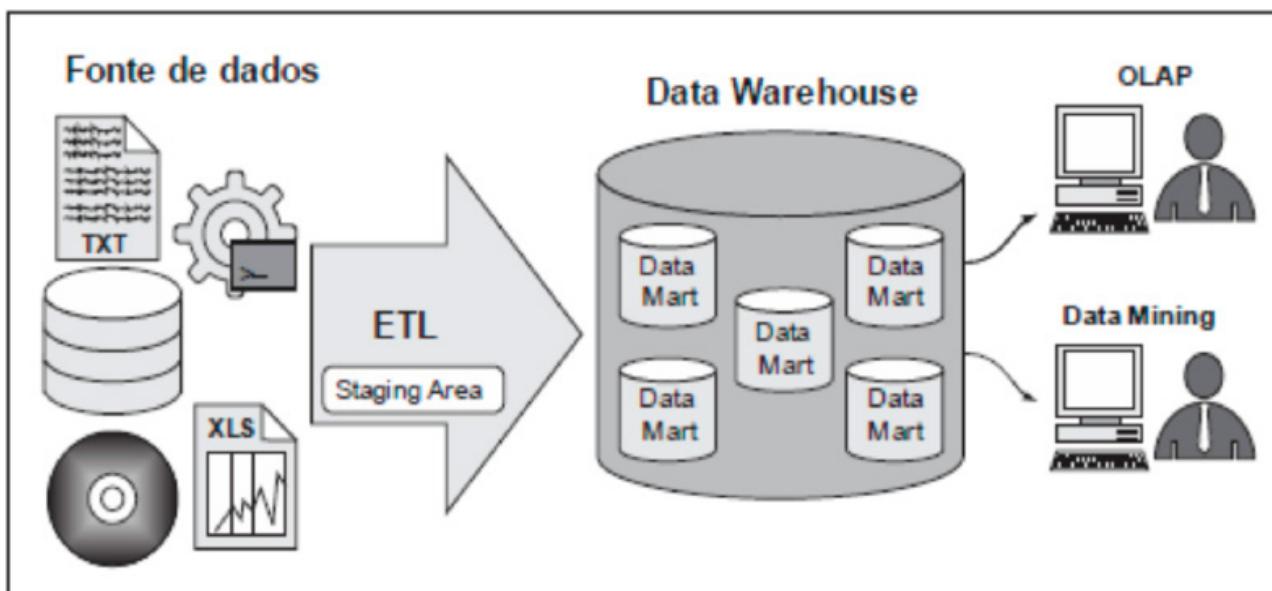
- a) classes e objetos.
- b) cubos e hipercubos.
- c) partições e índices.
- d) consultas materializadas e sumários.
- e) estrelas e constelações.

016. (CESPE/MPE-PI/ANALISTA MINISTERIAL/2012) Um cubo de dados é a representação multidimensional dos dados não agregados na qual é necessário que as dimensões tenham o mesmo tamanho.

017. (IDECAN/INMETRO/ANÁLISE E DESENVOLVIMENTO DE SISTEMAS/2015) A modelagem multidimensional é uma técnica de concepção e visualização de um modelo de dados de um conjunto de medidas que descrevem aspectos comuns de negócios. Um modelo multidimensional é formado por três elementos básicos. Assinale-os:

- a) Esquema, fatos e itens.
- b) Fatos, dimensões e itens.
- c) Medidas, esquema e fatos.
- d) Fatos, dimensões e medidas
- e) Dimensões, medidas e esquema.

018. (FCC/TRT-4^a REGIÃO/ANALISTA JUDICIÁRIO/TECNOLOGIA DA INFORMAÇÃO/2015) Considere a arquitetura geral de um sistema de BI- *Business intelligence* mostrada na figura abaixo.



Nesta arquitetura

- a) Data Marts representam áreas de armazenamento intermediário criadas a partir do processo de ETL. Auxiliam na transição dos dados das fontes OLTP para o destino final no Data Warehouse.
- b) OLAP é um subconjunto de informações extraído do Data Warehouse que pode ser identificado por assuntos ou departamentos específicos. Utiliza uma modelagem multidimensional conhecida como modelo estrela.
- c) os dados armazenados no Data Warehouse são integrados na base única mantendo as convenções de nomes, valores de variáveis e outros atributos físicos de dados como foram obtidos das bases de dados originais.
- d) o Data Warehouse não é volátil, permite apenas a carga inicial dos dados e consultas a estes dados. Além disso, os dados nele armazenados são precisos em relação ao tempo, não podendo ser atualizados.
- e) Data Mining se refere ao processo que, na construção do Data Warehouse, é utilizado para composição de análises e relatórios, armazenando dados descritivos e qualificando a respectiva métrica associada.

019. (FCC/TRT-MS/ANALISTA JUDICIÁRIO-TI/2017) Uma das formas de apresentação de um banco de dados multidimensional é através do modelo estrela. No centro de um modelo estrela encontra-se a tabela de:

- a) dimensão e, ao seu redor, as tabelas de fatos.
- b) dimensão, cuja chave primária deve ser composta.
- c) núcleo e, ao seu redor, as tabelas de nível.
- d) fatos, cuja chave primária deve ser simples.
- e) fatos e, ao seu redor, as tabelas de dimensões.

020. (FCC/SABESP/ANALISTA DE GESTÃO/SISTEMAS/2018) Um Analista está trabalhando em um Data Warehouse – DW que utiliza no centro do modelo uma única tabela que armazena as métricas e as chaves para as tabelas ao seu redor (que descrevem os dados que estão na tabela central) às quais está ligada. O esquema de modelagem utilizado pelo DW, a denominação da tabela central e a denominação das tabelas periféricas são, respectivamente,

- a) floco de neve, base, granulares.
- b) estrela, fato, dimensões.
- c) constelação, fato, granulares.
- d) atomic, base, branches.
- e) anel, base, dimensões.

021. (FCC/TRT-AL/ANALISTA JUDICIÁRIO/TI/2011) O modelo estrela, como estrutura básica de um modelo de dados multidimensional, possui uma configuração típica composta de uma entidade central:

- a) mining e um conjunto de entidades fatos.
- b) mining e um conjunto de entidades dimensões.
- c) mining e um conjunto de entidades roll-up.
- d) dimensão e um conjunto de entidades fatos.
- e) fato e um conjunto de entidades dimensões.

022. (UPENET/UPE/ANALISTA DE SISTEMAS/ENGENHARIA DE SOFTWARE/2017) O modelo dimensional de um data warehouse, no qual todas as tabelas relacionam-se diretamente com a tabela de fatos, de forma que as tabelas dimensionais devem conter todas as descrições que são necessárias para se definir uma classe, é denominado de:

- a) Floco de neve.
- b) Estrela.
- c) Barramento.
- d) Árvore.
- e) Anel.

023. (CESPE/AUDITOR MUNICIPAL DE CONTROLE INTERNO/TECNOLOGIA DA INFORMAÇÃO/DESENVOLVIMENTO DE SISTEMAS/2018) Com relação à modelagem dimensional e à otimização de bases de dados para *business intelligence*, julgue o item subsequente.

Na modelagem multidimensional utilizada em *Data Warehouses* para se prover melhor desempenho, a tabela fato central deve relacionar-se às suas dimensões por meio da chave primária oriunda da fonte de dados original. O valor dessa chave deve ser idêntico ao da fonte, para que tenha valor semântico e garanta que o histórico das transações seja mantido.

QUESTÕES DE CONCURSO

024. (INÉDITA/2020) O Data Mining é uma coleção de dados orientados por assuntos, integrados, variáveis com o tempo e voláteis, para dar suporte ao processo de tomada de decisão.



Bill Inmon destaca que o “**Data Warehouse é uma coleção de dados orientados por assuntos, integrados, variáveis com o tempo e não voláteis, para dar suporte ao processo de tomada de decisão**”.

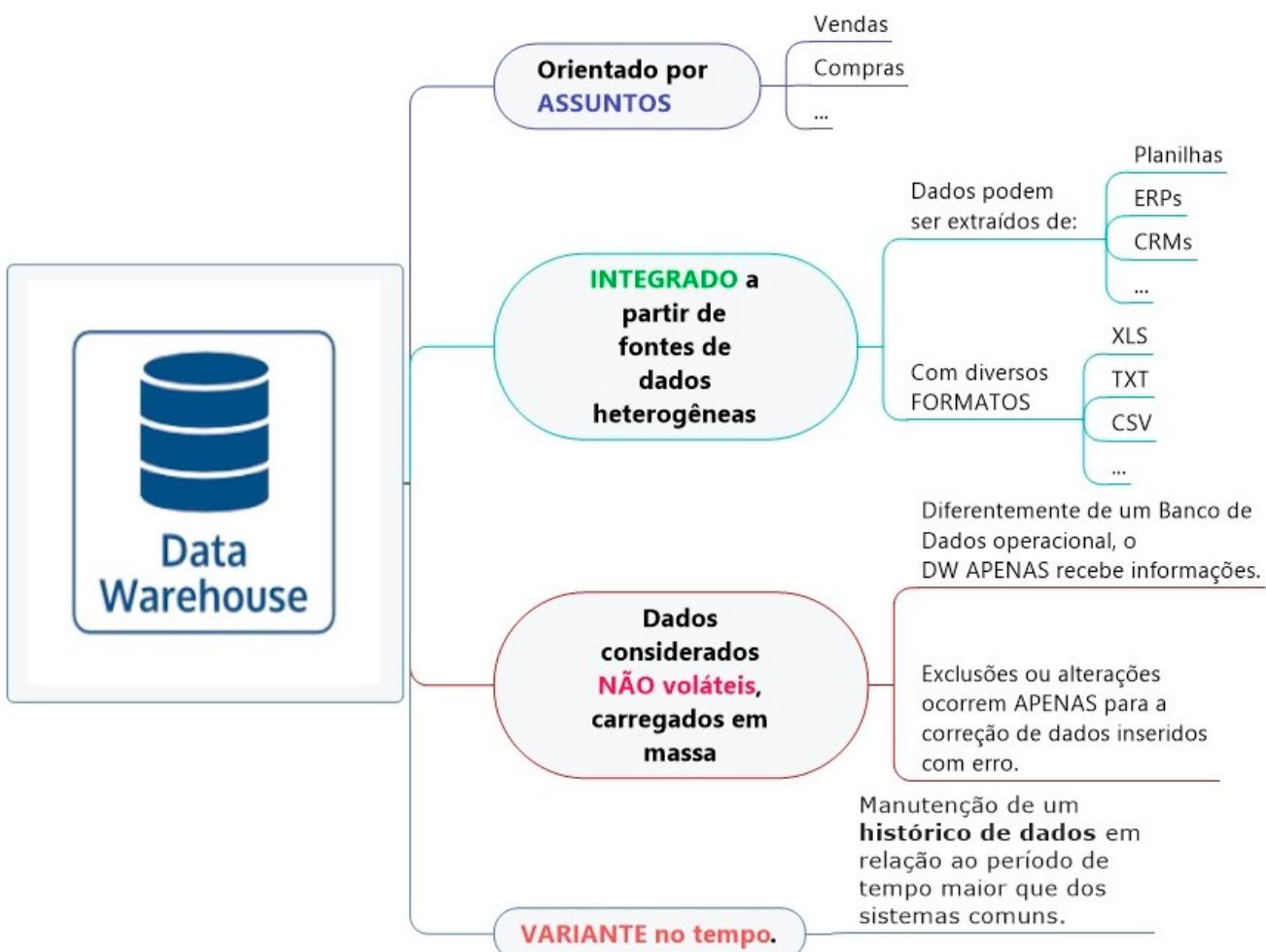


Figura. Características do Data Warehouse

Vejamos estas características de modo mais detalhado:

É orientado a assuntos:

- O DW é criado de forma orientada em torno dos principais assuntos da organização (clientes, vendas, produtos, estoques etc.).

É integrado:

- Na convenção consistente de nomes;
- Na estrutura consistente da codificação etc.

É não volátil:

- De modo geral, somente existem cargas e acessos (atualizações existem, mas **devido às manutenções**).

É variante no tempo:

- Qualquer unidade de dado deve ter precisão em um ponto bem definido do tempo.

A principal ideia do Data Warehouse é construir um depósito no qual será mantida a **memória histórica dos dados**, possibilitando a utilização dos mesmos para consulta e análise estratégica para a tomada de decisão!

Data Mining é o processo de **descoberta de novas correlações, padrões e tendências entre as informações de uma empresa**, por meio da análise de grandes quantidades de dados armazenados em bancos de dados usando técnicas de reconhecimento de padrões, estatísticas e matemáticas.

Errado.

025. (CESPE/EBSERH/ANALISTA DE TECNOLOGIA DA INFORMAÇÃO/2018) Julgue o item que se segue, a respeito de arquitetura e tecnologias de sistemas de informação.

A descoberta de novas regras e padrões em conjuntos de dados fornecidos, ou aquisição de conhecimento indutivo, é um dos objetivos de data mining.



O conhecimento é classificado em **indutivo** e **dedutivo**. O conhecimento dedutivo deduz novas informações baseadas na aplicação de regras lógicas predefinidas de dedução sobre dados existentes. **O Data Mining apoia o conhecimento indutivo**, que descobre novas regras e padrões nos dados fornecidos. (ELMASRI, NAVATHE, 2005).

Certo.

026. (CESPE/TCE-PA/AUDITOR DE CONTROLE EXTERNO/ÁREA INFORMÁTICA/ANALISTA DE SISTEMA/2016) Julgue o item a seguir, em relação a data warehouse e data mining.

No contexto de data mining, o processo de descoberta de conhecimento em base de dados consiste na extração não trivial de conhecimento previamente desconhecido e potencialmente útil.



O **Data Mining** é considerado uma parte de um processo maior conhecido como KDD (Knowledge Discovery in Databases) – em português, Descoberta de Conhecimento em Bases de Dados –, que, segundo Eduardo Gimenes, busca extrair informação válida, **previamente desconhecida** e de máxima abrangência a partir de grandes bases de dados, usando-as para efetuar decisões cruciais.

Addrians & Zantinge (1996) cita que o Data Mining, ou Mineração de Dados, permite a **extração não trivial de conhecimento previamente desconhecido e potencialmente útil de um banco de dados**.

Outra definição clássica para esse termo é a de FAYYAD et al. (1996): “processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados”, ou seja, buscar por informações em uma base de dados com os mais variados atributos e assuntos.

Já Laudon&Laudon destaca que Data Mining é a análise de grandes quantidades de dados a fim de encontrar padrões e regras que possam ser usadas para orientar a tomada de decisões e prever o comportamento futuro.

Mineração de dados, ou data mining, é o processo de análise de conjuntos de dados que tem por objetivo a descoberta de padrões interessantes e que possam representar informações úteis.

Certo.

027. (CESPE/TCE-PA/AUDITOR DE CONTROLE EXTERNO/ÁREA INFORMÁTICA/ANALISTA DE SISTEMA/2016) Julgue o item a seguir, em relação a data warehouse e data mining. [Data warehouse é um repositório de dados dinâmico, que sofre alterações frequentes, de modo a permitir que sejam feitas consultas em base de dados constantemente atualizada].



A questão tornou-se inadequada ao destacar que o Data Warehouse sofre alterações frequentes. **Os dados em um Data Warehouse não são voláteis**, ou seja, eles não mudam, salvo quando é necessário fazer correções de dados previamente carregados. No ambiente operacional, ao contrário, os dados sofrem alterações frequentes, pois são, em geral, atualizados registro a registro, em múltiplas transações.

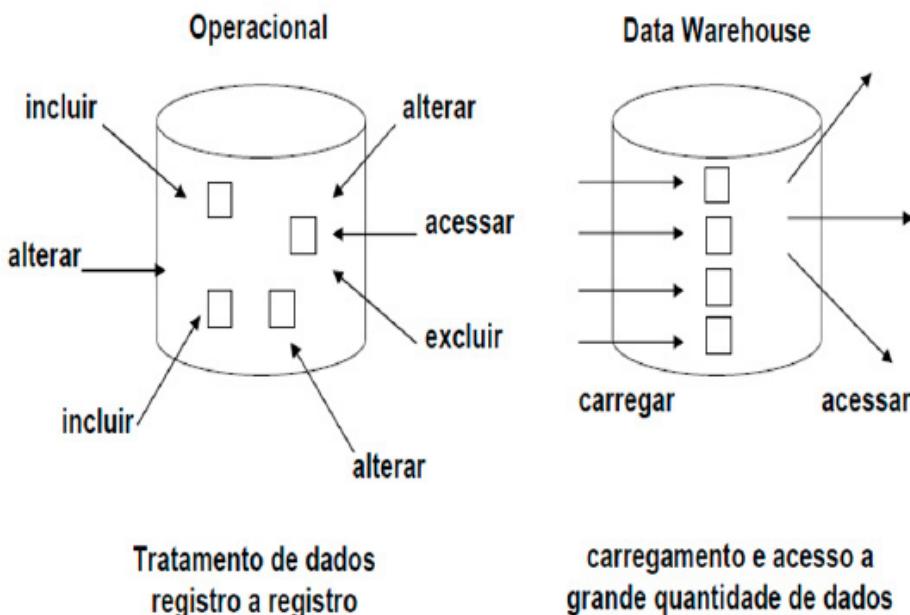


Figura. Carregamento e acesso a dados no DW

Errado.

028. (CESPE/TCU/AUDITOR FEDERAL DE CONTROLE EXTERNO/2015) O uso prático de data mining envolve o emprego de processos, ferramentas, técnicas e métodos oriundos da matemática, da estatística e da computação, inclusive de inteligência artificial.



Data Mining é a etapa do KDD (Knowledge Discovery in Databases - Descoberta de Conhecimento em Bases de Dados) em que os dados preparados são processados, ou seja, é onde se faz a mineração dos dados propriamente dita.

Nesse contexto, são aplicadas **técnicas** para identificação de padrões sobre os dados disponíveis. Esse processo tem o intuito de trabalhar os dados registrados ao longo do tempo de vida de um negócio a fim de se identificar **padrões** que representam alguma informação sobre o comportamento do negócio.

A figura a seguir apresenta um esquema que descreve o KDD.



O processo de KDD envolve os seguintes passos:

- **entendimento do domínio da aplicação** e identificação do objetivo do processo de KDD;
- **seleção**: criação de um conjunto-alvo de dados;
- **pré-processamento**: limpeza de dados e operações básicas como remoção de ruído, tratamento para a falta de dados etc.;
- **transformação**: encontrar características úteis para representar os dados, conforme o objetivo definido e realizar a redução ou transformação da dimensionalidade;
- **mineração de dados**: **casar os objetivos do processo de KDD com um método particular de mineração de dados e realizar a análise exploratória e seleção de modelo e hipótese, buscando padrões de interesse**;
- **interpretação**: interpretar e avaliar os padrões minerados, podendo retornar a passos anteriores caso seja necessário;
- **agir a partir do conhecimento descoberto**.

No **Data Mining** são aplicadas **técnicas** e ferramentas **matemáticas, estatísticas e computacionais** visando a extração de informações implícitas, **previamente desconhecidas**, e potencialmente **úteis**, a partir de dados. Para isso é necessário um programa de computador para detectar padrões e regularidades nos dados. Os termos destacados na questão estão de acordo com essa teoria, e a assertiva é correta.

Certo.

029. (CESPE/TCU/AUDITOR FEDERAL DE CONTROLE EXTERNO/2015) No ambiente organizacional, devido à grande quantidade de dados, não é recomendado o emprego de data mining para atividades ligadas a marketing.



Não existe restrição de setor para o emprego das técnicas de Data mining. Em relação à quantidade de dados também não há restrição desde que a empresa possua capacidade de tratamento (armazenagem e processamento) dos dados.

Errado.

030. (CESPE/TCU/AUDITOR FEDERAL DE CONTROLE EXTERNO/2015) A finalidade do uso do data mining em uma organização é subsidiar a produção de afirmações conclusivas acerca do padrão de comportamento exibido por agentes de interesse dessa organização.



A **mineração de dados (Data Mining)** pode ser utilizada para encontrar novos padrões, visando:

- **Predição**: o sistema encontra padrões para predizer (prever) o comportamento futuro de algumas entidades;

- **Descrição:** o sistema encontra padrões relevantes sobre os dados, em uma forma compreensível para o usuário.

A mineração de dados envolve o ajuste de modelos aos dados observados ou a determinação de padrões a partir destes dados. Os modelos ajustados fazem o papel do conhecimento inferido, sobre o qual o julgamento humano subjetivo é exigido na verificação se os modelos refletem conhecimento útil ou interessante.

Certo.

031. (CESPE/DEPEN/AGENTE PENITENCIÁRIO FEDERAL/TECNOLOGIA DA INFORMAÇÃO/2015) Acerca de datawarehouse e datamining, julgue o item subsequente: [O datawarehouse possibilita a análise de grandes volumes de dados, que, por sua vez, permitem a realização de uma melhor análise de eventos futuros].



Um **Data Warehouse** é um **repositório de informações colhidas de várias origens, armazenadas sob um esquema unificado, em um único local**.

Quando reunidos, os dados são armazenados por muito tempo, permitindo o acesso a dados históricos.

Ainda, o desenho da base de dados favorece os relatórios, a análise de grandes volumes de dados e a obtenção de informações estratégicas que podem **facilitar a tomada de decisão**.

O **Data Warehouse** **possibilita a análise de grandes volumes de dados**, coletados dos sistemas transacionais (OLTP). São as chamadas **séries históricas que possibilitem uma melhor análise de eventos passados, oferecendo suporte às tomadas de decisões presentes e a previsão de eventos futuros**.

Errado.

032. (CESPE/TJ-SE/ANALISTA JUDICIÁRIO/ANÁLISE DE SISTEMAS/2014) Com relação a sistemas de suporte à decisão, julgue os seguintes itens.

Assim como o Data Mining, os Data Marts são voltados para a obtenção de informações estratégicas de maneira automática, ou seja, com o mínimo de intervenção humana a partir da análise de dados oriundos de *Data Warehouses*.



O **Data Warehouse** é um conjunto de dados orientado por assuntos, não volátil, variável com o tempo e integrado, criado para dar suporte à decisão.

Data Marts são subconjuntos de dados de um Data Warehouse.

E, por fim, é o **Data Mining (Mineração de Dados)** o processo de identificar informações relevantes, tais como padrões, associações, mudanças, anomalias e estruturas, em grandes conglomerados de dados que estejam em banco de dados ou outros repositórios de informações.

Errado.

033. (CESPE/TJ-SE/PROGRAMAÇÃO DE SISTEMAS/2014) Os dados armazenados em um DataWarehouse devem estar integrados, temporalmente identificados, orientados ao assunto e devem ser protegidos, de modo a se evitar o acesso do usuário.



Segundo Bill Inmon (1997), um **data warehouse** é uma coleção de dados **orientada a assuntos, integrada, não volátil e variante ao longo do tempo**, para suporte à tomada de decisão.

É orientado a assuntos

- o DW é criado de forma orientada em torno dos principais assuntos da organização (clientes, vendas, produtos, estoques, etc.).

É integrado

- Na convenção consistente de nomes;
- Na estrutura consistente da codificação; etc.

É não volátil

- Somente existem cargas e acessos (atualizações existe, mas devido às manutenções).

É variante no tempo

- Qualquer unidade de dado deve ter precisão em um ponto bem definido do tempo.

Por fim, deve-se definir o nível de segurança a ser implementado no DW, de modo a que se permita o **acesso devido aos usuários autorizados**. A questão afirma que o acesso do usuário deve ser evitado, o que está errado!

Errado.

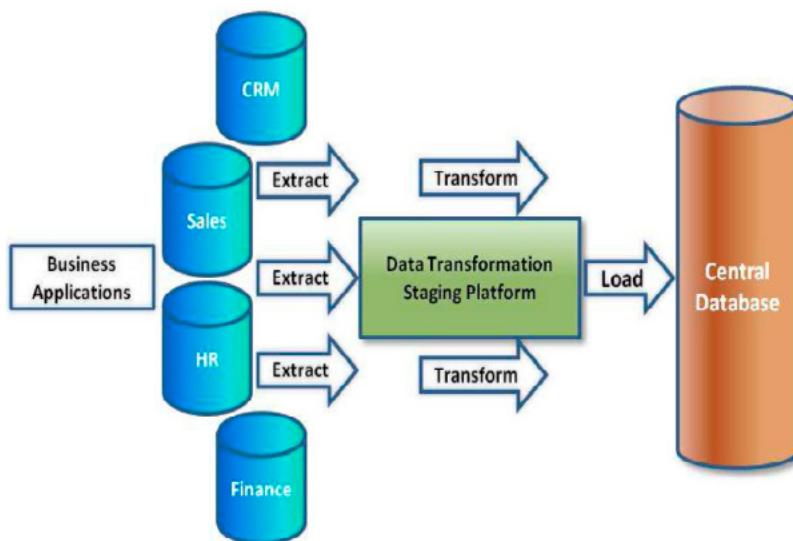
034. (CESPE/CORREIOS/ANALISTA DE CORREIOS/ANALISTA DE SISTEMAS/DESENVOLVIMENTO DE SISTEMAS/2011) Acerca de soluções de suporte a decisão, julgue os dois itens seguintes.

As ferramentas de software ETL (extract transform load) têm como função a extração de dados de diversos sistemas, a transformação desses dados de acordo com as regras de negócio e a carga dos dados em um data mart ou um DW.



As ferramentas de software **ETL (Extract Transform Load - Extração Transformação Carga)**, conforme visto na figura seguinte, têm como função a **extração** de dados a partir dos sistemas de origem e, geralmente, gravação em disco no ambiente de ETL antes de qualquer reestruturação dos dados.

Em seguida, é realizada a **transformação** desses dados de acordo com as regras de negócio e, por fim, a **carga** dos dados em um **Data Mart** ou em um **Data Warehouse (DW)**.



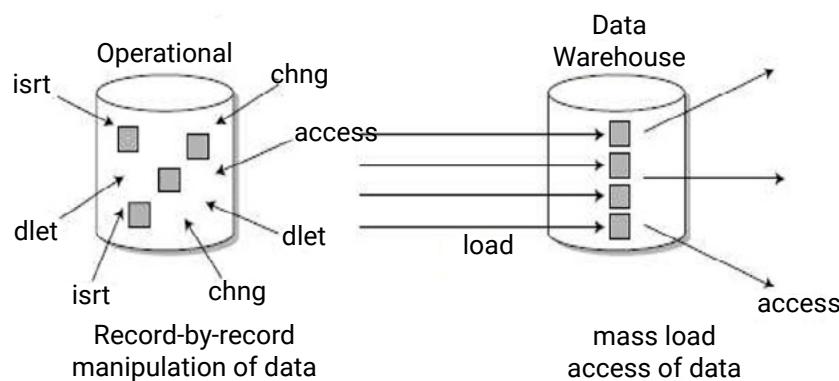
Certo.

035. (CESPE/CORREIOS/ANALISTA DE CORREIOS/ANALISTA DE SISTEMAS/DESENVOLVIMENTO DE SISTEMAS/2011) Em um ambiente data warehouse (DW), é possível a análise de grandes volumes de dados, os quais ficam disponíveis para serem alterados e manipulados pelo usuário.



Segundo Laudon & Laudon, um Data Warehouse (DW) é um banco de dados, com ferramentas de consulta e relatório, que **armazena dados atuais e históricos** extraídos de vários sistemas operacionais e consolidados para fins de análises e relatórios administrativos.

Em um ambiente DW, é possível a análise de grandes volumes de dados. No entanto, diferentemente do **ambiente operacional** em que os dados são, em geral, **atualizados registro a registro**, em múltiplas transações, o DW **permite apenas a carga inicial dos dados e consultas** a estes dados. Após serem **integrados e transformados**, os dados são **carregados em bloco** para o Data Warehouse, para que estejam disponíveis aos usuários **para acesso**.



Errado.

036. (FCC/DPE-AM/ANALISTA EM GESTÃO ESPECIALIZADO DE DEFENSORIA/ANALISTA DE BANCO DE DADOS/2018) Uma das características fundamentais de um ambiente de data warehouse está em

- a) servir como substituto aos bancos de dados operacionais de uma empresa, na eventualidade da ocorrência de problemas com tais bancos de dados.
- b) ser de utilização exclusiva da área de aplicações financeiras das empresas.
- c) proporcionar um ambiente que permita realizar análise dos negócios de uma empresa com base nos dados por ela armazenados.
- d) ser de uso prioritário de funcionários responsáveis pela área de telemarketing das empresas.
- e) armazenar apenas os dados mais atuais (máximo de 3 meses de criação), independentemente da área de atuação de cada empresa.



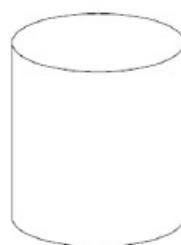
Os **Data Warehouse (DW)** podem ser utilizados por todas as áreas da empresa, não sendo de uso restrito dos funcionários do telemarketing, como citado na questão.

Bill Inmon destaca que o **Data Warehouse (DW)** é uma coleção de dados orientados por **assuntos, integrados, variáveis com o tempo e não voláteis**, para dar suporte ao processo de tomada de decisão com dados.

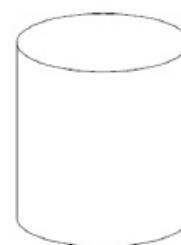
Por fim, cabe destacar que o **DW** armazena posições históricas das atividades no tempo. Enquanto o **dado de produção é atualizado de acordo com mudanças de estado do objeto em questão**, refletindo, em geral, o estado do objeto no momento do acesso. Em um DW, a cada ocorrência de uma mudança, uma nova entrada é criada, para marcar esta mudança.

É Variante no Tempo

Operacional



Data Warehouse



- Horizonte de tempo – atual a 60-90 dias
- Atualização de registros
- A estrutura chave pode ou não conter um elemento de tempo

- Horizonte de tempo – 5-10 anos
- Instantâneos sofisticados dos dados
- A estrutura chave contém um elemento de tempo


Figura. Variação de preços no tempo
Letra c.

037. (FCC/TRE-SP/ANALISTA JUDICIÁRIO/ANÁLISE DE SISTEMAS/2017) No processo ETL de um data warehouse:

- (I) a tradução de valores codificados conhecida como limpeza de dados (por exemplo, supondo que o sistema de origem armazena 1 para sexo masculino e 2 para feminino, mas o data warehouse armazena M para masculino e F para feminino) e
 - (II) a geração de valores de chaves substitutas (surrogate keys) **são**
- integrantes da etapa de transformação
 - integrantes das etapas de transformação e carga, respectivamente
 - manipulações de dados, aplicáveis no processo pivoting
 - elementos que possibilitam a conversão de drill-down para roll up
 - integrantes da etapa de carga



ETL, vem do inglês **Extract Transform Load**, ou seja, Extração Transformação Carga. O ETL visa trabalhar com toda a parte de extração de dados de fontes externas, transformação para atender às necessidades de negócios e carga dos dados dentro do Data Warehouse. Os projetos de Data Warehouse consolidam dados de diferentes fontes. A maioria dessas fontes tendem a ser bancos de dados relacionais ou flat files, mas podem existir outros tipos de fontes também. Um sistema ETL precisa ser capaz de se comunicar com bases de dados e ler diversos formatos de arquivos utilizados por toda a organização. E para que seja possível entender as etapas do processo de ETL faz-se necessário a compreensão de seus componentes que serão descritos a seguir.

Componentes do ETL:

- **Extração**: é a coleta de dados dos sistemas de origem (também chamados Data Sources ou sistemas operacionais), extraíndo-os e transferindo-os para o ambiente de DW, onde o sistema de ETL pode operar independente dos sistemas operacionais;
- **Limpeza, Ajustes e Consolidação (ou também chamada transformação)**: é nesta etapa que realizamos os devidos ajustes, podendo assim melhorar a qualidade dos dados e

consolidar dados de duas ou mais fontes. O estágio de transformação aplica uma série de regras ou funções aos dados extraídos para ajustar os dados a serem carregados. Algumas fontes de dados necessitarão de muito pouca manipulação de dados. Em outros casos, pode ser necessário trabalhar algumas transformações, como por exemplo, junção de dados provenientes de diversas fontes, seleção de apenas determinadas colunas e tradução de valores codificados (se o sistema de origem armazena 1 para sexo masculino e 2 para feminino, mas o *Data Warehouse* armazena M para masculino e F para feminino, por exemplo);

- **Entrega ou Carga dos dados:** consiste em fisicamente estruturar e carregar os dados para dentro da camada de apresentação seguindo o modelo dimensional. Dependendo das necessidades da organização, este processo varia amplamente. Alguns *Data Warehouses* podem substituir as informações existentes semanalmente, com dados cumulativos e atualizados, ao passo que outro DW (ou até mesmo outras partes do mesmo DW) podem adicionar dados a cada hora. A latência e o alcance de reposição ou acréscimo constituem opções de projeto estratégicas que dependem do tempo disponível e das necessidades de negócios;
- A parte de **Gerenciamento** é composta por serviços para auxiliar no gerenciamento do DataWarehouse. Aqui nós temos tasks específicas para gerenciamento de jobs, planos de backup, verificação de itens de segurança e compliance.

Fonte: Novais, Ramon Ramos de Castro. **MODELAGEM DIMENSIONAL**. <http://www.fatecsp.br/dti/tcc/tcc00071.pdf>. Último acesso 01/08/2017.

Letra a.

038. (FCC/DPE-RS/ANALISTA/BANCO DE DADOS/2017) Uma das técnicas bastante utilizadas em sistemas de apoio à decisão é o Data Mining, que se constitui em uma técnica

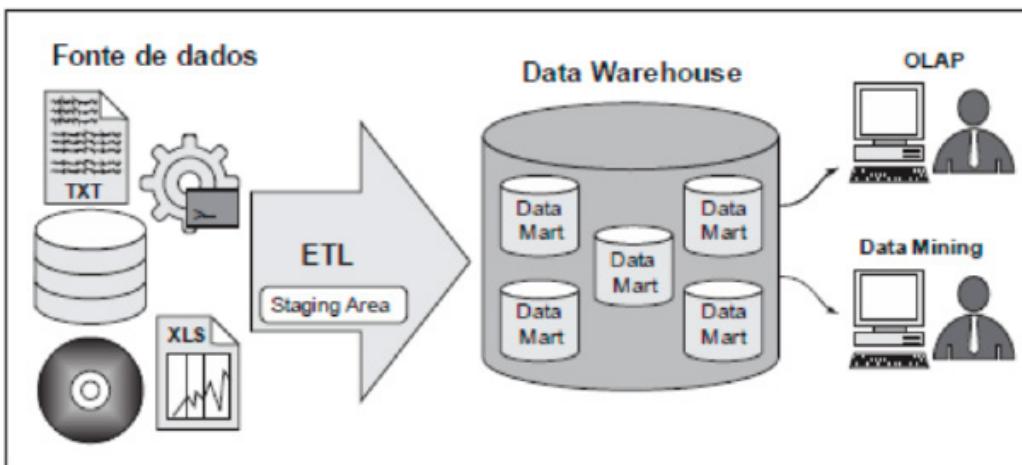
- a) para a exploração e análise de dados, visando descobrir padrões e regras, a princípio ocultos, importantes à aplicação.
- b) para se realizar a criptografia inteligente de dados, objetivando a proteção da informação.
- c) que visa sua distribuição e replicação em um cluster de servidores, visando aprimorar a disponibilidade de dados.
- d) de compactação de dados, normalmente bastante eficiente, permitindo grande desempenho no armazenamento de dados.
- e) de transmissão e recepção de dados que permite a comunicação entre servidores, em tempo real.



A **Mineração de dados** (ou **Data Mining**) é o nome dado ao processo de descoberta de novas correlações, padrões e tendências entre as informações de uma empresa, por meio da análise de grandes quantidades de dados armazenados em bancos de dados usando técnicas de reconhecimento de padrões, estatísticas e matemáticas.

Letra a.

039. (FCC/TRT-4^a REGIÃO/RS/ANALISTA JUDICIÁRIO/TECNOLOGIA DA INFORMAÇÃO/2015) Considere a arquitetura geral de um sistema de BI- *Business intelligence* mostrada na figura abaixo.



Nesta arquitetura

- a)** Data Marts representam áreas de armazenamento intermediário criadas a partir do processo de ETL. Auxiliam na transição dos dados das fontes OLTP para o destino final no Data Warehouse.
- b)** OLAP é um subconjunto de informações extraído do Data Warehouse que pode ser identificado por assuntos ou departamentos específicos. Utiliza uma modelagem multidimensional conhecida como modelo estrela.
- c)** os dados armazenados no Data Warehouse são integrados na base única mantendo as convenções de nomes, valores de variáveis e outros atributos físicos de dados como foram obtidos das bases de dados originais.
- d)** o Data Warehouse não é volátil, permite apenas a carga inicial dos dados e consultas a estes dados. Além disso, os dados nele armazenados são precisos em relação ao tempo, não podendo ser atualizados.
- e)** Data Mining se refere ao processo que, na construção do Data Warehouse, é utilizado para composição de análises e relatórios, armazenando dados descritivos e qualificando a respectiva métrica associada.



- a) Errada. **Staging Area** representa a área de armazenamento intermediário criada a partir do processo de ETL. Auxilia na transição dos dados das fontes OLTP para o destino final no Data Warehouse. **Data Mart** é um subconjunto de informações do DW que podem ser identificados por assuntos ou departamentos específicos.
- b) Errada. **Data Mart** é um subconjunto de informações extraído do Data Warehouse que pode ser identificado por assuntos ou departamentos específicos. Utiliza uma modelagem

multidimensional conhecida como modelo estrela. **OLAP** (On-line Analytical Processing) são ferramentas com capacidade de análise em múltiplas perspectivas das informações armazenadas.

- c) Errada. Os dados armazenados no Data Warehouse são integrados na base única mantendo as convenções de nomes, valores de variáveis e outros atributos físicos de dados. O **processo ETL** (do inglês **Extract Transform and Load**) faz a extração, tratamento e limpeza dos dados para inserção no DW.
- d) Certa. O **Data Warehouse** não é volátil, permite apenas a carga inicial dos dados e consultas a estes dados. Além disso, os dados nele armazenados são precisos em relação ao tempo, não podendo ser atualizados.

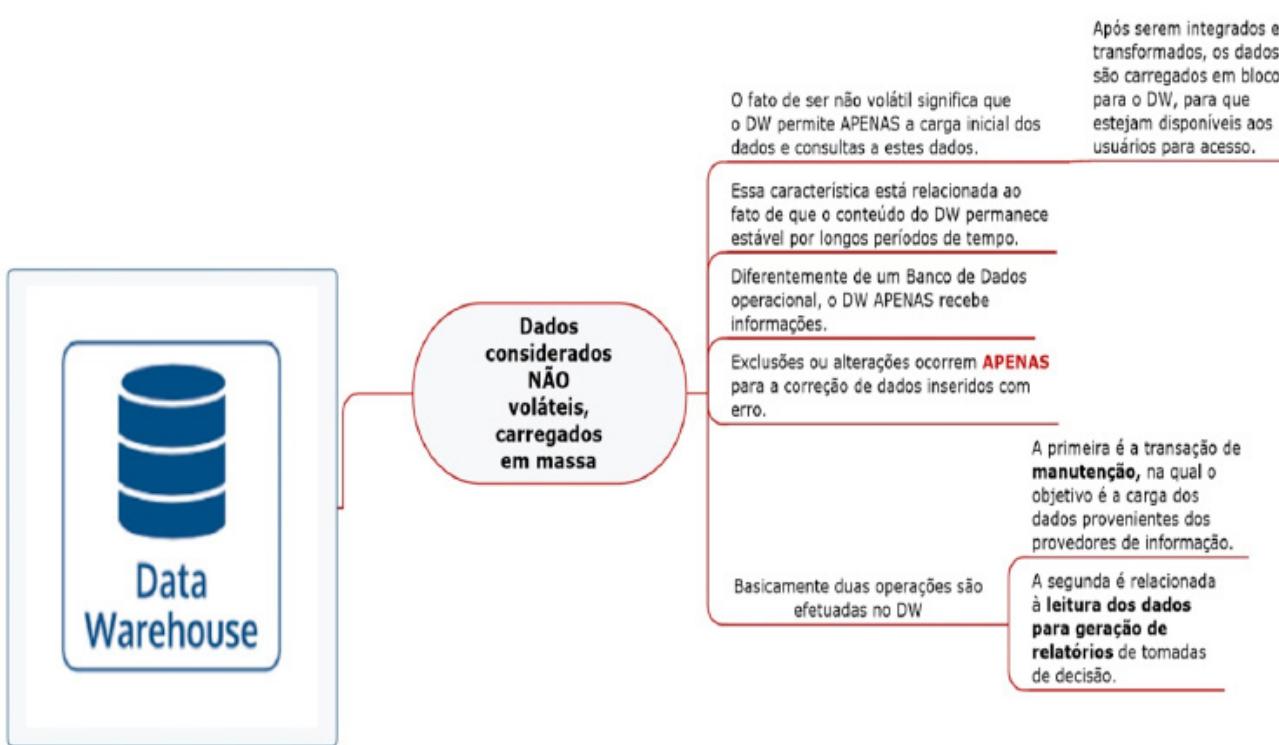


Figura. Dados não voláteis

- e) Errada. **Dimensão** se refere à tabela que, na construção do Data Warehouse, é utilizada para composição de análises e relatórios, armazenando dados descritivos e qualificando a respectiva métrica associada.

Letra d.

040. (FCC/ICMS-RJ/AUDITOR-FISCAL DA RECEITA ESTADUAL/2014) Com o advento da tecnologia de *Data Warehousing*, os ambientes de apoio à decisão passaram a ser denominados ambientes de Data Warehouse (DW).

Em relação à tecnologia DW, é correto afirmar:

- a) Um DW tem duas operações básicas: a carga dos dados (inicial e incremental) e o acesso a estes dados em modo leitura. Depois de carregado, um DW não necessita de operações de bloqueio por concorrência de usuários no acesso aos seus dados.

- b) Em um DW as convenções de nomes, valores de variáveis e outros atributos físicos de dados como *data types* são bastante flexíveis. Para facilitar a tomada de decisões, as informações são apresentadas de diferentes formas, da mesma maneira que foram carregadas dos sistemas legados.
- c) Um projetista de DW deve ter seu foco na modelagem dos dados e no projeto de banco de dados. Um sistema transacional armazena as informações agrupadas por assuntos de interesse da empresa que são mais importantes, enquanto um DW é orientado a processos e deve ser desenvolvido para manter disponíveis as transações realizadas diariamente.
- d) Os dados de um DW são um conjunto dinâmico de registros de uma ou mais tabelas, capturados em um momento de tempo predeterminado, por isso têm que ser sempre atualizados.
- e) Um sistema multidimensional, como o DW, deve atualizar o valor corrente das informações e sua exatidão é válida por um tempo curto, por exemplo, o valor total das notas fiscais processadas pela Receita às 12:00 de um dia pode ser diferente às 18:00 do mesmo dia.



Bill Inmon destaca que o **Data Warehouse (DW)** “é uma coleção de dados orientados por assuntos, integrados, variáveis com o tempo e não voláteis, para dar suporte ao processo de tomada de decisão”. Essas características podem ser descritas da seguinte forma:

- **Orientado a assunto:** refere-se ao fato do Data Warehouse (DW) ser organizado conforme diferentes visões de negócio, ou seja, armazena informações sobre temas específicos importantes para o negócio da empresa. Ex.: Vendas, Compras etc.;
- **Integrado** a partir de fontes de dados heterogêneas;
- **Não volátil:** os dados são sempre inseridos, nunca excluídos. **Em um DW não existem alterações de dados, somente a carga inicial e as consultas posteriores;**
- **Variável com tempo:** posições históricas das atividades no tempo.

Conforme visto, a restrição de não volatilidade permite basicamente duas operações em um DW, que são: a carga (inicial ou incremental) e consulta dos dados. Via de regra não há operações de atualizações de registros, nem necessidade de operações de bloqueio por concorrência de usuários no acesso aos seus dados (o que acontece tipicamente em operações de escrita de dados), pois os dados em um DW são apenas para leitura. Assim, a letra A é a resposta da questão.

Letra a.

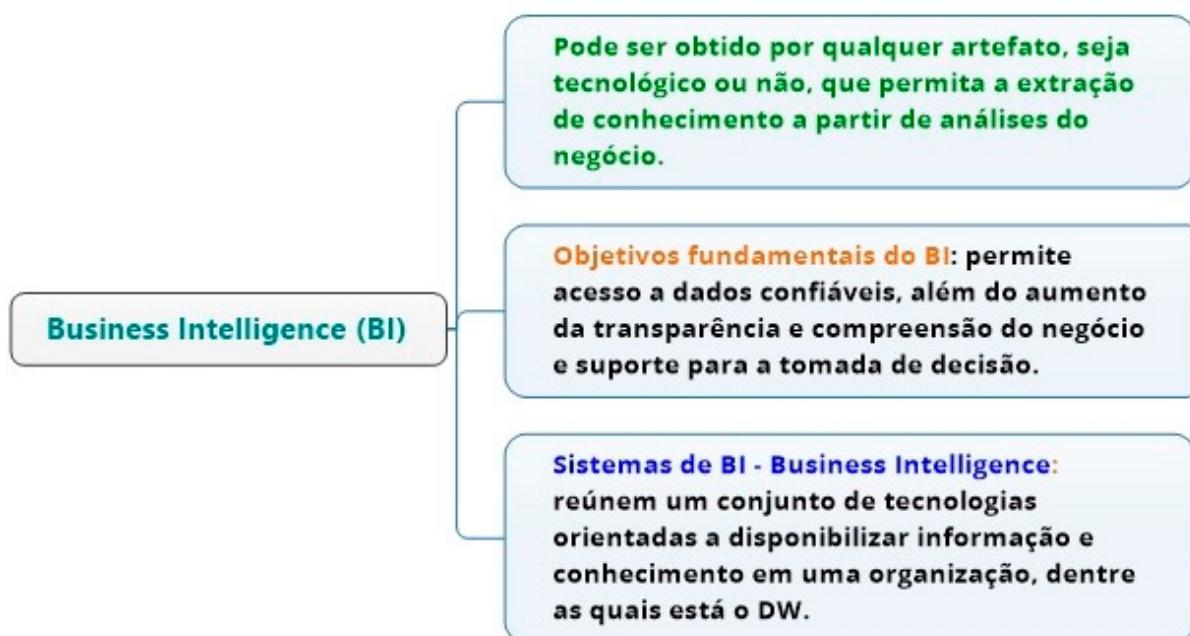
041. (ESAF/STN/DESENV SISTEMAS/2008) Um depósito de dados organizado por assunto, não-volátil, integrado e variável em função do tempo, utilizado para apoiar decisões de gerenciamento, é denominado

- a) *datawarehouse*.
- b) *gestão do conhecimento*.
- c) *business intelligence*.

- d) mineração de dados.
 e) OLAP (*OnLine Analytical Processing*).



- a) Certa. Um **Data Warehouse (Armazém ou Depósito de Dados)** é um sistema utilizado para armazenar informações consolidadas de um banco de dados, possibilitando a análise de grandes volumes de dados, coletados a partir de sistemas transacionais (OLTP).
- b) Errada. Define-se **Gestão do Conhecimento** como a busca da melhoria de desempenho das instituições por meio de processos de procura, extração, compartilhamento e criação de conhecimento, aplicando diferentes ferramentas e tecnologias de informação e de comunicação.
- c) Errada. Define-se **Business intelligence (Inteligência de negócios – BI)** como o processo de coleta, estruturação, avaliação, disponibilização e monitoramento de informações para suporte ao gerenciamento de negócios.



- d) Errada. Etapa do processo de Descoberta de Conhecimento em Bases de Dados (**KDD – Knowledge Discovery in Databases**) que corresponde à execução de um algoritmo particular que, sob algumas limitações aceitáveis de eficiência computacional, encontra padrões ou modelos nos dados.
- e) Errada. **OLAP (On-line Analytical Processing - Processamento Analítico On-Line)** é o processo de manipulação e avaliação de um grande volume de dados sob múltiplas aspectos. A seguir mais algumas observações que merecem destaque neste momento.

No contexto das tecnologias relacionadas ao Business intelligence, duas aplicações são identificadas e muito cobradas em provas:

- a primeira, que sustenta o negócio por meio de **ferramentas OLTP** (*On Line Transaction Processing*); e

- a segunda, que analisa o negócio por meio de **ferramentas OLAP** (On Line Analytical Processing).

Vamos ao detalhamento dessas aplicações!

Sistema Transacional (OLTP)	<p>Os sistemas OLTP (Online Transaction Processing) são os sistemas que capturam as transações de um negócio e as mantêm em estruturas relacionais chamadas Banco de Dados.</p> <p>As principais características dos sistemas OLTP são:</p> <ul style="list-style-type: none"> • Realizar transações em tempo real do processo de um negócio, motivo pelo qual os dados armazenados mudam continuamente. Os sistemas OLTP, nas suas transações, controlam processos essenciais do negócio. • Os sistemas OLTP são os responsáveis pela manutenção dos dados, acrescentando dados, realizando atualizações ou eliminando-os. • Para a tomada de decisões, os sistemas OLTP possuem capacidades limitadas, pois não é seu objetivo e, portanto, não é uma prioridade no seu desenvolvimento. Se desejasse obter uma determinada informação histórica relativa ao negócio consultando um sistema OLTP, seria produzido um impacto negativo no funcionamento do sistema.
Sistemas OLAP	<p>Os sistemas OLAP (Online Analytical Processing, ou Processamento Analítico Online) oferecem uma alternativa aos sistemas transacionais, proporcionando uma visão dos dados orientada à análise, além de uma navegação rápida e flexível.</p> <p>A tecnologia OLAP apresenta as seguintes características:</p> <ul style="list-style-type: none"> • Os bancos de dados OLAP apresentam um esquema otimizado para que as perguntas realizadas pelos usuários sejam respondidas rapidamente. • As perguntas realizadas a um OLAP devem permitir a utilização interativa com os usuários.

Finalizando, **cabe destacar que os sistemas OLTP registram as transações, enquanto que os sistemas OLAP realizam uma análise minuciosa dos dados brutos, extraíndo informações variadas para a tomada de decisões.**

Letra a.

042. (INÉDITA/2020) Julgue o item seguinte: A granularidade diz respeito ao nível de detalhe ou de resumo contido nas unidades de dados existentes no data warehouse. Quanto menos detalhe, mais baixo o nível de granularidade. Quanto mais detalhe, mais alto o nível de granularidade.



A **granularidade** diz respeito ao nível de detalhe ou de resumo contido nas unidades de dados existentes no data warehouse.



Quanto mais detalhe, mais baixo o nível de granularidade. Quanto menos detalhe, mais alto o nível de granularidade. (INMON, 1997, p. 45).

Veja a figura seguinte, extraída de (Machado, 2000, p. 45) bem interessante para visualização desse contexto.

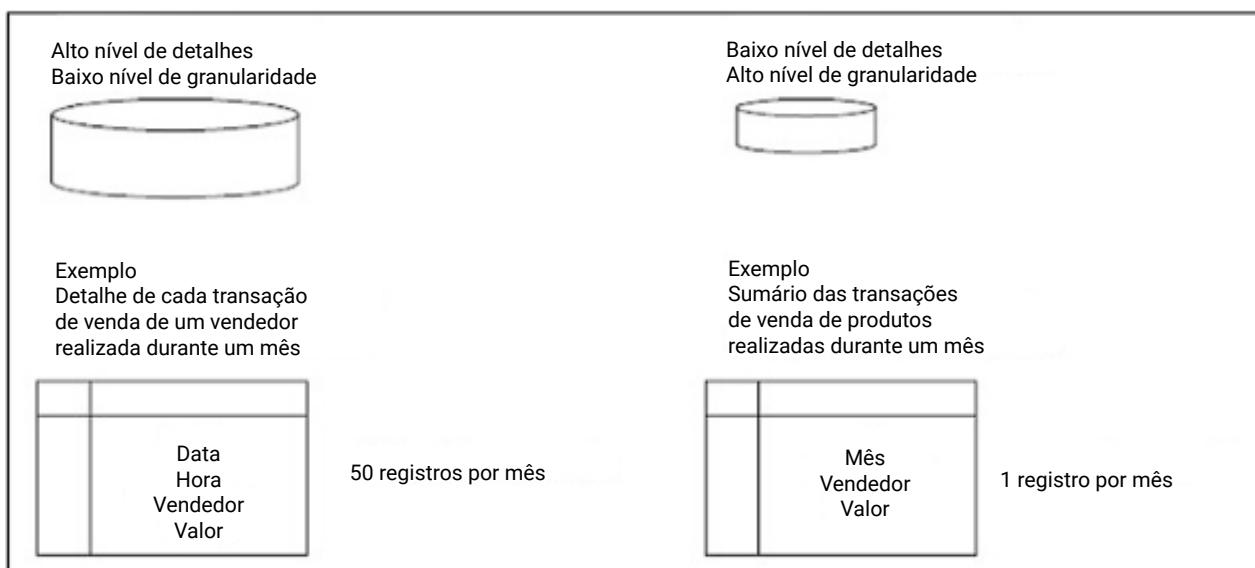


Figura. Granularidade de Dados

INMON (1997) destaca que existe a possibilidade de utilizar um nível duplo de granularidade (níveis duais de granularidade). Esta técnica se enquadra nos requisitos da maioria das empresas. **São criadas duas camadas: uma camada para os dados levemente resumidos e outra para os dados históricos.**

Com a criação de dois níveis de granularidade, é possível atender a todos os tipos de consultas, com um melhor desempenho, visto que a maior parte do processamento analítico se dirige aos dados levemente resumidos que são compactos e de fácil acesso e para as ocasiões em que um maior nível de detalhe deve ser analisado, existe o nível de dados históricos, o qual é complexo e de alto custo (INMON, 1997).

Errado.

043. (INÉDITA/2020) Julgue o item seguinte: Data Mining é um ambiente de suporte à decisão que alavanca dados armazenados em diferentes fontes e os organiza e entrega aos tomadores de decisões. Resumindo, é uma tecnologia de gestão e análise de dados.



Data Warehouse é um ambiente de **suporte à decisão** que alavanca dados armazenados em diferentes fontes e os organiza e entrega aos tomadores de decisões. Resumindo, é **uma tecnologia de gestão e análise de dados**.

Errado.

044. (INÉDITA/2020) Julgue o item seguinte: Data Warehouse (DW) propõe sustentar a tomada de decisão com dados. Trata-se de uma coleção de dados orientada por assunto, integrada, volátil, variante no tempo, que dá apoio às decisões da administração.



Data Warehouse (DW) propõe sustentar a tomada de decisão com dados.

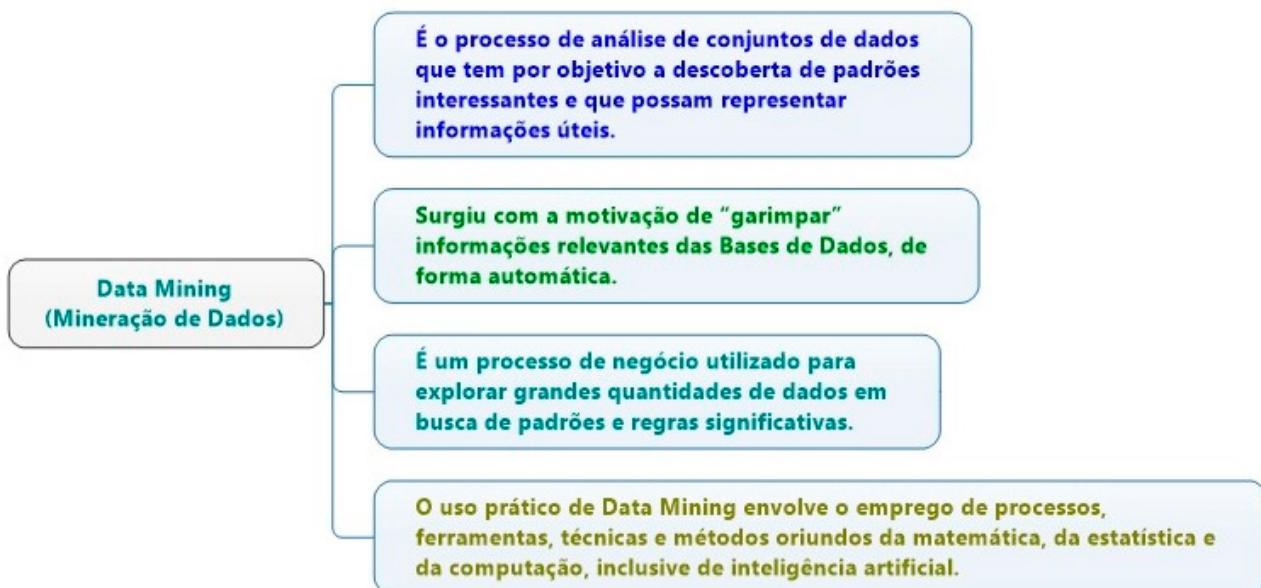
Trata-se de **uma coleção de dados orientada por assunto, integrada, não volátil, variante no tempo**, que dá apoio às decisões da administração.

Errado.

045. (INÉDITA/2020) Julgue o item seguinte: Mineração de Dados é o processo de análise de conjuntos de dados que tem por objetivo a descoberta de padrões interessantes e que possam representar informações úteis.



A **mineração de dados** é **um processo de negócio** para explorar grandes quantidades de **dados** para descobrir **padrões e regras significativas** (2011).



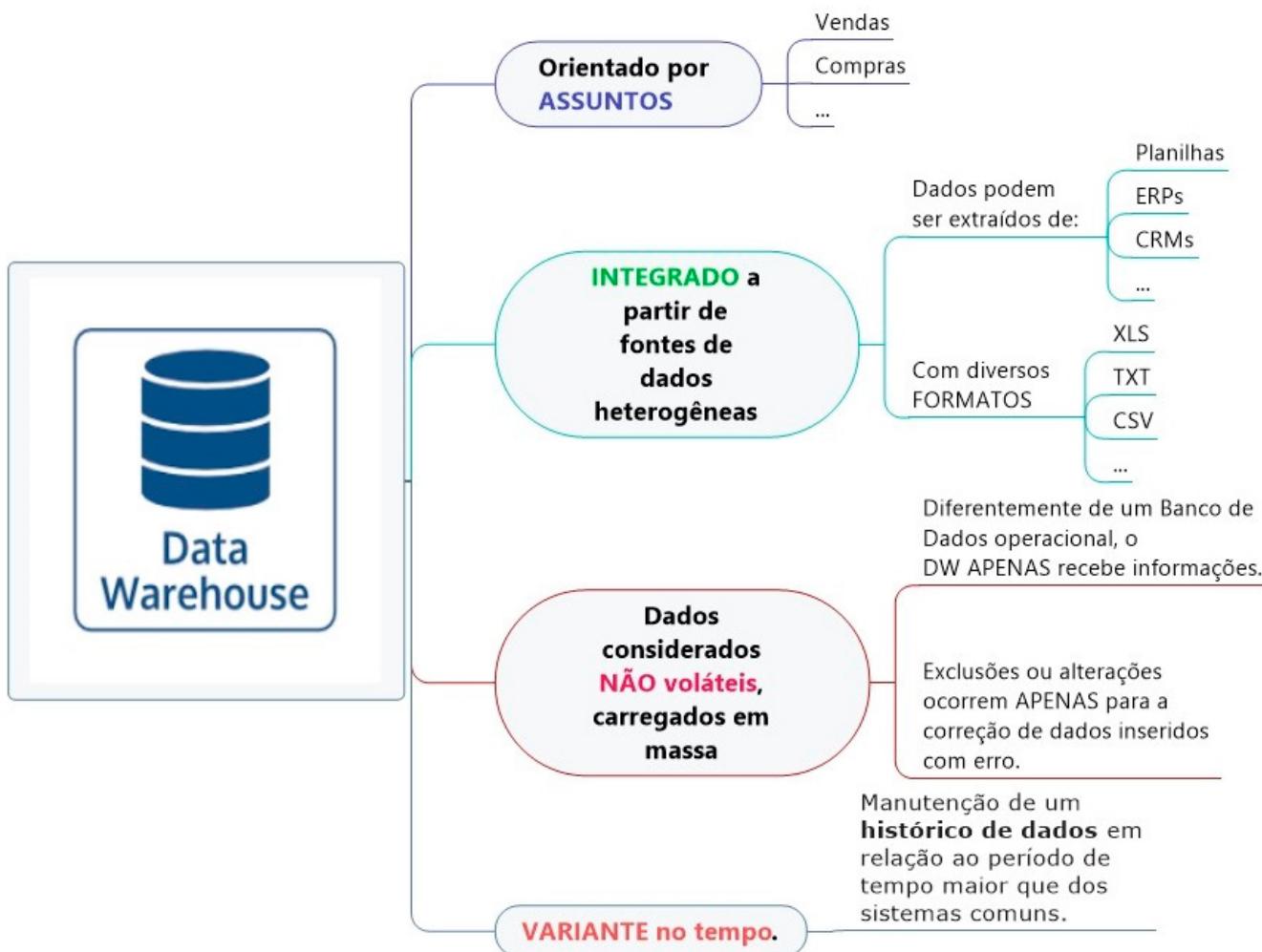
Certo.

046. (ESAF/MPOG/ANALISTA DE TI/2010/ADAPTADA) Um ambiente Datawarehouse:

- a) é uma coleção de dados voláteis, invariantes em termos locais, integrados e orientados a um assunto, direcionados a decisões operacionais.
- b) ao ser projetado segundo a abordagem *bottom-up*, fundamenta-se nas necessidades de informação de todos os níveis gerenciais da empresa.
- c) para ser implementado, requer a remoção de metadados e a integração de dados existentes em um banco de dados.
- d) ao ser projetado, segundo a abordagem *top-down*, não comporta necessidades de informações gerenciais.
- e) é uma coleção de dados não voláteis, variantes em termos temporais, integrados e orientados a assuntos, utilizados no suporte a decisões gerenciais.



A questão original foi anulada e fiz uma alteração na atual. Segundo Inmon, o **Data Warehouse** (DW) é uma coleção de dados orientada a assunto, integrada, não volátil, variável no tempo para suporte às decisões da gerência.



Mais detalhes a seguir.

Orientado a assunto: os dados armazenados em um DW são organizados por assunto, tratando de temas específicos e importantes para o negócio da organização.

Integrados: o DW é um depósito de dados integrados de múltiplas fontes, processados para armazenamento em um modelo multidimensional.

Variante no tempo: um DW mantém os dados históricos, permitindo a detecção de tendências, desvios e relações de longo prazo para previsões e comparações.

Não voláteis: depois que os dados são inseridos em um DW, os usuários não podem modificá-los ou atualizá-los. De modo geral, são permitidas apenas operações de carga e consulta.

Letra e.

047. (CESPE/MPU/ANALISTA DE INFORMÁTICA/PERITO/2010) Dados para a formação de *Data Warehouses* podem ser provenientes de e-mails, conversas de telefone, documentos, planilhas. A abordagem linguística utiliza a análise de strings de caracteres para classificar o texto em determinadas categorias, que são definidas previamente de acordo com o objetivo da empresa.



A primeira parte da questão, descrita como: “dados para a formação de *Data Warehouses* podem ser provenientes de e-mails, conversas de telefone, documentos, planilhas”, está correta! Conforme destaca Heuseler (2010), **dados não-estruturados** podem ser incorporados ao processo de *Data Warehousing*. “Esses dados podem ser provenientes de e-mails, conversas de telefone, documentos, planilhas etc. Devido à natureza destes dados, que têm como essência o texto livre, incorporá-los simplesmente no mundo analítico, segundo Inmon, criaria um universo de análise muito pouco produtivo, sendo necessário que estes textos passem por uma série de tratamentos antes de serem incorporados efetivamente. É preciso separar o conteúdo realmente relevante para a empresa.”

Ainda, segundo Heuseler (2010), “no processo de obtenção de e-mails, por exemplo, devem-se separar os conteúdos trocados em razão de simples conversas entre amigos daqueles que tratam de temas que sejam realmente de interesse da corporação. Além de sua relevância, é importante determinar qual o conteúdo deste texto, ou seja, dar significado a este para que a análise possa ganhar em riqueza. Nota-se claramente que a grande preocupação neste momento de obtenção dos dados é o do tratamento do conteúdo adquirido, de maneira a tornar este significativo para análises”.

Para “entender” estes textos, existem duas abordagens a serem consideradas: linguística e temática.

A **abordagem linguística** trata o texto de acordo com a língua na qual ele está escrito, sendo sua utilização muito complexa e pouco flexível (uma análise deve ser feita para cada língua).

A **abordagem temática**, destacada na questão, utiliza a análise de **strings de caracteres** para classificar o texto em determinadas categorias, que são definidas previamente de acordo com o objetivo da empresa.

Assim, conforme visto, a questão trouxe a descrição da abordagem temática, ao invés da abordagem linguística, e a assertiva é falsa.

Referência utilizada: Heuseler,F. M. Uma abordagem multifacetada para exploração integrada de dados estruturados e não-estruturados em ambientes OLAP. 2010. Disponível em: <http://teses2.ufrj.br/15/teses/735155.pdf>

Errado.

048. (FCC/TRT-14^a REGIÃO/RO E AC/TÉCNICO JUDICIÁRIO/TECNOLOGIA DA INFORMAÇÃO/2016) Quando uma empresa utiliza Data Warehouse (DW) é necessário carregá-lo para permitir a análise comercial. Para isso, os dados de um ou mais sistemas devem ser extraídos e copiados para o DW em um processo conhecido como

- a)** ERP.
- b)** BI.
- c)** CRM.
- d)** ETL.
- e)** Data Mart.



ETL (Extract Transform Load - Extração Transformação Carga) visa trabalhar com toda a parte de **extração de dados de fontes externas, transformação para atender às necessidades de negócios e carga dos dados dentro do Data Warehouse**.

Os projetos de Data Warehouse consolidam dados de diferentes fontes. A maioria dessas fontes tendem a ser bancos de dados relacionais ou flat files (arquivo de texto, geralmente contendo um registro por linha, sendo chamado de arquivo simples por não conter formatação adicional), mas podem existir outros tipos de fontes também. Um sistema ETL precisa ser capaz de se comunicar com bases de dados e ler diversos formatos de arquivos utilizados por toda a organização.

Letra d.

049. (FCC/TRT-20^a/TÉCNICO DE TI/2016) Considere, por hipótese, que o Tribunal Regional do Trabalho da 20^a Região tenha optado pela implementação de um DW (Data Warehouse) que inicia com a extração, transformação e integração dos dados para vários DMs (Data Marts) antes que seja definida uma infraestrutura corporativa para o DW. Esta implementação

- a)** tem como vantagem a criação de legamarts ou DMs legados que facilitam e agilizam futuras integrações.
- b)** é conhecida como top down.
- c)** permite um retorno de investimento apenas em longo prazo, ou seja, um slower pay back.
- d)** tem como objetivo a construção de um sistema OLAP incremental a partir de DMs independentes.

- e) não garante padronização dos metadados, podendo criar inconsistências de dados entre os DMs.



a) Errada. Um dos maiores perigos na implementação de um DW (Data Warehouse) é a criação de DMs (Data Marts) independentes. O advento de ferramentas de drag-and-drop facilitou o desenvolvimento de soluções individuais, de acordo com as necessidades da empresa. Estas soluções podem não considerar a arquitetura de forma global. Assim, os DMs independentes transformam-se em DMs legados, ou legamarts, que dificultam, quando não inviabilizam futuras integrações. Eles são parte do problema e não da solução!

b) Errada. A estratégia adotada na questão é conhecida como **Bottom Up** pois considera que **um DW pode ser composto a partir de DMs (Data Marts) previamente desenvolvidos**.

Na estratégia conhecida como **Top Down** todo o DW (Data Warehouse) estará construído para então disponibilizar os DMs (Data Marts), ou seja, parte-se de uma visão geral de alto nível para uma visão mais detalhada. Em outras palavras, nesse contexto é realizada a modelagem integral do DW, seguida pelas extrações de dados. **Considera que se deve desenvolver um DW completo e centralizado antes que partes dele, summarizadas, possam ser derivadas na forma de DMs**. A principal **vantagem** é a criação de um modelo único. O **revés** fica por conta do maior tempo de projeto.

c) Errada. A estratégia adotada na questão, conhecida como **Bottom Up**, possui uma estrutura que já pode ser utilizada pela empresa antes de se finalizar a construção completa do DW. Assim, **Bottom Up** permite um retorno de investimento a curto prazo, mais rápido, quando comparado à estratégia **Top Down**.

d) Errada. A estratégia **Top down** tem como objetivo a construção de um sistema OLAP incremental a partir de DMs independentes. Nesse caso, o DW abastece os DMs que abastecem OLAP. Na estratégia **Bottom Up**, os DMs abastecem o DW e OLAP.

e) Certa. No que se refere ao **método de construção de Data Warehouse** conhecido como **Bottom-up**, destacado na questão, tem-se que o foco é em uma área por vez, com o crescimento gradual do DW (Data Warehouse). **Considera que um DW possa ser composto a partir de DMs (Data Marts) previamente desenvolvidos**. A **vantagem** é a obtenção de resultados a intervalos mais curtos, garantindo muitas vezes sustentação ao projeto. A **desvantagem** é a maior dificuldade de se consolidar informações entre as diversas áreas, não garantindo a padronização dos metadados, e podendo criar inconsistências de dados entre os DMs.

Fonte: <https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/85434/224374.pdf?sequence=1&isAllowed=y>. Acesso em: abril, 2020.

Letra e.

050. (CESPE/TCE-SC/AUDITOR-FISCAL DE CONTROLE EXTERNO/INFORMÁTICA/2016)
No que concerne à modelagem dimensional, julgue o item que se segue.

Um modelo dimensional é composto por uma tabela com uma chave simples, denominada tabela de fatos, e um conjunto de tabelas maiores, que contém chaves compostas, conhecidas como tabelas de dimensão.



Um **modelo dimensional** é composto por **uma tabela central (tabela de fatos)** e várias outras a ela interligadas (**tabelas de dimensões**).

A tabela de fatos sintetiza o relacionamento existente entre as diversas dimensões. Isto ocorre porque **a chave da tabela de fatos é a associação das chaves primárias das tabelas de dimensões**, o que já invalida a questão!

As **tabelas de dimensões** são as companheiras das tabelas de fatos. **Cada dimensão é definida pela sua chave primária**, que serve para manter a integridade referencial na tabela de fatos à qual está relacionada.

Errado.

051. (CESPE/SERPRO/ANALISTA/DESENVOLVIMENTO DE SISTEMAS/2013) Julgue os próximos itens, a respeito de conceitos e aplicações de técnicas de modelagem dimensional e otimização de bases de dados.

Na modelagem multidimensional, a tabela de fatos deve ser normalizada até a terceira forma normal (3FN), a fim de otimizar o armazenamento e permitir que as consultas sejam respondidas rapidamente.



Na modelagem multidimensional, a **desnormalização** é uma prática. **O processo de normalização, separando os atributos em várias tabelas diferentes faz com que as consultas fiquem bem mais complexas e o banco de dados gaste mais tempo para recuperar informações.**

Segundo o autor Mannino, o que gera **tempo de resposta otimizado** na modelagem multidimensional é a **não normalização das tabelas dimensão**, portanto, a tabela fato não garante velocidade no tempo de resposta.

No Modelo Estrela, é recomendada a desnormalização das tabelas dimensão. O esquema floco de neve é uma variação do esquema estrela, no qual todas as tabelas de dimensão são normalizadas, geralmente, até a terceira forma normal (3FN).

Fonte: Mannino, Michael V.. Projeto, Desenvolvimento de Aplicações e Administração de Banco de Dados.

Errado.

052. (CESPE/SUFRAMA/ANALISTA/TECNOLOGIA DA INFORMAÇÃO/2014) Na modelagem, o esquema estrela é um refinamento em que parte da hierarquia dimensional é normalizada em um conjunto de tabelas dimensão menores, de forma similar a um floco de neve.



A assertiva refere-se ao esquema **floco de neve (snowflake)**. Nesse modelo as tabelas dimensionais relacionam-se com a tabela de fatos, mas **algumas dimensões relacionam-se apenas**

entre elas. Isto ocorre para fins de normalização das tabelas dimensionais, visando diminuir o espaço ocupado por estas tabelas. Informações como Categoria, Departamento e Marca tornar-se-ão, no exemplo, tabelas de dimensões auxiliares.

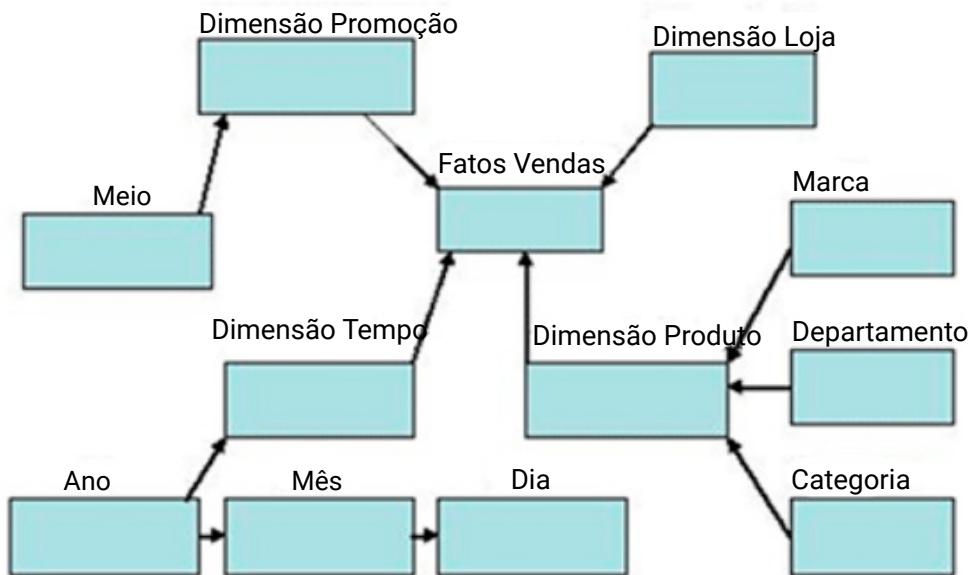
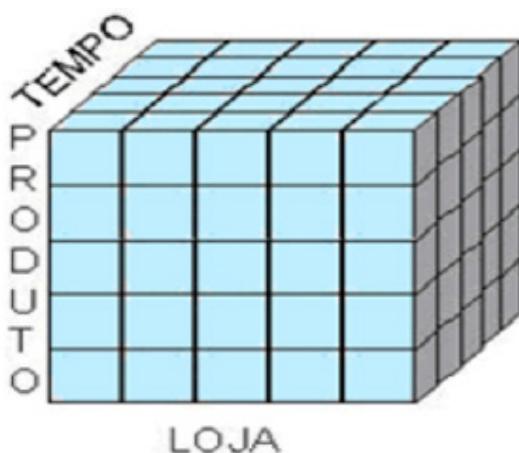


Figura. Modelo Floco de Neve

Errado.

053. (CESPE/SERPRO/TÉCNICO/PROGRAMAÇÃO E CONTROLE DE SERVIÇOS DE TECNOLOGIA DA INFORMAÇÃO/2013) Com relação a gerenciamento eletrônico de documentos (GED) e *datawarehouse*, julgue o item que se segue.

Denomina-se de análise dimensional o processo de análise mediante o qual dados originários são selecionados das bases operacionais e modelados no ambiente de *datamining*, de acordo com as perspectivas ou visões elegidas para a realização.



A **análise multidimensional** é uma das grandes utilidades da **tecnologia OLAP**, consistindo em ver determinados **cubos de informações** de diferentes ângulos e de vários **níveis de agregação**.

Veja, por exemplo, o caso de uma empresa que possui várias lojas filiais e que deseja acompanhar o desempenho de suas vendas ao longo do tempo.

Um desenhista de Data Warehouse visualiza estas informações de uma forma como um cubo que pode ser descrito com três **dimensões** principais que são: tempo, loja e produto. Na intersecção destas três dimensões está a quantidade de produtos que foi vendida.

Essa questão foi alvo de inúmeros recursos, no entanto, ao final, a banca a considerou como correta. Vide justificativa:

Comentários da banca para a alteração do gabarito de E para C:

A modelagem dos dados, no contexto de Data Warehouse, chamada de **análise multidimensional**, é um processo de análise que envolve selecionar dados originários das bases operacionais e modelá-los no ambiente de **Data Warehouse**, de acordo com as **dimensões (perspectivas ou visões)** sobre as quais se deseja tratá-los. Portanto o item está correto e opta-se por alterá-lo.

Ficou incoerente o comentário com a resposta final, mas o gabarito final foi mesmo correto!

Certo.

054. (CESPE/SE-DF/ANALISTA DE GESTÃO EDUCACIONAL/TECNOLOGIA DA INFORMAÇÃO/2017) Com relação aos conceitos de modelagem multidimensional de dados para inteligência computacional, julgue o seguinte item.

Ao se modelar uma tabela-fato, deve-se considerar que a chave primária é composta e que a dimensão tempo sempre será parte integrante dessa chave.



Ao se modelar uma tabela-fato, **o conjunto das chaves primárias de cada tabela-dimensão formará a chave primária da tabela-fato**. É disso que a questão “fala” quando é afirmado “deve-se considerar que a chave primária é composta”.

Com relação à afirmação “a dimensão tempo sempre será parte integrante dessa chave.”, o próprio autor Ralph Kimball, um dos precursores dos conceitos de *data warehouse* afirma: “*The time dimension is a unique and powerful dimension in every data mart and enterprise data warehouse.*” (em tradução livre: **A dimensão tempo é uma dimensão única e poderosa em cada data mart e data warehouse empresarial**).

Portanto, questão correta.

Obs.: Essa última afirmação sobre a dimensão tempo é um ótimo lembrete para ler antes da prova.

Certo.

055. (CESPE/TCE-PE/AUDITOR DE OBRAS PÚBLICAS/2017) A respeito da modelagem dimensional, julgue os próximos itens.

Todas as tabelas de um modelo apresentam um elemento de tempo como parte da estrutura da chave.



As tabelas-fato contêm referências para as tabelas dimensionais (ou também tabelas de dimensão), por meio de sua chave primária. Esta chave é composta das chaves que cada tabela dimensional apresenta.

Uma tabela dimensional importante em um modelo é a dimensão Tempo. **Ela deve ser parte da chave primária da tabela-fato, mas não necessariamente, parte da chave das tabelas dimensionais.**

Em momento algum, podemos dizer que todas as tabelas (fato ou dimensionais) devem ter a dimensão Tempo como parte da chave primária. Isso significa que as tabelas dimensionais, ou também tabelas de dimensão, não precisam ter, necessariamente, o elemento de tempo como parte da estrutura da chave.

Por fim, cabe destacar que apenas a tabela de dimensão tempo e a tabela fato possuem elementos de tempo como parte da sua estrutura.

Portanto, item errado.

Errado.

056. (CESPE/ANTAQ/ANALISTA/ SISTEMAS E NEGÓCIOS/2014) Em relação aos sistemas de suporte à decisão, julgue o item a seguir.

Em uma modelagem multidimensional, as métricas são armazenadas na tabela fato, independentemente de estarem em um modelo **Estrela** ou **Star Schema**, podendo ser aditiva, correspondente a valores que podem ser aplicados às operações de soma, subtração e média, ou não aditiva, correspondente a valores percentuais, ou relativos, que não podem ser manipulados livremente.



Em uma modelagem multidimensional, **métricas** são as informações armazenadas nas tabelas **fato** que permitem medir o desempenho dos processos do negócio.

“Em uma **tabela de fatos**, uma linha corresponde a uma **medição**. Uma medição é uma linha em uma tabela de fatos. Todas as medições em uma tabela de fatos devem estar alinhadas na mesma granularidade” (KIMBALL; ROSS, 2002, p. 21).

Um aspecto importante é que a tabela de fatos deve representar uma unidade do processo do negócio, não devendo misturar assuntos diferentes numa mesma tabela de fatos.

Os atributos mais comuns em uma tabela de fatos são valores numéricos. Estes valores são, em sua maioria, aditivos. As **métricas aditivas** são as que permitem operações como adição, subtração e média de valores por todas as dimensões, em quaisquer combinações de registros, como “total de itens vendidos” por combinação de data, produto e loja. Métricas aditivas

são importantes porque normalmente as aplicações de DW não retornam uma linha da tabela de fatos, mas sim centenas, milhares e até milhões (HOKAMA et al, 2004, p. 36).

As **métricas não aditivas** são valores que não podem ser manipulados livremente, como valores percentuais ou relativos. Exemplos de métricas não aditivas são preço de custo e preço de venda de um produto em uma venda.

Já as **métricas semiaditivas** são valores que não podem ser somados em todas as dimensões. Por exemplo: numa tabela com o registro diário do saldo bancário dos clientes de uma agência, não faz sentido somar os saldos bancários diários de um cliente durante um mês, mas pode-se somar os saldos de todos os clientes de uma agência em determinada data.

Veja mais: <http://meusite.mackenzie.com.br/rogerio/tgi/2004ModelagemDW.pdf>

<http://www.fatecsp.br/dti/tcc/tcc00071.pdf>.

Certo.

057. (VUNESP/TJ-PA/ANALISTA JUDICIÁRIO/ANÁLISE DE SISTEMA/DESENVOLVIMENTO/2014) Considerando o processo de extração, transformação e carga utilizado na montagem de um data warehouse, é correto afirmar que

- a)** a carga de dados corresponde à transferência de dados do ambiente de data warehouse para o ambiente de produção.
- b)** um exemplo de uma transformação de dados consiste na conversão de diversas unidades de medida para uma única unidade (por exemplo, cm, mm e m para cm).
- c)** a fase de extração de dados considera apenas os dados existentes no ambiente operacional há pelo menos um mês
- d)** a transformação de dados se restringe a inserir marcas de tempo em cada registro extraído do ambiente operacional.
- e)** a carga de dados do ambiente de produção para o data warehouse só pode ser feita uma vez por semana.



O **processo ETL (Extract Transform Load - Extração Transformação Carga)** consiste na **extração** (leitura de dados de uma ou mais bases de dados), **transformação** (conversão dos dados extraídos de sua forma anterior para a forma em que precisa estar para que possa ser colocado em um Data Warehouse ou simplesmente outro banco de dados), e **carga** (colocar os dados no Data Warehouse).

Um exemplo de uma transformação de dados consiste na conversão de diversas unidades de medida para uma única unidade (por exemplo, cm, mm e m para cm).

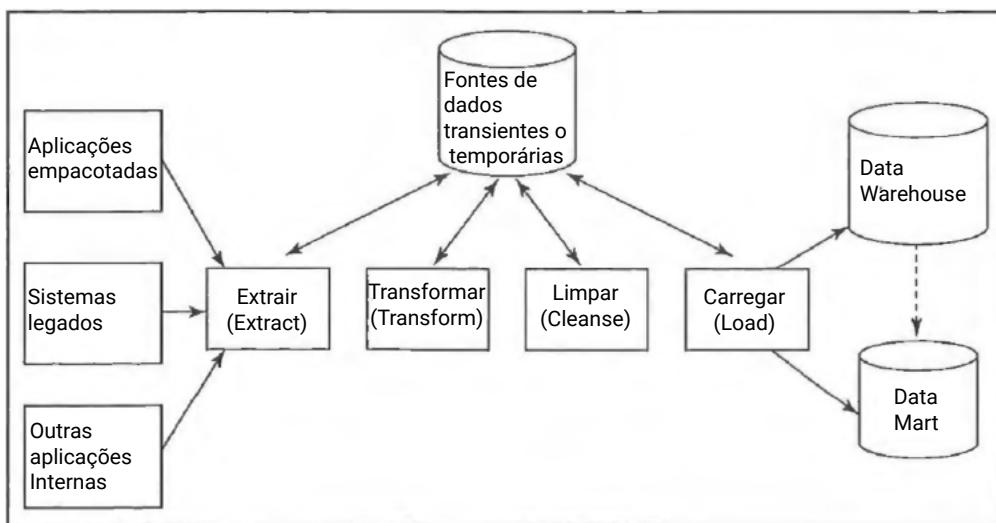


Figura. ETL

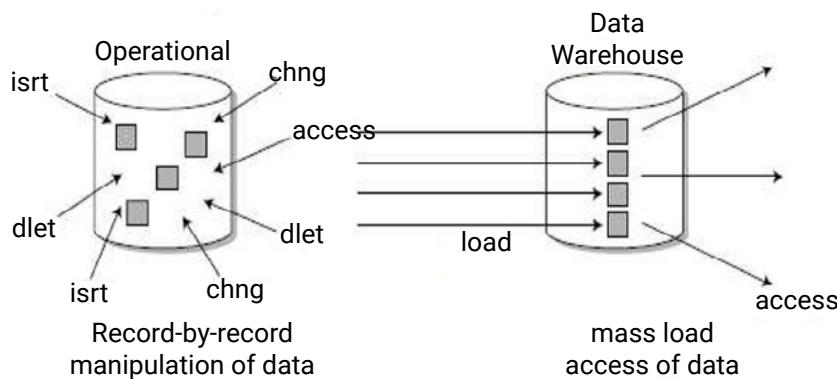
Letra b.

058. (CESPE/CORREIOS/ANALISTA DE CORREIOS/ANALISTA DE SISTEMAS/DESENVOLVIMENTO DE SISTEMAS/2011) Em um ambiente data warehouse (DW), é possível a análise de grandes volumes de dados, os quais ficam disponíveis para serem alterados e manipulados pelo usuário.



Segundo Laudon & Laudon, um **Data Warehouse (DW)** é um banco de dados, com ferramentas de consulta e relatório, que armazena dados atuais e históricos extraídos de vários sistemas operacionais e consolidados para fins de análises e relatórios administrativos.

Em um ambiente DW, é possível a análise de grandes volumes de dados. No entanto, diferentemente do **ambiente operacional** em que os dados são, em geral, **atualizados registro a registro**, em múltiplas transações, o DW **permite apenas a carga inicial dos dados e consultas** a estes dados. Após serem **integrados e transformados**, os dados são **carregados em bloco** para o Data Warehouse, para que estejam disponíveis aos usuários **para acesso**.



Errado.

059. (CESPE/SERPRO/ANALISTA/ADMINISTRAÇÃO DE SERVIÇOS DE TECNOLOGIA DA INFORMAÇÃO/2013) Com relação ao *datawarehouse*, julgue o item subsequente. [Os dados armazenados em um *datawarehouse* devem ser normatizados].



Em geral, os *Data Warehouses* são desnormalizados (não seguem todas as regras de normalização).

Errado.

060. (FCC/TCE-RS/ANÁLISE DE INFORMAÇÕES/2018) Considerando a teoria da modelagem dimensional, composta por tabelas dimensão e tabela fato, utilizada em *Data Warehouses*,

- a) todas as tabelas dimensão devem possuir o mesmo número de atributos.
- b) o grau de relacionamento da tabela fato para as tabelas dimensão é de muitos para muitos.
- c) a tabela fato não deve possuir atributos do tipo numérico.
- d) não há relacionamento entre as tabelas dimensão e a tabela fato.
- e) não há limitação quanto ao número de tabelas dimensão.



A **modelagem multidimensional**, ou **dimensional** como às vezes é chamada, é a técnica de modelagem de banco de dados para o auxílio às consultas em um Data Warehouse.

Toda modelagem dimensional possui dois elementos imprescindíveis, que são: as **tabelas Fatos** e as **tabelas Dimensões**. Ambas são obrigatórias e possuem característica complementares dentro de um Data Warehouse.

As **Tabelas Fato** servem para o armazenamento, medidas (quase sempre) numéricas associadas a eventos de negócio. Uma tabela fato armazena as medições de desempenho decorrentes de eventos dos processos de negócios de uma organização. Basicamente representa uma medida de negócios. Todas as tabelas fato têm duas ou mais chaves estrangeiras que ligam para as chaves primárias das tabelas de dimensão.

As **dimensões** organizam os dados em função de uma área de interesse para os usuários. Cada dimensão descreve um aspecto do negócio e proporciona o acesso intuitivo e simples aos dados. Uma dimensão oferece ao usuário um grande número de combinações e intersecções para analisar dados.

As **tabelas dimensões** apresentam o contexto descritivo. São companheiros integrais para uma tabela de fatos e contém o contexto textual associado a um evento de medição dos processos de negócios. Elas descrevem o “quem, o que, onde, quando, como e porquê” associado ao evento. As tabelas dimensão têm uma relação 1:N com a tabela fato. Finalizando, cabe destacar que **não existe um número limitado para a quantidade de dimensões presentes no modelo**.

O gabarito da questão encontra-se na alternativa E.

Letra e.

061. (FCC/TRT-3^a REGIÃO/MG/TÉCNICO JUDICIÁRIO/TECNOLOGIA DA INFORMAÇÃO/2015) Um técnico de TI precisa utilizar um subconjunto de dados de um Data Warehouse direcionado à área administrativa de um Tribunal. Esses dados serão armazenados em um banco de dado modelado multidimensionalmente, que será criado capturando-se dados diretamente de sistemas transacionais, buscando as informações relevantes para os processos de negócio da área administrativa. Esse banco de dados será um

- a)** Big Data.
- b)** Data Mart.
- c)** OLAP.
- d)** MOLAP.
- e)** Data Mining.



O **Data Mart (DM)** nada mais é que um subconjunto de dados de um Data Warehouse, em que tipicamente desempenham o papel de um DW departamental, regional ou funcional.

- Segundo Ralph Kimball (2002)
 - Um Data Mart é um subconjunto lógico do Data Warehouse, normalmente visto como um Data Warehouse setorial.

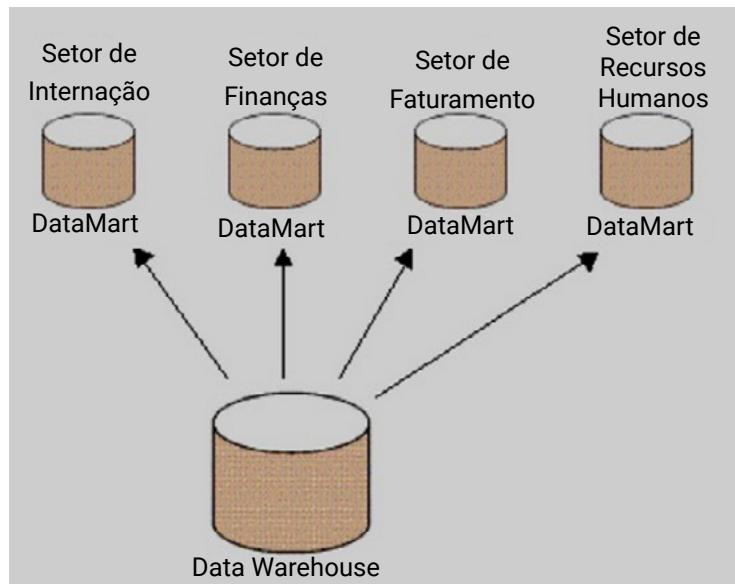


Figura. Data Warehouse e Data Marts

Letra b.

062. (FCC/MPE-MA/ANALISTA MINISTERIAL/BANCO DE DADOS/2013) Na modelagem dimensional de um data warehouse voltado para vendas, o tipo de tabela fato que inclui pares de produtos adquiridos em uma mesma compra recebe a denominação de:
 a) cesta de mercado.
 b) tabela de degeneração.
 c) data mart.
 d) outrigger.
 e) pacote de integralização.



Cesta de Mercado (ou Market basket analysis), isso é em relação a associação entre os produtos que são adquiridos e venda, auxiliando na escolha dos melhores lugares para se colocar produtos aumentando assim a venda dos mesmos.

Letra a.

063. (FUNDATEC/PROCERGS/TÉCNICO COMPUTAÇÃO/ANALISTA DE SUPORTE ÊNFASE EM METODOLOGIA/2012) Um data warehouse é um repositório de dados históricos orientados a assuntos, sendo organizados para serem acessíveis em uma forma prontamente aceitável para atividades de processamento analítico. Assim sendo, analise as assertivas abaixo, no que se refere às características e à definição de um data warehouse:

- I – Segmentados - são vários bancos de dados pequenos contendo informações divergentes.
- II – Organização - os dados são organizados por assunto, e contêm informações relevantes apenas para o apoio à decisão.
- III – Consistência - os dados nos diferentes bancos de dados operacionais podem ser codificados de forma diferente.
- IV – Variante tempo - os dados são mantidos por muitos anos, de modo que possam ser usados para tendências, previsões e comparações com o tempo.
- V – Desestruturados - possuem uma quantidade grande de dados que dificultam a tomada de decisão.

Quais estão corretas?

- a) Apenas I e III.
- b) Apenas I, II e III.
- c) Apenas II, III e IV.
- d) Apenas III, IV e V.
- e) I, II, III, IV e V.

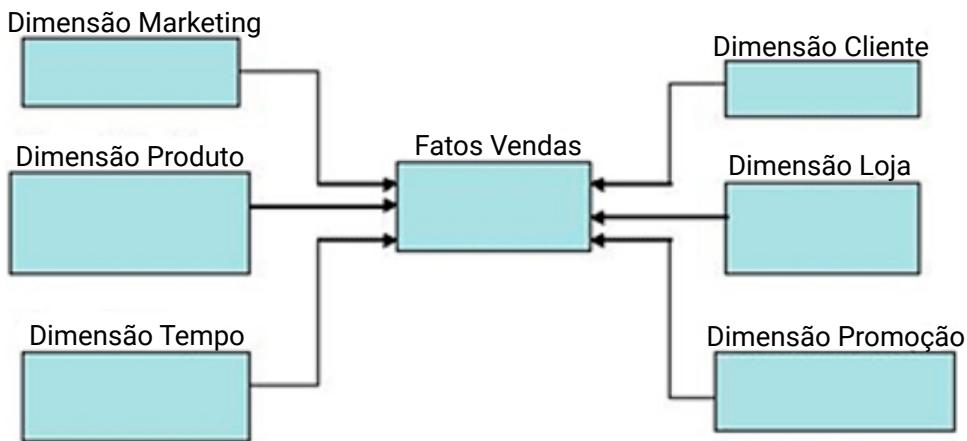


Um **Data Warehouse (DW)** é uma coleção de dados orientada por assunto, integrada, não volátil e variável no tempo. Em geral, eles consolidam uma grande quantidade de dados da empresa para oferecer apoio à decisão.

Logo as alternativas I (Segmentados) e V (Desestruturados) não podem estar corretas. Em geral, os DW são desnormalizados (não seguem todas as regras de normalização), mas não são desestruturados. Sendo assim, a resposta correta é a letra C.

Letra c.

064. (FGV/FISCAL DE RENDAS/SEFAZ-RJ/2007) Observe a figura abaixo, que ilustra um exemplo de um modelo dimensional para construção de banco de dados para DataWarehouse.



São características desse modelo:

- I – Todas as tabelas relacionam-se diretamente com a tabela de fatos, e assim as tabelas dimensionais devem conter todas as descrições que são necessárias para definir uma classe como Produto, Tempo ou Loja nela mesma.
- II – As tabelas de dimensões não são normalizadas; logo, campos como Categoria, Departamento, Marca contêm sua descrição repetida em cada registro, assim aumentando o tamanho das tabelas de dimensão por repetirem essas descrições de forma textual em todos os registros.

O modelo é do tipo conhecido por:

- a) hierárquico.
- b) distribuído.
- c) relacional.
- d) estrela.
- e) anel.



Nesse contexto foi utilizado o modelo **estrela**, cuja estrutura está formada por uma tabela central – **tabela de fatos** – e um conjunto de tabelas organizadas ao redor dela, as **tabelas de dimensões** (cada dimensão descreve um aspecto do negócio).

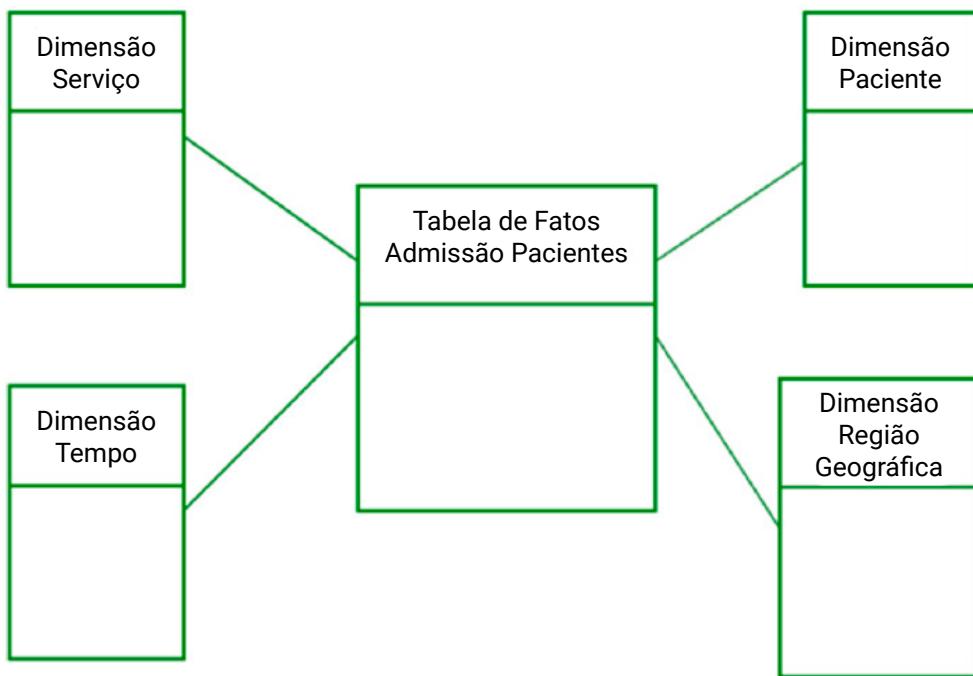
Nas pontas da estrela estão as tabelas de dimensões que contém os atributos das aberturas que interessam ao negócio e que podem ser utilizadas como critérios de filtro e são relativamente pequenas. Cada tabela de dimensão está relacionada com a tabela de fatos por um identificador.

O centro da estrela é a tabela de fatos.

As pontas da estrela são as tabelas de dimensões.

Cada esquema está formado por apenas uma tabela de fatos.

Geralmente é um esquema totalmente não padronizado e pode estar parcialmente padronizado nas tabelas de dimensões.



No exemplo é apresentado um esquema estrela considerando a necessidade de analisar como evolui a Admissão de Pacientes (Fato) por serviço, pacientes e região geográfica ao longo do tempo.

Letra d.

065. (INÉDITA/2020) Julgue o item seguinte: Dentro da modelagem multidimensional fatos estabelecem a organização dos dados, determinando possíveis consultas/cruzamentos. Por exemplo: região, tempo, canal de venda, ...



Dimensões: estabelecem a organização dos dados, determinando possíveis consultas/cruzamentos. Por exemplo: região, tempo, canal de venda,...

Errado.

066. (INÉDITA/2020) Julgue o item seguinte: dentro da modelagem multidimensional fatos são os valores a serem analisados, como médias, totais e quantidades.



Medidas são os valores a serem analisados, como médias, totais e quantidades.

Errado.

067. (INÉDITA/2020) Julgue o item seguinte: Data Mining é um ambiente de suporte à decisão que alavanca dados armazenados em diferentes fontes e os organiza e entrega aos tomadores de decisões. Resumindo, é uma tecnologia de gestão e análise de dados.



Data Warehouse é um ambiente de **suporte à decisão** que alavanca dados armazenados em diferentes fontes e os organiza e entrega aos tomadores de decisões. Resumindo, é **uma tecnologia de gestão e análise de dados**.

Errado.

068. (INÉDITA/2020) Julgue o item seguinte: Data Warehouse (DW) propõe sustentar a tomada de decisão com dados. Trata-se de uma coleção de dados orientada por assunto, integrada, volátil, variante no tempo, que dá apoio às decisões da administração.



Data Warehouse (DW) propõe sustentar a tomada de decisão com dados. Trata-se de **uma coleção de dados orientada por assunto, integrada, não volátil, variante no tempo**, que dá apoio às decisões da administração.

Errado.

069. (INÉDITA/2020) Julgue o item seguinte: Mineração de Dados é o processo de análise de conjuntos de dados que tem por objetivo a descoberta de padrões interessantes e que possam representar informações úteis.



A **mineração de dados** é **um processo de negócio** para explorar grandes quantidades de **dados** para descobrir **padrões e regras significativas** (2011).

Certo.

070. (VUNESP/TCE-SP/AGENTE DA FISCALIZAÇÃO FINANCEIRA - SISTEMAS, GESTÃO DE PROJETOS E GOVERNANÇA DE TI/2015) Uma das formas de modelagem utilizada no desenvolvimento de *Data Warehouses* é a modelagem multidimensional. Nesse tipo de modelagem,

- a)** sempre há diversas tabelas fato representadas e apenas uma tabela dimensão.
- b)** não há utilização de atributos com a função de chave primária nas tabelas fato e dimensão.
- c)** as medições numéricas que representam o desempenho do negócio modelado são armazenadas na tabela fato.
- d)** o relacionamento estabelecido entre as tabelas fato e dimensão é de um para um.
- e)** as tabelas dimensão não admitem o uso de atributos do tipo numérico.



A **modelagem dimensional ou multidimensional** é a **técnica de modelagem de banco de dados utilizada para a representação dos dados nos bancos de dados multidimensionais**.

A **modelagem dimensional** é usada para o **auxílio às consultas em um Data Warehouse** nas mais diferentes perspectivas, definindo a representação e armazenamento dos dados nestes repositórios, bem como promovendo o processamento de consultas multidimensionais complexas. A visão multidimensional permite, ainda, o **uso mais intuitivo para o processamento analítico pelas ferramentas OLAP** (Online Analytical Processing), que possibilitam a análise de dados complexos do Data Warehouse.

Uma modelagem dimensional deve ser planejada de forma atômica, ou seja, deve suportar filtros, consultas agrupadas, categorizadas e satisfazer as regras de negócio. Os usuários devem poder analisar tanto a informação totalizada quanto chegar aos detalhes mínimos desejados. Para armazenar a informação, como em um modelo relacional, é necessária a criação de entidades (tabelas) com atributos (colunas). A base para a modelagem dimensional são as **tabelas de fatos**, que possuem uma história de uma ação relacionada diretamente com os **indicadores numéricos da mesma (medidas)**:

- **Tabela Fato:** principal tabela no DW, em que são encontradas as informações de um “fato” ocorrido, ou seja, o histórico de uma ação armazenada;
- **Tabela Dimensão:** contém as informações descritivas e qualificadoras do negócio. Exemplo: Raças, Categorias, Cores, Produtos, Cidades, Tempo etc.

Além das tabelas Fato e Dimensão, a modelagem dimensional possui outros termos comuns, como, por exemplo:

- **Medidas (measures):** **um valor numérico, encontrado nas Tabelas Fato, que representam a contabilização de uma ação armazenada.** Exemplo: Quantidade, Valor Unitário e Valor Total de um produto em uma Nota Fiscal. Ao ser consultada, é agregada com uma função: soma, média, mínimo, máximo etc.;
- **Dimensão (dimension):** uma lista independente que categoriza uma informação do Fato. Exemplo: uma dimensão de categoria de um produto, uma dimensão de qual sexo o cliente pertence, uma dimensão que caracteriza o tipo de um animal etc. Uma dimensão pode ser analisada tanto como coluna quanto níveis com subtotais etc.

Conforme visto, a letra C é a resposta. As **medições numéricas** que representam o desempenho do negócio modelado são armazenadas na **tabela fato**.

As tabelas de Fatos contêm as métricas. Possui o caráter quantitativo das informações descriptivas armazenadas nas Dimensões. **É onde estão armazenadas as ocorrências do negócio** e possui relacionamento de “muitos para um” com as tabelas periféricas (Dimensão).

Veja mais:

<http://corporate.canaltech.com.br/materia/business-intelligence/entendendo-a-modelagem-multidimensional-19988>

Letra c.

071. (VUNESP/TJ-PA/ANALISTA JUDICIÁRIO/ANÁLISE DE SISTEMA/DESENVOLVIMENTO/2014) Um dos modelos mais utilizados na modelagem de *Data Warehouses* é o modelo dimensional, sobre o qual é correto afirmar que

- a) o número mínimo de tabelas dimensão em um modelo dimensional é 4.
- b) cada modelo dimensional contém, pelo menos, 2 tabelas fato
- c) as tabelas dimensão contêm atributos que visam descrever características de cada dimensão.
- d) cada tabela fato não pode conter mais do que 10 atributos.
- e) a cardinalidade do relacionamento entre tabelas dimensão e tabelas fato é de 1 para 1.



As **tabelas de dimensões**, ou simplesmente **tabelas-dimensões**, contêm as informações descriptivas e qualificadoras do negócio. Exemplo: Raças, Categorias, Cores, Produtos, Cidades, Tempo etc. Quando a ausência de informação na tabela Fato for constatada, a mesma deve ser corrigida em suas dimensões por meio de uma nova categoria.

As **tabelas de dimensão** são acompanhantes integrais de uma tabela de fatos, contendo os **descritores textuais do negócio**. Em um modelo dimensional bem concebido, as **tabelas de dimensão têm muitas colunas ou atributos**. Esses atributos descrevem as linhas na tabela de dimensões.

Letra c.

072. (ESAF/RECEITA FEDERAL/ANALISTA TRIBUTÁRIO DA RF/PROVA 2/ÁREA INFORMÁTICA/2012) São componentes principais de um processo de *Data Warehousing*:

- a) Fontes de dados. Extração de dados. Carregamento de dados. Banco de dados abrangente. Metadados. Ferramentas de middleware.
- b) Fontes de usuários. Extração de relações. Carregamento de dados. Banco de dados focalizado. Metadados. Ferramentas de upperware.
- c) Fontes de dados. Extração de dados. Monitoramento de dados. Banco de dados abrangente. Métodos de dados. Usuários de middleware.
- d) Fontes de dados homogêneos. Manipulação de dados. Carregamento de programas. Banco de dados equivalente. Polidados. Ferramentas de middleware.

e) Fontes de dados. Transformação de dados. Carregamento de sites de dados. Banco de dados abrangente. Meta-relações. Ferramentas de middlehousing.



São componentes principais de um processo de *Data Warehousing*:

- As fontes de dados (conhecidas como Data sources);
- Extração e transformação de dados (Data extraction and transformation);
- Carga de dados (Data loading);
- Banco de dados abrangente (Comprehensive database);
- Metadados (Metadata);
- Ferramentas de middleware (Middleware tools).

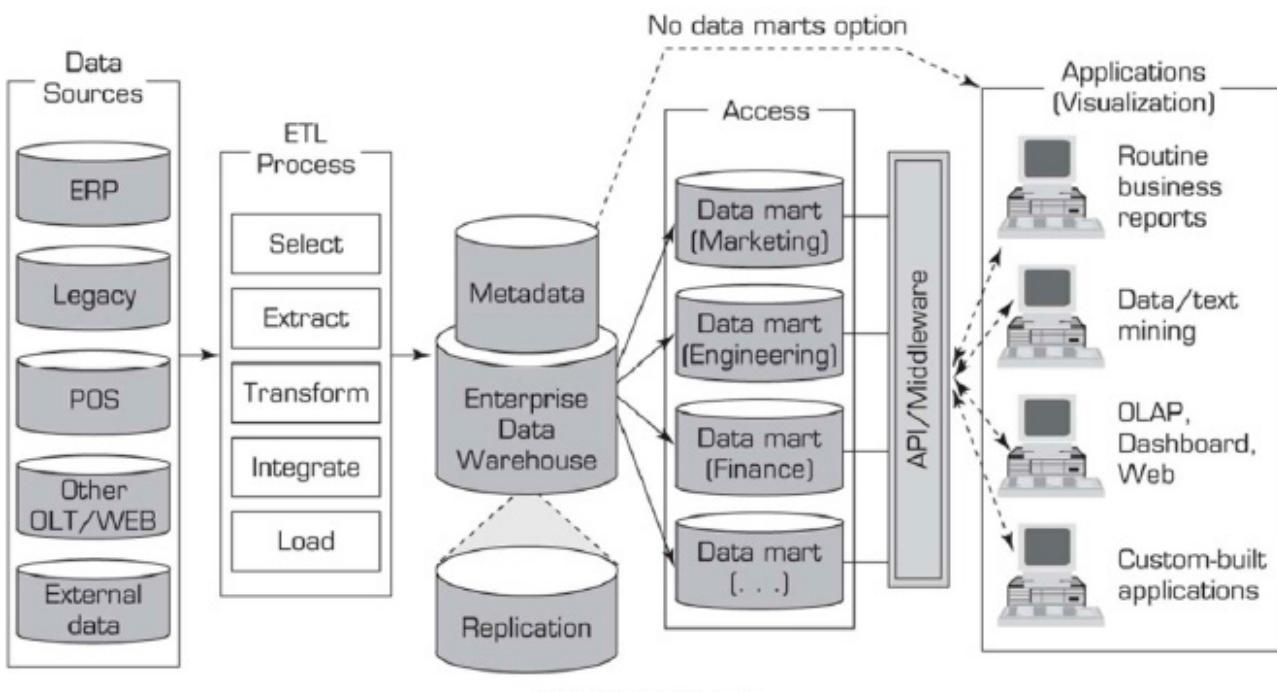


Figura. Processo de Datawarehousing

Letra a.

GABARITO

- | | |
|-------|-------|
| 1. b | 38. a |
| 2. E | 39. d |
| 3. E | 40. a |
| 4. e | 41. a |
| 5. b | 42. E |
| 6. a | 43. E |
| 7. e | 44. E |
| 8. e | 45. C |
| 9. a | 46. e |
| 10. a | 47. E |
| 11. e | 48. d |
| 12. E | 49. e |
| 13. d | 50. E |
| 14. c | 51. E |
| 15. b | 52. E |
| 16. E | 53. C |
| 17. d | 54. C |
| 18. d | 55. E |
| 19. e | 56. C |
| 20. b | 57. b |
| 21. e | 58. E |
| 22. b | 59. E |
| 23. E | 60. e |
| 24. E | 61. b |
| 25. C | 62. a |
| 26. C | 63. c |
| 27. E | 64. d |
| 28. C | 65. E |
| 29. E | 66. E |
| 30. C | 67. E |
| 31. E | 68. E |
| 32. E | 69. C |
| 33. E | 70. c |
| 34. C | 71. c |
| 35. E | 72. a |
| 36. c | |
| 37. a | |

REFERÊNCIAS

BEAL, Adriana. **Gestão estratégica da informação: como transformar a informação e a tecnologia da informação em fatores de crescimento e de alto desempenho nas organizações.** Atlas, 2012.

BIO, S. R. **Sistemas de informação: um enfoque gerencial.** São Paulo: Atlas.

CHAUDHURI, S., DAYAL, U. **An Overview of Data Warehousing and OLAP Technology**, ACM SIGMOD Record, vol.26, 65-74, 1997.

CLARO, Daniela Barreiro. **Dados da Web Atual**. Disponível em: < <http://walderson.com/site/wp-content/uploads/2017/02/CCI.Aula-04.Dados-da-Web-Atual.pdf> > Acesso em: 1 mar. 2020.

ELIAS, D. **As Melhores Ferramentas do Mercado para Business intelligence**. 2014. Disponível em: <<http://corporate.canaltech.com.br/noticia/business-intelligence/As-melhores-ferramentas-do-mercado-para-Business-Intelligence/>>.

>. Acesso em: 20 set. 2020.

GARTNER. 2020. **Magic Quadrant for Analytics and Business intelligence Platforms** Disponível em: <<https://www.gartner.com/doc/reprints?id=1-1-YA9EA1Y&ct200206=&st=sb&ref=lp&signind3bc9455788b4ae25dceff374a965001>>. Acesso em: 11 fev. 2020.

LOPES, Adriano; VILLA, Cláudia Regina. **A Gestão da Qualidade como facilitadora da Gestão da Informação e do Conhecimento em ambiente jurídico**. XI Seminário Latinolberoamericano de Gestión Tecnológica. 2005.

HOSCO. **Recuperação de dados**. Disponível em: <recuperacao-de-dados.eti.br>. Acesso em: 30 set. 2017.

MACHADO, F. N. R. **Tecnologia e Projeto de Data Warehouse**. 2ª edição. São Paulo: Érica, 2004.

MELLO, Ronaldo dos Santos et al. **Dados semiestruturados**. XV Simpósio Brasileiro de Banco de Dados, 2000.

REZENDE, Denis Alcides. **Planejamento de sistemas de informação e informática: guia prático para planejar a tecnologia da informação integrada ao planejamento estratégico das organizações**. 5ed. São Paulo: Atlas, 2016.

SANT'ANA, Ricardo César Gonçalves. **Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação.** Informação & Informação, v. 21, n. 2, p. 116-142, 2016.

TAKEUCHI, Hirotaka; NONAKA, Ikujiro. **Gestão do conhecimento.** Bookman, 2009.

Patrícia Quintão



Mestre em Engenharia de Sistemas e computação pela COPPE/UFRJ, Especialista em Gerência de Informática e Bacharel em Informática pela UFV. Atualmente é professora no Gran Cursos Online; Analista Legislativo (Área de Governança de TI), na Assembleia Legislativa de MG; Escritora e Personal & Professional Coach.

Atua como professora de Cursinhos e Faculdades, na área de Tecnologia da Informação, desde 2008. É membro: da Sociedade Brasileira de Coaching, do PMI, da ISACA, da Comissão de Estudo de Técnicas de Segurança (CE-21:027.00) da ABNT, responsável pela elaboração das normas brasileiras sobre gestão da Segurança da Informação.

Autora dos livros: Informática FCC - Questões comentadas e organizadas por assunto, 3^a. edição e 1001 questões comentadas de informática (Cespe/UnB), 2^a. edição, pela Editora Gen/Método.

Foi aprovada nos seguintes concursos: Analista Legislativo, na especialidade de Administração de Rede, na Assembleia Legislativa do Estado de MG; Professora titular do Departamento de Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia; Professora substituta do DCC da UFJF; Analista de TI/Suporte, PRODABEL; Analista do Ministério Público MG; Analista de Sistemas, DATAPREV, Segurança da Informação; Analista de Sistemas, INFRAERO; Analista - TIC, PRODEMGE; Analista de Sistemas, Prefeitura de Juiz de Fora; Analista de Sistemas, SERPRO; Analista Judiciário (Informática), TRF 2^a Região RJ/ES, etc.

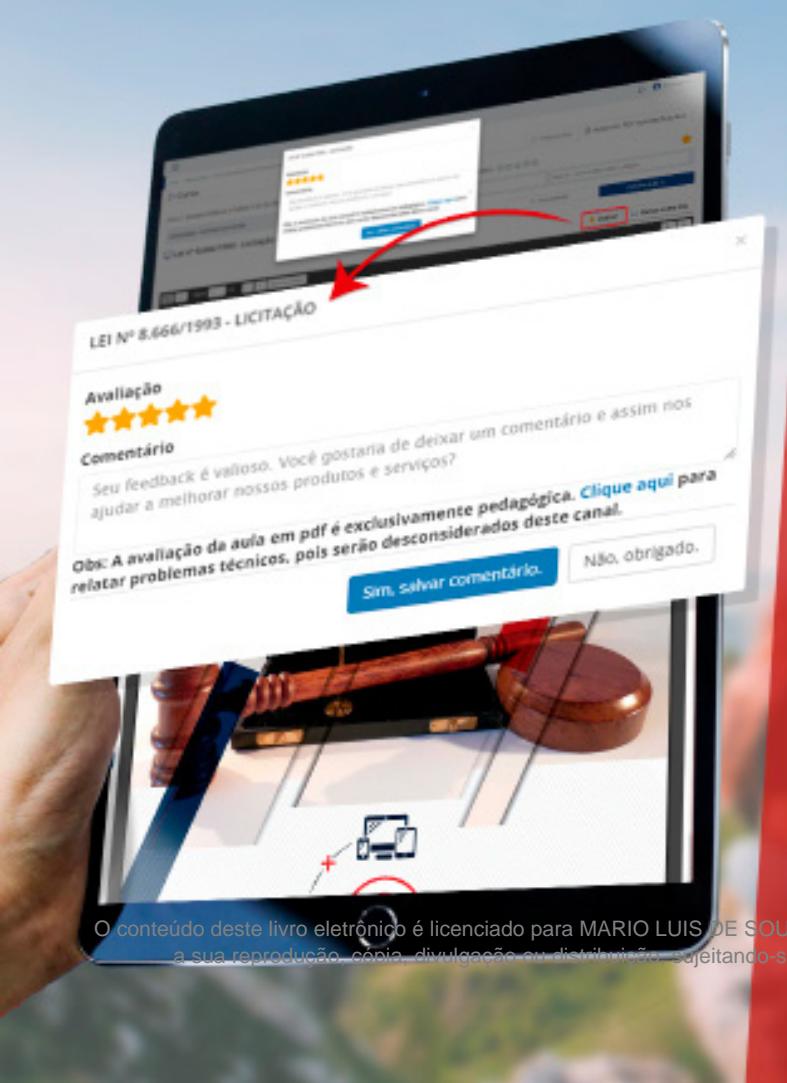
@coachpatriciaquintao

/profapatriciaquintao

@plquintao

t.me/coachpatriciaquintao

O conteúdo deste livro eletrônico é licenciado para MARIO LUIS DE SOUZA - 41250799864, vedada, por quaisquer meios e a qualquer título, a sua reprodução, cópia, divulgação ou distribuição, sujeitando-se aos infratores à responsabilização civil e criminal.



NÃO SE ESQUEÇA DE AVALIAR ESTA AULA!

SUA OPINIÃO É MUITO IMPORTANTE
PARA MELHORARMOS AINDA MAIS
NOSSOS MATERIAIS.

ESPERAMOS QUE TENHA GOSTADO
DESTA AULA!

PARA AVALIAR, BASTA CLICAR EM LER
A AULA E, DEPOIS, EM AVALIAR AULA.

AVALIAR 