



By @kakashi_copiador

Aula 11

Caixa Econômica Federal (CEF) (Técnico Bancário) Passo Estratégico de Informática - 2023 (Pré-Edital)

Autor:

Thiago Rodrigues Cavalcanti

23 de Fevereiro de 2023

VISÃO GERAL SOBRE SISTEMAS DE SUPORTE À DECISÃO E INTELIGÊNCIA DE NEGÓCIO

Sumário

Análise Estatística.....	1
Roteiro de revisão e pontos do assunto que merecem destaque	2
Visão geral sobre sistemas de suporte à decisão e inteligência de negócio	2
Business Intelligence.....	2
Data warehouse	4
Modelagem dimensional	12
Aposta estratégica.....	24
Questões estratégicas	26

ANÁLISE ESTATÍSTICA

Inicialmente, convém destacar os percentuais de incidência de todos os assuntos previstos no nosso curso – quanto maior o percentual de cobrança de um dado assunto, maior sua importância:

Assunto	Grau de incidência em concursos similares
	CESGRANRIO
8 - Visão geral sobre sistemas de suporte à decisão e inteligência de negócio.	23,49%
1 - Edição de planilhas (ambientes Microsoft Office - Excel - versão O365).	20,13%
4 - Redes de computadores: Conceitos básicos, ferramentas, aplicativos e procedimentos de Internet e intranet.	14,09%
1 - Edição de textos (ambientes Microsoft Office - Word - versão O365).	9,40%
2 - Segurança da informação: fundamentos, conceitos e mecanismos de segurança; Segurança cibernética: Resolução CMN nº 4893, de 26 de fevereiro de 2021.	9,40%
3 - Conceitos de organização e de gerenciamento de informações, arquivos, pastas e programas.	7,38%



1 - Edição de apresentações (ambientes Microsoft Office - PowerPoint - versão O365).	6,71%
6 - Correio eletrônico, grupos de discussão, fóruns e wikis. 7 - Redes Sociais (Twitter, Facebook, LinkedIn, WhatsApp, YouTube, Instagram e Telegram).	6,04%
5 - Navegador Web (Microsoft Edge versão 91 e Mozilla Firefox versão 78 ESR).	3,36%
9 - Conceitos de tecnologias e ferramentas multimídia, de reprodução de áudio e vídeo. 10 - Ferramentas de produtividade e trabalho a distância (Microsoft Teams, Cisco Webex, Google Hangout, Zoom, Google Drive e Skype).	0,00%

ROTEIRO DE REVISÃO E PONTOS DO ASSUNTO QUE MERECEM DESTAQUE

A ideia desta seção é apresentar um roteiro para que você realize uma revisão completa do assunto e, ao mesmo tempo, destacar aspectos do conteúdo que merecem atenção.

Para revisar e ficar bem preparado no assunto, você precisa, basicamente, seguir os passos a seguir:

Visão geral sobre sistemas de suporte à decisão e inteligência de negócio

O conceito de sistemas de suporte à decisão e inteligência de negócio é desenvolvido na área da Administração.

De acordo com Chiavenato, os sistemas de informação gerencial (SIG) constituem sistemas computacionais capazes de proporcionar informação como matéria prima para todas as decisões a serem tomadas pelos participantes tomadores de decisão dentro da organização. Os sistemas de informação gerencial constituem uma combinação de sistemas de computação, de procedimentos e de pessoas e têm como base um banco de dados. Todo SIG possui três tipos de componentes: dados, sistemas de processamento de dados e canais de comunicação.

Diante desse conceito, podemos correlacionar os sistemas de suporte à decisão e inteligência de negócio com os seguintes temas da área de Tecnologia: *Business Intelligence* e *Data Warehouse*.

Esse conteúdo está diretamente interligado aos FUNDAMENTOS SOBRE ANÁLISE DE DADOS.

Business Intelligence

Business Intelligence (BI) é um termo abrangente que combina **arquiteturas, ferramentas, bancos de dados, ferramentas de análise, aplicações e metodologias**. Seus principais objetivos incluem:



1. Permitir o acesso interativo, por vezes em tempo real, aos dados.
1. Permitir a manipulação de dados.
2. Dar aos gestores e analistas a capacidade de realizar análise adequada.

Segundo o Turban¹, BI é baseado na transformação dos dados em informação, em seguida, informações em decisões e, finalmente, em ações. Como você pode observar, BI é um termo “guarda-chuva” que engloba diversos componentes. Para visualizar melhor essa estrutura vamos descrever a arquitetura e os componentes de BI.

O BI tem **quatro grandes componentes**:

1. Um **data warehouse** (DW) com seus dados-fonte utilizados para a análise de negócios.
2. A **análise de negócio ou business analytics**, uma coleção de ferramentas para manipular e analisar os dados no data warehouse, incluindo data mining.
3. **Business performance management (BPM)** para monitorar e analisar o desempenho
4. Uma **interface de usuário** (como o dashboard).

Observe que o ambiente de data warehousing é sobretudo de responsabilidade de uma equipe técnica, e o ambiente de análise (também conhecido como análise de negócios) está no âmbito dos usuários de negócios. A figura a seguir tenta apresentar esses conceitos de forma mais organizada dentro da arquitetura de BI.

Dentro das ferramentas de interface com os usuários temos o dashboard. Os **dashboards** fornecem **uma visão abrangente e visual das medidas** (indicadores-chave de desempenho), tendências e exceções do desempenho corporativo provenientes de múltiplas áreas do negócio. Os **gráficos** mostram o desempenho real em comparação às métricas desejadas, propiciando **uma visão imediata da saúde da organização**. Outras ferramentas que “transmitem” informações são **portais corporativos, cockpits digitais e outras ferramentas de visualização**.

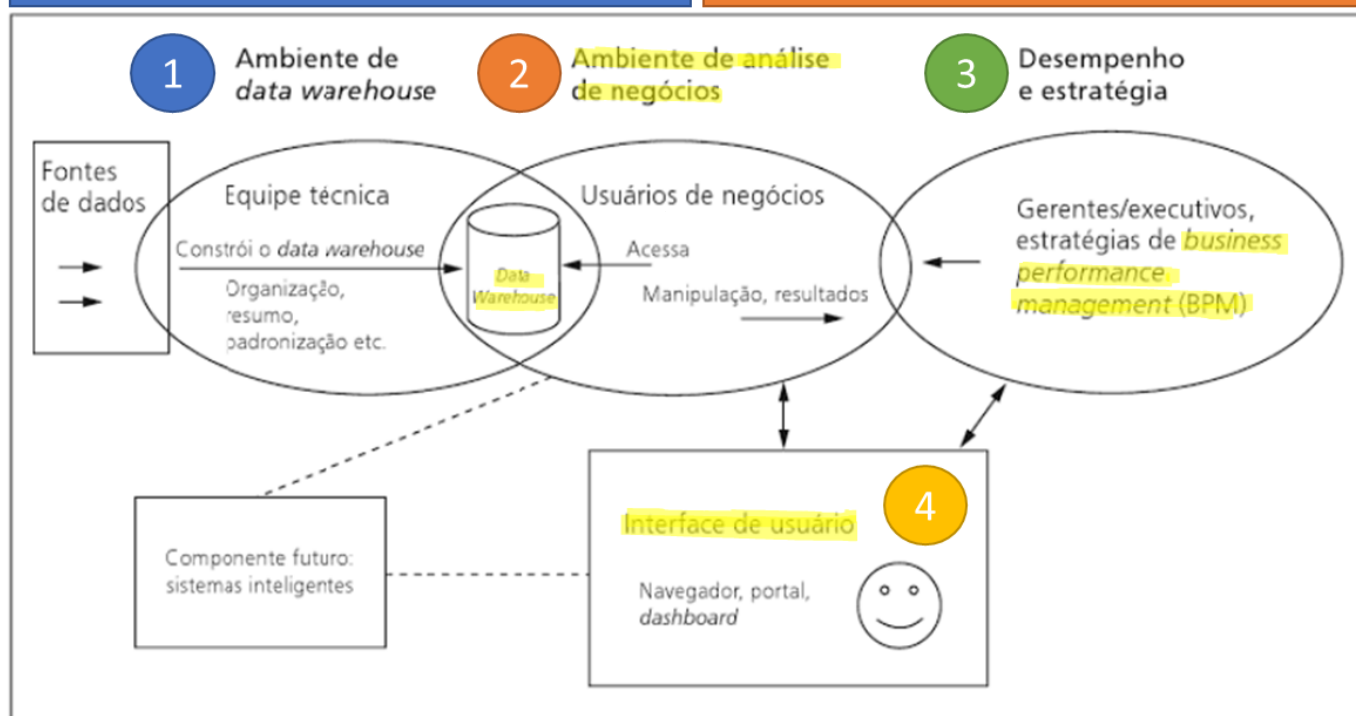
¹ Efraim Turban (M.B.A., Ph.D, Universidade da Califórnia, Berkeley) é professor convidado no Pacific Institute for Information System Management, na Universidade do Havaí. Autor de vários livros, incluindo Business Intelligence: Um enfoque gerencial para a inteligência do negócio.



1. Data Warehouse - repositório de dados especial preparado para dar suporte a aplicações de tomada de decisão.

2. Análise de negócios - ferramentas de software que permitem aos usuários realizarem análises de dados.

OLAP, Mineração de dados



3. Business (ou corporate) performance management (BPM) baseia no BSC, uma estrutura para definir, implementar e gerenciar a estratégia de negócios conectando objetivos a medidas factuais.

4. Interface de usuário - dashboards e outras ferramentas de transmissão de informações.

Figura 1 - Arquitetura de alto nível do BI

Data warehouse

Definição

Vamos apresentar a definição de DW de três diferentes autores:

Kimball: É um conjunto de ferramentas e técnicas de projeto, que quando aplicadas às necessidades específicas dos usuários e aos bancos de dados específicos permitirá que planejem e construam um Data Warehouse.



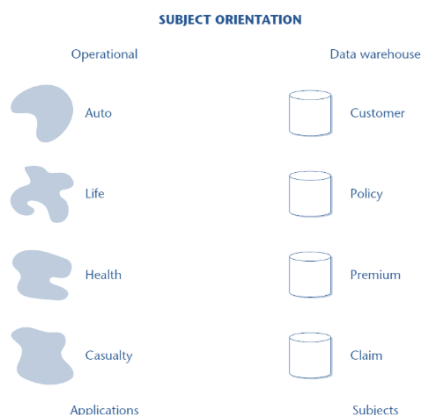
Laudon&Laudon: Banco de dados, com ferramentas de consulta e relatório, que armazena dados atuais e históricos extraídos de vários sistemas operacionais e consolidados para fins de análises e relatórios administrativos.

Inmon: É uma coleção de dados orientados por assunto, integrados, variáveis com o tempo e não voláteis, para dar suporte ao processo de tomada de decisão.

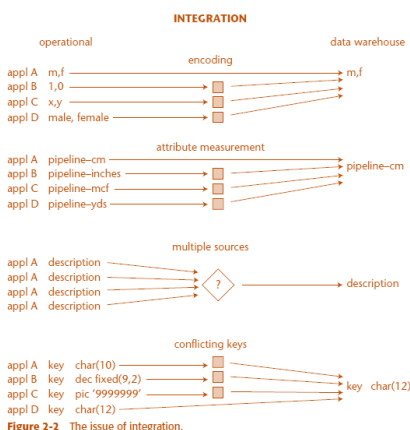
Um comentário interessante sobre essas definições, em especial sobre as definições do Inmon e do Kimball: perceba que **o Kimball** apresenta dentro do conceito um conjunto de ferramentas e técnicas, ou seja, **o processo de estruturar um data warehouse** está incluso dentro da definição. O Inmon, por sua vez, considera **apenas a coleção de dados**. Ou seja, o DW é um repositório de dados **atuais e históricos** que podem ser usados por gestores da organização. Esse conjunto de dados será organizados de acordo com **algumas características** e dará suporte **a tomada de decisão**.

Característica de um DW

Na sessão anterior, a definição do Inmon traz consigo quatro características que são de suma importante para o entendimento do assunto. Juntam-se a elas a granularidade e a credibilidades dos dados. Essas características precisam ser entendidas para a compreensão do assunto.

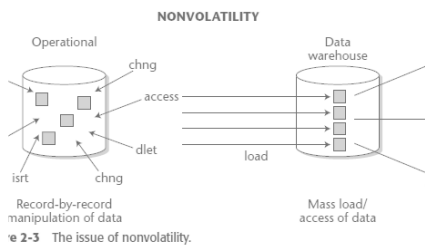


Ser **orientado por assunto** refere-se ao fato do *Data Warehouse* armazenar informações sobre temas específicos importantes para o negócio da empresa. São exemplos típicos de temas: produtos, atividades, contas, clientes. Em contrapartida, quando observamos o ambiente operacional percebemos que ele é organizado por aplicações funcionais. Por exemplo, em uma organização bancária, estas aplicações incluem empréstimos, investimentos e seguros. Observe um exemplo na figura abaixo retirada do livro do Inmon

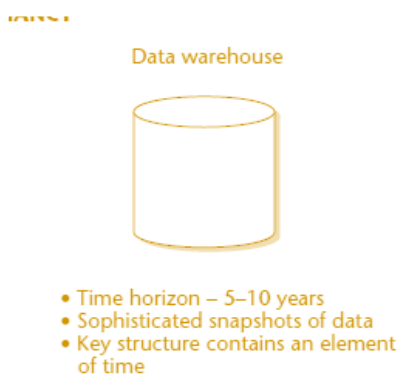


Ser **integrado** refere-se à consistência de nomes, das unidades, das variáveis, etc. É importante que os dados armazenados sejam transformados até um estado uniforme. Por exemplo, considere sexo como um elemento de dado. Uma aplicação pode codificar sexo como M/F, outra como 1/0 e uma terceira como H/M. Conforme os dados são inseridos ou repassados para o *Data Warehouse*, eles são convertidos para um mesmo padrão. O atributo Sexo, portanto, seria codificado apenas de uma forma.

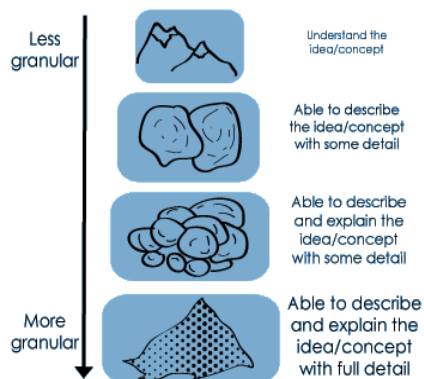




O fato de ser **não volátil** significa que o *Data Warehouse* permite apenas a carga inicial dos dados e consultas a estes dados. Após serem integrados e transformados, os dados são carregados em bloco para o DW, para que estejam disponíveis aos usuários para acesso.



Ser **variante no tempo** trata do fato de um registro 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 DW, a cada ocorrência de uma mudança, uma nova entrada é criada para marcar esta mudança.



A **granularidade de dados** refere-se ao nível de sumarização dos elementos e de detalhe disponíveis nos dados, considerado **o mais importante aspecto** do projeto de um *Data Warehouse*. Em um nível de granularidade muito alto, **mais espaço demandado e mais precisão**. Em contrapartida, quanto menor a granularidade, maior é o nível de detalhes e maior será o seu volume de dados armazenados.

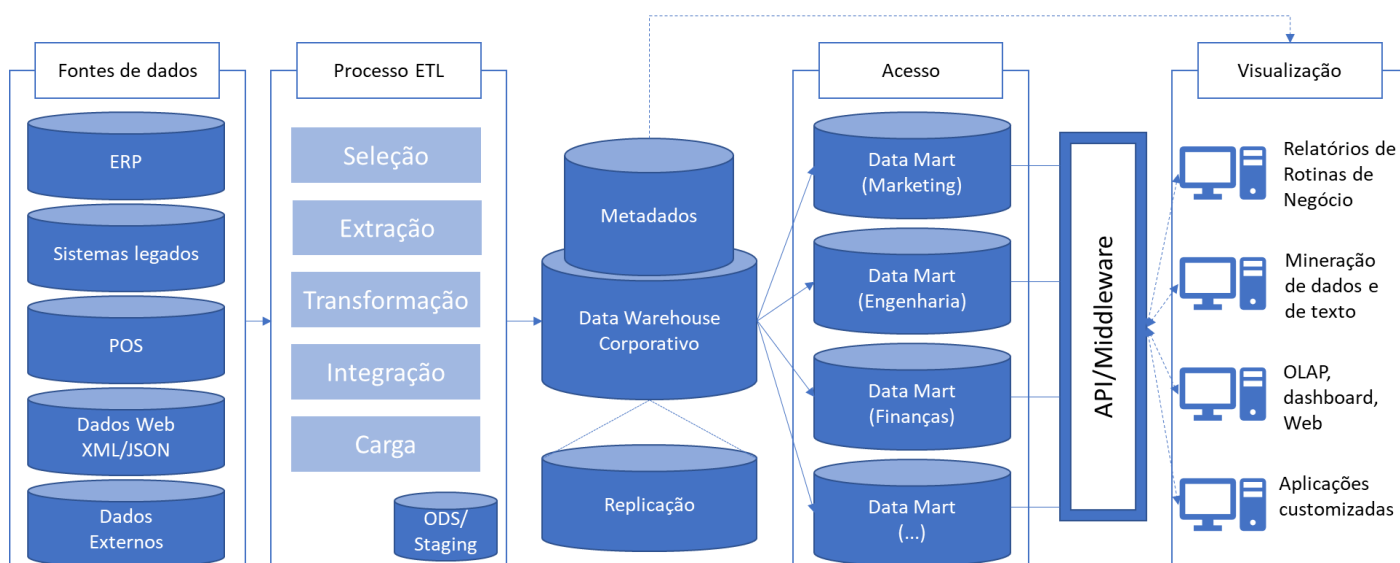




É necessário, para que as análises sejam consideradas corretas que os dados tenham baixa dispersão e estejam consistentes com os valores observados. É importante lembrar que serão feitas várias manipulações nos dados, inclusive agregações e cálculos estatísticos. A depender dos desvios presentes nos dados as análises podem ser inócuas ou incoerentes com a realidade.

Processo de Data Warehousing

Apresentamos abaixo uma figura que descreve o processo de DW/BI. Essa figura exhibe os **componentes** que fazem parte do sistema.



Vejamos a descrição sucinta de cada um dos principais **componentes**:

3. **As fontes de dados (*Data source*)** - Os dados são provenientes de vários sistemas legados² independentes e, possivelmente, a partir de provedores de dados externos. Os dados também podem vir de um processamento de transações online (OLTP) ou sistema de ERP. Dados da Web na forma de logs também podem alimentar um data warehouse.
4. **Extração e transformação de dados (*Data extraction and transformation*)** - Os dados são extraídos e devidamente transformados usando software personalizado ou comercial chamado de ferramenta de ETL.

² Sistema legado é o termo utilizado em referência aos sistemas computacionais de uma organização que, apesar de serem bastante antigos, fornecem serviços essenciais. (Wikipédia)



5. **Carregamento dos dados. (Data loading)** – Depois que os dados são carregados para uma área de preparação (*staging area*), onde eles são transformados e limpos, eles estão, então, prontos para serem carregados no *data warehouse (EDW)* e/ou *data marts*.
6. **Banco de dados abrangente.** - Essencialmente, este é o EDW para apoiar todas as análises de suporte à decisão, fornecendo uma visão resumida relevante e informações detalhadas proveniente de fontes diferentes.
7. **Metadados** - Os metadados são mantidos para que possam serem avaliados pelo pessoal de TI e pelos usuários. Metadados incluem programas sobre os dados e as regras para a organização de resumos de dados que são fáceis de indexar e pesquisar, especialmente com as ferramentas da Web.
8. **Ferramentas de middleware** - Ferramentas de middleware habilitam o acesso ao DW. Usuários avançados, como os analistas de BI, podem escrever suas próprias consultas SQL. Outros podem empregar um ambiente gerenciador de consulta, como *Business Objects*, para acessar os dados. Existem muitas aplicações *front-end* que os usuários podem usar para interagir com os dados armazenados nos repositórios de dados, incluindo a mineração de dados, OLAP, ferramentas de relatórios e ferramentas de visualização de dados.

Duas observações são importantes sobre a figura anterior. Primeiramente, você deve ter percebido a presença de um elemento denominado **ODS (operational data storage)/Staging**. Essa base de dados ajuda as ferramentas ETL a trabalharem com os dados no processo de transformação e integração dos dados. O outro ponto seria o **Data Mart (DM)**.

Considerando que um armazém de dados (DW) combina bases de dados de toda a empresa, um **Data Mart (DM)** é geralmente menor e se concentra em um assunto ou departamento específico. Um DM é um subconjunto de um armazém de dados, geralmente constituídos por uma única área temática (por exemplo, marketing, operações). O DM pode ser **dependente** ou **independente**.

Ser **dependente** é ser um subconjunto que é criado diretamente a partir do armazém de dados. Esse modelo tem como vantagens de usar um modelo de dados consistente e o fornecimento de dados de qualidade. DM dependentes apoiam o conceito de um único modelo de dados em toda a empresa, mas o *data warehouse* deve ser construído primeiro. O DM **dependente** garante que o usuário final está vendo a mesma versão dos dados que é acessada por todos os outros usuários de data warehouse.

O alto custo de dados armazéns limita seu uso para grandes empresas. Como uma alternativa, muitas empresas usam uma versão em escala reduzida de um armazém de dados referida como uma DM **independente**. Um *Data Mart independente* é um pequeno armazém concebido para uma unidade estratégica de negócios ou um departamento, mas sua origem não é um EDW. Percebam que o DM tem como sua principal característica o escopo reduzido do projeto.

Arquitetura de Data Warehouse

Existem diversas arquiteturas de DW. Contudo, é importante perceber que algumas partes aparecem em todas as arquiteturas. As três partes de uma arquitetura são:

1. O **próprio DW**, que contém os dados e o software associados.
2. **Software de aquisição de dados (back-end)**, que extrai dados de sistemas legados e fontes externas, os consolida e resume, e depois os carrega.



3. Software cliente (front-end), que permite aos usuários acessar e analisar dados a partir do DW.

As figuras a seguir mostram os diferentes tipos de arquitetura para sistemas de DW. Em seguida, à cada uma delas, apresentaremos as definições relacionadas.

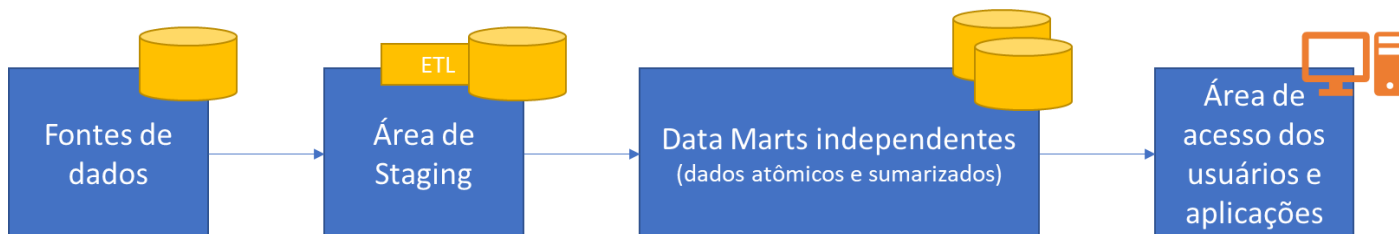


Figura 2 - Arquitetura de data marts independentes

Data Marts (DM) independentes. Esta é sem dúvida a arquitetura **mais simples e menos onerosa**. Os DM são desenvolvidos para operar de forma independente visando atender às necessidades das unidades organizacionais individuais. Por causa da independência, eles podem ter definições de dados inconsistentes, diferentes dimensões e medidas, o que torna difícil analisar os dados cruzando informações através de *Data Marts* distintos.

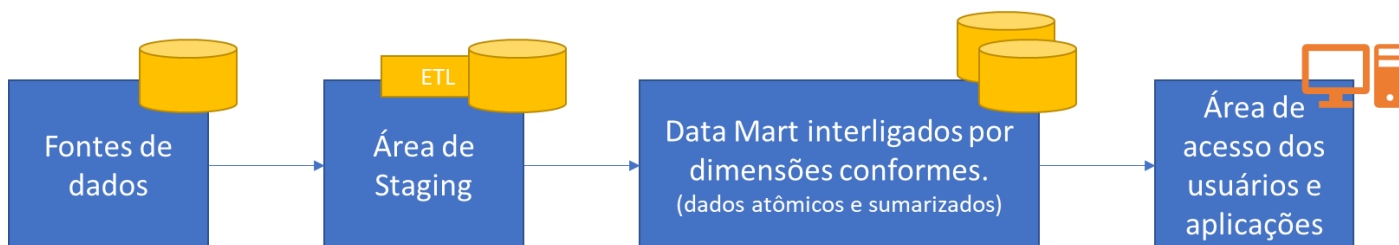


Figura 3 - Arquitetura de barramento de Data Mart (KIMBALL)

Arquitetura de barramento de Data Mart (KIMBALL). Esta arquitetura é **uma alternativa viável para os data marts independentes**. Nela os *Data Marts* individuais são interligados através de algum tipo de middleware. Como os dados são interligados entre os DM individuais, há uma chance alta de que a consistência entre os dados de toda a empresa (pelo menos no nível de metadados) seja mantida. Mesmo que consultas complexas utilizem dados de diversos DMs, o desempenho destes tipos de análise pode não ser satisfatório.



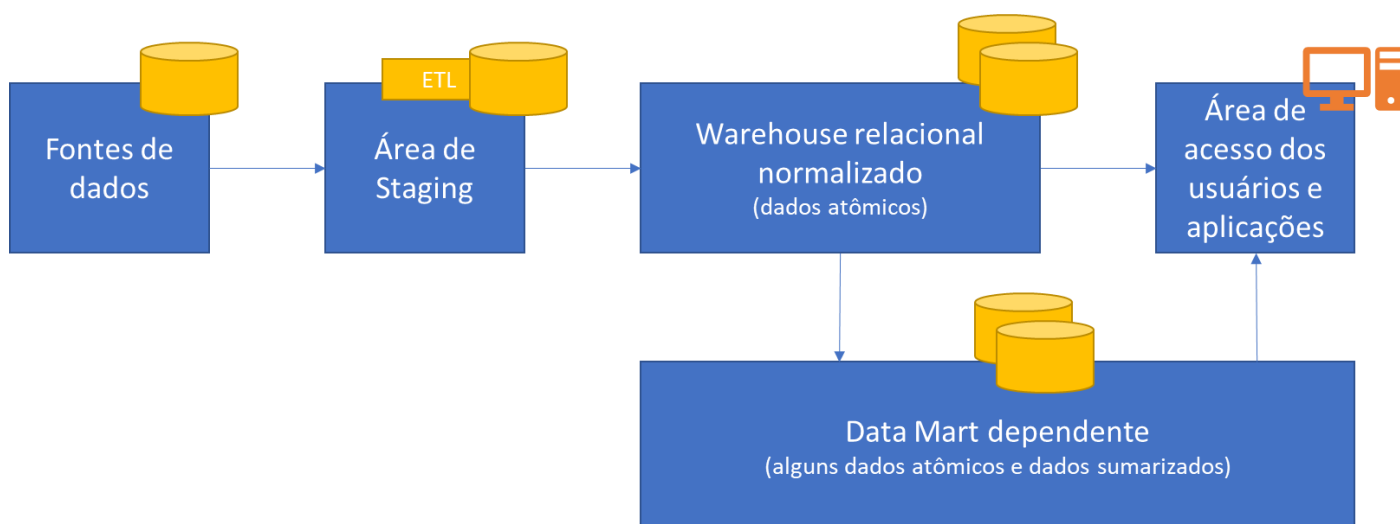


Figura 4 - Arquitetura Hub-and-spoke (INMON)

Hub-and-spoke (Inmon). Esta é talvez a mais famosa de arquitetura de DW. Aqui, a atenção está focada **na construção de uma infraestrutura escalável e sustentável** (muitas vezes desenvolvidas de forma iterativa, assunto por assunto) que inclui **um data warehouse centralizado e vários data marts dependentes** (um para cada unidade organizacional). Esta arquitetura permite uma customização fácil de interfaces de usuário e relatórios. No lado negativo, esta arquitetura não tem a visão holística, e pode levar a redundância e a latência de dados. A latência dos dados refere-se ao tempo máximo permitido para disponibilização dos dados através do sistema de BI.

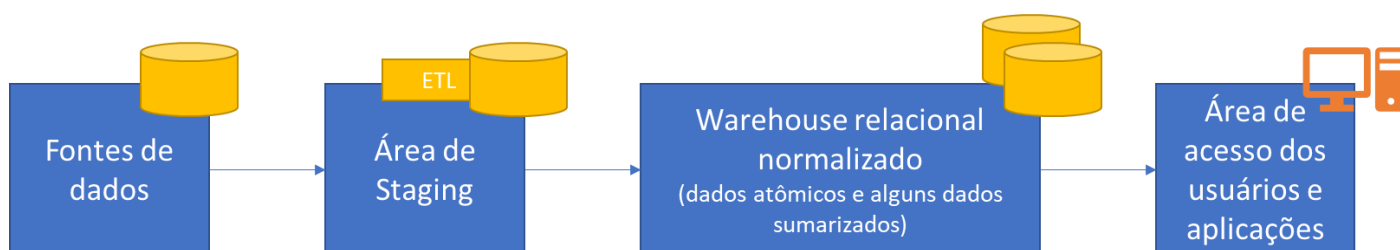


Figura 5 - Arquitetura de Data Warehouse centralizado

Data warehouse centralizado. A arquitetura centralizada de *data warehouse* é similar à arquitetura *hub-and-spoke*, exceto que **não há Data Marts dependentes**, em vez disso, há **um data warehouse da empresa inteira que serve para a necessidade de todas as unidades organizacionais**. Esta abordagem centralizada fornece aos usuários acesso a todos os dados no warehouse em vez de limitar-lhes a DM de dados. Adicionalmente, reduz-se a quantidade de dados que a equipe técnica tem de transferir, portanto, **simplifica-se o gerenciamento e administração de dados**. Se projetado e implementado corretamente, essa arquitetura oferece uma visão oportuna e holística da empresa quem, quando e onde quer que esteja dentro da organização. A arquitetura de armazéns de dados central, que **é defendida principalmente pela Teradata Corporation**, aconselha usar armazéns de dados sem data marts.



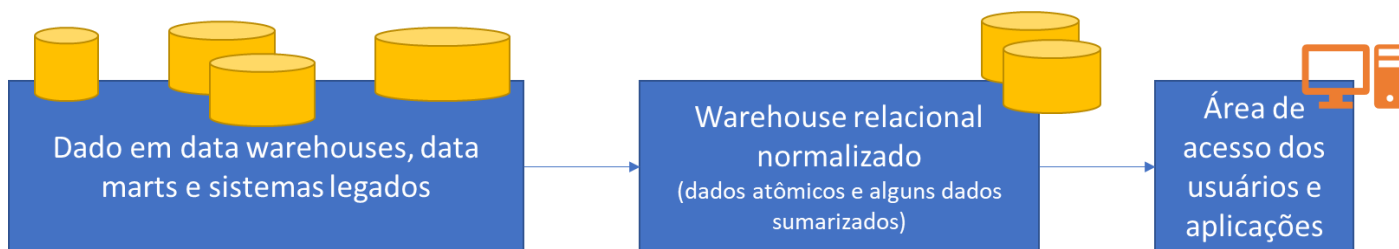


Figura 6 - Arquitetura de armazém de dados federada

Armazém de dados federado. A abordagem federada é uma concessão as forças naturais que minam os melhores planos para o desenvolvimento de um sistema perfeito. Ela usa todos os meios possíveis para integrar recursos analíticos de várias fontes para atender evolução das necessidades e condições do negócio. Essencialmente, a abordagem **envolve a integração de sistemas distintos**. Em uma arquitetura federada de apoio à decisão, **estruturas existentes são deixadas no lugar**, e os dados são cedidos a partir dessas fontes, conforme necessário. A abordagem federada é suportada por fornecedores de middleware distribuídos que propõem consultar e juntar capacidades.

Kimball x Inmon

Considerados os dois principais autores de *Data Warehouse*, **Bill Inmon** e **Ralph Kimball** travam há décadas uma batalha teórica no campo de BI.

Ralph Kimball é um defensor da teoria de que o DW deve ser dividido para depois ser conquistado, ou seja, que o mais viável para as empresas é desenvolver vários *Data Marts* para posteriormente integrá-los e, assim, chegar-se ao EDW. Na sua avaliação, as empresas devem construir *Data Marts* orientados por assuntos. Ao final, teríamos uma série de pontos de conexão entre eles, que seriam as tabelas Fato e Dimensões em conformidade. Dessa forma, informações entre os diferentes *Data Marts* poderiam ser geradas de maneira íntegra e segura. Kimball batizou esse conceito de *Data Warehouse Bus Architecture*.

Bill Inmon rebate essa teoria e propõe justamente o contrário. Na sua avaliação deve-se **construir primeiramente um Data Warehouse**, modelando toda a empresa para se chegar a um único modelo corporativo, partindo posteriormente para os *Data Marts* construídos por assuntos ou departamentais. Inmon defende a ideia de que o ponto de partida seriam os *CIF – Corporate Information Factory* – uma infraestrutura ideal para ambientar os dados da empresa. O CIF seria alimentado pelos sistemas transacionais. A construção de um ODS (*Operational Data Store*) seria facultativa, mas essa iniciativa ajudaria a reduzir a complexidade da construção de um DW, uma vez que todo o esforço de integração entre os sistemas transacionais da empresa seria depositado nele.

A tabela abaixo resume os principais pontos divergentes entre os dois autores nessa batalha de gigantes:



Característica	Inmon	Kimball
Metodologia e arquitetura	De cima para baixo (TOP DOWN)	De baixo para cima (BOTTOM UP)
Estrutura arquitetônica	Data warehouse para toda a empresa (atômico) alimenta os bancos de dados departamentais. (DW → DM)	Os data marts modelam um único processo de negócios, e chega-se à consistência da empresa por meio de um barramento de dados. (DM → DW)
Modelagem dos dados	Normalizados até a terceira forma normal (3FN) → Snow flake schema	Não normalizado (Star schema)
Orientação dos dados	Por assunto ou por dados	Por processos
Acessibilidade ao usuário final	Baixa	Alta
Público alvo	Profissionais de TI	Usuários finais
Ferramentas de modelagem	Tradicionais (diagrama entidade-relacionamento e diagrama de fluxo de dados)	Modelagem dimensional

Inmon é considerado o **pai do conceito de DW** e sustenta a tese de que a melhor estratégia seria a construção de um DW de forma **TOP-DOWN**. A sua ênfase sempre foi em um grande depósito central de informações. O **Kimball** é considerado o criador do conceito de star schema. Ele propõe uma abordagem **BOTTOM-UP** para construção do DW, sendo considerado um estilo mais simples com uma abordagem incremental. Veja que TOP-DOWN termina com N assim como Inmon. 😊

Modelagem dimensional

Tendo compreendido os requisitos de dados exclusivos para o suporte à decisão, você está preparado para conhecer mais de perto a tecnologia que permite cumprir esses requisitos. O modelo de dados multidimensional é **adequado para a representação de dados e as operações concebidas especificamente para o processamento de suporte à decisão nos data warehouses**.

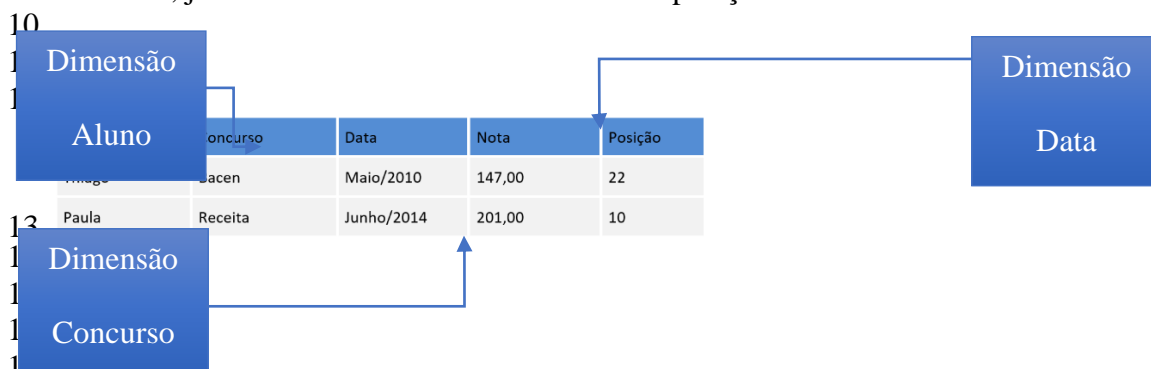
Dados brutos e concisos podem ser organizados de maneiras diferentes para análise e apresentação. Uma maneira eficiente de fazer isso é chamada de multidimensionalidade. **A maior vantagem da multidimensionalidade é que ela permite que os dados sejam organizados como cada usuário gosta de vê-los**, e não como os analistas de sistemas gostam. Apresentações diferentes dos mesmos dados podem ser providenciadas de modo rápido e fácil.

A modelagem multidimensional **oferece um conjunto de técnicas** para projetar a estrutura de um banco de dados **que seja fácil e rápido na recuperação dos dados**. Seus benefícios são simplicidade e performance para trabalhar com dados agregados ou sumarizados. Neste sentido precisamos nos preocupar com 2 conceitos importantes: **dimensões** e **medidas**. As dimensões são as perspectivas ou atributos que serão usados para representar um evento de negócio. Um evento de negócio pode ter uma ou mais medidas. Vamos construir um exemplo.

- Imagine que Ricardo Vale e Heber Carvalho resolveram criar uma base de dados para organizar os resultados dos alunos do Estratégia ao longo dos anos. Para isso, eles resolveram analisar os dados em 3 dimensões: Alunos, Concursos e Tempo. Agora, eles

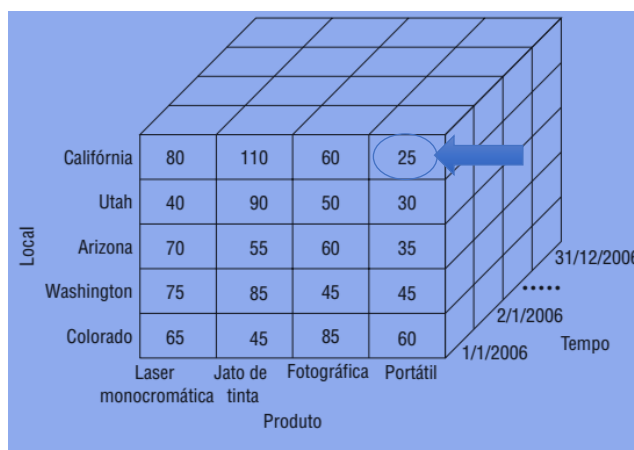


vão construir uma tabela central Aprovação que lista o nome do aluno, o ano e mês do concurso, juntamente com as medidas de nota e posição do aluno nos concursos.



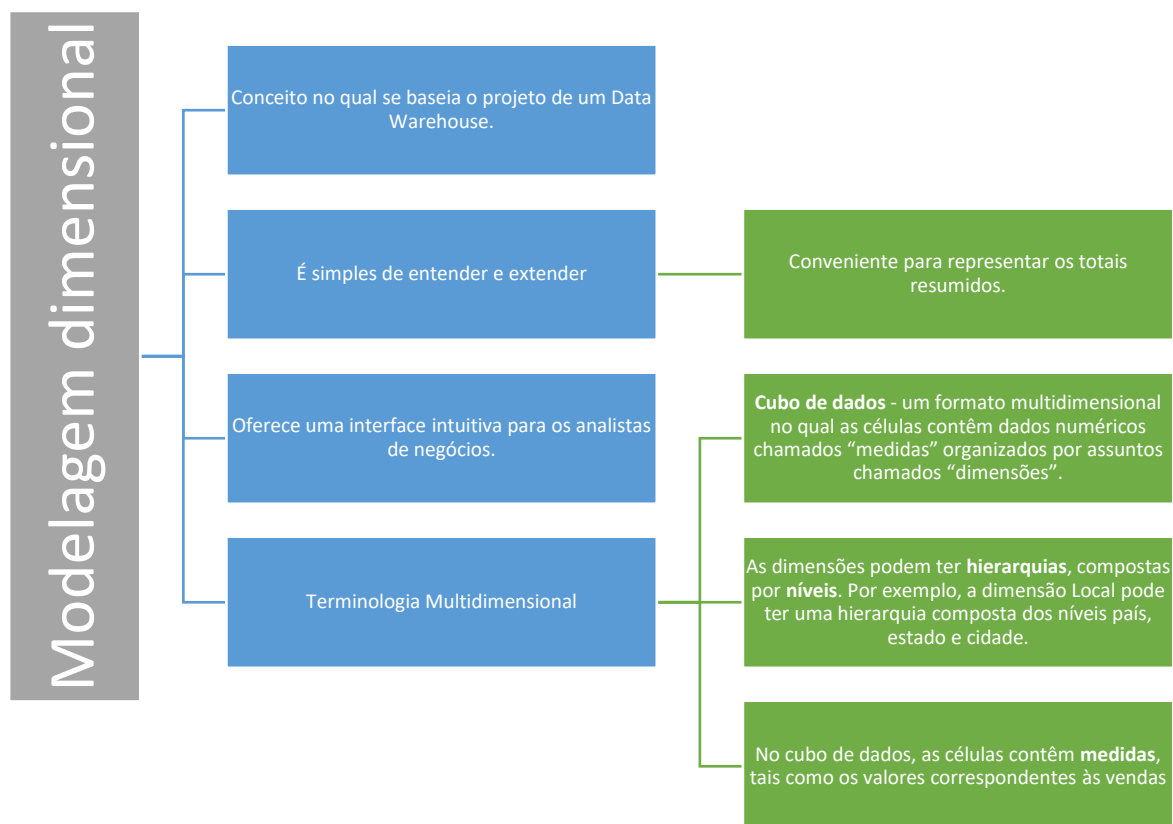
18. Você consegue perceber que existe uma simplicidade no modelo que é adequada a forma como os gestores querem visualizar os dados em determinado momento? No cenário criado optamos por apresentar os dados em tabelas (banco de dados relacionais), mas existe outra forma de armazenamento dos dados em disco: nos bancos de dados multidimensionais.

Uma outra forma de descrever os dados é por meio do **cubo de dados**. O cubo de dados ou hipercubo generaliza as representações bidimensional. O cubo de dados é formado por células, as quais contêm medidas (valores numéricos como a quantidade de vendas unitárias) e dimensões para classificar ou agrupar dados numéricos (por exemplo, Produto, Local e Tempo). Vejamos um exemplo para não ficarmos num universo muito abstrato:



Cada dimensão contém valores denominados **membros**. Por exemplo, a dimensão Local tem cinco membros (Califórnia, Washington, Utah, Arizona e Colorado). Se eu pedir para você me dizer quantas impressoras (Produto) foram vendidas no dia 1º de janeiro de 2006 na Califórnia, qual seria sua resposta? Se você respondeu 25? Parabéns! Você está acompanhando bem a nossa aula. Antes de continuarmos vamos montar um diagrama com os conceitos que vimos até aqui:





Resumindo: 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* nas mais diferentes perspectivas. A visão multidimensional permite o uso mais intuitivo para o processamento analítico pelas ferramentas OLAP (*On-line Analytical Processing*).

Modelagem de dados relacional para Dados Multidimensionais

Para usar um banco de dados relacional em um data warehouse, uma nova técnica de modelagem de dados é necessária para representar os dados multidimensionais. O **esquema estrela** é uma representação de modelagem própria dos cubos de dados multidimensionais. No banco de dados relacional, o diagrama desse esquema **tem a aparência de uma estrela**, em cujo **centro** encontra-se uma grande tabela, denominada **tabela de fatos**, que está vinculada a várias **tabelas dimensão**, **dispostas radialmente**.

A tabela de **fatos armazena dados numéricos (fatos)**, como resultados de vendas, ao passo que as **tabelas dimensão armazenam dados descritivos** correspondentes a dimensões individuais do cubo de dados, como produto, local e tempo. Existe um relacionamento 1-N entre cada tabela de dimensão e a tabela de fatos. Cada **processo de negócio** é representado por um modelo dimensional.

Uma tabela fato contém vários fatos, correspondentes a cada uma das suas linhas. Cada linha corresponde a um **evento de medição**. Os dados em cada linha estão a um nível específico de detalhe, referido como o **grão**, por exemplo, uma linha por produto vendido numa operação de venda.

Uma **única linha da tabela fato** tem uma relação **um-para-um com o evento de medição**, como descrito pela **granularidade** da tabela fato. Uma **tabela fato corresponde a um evento físico observável, e não às exigências de um relatório específico**. Dentro de uma tabela fatos, apenas fatos consistentes com a



granularidade definida são permitidos. Por exemplo, em uma transação de vendas no varejo, a quantidade de um produto vendido e seu preço são bons fatos.

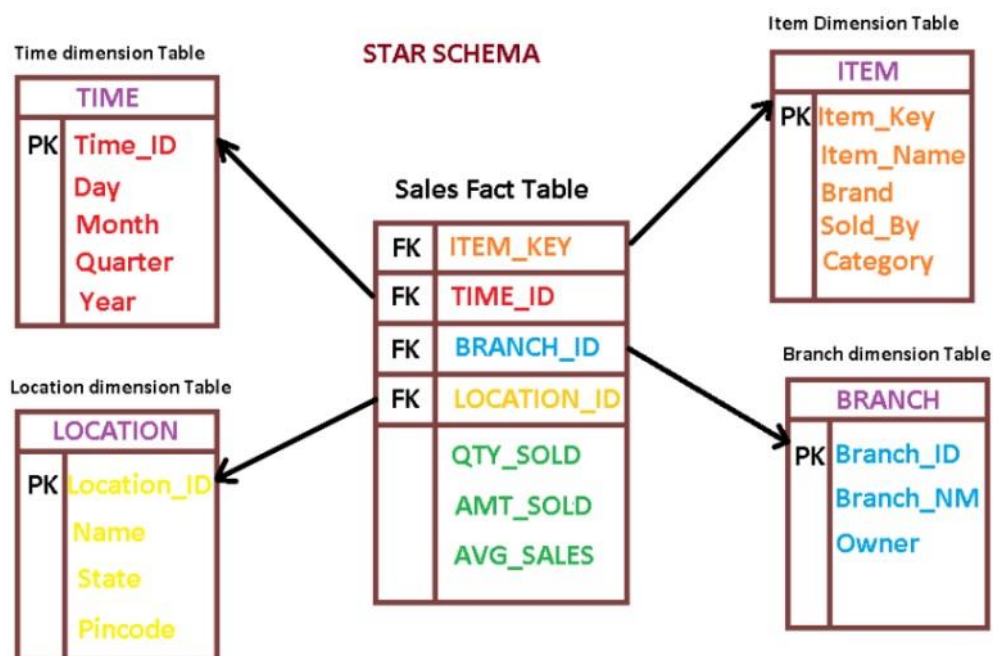
A ideia de que **um evento de medição no mundo físico** tenha uma relação de um-para-um com uma única linha na tabela fato é um princípio fundamental para a modelagem dimensional. Tudo o resto constrói a partir desta fundação. Você vai ver que os **fatos** são, por vezes **aditivos**, **semi-aditivos** ou mesmo **não-aditivos**.

Aditivos	Semiaditivos	Não aditivos
<ul style="list-style-type: none">• Podem ser agrupadas em uma qualquer das dimensões associadas à tabela de fatos• Os fatos mais flexíveis e úteis.• Lucro líquido	<ul style="list-style-type: none">• Podem ser agrupadas em algumas dimensões, mas não todas.• Ex.: Saldo em conta - não podem ser resumidos por meio da dimensão de tempo	<ul style="list-style-type: none">• Nunca podem ser adicionados ou somados• Taxas e percentuais

Perceba que alguns atributos, mesmo sendo numéricos, não fazem sentido serem agregados ou somados. Outro ponto é que algumas dimensões **não numéricas** podem ser **eventualmente um fato**. Nestes casos as **informações textuais** só permitem contagem e estatísticas associadas a quantidade de eventos com a mesma descrição.

Todas as **tabelas fato** têm **duas ou mais chaves estrangeiras** que ligam para as chaves primárias das tabelas de dimensão. Por exemplo, a chave do produto na tabela fato deve ser sempre associada a um produto específico na tabela de dimensão de produto. Quando todas as chaves na tabela fato corretamente coincidirem com as respectivas chaves primárias das tabelas dimensão correspondentes, as tabelas satisfazem a integridade referencial. Você pode acessar a tabela de fatos através das tabelas de dimensões por meio de uma junção. A relação entre as chaves das tabelas fato e dimensão pode ser vista na figura abaixo:





Chave **primária** da tabela fato:

Composta pelos atributos que “apontam” para as chaves primárias das tabelas dimensão.

No exemplo temos:
ITEM_KEY,
BRANCH_ID,
Location_ID e
Time_ID

As **tabelas dimensões** apresentam o contexto **descritivo**. São companheiros integrais para uma tabela de fatos e contém o aspecto textual e discreto associados a um evento de medição dos processos de negócios. Elas descrevem o "**quem, o que, onde, quando, como e porque**" associado ao evento. As tabelas dimensões representam **aspectos qualitativos** de seu data warehouse.

A tabela de dimensão é projetada com **uma coluna que funciona como uma chave primária única**. Esta chave primária **não** pode ser a chave natural da base de dados operacional, pois haverá várias linhas na dimensão referenciando a chave natural. Por exemplo, quando alterações são feitas no preço de um produto no modelo operacional, esse registro precisa ser representado como uma nova linha no DW. Além disso, chaves naturais para uma dimensão podem ser criadas por mais do que um sistema, e estas chaves naturais podem ser incompatíveis ou mal administradas.

O sistema de DW/BI precisa reivindicar o controle das chaves primárias de todas as dimensões, ao invés de usar chaves naturais explícitas ou chaves naturais com datas concatenadas, você deve criar **as chaves primárias inteiras (numéricas) anônimas para cada dimensão**. Essas chaves são conhecidas como **chaves artificiais ou substitutas**. (Faz sentido! Elas substituem as chaves naturais.)

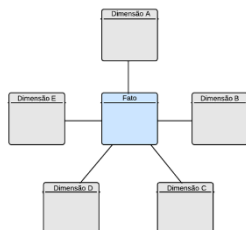
Estas chaves substitutas para dimensão são **números inteiros simples, atribuídos em sequência**, a cada vez que uma nova chave é necessária, começando com o valor um. Uma observação: a **dimensão data é isenta da regra fundamental da chave substituta**. Esta dimensão altamente previsível e estável pode usar uma chave primária mais significativa.

Antes de continuarmos vamos tentar resumir os pontos associados as tabelas presentes no modelo dimensional implementado em bancos de dados relacionais.



Tabelas

Fato



Dimensão

As linhas apresentam fatos associados a um processo de negócio específico.

Cada tupla apresenta as medidas ou métricas que podem ser usados para análise de decisões e relatórios de consulta.

Aspectos quantitativo e chave primária composta.

Apresentam os níveis das hierarquias.

Descreve como os dados serão acessados e analisados.

Contém informações de classificação e agregação.

Tem uma relação de 1-N com as linhas da tabela fato central.

A figura abaixo apresenta um exemplo de um esquema estrela. Neste esquema uma **única tabela fato** aparece no modelo e está **ligada diretamente a todas as dimensões**. As dimensões, por sua vez não são normalizadas, o que leva a **um menor número de tabelas** quando comparado com o esquema floco de neve (falaremos sobre ele a seguir). Esse tipo de esquema baseado nos princípios de **simplicidade e performance** é mais eficiente em termos de tempo de resposta bem como mais fácil de ser entendido pelos usuários.

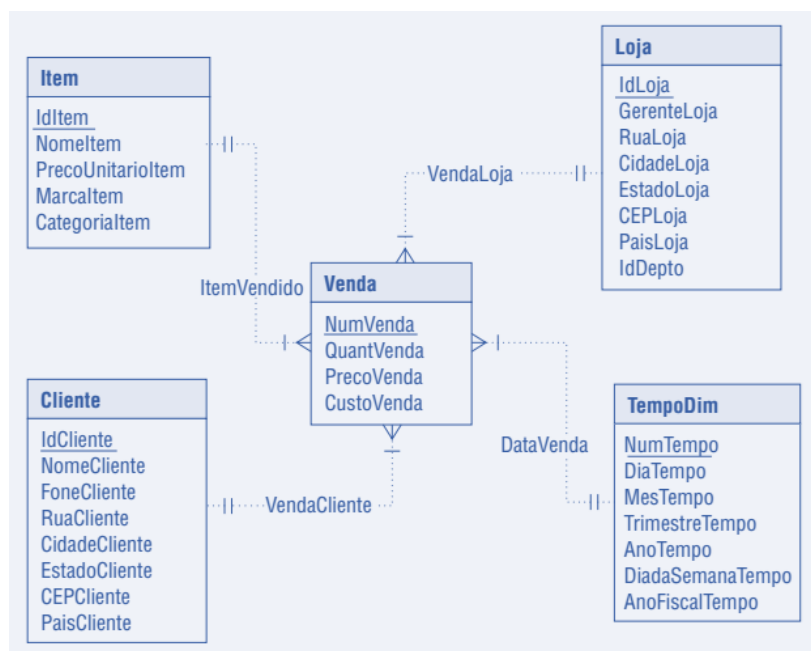


Figura 7 - Esquema Estrela de vendas

Variações do Esquema Estrela

Esquema constelação ou multiestrela

O esquema estrela da figura acima representa **um único processo de negócio** para o rastreamento de vendas. Outros esquemas podem ser necessários para outros processos, como **entregas e compras**. No caso de processos de negócio relacionados que **compartilhem algumas das tabelas dimensionais**, o esquema estrela



pode ser estendido para um **esquema constelação**, com vários tipos de entidades de fatos, como mostra a figura a seguir.

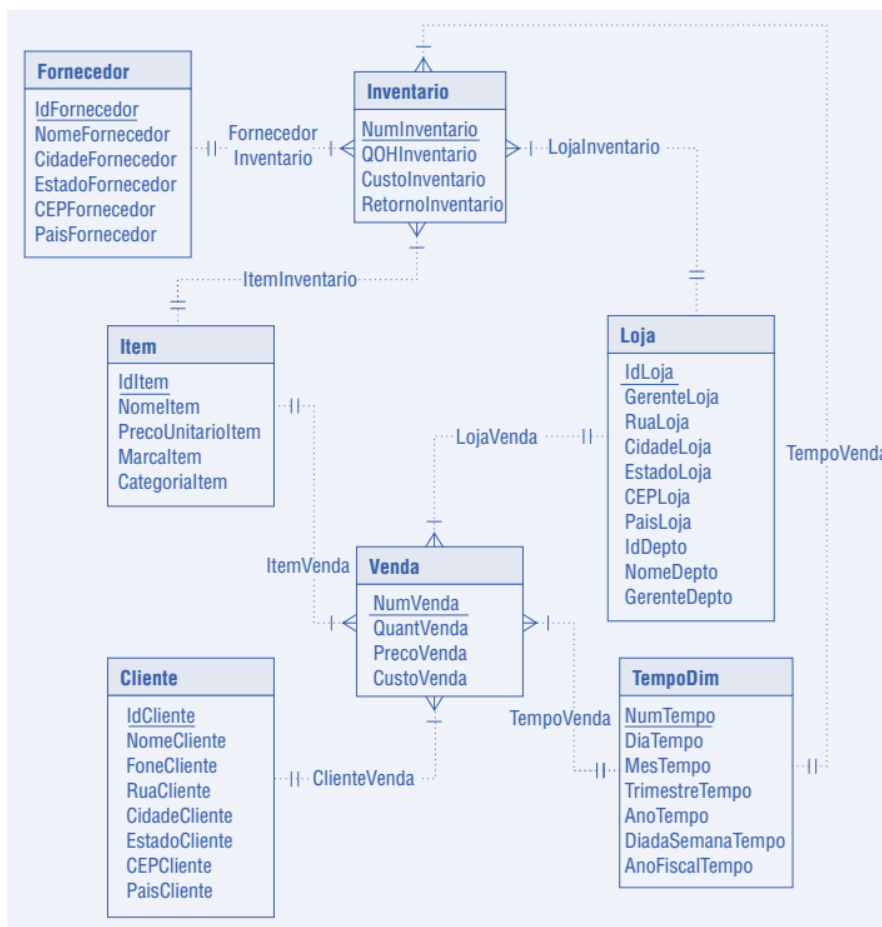


Figura 8 - Constelação com 2 tabelas fato: vendas e inventário.

Veja que **Inventário se torna uma tabela de fatos** e os relacionamentos 1-N tornam-se chaves estrangeiras na tabela de fatos. O tipo de entidade Inventário adiciona inúmeras medidas, incluindo a quantidade ofertada de um item, o custo desse item e a quantidade devolvida. **Todas as tabelas dimensionais são compartilhadas entre ambas as tabelas de fatos, exceto as tabelas Fornecedor e Cliente.** A essas tabelas compartilhadas dar-se o nome de **dimensões conformes**. Uma desvantagens que podemos observar no modelo multiestrela é que ele acaba limitando as consultas feitas sobre o data warehouse.

Esquema floco de neve

Neste tipo de modelo de dados algumas dimensões podem ser normalizadas até a terceira forma normal (3FN). Essa normalização cria uma hierarquia entre as tabelas dimensões, fazendo com que algumas tabelas não estejam conectadas diretamente a tabela fato. Esse modelo reduz o espaço de armazenamento dos dados, por a normalização elimina parte da redundância do modelo. Contudo, esse benefício não melhora a performance e acaba criando um modelo mais complexo para o usuário final. Essa complexidade é refletida em consultas mais complexas e difíceis de entender.



Em geral não é necessário normalizar as tabelas dimensão para impedir anomalias no armazenamento porque normalmente elas são estáveis e pequenas. A natureza de um datawarehouse indica que as **tabelas dimensão devem ser projetadas para recuperação, não para atualização**. O desempenho associado à recuperação de dados é melhorado eliminando-se as operações de junção que seriam necessárias para combinar tabelas dimensionais totalmente normalizadas.

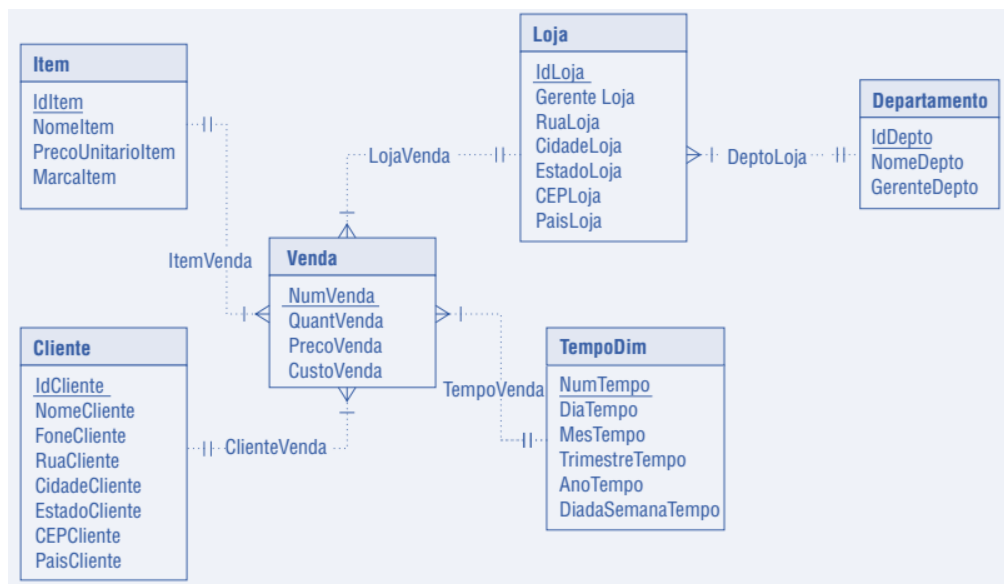


Figura 9 - Esquema floco de neve para vendas

Resumido de forma rápida os esquemas que estudamos até agora temos:

Esquema estrela

Em um banco de dados relacional, um esquema estrela tem uma tabela de fatos no centro relacionada às várias tabelas dimensão nos relacionamentos 1-N.

Esquema constelação

Em um banco de dados relacional, um esquema constelação contém várias tabelas de fatos no centro relacionado a tabelas dimensão. Em geral, as tabelas de fatos compartilham algumas tabelas dimensão.

Esquema floco de neve

Em um banco de dados relacional, o esquema floco de neve tem níveis múltiplos de tabela de dimensão relacionados com uma ou mais tabelas de fatos. Tabelas de dimensão (em geral pequenas) são normalizadas à terceira forma normal (3FN).

Tipos de tabela fato

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

1. Fato transacional
2. Fato agregada



3. Fato consolidada
4. Fato snapshot periódico
5. Fato de snapshot acumulado
6. Fato sem fato

Tabela Fato Transacional

A linha em uma tabela de fatos de transações corresponde a **um evento de medição em um ponto no espaço e no tempo**. As tabelas fato de transações de **granularidade atômicas são as tabelas mais expressivas e dimensionais**. Essa dimensionalidade robusta permite o máximo de **operações de slice and dice sobre os dados**. Segundo o especialista Rafael Piton, a maioria dos bilhões de linhas que temos no Data Warehouse são de tabelas fato transacional. Elas geralmente utilizam métricas aditivas, aquelas métricas que a gente pode somar por todas as dimensões. A figura abaixo apresenta um exemplo de tabela fato transacional que armazena informações sobre vendas (sales fact).

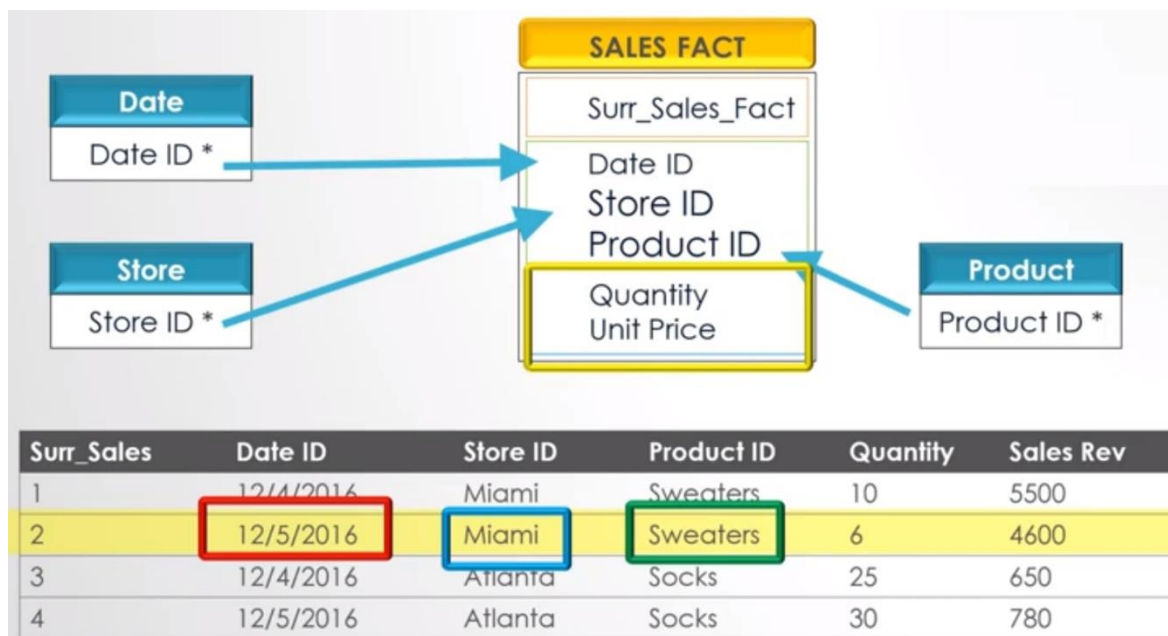


Tabela Fato Agregada

A tabela fato agregada tem a **função de acelerar o desempenho das consultas**. Ela sumariza os dados de outra tabela fato. Geralmente é construída para **armazenar o resultado de consultas agregadas muito utilizadas**. Por exemplo, se você consulta todos os dias o valor de vendas nos meses anteriores esse pode ser um bom dado para ser armazenado em uma tabela agregada. Assim, a consulta a essa informação será bem mais eficiente. Entretanto, temos que perceber que existe um esforço adicional para construção e manutenção dos dados nesta tabela.

Tabela Fato Consolidada



Muitas vezes é conveniente **combinar fatos de vários processos em uma única tabela fato consolidada** caso possam ser expressos na mesma granularidade. Por exemplo, os valores reais de vendas podem ser consolidados com as previsões de vendas em uma única tabela fato para tornar a tarefa de analisar os valores reais contra previsões simples e rápida.

Tabela Fato de Snapshot Periódico

A linha em **uma tabela fato de snapshot periódico** resume muitos eventos de medição ou estados atuais que ocorrem ao longo de **um período normal**, como um dia, uma semana ou um mês. A **granularidade é o período**, e não a transação individual. **Snapshots periódicos** muitas vezes contêm muitos fatos porque qualquer evento de medição de acordo com a granularidade da tabela de fatos é permitido. Estas tabelas de fatos são **uniformemente densas** em suas chaves estrangeiras, porque mesmo que nenhuma atividade ocorra durante o período, uma linha é tipicamente inserida na tabela fato contendo um zero ou o valor nulo para cada fato.

Tabela Fato Snapshot Acumulado

Uma linha numa **tabela de snapshot de fatos acumulando** resume os eventos que ocorrem em etapas de medição previsíveis entre o início e o fim de um processo, pipeline ou fluxo de trabalho, tais como cumprimento de ordem ou processamento de pedidos, que têm um ponto definido de início, etapas intermediárias padrões, e um ponto final definido podem ser modelados com este tipo de tabela de fatos. Existe uma chave estrangeira data na tabela original para cada etapa crítica no processo. Uma linha individual em uma tabela de snapshot acumulando, correspondente, por exemplo, a uma ordem ou pedido de um determinado produto. Veja a figura abaixo.

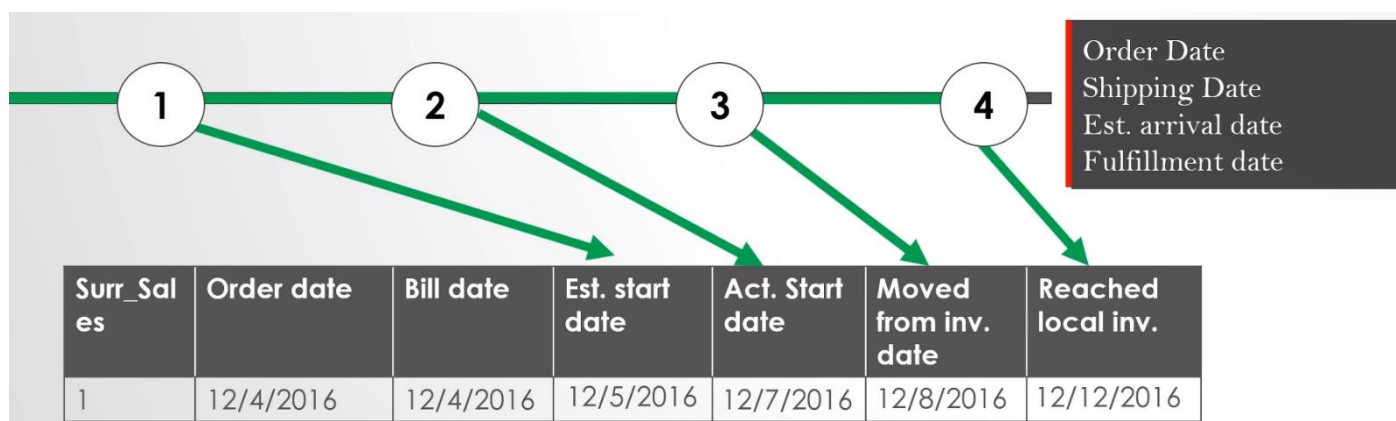


Tabela Fato sem Fato

Tabelas **fato sem fato** também podem ser usadas para analisar o que não aconteceu. Essas consultas têm sempre duas partes: uma tabela fato de cobertura que contém todas as possibilidades de eventos que podem acontecer e uma tabela fato de atividade que contém os eventos que aconteceram. Quando a atividade é subtraída da cobertura, o resultado é o conjunto de acontecimentos que não aconteceram.

10 regras para modelagem dimensional



Podemos dizer que as regras a seguir não devem ser quebradas e também existem algumas regras de ouro não tão severas a serem seguidas.

Regra #1: Carregue dados detalhados para as estruturas dimensionais.

Modelos dimensionais devem ser populados com dados detalhados o suficiente para **suportar os requisitos imprevisíveis de filtragem e agrupamento** necessárias para atender as consultas dos usuários de negócios. Usuários normalmente não precisam visualizar um registro por vez, mas é impossível prever as diferentes maneiras que eles pretendem ver os dados e acessar os detalhes.

Regra #2: Estructure os modelos dimensionais em torno dos processos de negócios.

Os processos de negócios são as atividades desenvolvidas por sua empresa que representam eventos mensuráveis, como registrar um pedido ou emitir uma fatura para um consumidor. Processos de negócios normalmente capturam ou geram métricas únicas de desempenho associadas a cada evento. **Estas métricas traduzidas em fatos, com cada processo de negócios representado por uma única tabela fato.** Além destas tabelas fato para cada processo, tabelas fato consolidadas às vezes são criadas para combinar métricas de vários processos em uma tabela fato com um nível padronizado de detalhe.

Regra #3: Tenha certeza de que cada tabela fato tenha uma dimensão de data associada.

Os eventos mensuráveis descritos na Regra #2 **sempre tem uma data de algum tipo associada a eles**, sejam eles um balancete mensal ou uma transferência de dinheiro registrada em seu centésimo de segundo. **Cada tabela fato deve ter ao menos uma chave estrangeira associada a uma tabela de dimensão data**, cuja granularidade é cada único dia, com os atributos de calendário e suas características não padronizadas relacionadas a data do evento, como o período fiscal. Às vezes múltiplas chaves estrangeiras de data estão ligadas em uma única tabela fato. Vide o exemplo de tabela fato acumulado vista na seção anterior.

Regra #4: Certifique-se que todos os fatos em uma única tabela fato estão na mesma granularidade ou nível de detalhe.

Existem **três granularidades fundamentais** para classificar todas as tabelas fato: **transacional, snapshot periódico** ou **snapshot acumulado**. Falamos sobre cada uma delas anteriormente. Independentemente de sua granularidade, **cada métrica em uma tabela fato deve estar exatamente no mesmo nível de detalhe**. Quando você mistura fatos representando muitos níveis de granularidade em uma mesma tabela fato, você estará criando confusão para os usuários de negócios e tornando as aplicações de BI vulneráveis a erros de valores ou outros resultados incorretos.

Regra #5: Resolva relacionamentos muitos-para-muitos em tabelas fato.

Como uma tabela fato guarda os resultados de um evento de um processo de negócios, existem inerentemente relacionamentos muitos-para-muitos (N:N) entre suas chaves estrangeiras, como diferentes produtos vendidos em diferentes lojas em diferentes dias. **Estes campos de chaves estrangeiras nunca devem conter valores nulos.**

Regra #6: Resolva os relacionamentos muitos-para-um nas tabelas de dimensões.



Hierarquicamente, relacionamentos muitos-para-um (N:1) entre atributos são normalmente desnormalizados ou concentrados em uma única tabela dimensão. Caso você não queira passar a maior parte de sua carreira desenhando modelos entidade-relacionamento para sistemas transacionais, você precisa resistir a sua instintiva tendência a normalizar ou criar um snow flake com subdimensões menores para cada relacionamento N:1; **desnormalização de dimensões é a regra do jogo na modelagem dimensional.**

Regra #7: Gravar nomes de relatórios e valores de domínios de filtros em tabelas dimensão.

Os códigos e, mais importante ainda, **as decodificações e descrições associadas a eles usadas como nomes de colunas em relatórios e como filtros em consultas devem ser gravadas em tabelas dimensionais.** Evite gravar campos com códigos criptográficos ou volumosos campos descritivos na própria tabela fato; da mesma forma, não grave apenas o código na tabela de dimensão e assuma que os usuários não precisam das decodificações descritivas ou que elas podem ser resolvidas na aplicação de BI. Se a informação for um nome de linha/coluna ou um filtro de menu, então ela deve ser tratada como um atributo de dimensão.

Regra #8: Tenha certeza de que as tabelas dimensão usam uma chave artificial.

Chaves artificiais, sem significado e sequenciais (exceto para a dimensão data, onde chaves cronologicamente definidas e mais inteligíveis são aceitáveis) **provém um grande número de benefícios operacionais**; chaves menores significam menores tabelas fato, menores índices, e desempenho melhorado. Chaves artificiais são absolutamente necessárias no caso de você estar registrando as alterações dos atributos da dimensão com uma nova linha para cada mudança. Mesmo que seus usuários de negócios inicialmente não visualizem o valor de registrar as alterações nos atributos, **usar chaves artificiais tornará uma futura alteração de política menos onerosa.**

Regra #9: Crie dimensões padronizadas para integrar os dados na empresa.

Dimensões padronizadas (também conhecidas por dimensões comuns, principais, ou de referência) são essenciais para o data warehousing empresarial. Gerenciadas no sistema de ETL e então reutilizadas associadas a diversas tabelas fato; **dimensões padronizadas ou conformes trazem atributos descritivos consistentes para os modelos dimensionais e permitem a habilidade de navegar através dos dados integrados de diferentes processos de negócios.** A reutilização das dimensões padronizadas diminui o tempo de desenvolvimento eliminando o desenho redundante e o esforço de desenvolvimento.

Regra #10: Avalie requisitos e realidade continuamente para desenvolver uma solução de DW/BI que seja aceita pelos usuários de negócios e suporte seu processo de tomada de decisões.

Os responsáveis pela modelagem dimensional devem constantemente balancear os requisitos dos usuários de negócios com as realidades inerentes aos dados de origem associados para desenvolver um modelo que possa ser implantado, e que, mais importante ainda; tenha uma boa chance de ser útil aos negócios.



APOSTA ESTRATÉGICA

A ideia desta seção é apresentar os pontos do conteúdo que mais possuem chances de serem cobrados em prova, considerando o histórico de questões da banca em provas de nível semelhante à nossa, bem como as inovações no conteúdo, na legislação e nos entendimentos doutrinários e jurisprudenciais³.



Características do Data Warehouse

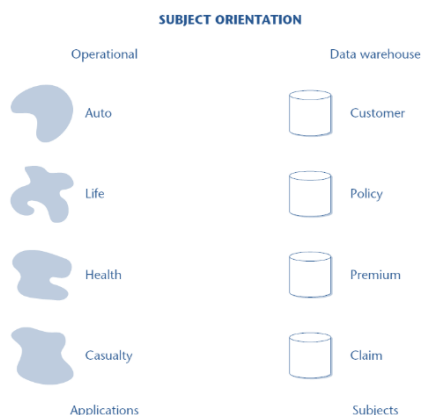


Figure 2-1 An example of a subject orientation of data.

Ser **orientado por assunto** refere-se ao fato do *Data Warehouse* armazenar informações sobre temas específicos importantes para o negócio da empresa. São exemplos típicos de temas: produtos, atividades, contas, clientes. Em contrapartida, quando observamos o ambiente operacional percebemos que ele é organizado por aplicações funcionais. Por exemplo, em uma organização bancária, estas aplicações incluem empréstimos, investimentos e seguros. Observe um exemplo na figura abaixo retirada do livro do Inmon

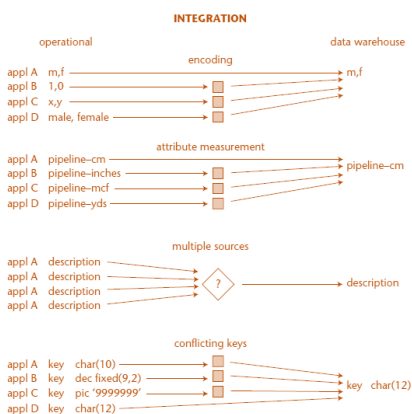
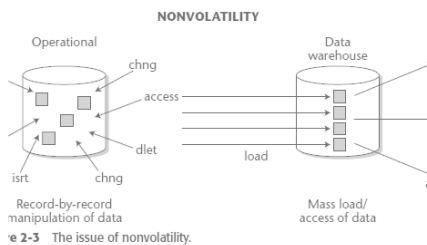


Figure 2-2 The issue of integration.

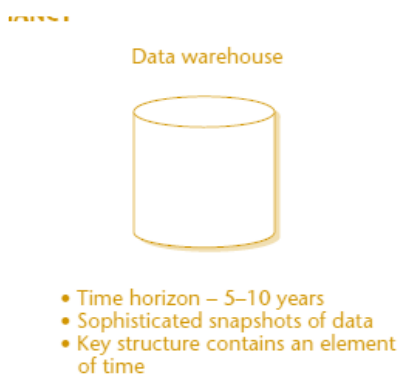
Ser **integrado** refere-se à consistência de nomes, das unidades, das variáveis, etc. É importante que os dados armazenados sejam transformados até um estado uniforme. Por exemplo, considere sexo como um elemento de dado. Uma aplicação pode codificar sexo como M/F, outra como 1/0 e uma terceira como H/M. Conforme os dados são inseridos ou repassados para o *Data Warehouse*, eles são convertidos para um mesmo padrão. O atributo Sexo, portanto, seria codificado apenas de uma forma.

³ Vale deixar claro que nem sempre será possível realizar uma aposta estratégica para um determinado assunto, considerando que às vezes não é viável identificar os pontos mais prováveis de serem cobrados a partir de critérios objetivos ou minimamente razoáveis.

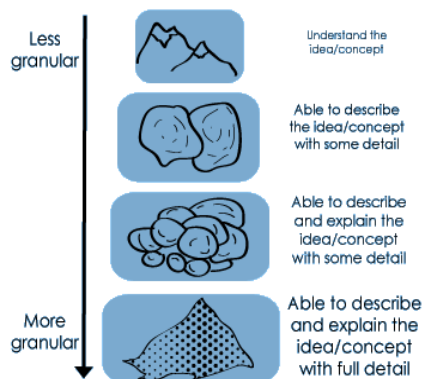




O fato de ser **não volátil** significa que o *Data Warehouse* permite apenas a carga inicial dos dados e consultas a estes dados. Após serem integrados e transformados, os dados são carregados em bloco para o DW, para que estejam disponíveis aos usuários para acesso.



Ser **variante no tempo** trata do fato de um registro 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 DW, a cada ocorrência de uma mudança, uma nova entrada é criada para marcar esta mudança.



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





É necessário, para que as análises sejam consideradas corretas que os dados tenham baixa dispersão e estejam consistentes com os valores observados. É importante lembrar que serão feitas várias manipulações nos dados, inclusive agregações e cálculos estatísticos. A depender dos desvios presentes nos dados as análises podem ser inócuas ou incoerentes com a realidade.

Imprima o capítulo Aposta Estratégica separadamente e dedique um tempo para absolver tudo o que está destacado nessas duas páginas. Caso tenha alguma dúvida, volte ao Roteiro de Revisão e Pontos do Assunto que Merecem Destaque. Se ainda assim restar alguma dúvida, não hesite em me perguntar no fórum.

QUESTÕES ESTRATÉGICAS

Nesta seção, apresentamos e comentamos uma amostra de questões objetivas selecionadas estrategicamente: são questões com nível de dificuldade semelhante ao que você deve esperar para a sua prova e que, em conjunto, abordam os principais pontos do assunto.

A ideia, aqui, não é que você fixe o conteúdo por meio de uma bateria extensa de questões, mas que você faça uma boa revisão global do assunto a partir de, relativamente, poucas questões.



1. CESGRANRIO - Profissional (LIQUIGÁS)/Arquiteto de Soluções/Júnior TI/2018/Edital 02

Um modelo multidimensional contempla dois tipos de tabelas, que são a

- A) tabela OLAP e a tabela OLTP
- B) tabela Principal e a tabela Secundária
- C) tabela Física e a tabela Lógica
- D) tabela Dimensão e a tabela Fato
- E) tabela Visão e a tabela Agregação



Comentários

O modelo multidimensional possui as **tabelas de fato e as tabelas de dimensão**.

A tabela fato é a tabela principal do DW, porque é ela que interliga as tabelas de dimensão. Ela possui as métricas, o que se quer medir, e as chaves estrangeiras, que fazem referência às tabelas de dimensão.

As tabelas de dimensão são constituídas de tuplas que representam os atributos da dimensão, qualificando as informações da tabela fato. Elas armazenam de forma descritiva os registros da tabela fato.

Analisando as demais alternativas, não existem correlações das alternativas com as tabelas do modelo multidimensional. Na alternativa A, temos dois conceitos relacionados a Business Intelligence (BI): OLAP, que se refere a um tipo de ferramenta de análise de dados em um data warehouse; e OLTP, que se refere a sistemas transacionais, ou seja, usados no processamento de dados rotineiros de uma organização.

Gabarito: alternativa D.

2. CESGRANRIO - Profissional (LIQUIGÁS)/Arquiteto de Soluções/Júnior TI/2018/Edital 02

Uma empresa utiliza Bancos de Dados (BDs) para armazenar os dados necessários ao seu funcionamento.

Nesse contexto, um BD Multidimensional

- A) é útil para tomada de decisão na empresa.
- B) é útil para o controle detalhado das atividades da empresa.
- C) deve conter o histórico detalhado de todas as operações realizadas na empresa.
- D) deve ser materializado antes de quaisquer outros BDs Relacionais na empresa.
- E) pode substituir, após estar consolidado, todos os BDs Relacionais existentes na empresa.

Comentários

Bancos de dados multidimensionais são usados em data warehouses (DW), utilizando técnicas de mineração de dados, permitem que o usuário faça análise dos dados para tomada de decisão.

Sendo assim, chegamos à resposta já na alternativa A.

Gabarito: alternativa A.

3. CESGRANRIO - Analista (UNIRIO)/Tecnologia da Informação/2019



Um desenvolvedor recebeu a incumbência de construir uma Data Warehouse sobre as vendas da empresa. Para isso, teve acesso a um banco de dados relacional que incluía as seguintes tabelas, onde as chaves primárias estão sublinhadas:

ITEMDEVENDA(itemid,venda_id,produto_id,quantidade,desconto,valorvendido)

VENDA(venda_id,vendedor_id,cpfcliente,valor_total,valorimposto,data,hora)

PRODUTO(produto_id,fabricante_id,nomeproduto,valorunitario)

CLIENTE(cpfcliente,nomecliente,idadecliente,enderecocliente,telefonecliente)

VENDEDOR(vendedor_id,CPFvendedor,nomevendedor)

LOJA(loja_id,nomeloja,endereco_loja,cidade_id)

CIDADE(cidade_id,nomecidade)

FABRICANTE(fabricante_id,nomefabricante)

Os diretores desejam estudar relatórios sobre a venda de produtos, como a quantidade de cada produto vendido em um dia, semana ou mês, ou que produtos são vendidos juntos.

Para garantir a menor granularidade possível no modelo dimensional e alta flexibilidade nas consultas, em um modelo Estrela, a tabela fato deve registrar em cada linha o(a)

- A) valor agregado de cada venda por dia
- B) agregado de cada venda de produto por dia, semana ou mês
- C) agregado de venda de cada produto por dia
- D) equivalente ao item de venda
- E) quantidade de produtos vendidos por fabricante

Comentários

Menor granularidade, refere-se ao máximo de detalhes possíveis em uma consulta. Ao observar as tabelas informadas na questão, nota-se que a tabela ITEMDEVENDA tem atributos relevantes referentes aos itens das vendas, como o atributo quantidade (requisito apresentado na questão), além disso, ela se relaciona com as tabelas VENDA (venda_id) e PRODUTOS (produto_id), que contribuem para o relatório detalhado sobre a venda de produtos. Devido a isso, a tabela fato deve registrar em cada linha os itens da venda. Portanto, a alternativa correta é a letra D.

Gabarito: alternativa D.

4. CESGRANRIO - Analista (PETROBRAS)/Sistema Júnior/2018

Ao construir um modelo de dados para um data warehouse de sua empresa, um desenvolvedor viu-se às voltas com três tabelas relacionais: venda, cliente e vendedor.

Ao fazer uma transformação para o modelo estrela, ele deve organizar:



- A) venda, como tabela fato; cliente e vendedor, como tabelas dimensão
- B) cliente e vendedor, como tabelas fato; venda, como tabela dimensão
- C) cliente, como tabela fato; venda e vendedor, como tabelas dimensão
- D) vendedor e venda, como tabelas fato; cliente, como tabela dimensão
- E) vendedor, como tabela fato; cliente e venda, como tabelas dimensão

Comentários

O **esquema estrela** é uma representação de modelagem própria dos cubos de dados multidimensionais. No banco de dados relacional, o diagrama desse esquema **tem a aparência de uma estrela**, em cujo **centro** encontra-se uma grande tabela, denominada **tabela de fatos**, que está vinculada a várias **tabelas dimensão**, **dispostas radialmente**.

A tabela de **fatos armazena dados numéricos (fatos)**, como resultados de vendas, ao passo que as **tabelas dimensão armazenam dados descritivos** correspondentes a dimensões individuais do cubo de dados, como produto, local e tempo. Existe um relacionamento 1-N entre cada tabela de dimensão e a tabela de fatos. Cada **processo de negócio** é representado por um modelo dimensional.

Uma tabela fato contém vários fatos, correspondentes a cada uma das suas linhas. Cada linha corresponde a **um evento de medição**. Os dados em cada linha estão a um nível específico de detalhe, referido como o **grão**, por exemplo, uma linha por produto vendido numa operação de venda.

Uma única linha da tabela fato tem uma relação **um-para-um com o evento de medição**, como descrito pela **granularidade** da tabela fato. **Uma tabela fato corresponde a um evento físico observável, e não às exigências de um relatório específico**. Dentro de uma tabela fatos, apenas fatos consistentes com a granularidade definida são permitidos. Por exemplo, em uma transação de vendas no varejo, a quantidade de um produto vendido e seu preço são bons fatos.

Diante dessa explicação, as três tabelas da questão são: venda, cliente e vendedor. Das três, apenas venda pode ser medida, logo ela será a tabela fato nesse modelo, as outras duas serão tabelas dimensão.

Gabarito: alternativa A.

5. CESGRANRIO - Técnico Científico (BASA)/Tecnologia da Informação/2018

Um Data Warehouse é recomendado para armazenar dados

- A) sumarizados de um departamento.
- B) sumarizados 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.



Comentários

Os dados de um Data Warehouse (DW) são provenientes de diversos sistemas existentes dentro de uma organização. Eles são persistentes e sumarizados, possibilitando que se armazene um grande volume de dados.

Na alternativa B temos que é recomendado para armazenar dados sumarizados de toda a empresa para apoio à decisão e utilização de ferramentas OLAP. De fato, os dados de um Data Warehouse são sumarizados de todas as áreas da empresa. O DW é importante para a manutenção do negócio e para o auxílio à tomada de decisão, permitindo o uso de ferramentas OLAP para a análise dos dados através de múltiplas perspectivas.

Gabarito: alternativa B.

6. CESGRANRIO - Analista de Sistemas Júnior (TRANSPETRO)/Processos de Negócio/2018

Os sistemas de data warehouse diferem de várias formas dos sistemas transacionais das empresas, como, por exemplo, em seu modelo de dados. Para transferir e transformar os dados dos sistemas transacionais para os sistemas de data warehousing, é comum utilizar, como estratégia, a existência de uma camada especial da arquitetura conhecida como

- A) Data Marts
- B) Data Staging Area
- C) Dimensional Model Area
- D) Presentation Area
- E) Living Sample Area

Comentários

Durante o processo de ETL, é utilizado um repositório temporário de dados chamado Data Staging Area. Ao final do ETL, os dados são carregados no Data Warehouse, onde ficarão armazenados de forma persistente. Portanto, a alternativa correta é a letra B.

Gabarito: alternativa B.

7. CESGRANRIO - Analista de Sistemas Júnior (TRANSPETRO)/SAP/2018

No processo de extração, transformação e carga (ETL), uma característica da data staging área é a de

- A) aderir ao barramento de serviços da organização, de forma fisicamente descentralizada.
- B) apresentar serviços de consulta e apresentação para usuários finais dos servidores OLTP.



- C) armazenar os dados selecionados provenientes das diversas fontes para serem usados pelo processo.
- D) criar uma estrutura normalizada em enterprise data warehouses para suportar aplicações OLAP.
- E) suportar o processo de engenharia reversa interpretável pelo administrador de banco de dados.

Comentários

A ideia é exatamente essa: após serem extraídos das diversas fontes, os dados serão carregados temporariamente no data staging area para que possam passar pela etapa de transformação. Ao final do ETL, os dados irão ser carregados no Data Warehouse, onde ficarão armazenados de forma persistente. Portanto, a alternativa C está correta.

As demais alternativas estão erradas pois fogem completamente do escopo ao não trazer nada relacionado ao data staging area.

Gabarito: alternativa C.

8. CESGRANRIO - Profissional (LIQUIGÁS)/Analista de Sistemas/Júnior TI/2018/Edital 02

Uma empresa pretende modelar um banco de dados multidimensional para tomada de decisão na área de vendas. O gerente da área quer poder avaliar os dados históricos de vendas, por código ou nome de produto, por mês/ano, por loja e por cidade (há várias lojas em uma mesma cidade; das cidades é registrado apenas o seu nome; é necessário registrar em que cidade está cada loja).

Nesse banco de dados bem projetado, segundo um esquema estrela, haverá

- A) apenas uma tabela com atributos, como código do produto, nome do produto, mês/ano da venda, loja e cidade.
- B) três tabelas de dimensões (produto, tempo, loja) e uma tabela de fatos (vendas), uma vez que cidade é atributo da tabela loja.
- C) três tabelas de dimensões (produto, tempo, loja) e uma tabela de fatos (vendas), uma vez que cidade é atributo da tabela venda.
- D) quatro tabelas de dimensões (produto, tempo, loja, cidade) e uma tabela de fatos (vendas).
- E) quatro tabelas de fatos (produto, tempo, loja, cidade) e uma tabela de dimensões (vendas).

Comentários

Em um esquema estrela, a modelagem consiste em tabelas de dimensão ligadas a uma tabela fato.

As tabelas de dimensões possuem as características da venda. Neste caso, quando a venda foi realizada (dimensão tempo), qual produto foi vendido (dimensão produto) e onde ocorreu a venda (dimensão loja).



A tabela fato armazena o fato propriamente dito, o que ocorreu na venda. Possui as chaves das tabelas de dimensões e as métricas que caracterizam a venda. Na questão, o gerente quer avaliar o histórico das vendas, então a tabela fato será a tabela vendas.

O ponto chave da questão é saber se cidade será uma dimensão ou um atributo da dimensão loja. Como o enunciado da questão diz que apenas o nome da loja é registrado, e que é necessário registrar a cidade onde a loja se encontra, logo, como estamos usando o esquema estrela, o nome da cidade será um atributo da dimensão loja, para que exista essa relação entre loja e cidade. Além disso, toda dimensão deve ter um atributo chave, que torna a característica única.

Gabarito: alternativa B.

9. CESGRANRIO - Analista de Sistemas Júnior (TRANSPETRO)/SAP/2018

A tomada de decisão nas organizações pode ser suportada por um conjunto de ferramentas computacionais que refinarão as análises e darão mais segurança aos seus clientes gestores. Uma dessas ferramentas é o uso de bases de dados dimensionais, os armazéns de dados, ou data warehouses.

Uma das características inerentes a esse repositório de dados é verificada por

- A) diferenciar-se na estrutura de construção dos data marts, esses últimos construídos durante o processo de data mining.
- B) permitir operações de manutenção similares às bases transacionais, com operações de alteração e remoção de registros específicos.
- C) possibilitar a sua construção a partir de fontes de dados tanto internas quanto externas à organização.
- D) possuir características atemporais, pelo caráter consolidador de suas estruturas de dados.
- E) ser direcionada a aplicações, com foco nos processos de negócio organizacionais.

Comentários

Vamos analisar as alternativas para encontrar o nosso gabarito:

- A) ERRADA. Os data marts são repositórios de dados especializados em uma determinada área ou departamento do negócio. Dependendo da abordagem, são construídos antes ou depois dos data warehouses.
- B) ERRADA. Os data warehouses permitem operações OLAP, que possibilitam a análise dos dados sob múltiplas perspectivas. Operações OLTP são suportadas por bancos de dados transacionais, e não pelos data warehouses. Como regra, um dado inserido no DW não sofre alteração ou remoção.
- C) CERTA. Os dados armazenados no data warehouse são provenientes das mais diversas fontes. Essas fontes podem ser tanto internas quando externas à organização e são extraídas e transformadas durante o processo de ETL (Extraction, Transform e Load). Ao fim desse processo, são carregados nos DWs, onde serão armazenados de forma persistente.



D) ERRADA. O data warehouse armazena os dados de uma organização sob diversas dimensões, e uma das dimensões clássicas e mais comuns é a do tempo.

E) ERRADA. Os data warehouses não são direcionados a aplicações. Eles são repositórios de dados que permitem uma análise sob múltiplas perspectivas, auxiliando a tomada de decisão e direcionando a estratégia do negócio.

Gabarito: alternativa C.

10. CESGRANRIO - Analista de Sistemas Júnior (TRANSPETRO)/SAP/2018

Na construção de data warehouses é possível – e por vezes recomendado – que a equipe projetista considere a utilização de diversas fontes de dados. Com isso, espera-se melhorar a qualidade das análises a serem realizadas, a partir desse data warehouse.

Qual tarefa a seguir listada NÃO corresponde a uma ação de preparação de dados nessa etapa?

- A) Avaliação estocástica dos metadados
- B) Discretização de atributos numéricos
- C) Imputação de valores ausentes
- D) Seleção de atributos relevantes
- E) Verificação de cálculos inválidos

Comentários

Antes de serem carregados nos data warehouses, os dados das diversas fontes de uma organização passam pelo processo de ETL (Extraction, Transform and Load). Durante a etapa de transformação, são executadas diversas tarefas com o intuito de melhorar a qualidade dos dados, que muitas vezes podem apresentar inconsistências por serem provenientes de fontes diferentes. A) ERRADA. Os data marts são repositórios de dados especializados em uma determinada área ou departamento do negócio. Dependendo da abordagem, são construídos antes ou depois dos data warehouses.

Entre as alternativas, a única que NÃO corresponde a uma ação de preparação de dados nessa etapa é a alternativa A. A avaliação estocástica utiliza métodos probabilísticos para análise. Não é recomendado o uso de métodos baseados em probabilidades pois esses métodos poderiam deixar passar dados inconsistentes, prejudicando uma análise posterior.

Gabarito: alternativa A.

...

Forte abraço e bons estudos.

"Hoje, o 'Eu não sei', se tornou o 'Eu ainda não sei'"



(Bill Gates)

Thiago Cavalcanti



Face: www.facebook.com/profthiogocavalcanti

Insta: www.instagram.com/prof.thiago.cavalcanti

YouTube: youtube.com/profthiogocavalcanti



ESSA LEI TODO MUNDO CONHECE: PIRATARIA É CRIME.

Mas é sempre bom revisar o porquê e como você pode ser prejudicado com essa prática.



1 Professor investe seu tempo para elaborar os cursos e o site os coloca à venda.



2 Pirata divulga ilicitamente (grupos de rateio), utilizando-se do anonimato, nomes falsos ou laranjas (geralmente o pirata se anuncia como formador de "grupos solidários" de rateio que não visam lucro).



3 Pirata cria alunos fake praticando falsidade ideológica, comprando cursos do site em nome de pessoas aleatórias (usando nome, CPF, endereço e telefone de terceiros sem autorização).



4 Pirata compra, muitas vezes, clonando cartões de crédito (por vezes o sistema anti-fraude não consegue identificar o golpe a tempo).



5 Pirata fere os Termos de Uso, adultera as aulas e retira a identificação dos arquivos PDF (justamente porque a atividade é ilegal e ele não quer que seus fakes sejam identificados).



6 Pirata revende as aulas protegidas por direitos autorais, praticando concorrência desleal e em flagrante desrespeito à Lei de Direitos Autorais (Lei 9.610/98).



7 Concurseiro(a) desinformado participa de rateio, achando que nada disso está acontecendo e esperando se tornar servidor público para exigir o cumprimento das leis.



8 O professor que elaborou o curso não ganha nada, o site não recebe nada, e a pessoa que praticou todos os ilícitos anteriores (pirata) fica com o lucro.



Deixando de lado esse mar de sujeira, aproveitamos para agradecer a todos que adquirem os cursos honestamente e permitem que o site continue existindo.