

Aula 00

*CNU (Bloco 2 - Tecnologia, Dados e
Informação) Passo de Conhecimentos
Específicos - Eixo Temático 5 - Apoio à
Decisão e Inteligência Artificial - 2024
(Pós-Edital)*

Autor:

07 de Fevereiro de 2024

BUSINESS INTELLIGENCE E DATA WAREHOUSE

Sumário

CONTEÚDO	1
ANÁLISE ESTATÍSTICA	2
GLOSSÁRIO DE TERMOS.....	3
ROTEIRO DE REVISÃO.....	5
Introdução a Business Intelligence.....	6
Governança de Dados	14
Data Warehouse.....	18
Modelagem Multidimensional	30
QUESTÕES ESTRATÉGICAS.....	41
QUESTIONÁRIO DE REVISÃO E APERFEIÇOAMENTO	59
Perguntas.....	60
Perguntas e Respostas	62
LISTA DE QUESTÕES ESTRATÉGICAS	64
Gabaritos	74

CONTEÚDO

Business Intelligence. Conceitos básicos, histórico. Componentes. Governança de dados. Data Warehouse. Características. Tipos. Processo de DW. Arquitetura. Modelagem Dimensional. Processo de design dimensional.



ANÁLISE ESTATÍSTICA

Inicialmente, convém destacar o percentual de incidência do assunto, dentro da disciplina **Banco de Dados e Business Intelligence** em concursos/cargos similares. Quanto maior o percentual de cobrança de um dado assunto, maior sua importância.

Obs.: *um mesmo assunto pode ser classificado em mais de um tópico devido à multidisciplinaridade de conteúdo.*

Assunto	Relevância na disciplina em concursos similares
SQL	21.6 %
BI (Business Intelligence)	9.0 %
DW - Data Warehouse	7.2 %
SQL Server	7.2 %
Oracle	6.3 %
Banco de Dados Multidimensionais	5.4 %
Data Mining	5.4 %
Administração de banco de dados	3.6 %
Banco de Dados	2.7 %
Formas normais	2.7 %
ETL (Extract Transform Load)	2.7 %
Banco de Dados Relacionais	2.7 %
Arquitetura de Banco de Dados	1.8 %
SGBD - Sistema de Gerenciamento de Banco de Dados	1.8 %
OLAP (On-line Analytical Processing)	1.8 %
Segurança	1.8 %
MS-Access	1.8 %
Modelo relacional	1.8 %
Metadados e Metainformação	1.8 %



Álgebra relacional	0.9 %
Banco de Dados Paralelos e Distribuídos	0.9 %
Gerência de Transações	0.9 %
Modelagem de dados	0.9 %
Gatilhos (Triggers)	0.9 %
DER - Diagrama de Entidade e Relacionamento	0.9 %
Visão (View)	0.9 %
Banco de Dados Textuais	0.9 %
Índices	0.9 %
PostgreSQL	0.9 %
MySQL	0.9 %
Big Data	0.9 %

GLOSSÁRIO DE TERMOS

Faremos uma lista de termos que são relevantes ao entendimento do assunto desta aula. Caso tenha alguma dúvida durante a leitura, esta seção pode lhe ajudar a esclarecer.

Business Intelligence: Campo de análise de dados que visa fornecer suporte de tomada de decisão em negócios. Envolve a coleta, integração, análise e apresentação de dados de negócios.

Data Warehouse (DW): Sistema de armazenamento de dados projetado para análise e relatórios, em vez de processamento de transações. Armazena dados históricos e consolidados.

Data Mart: Subconjunto de um data warehouse orientado a um determinado tema ou área de negócios, como vendas ou marketing.

OLTP (Online Transaction Processing): Categoria de sistemas de processamento de dados que facilita e gerencia aplicações transacionais para dados de entrada e recuperação de transações.

OLAP (Online Analytical Processing): Categoria de software que permite aos usuários analisar facilmente informações de várias dimensões do banco de dados.



Self-Service BI: Práticas e ferramentas de BI que permitem aos usuários finais acessar e trabalhar com dados corporativos mesmo sem a ajuda do departamento de TI.

Governança de Dados: Conjunto de processos que garantem a qualidade, disponibilidade, integridade, segurança e usabilidade dos dados gerenciados em uma organização.

Curadoria de Dados: Processo de gestão de dados que inclui coleta, organização, administração, proteção e manutenção de dados.

Orientação por Assunto: Uma característica de um DW, refere-se à organização dos dados em torno de temas ou assuntos específicos para análise.

Integração: Em um DW, integração se refere à consolidação de dados de várias fontes em uma estrutura unificada.

Variação no Tempo: Um DW é orientado no tempo, armazenando dados históricos para permitir análises de tendências.

Não Volatilidade: Uma vez que os dados são inseridos em um DW, eles não são alterados. Isso assegura a consistência dos relatórios ao longo do tempo.

Granularidade: Refere-se ao nível de detalhe ou resumo contido nos dados de um DW.

Credibilidade dos Dados: Refere-se à qualidade e confiabilidade dos dados em um DW.

Data Warehouse Empresarial: Um DW que serve toda a organização com uma estrutura de dados integrada e consolidada.

Armazenamento de Dados Operacionais (ODS): Repositório de dados de curto prazo que é usado para tarefas operacionais.

Arquitetura em Camadas de DW: A estruturação do DW em camadas, geralmente incluindo camadas de fonte de dados, armazenamento de dados, processamento de dados e apresentação de dados.

Arquitetura de Data Marts Independentes: Arquitetura onde data marts são criados sem um DW centralizado, cada um com seus próprios dados.

Arquitetura de Barramento de Data Marts: Arquitetura que utiliza um modelo de dados consistente para todos os data marts na organização.



Arquitetura Hub-and-spoke: Arquitetura DW onde os data marts são alimentados a partir de um DW centralizado

Arquitetura de DW Centralizado: Arquitetura onde todos os dados são armazenados em um único DW.

Arquitetura de DW Federado: Arquitetura onde múltiplos DWs ou data marts são coordenados e parecem um único DW para os usuários.

Abordagem Top-Down: Abordagem de design que começa com uma visão abrangente e detalhada e depois a desagrega em componentes menores.

Abordagem Bottom-Up: Abordagem de design que começa com os componentes menores e mais detalhados e os combina em um todo mais abrangente.

Modelagem Dimensional: Método de design de banco de dados que visa melhorar a legibilidade e o desempenho de bases de dados orientadas a consultas.

Dimensão: Categoria de informação usada para análise em um DW, como tempo, localização ou produto.

Fato: Medida quantitativa ou qualitativa registrada em uma tabela de fatos em um DW.

Esquema em Estrela: Modelo de banco de dados onde uma tabela de fatos está no centro, cercada por tabelas de dimensão.

Esquema Floco de Neve: Modelo de banco de dados onde tabelas de dimensão estão normalizadas, resultando em uma estrutura que se parece com um floco de neve.

Esquema em Constelação: Modelo de banco de dados com várias tabelas de fatos que compartilham tabelas de dimensão.

ROTEIRO DE REVISÃO

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.



Introdução a Business Intelligence

Business Intelligence (BI) é um termo abrangente que se refere ao uso de dados, informações e conhecimento no **processo de tomada de decisões nas organizações**. A BI compreende estratégias e tecnologias usadas pelas empresas para a análise de dados de negócios, proporcionando uma visão abrangente das operações comerciais.

No mundo empresarial moderno, a quantidade de dados produzidos é imensa. Para tomar decisões informadas, as empresas precisam não apenas de acesso a esses dados, mas também de meios para organizar, analisar e entender o que os dados significam. Aqui é onde a Business Intelligence desempenha um papel crucial.

As soluções de BI permitem às empresas transformar dados brutos em informações significativas. Elas fazem isso ao coletar dados de várias fontes, organizá-los em um formato acessível (muitas vezes através de um data warehouse), e então usar ferramentas de análise e visualização para extrair insights que podem informar estratégias de negócios e tomada de decisões.

Sistemas de BI podem fornecer informações históricas, atuais e preditivas através de relatórios, sumarizações, dashboards, gráficos, mapas e outros métodos visuais. Além disso, eles ajudam a identificar tendências de mercado, padrões de consumo, e potenciais problemas operacionais, todos os quais são vitais para a saúde e o sucesso a longo prazo de uma empresa.

Histórico e Evolução

A ideia de usar dados para tomar decisões de negócios existe há séculos, mas o conceito de Business Intelligence, como o conhecemos hoje, começou a se formar na segunda metade do século 20.

Anos 1960 e 1970: A origem da BI como uma disciplina formal pode ser rastreada até os anos 1960 e 1970, quando os primeiros sistemas de informação de gestão foram desenvolvidos. Esses sistemas eram bastante rudimentares e estavam mais focados em processamento de transações do que em análise de dados. Eles eram usados principalmente para tarefas como folha de pagamento e contabilidade. Durante essa época, o conceito de armazenamento de dados começou a se desenvolver, permitindo que as empresas armazenassem grandes quantidades de dados para futuras análises.

Anos 1980: Durante os anos 1980, o termo "Business Intelligence" foi popularizado por Howard Dresner, um analista do Gartner, para descrever o processo de usar dados para melhorar as decisões de negócios. Os sistemas de informação executiva (EIS) tornaram-se populares durante



esta época, permitindo que os gerentes visualizassem os dados da empresa em um formato fácil de entender.

Anos 1990: Na década de 1990, a BI começou a evoluir rapidamente graças ao advento dos Data Warehouses e das ferramentas de extração, transformação e carga (ETL). Estas tecnologias permitiram às empresas recolher, limpar e armazenar dados de múltiplas fontes em um único local centralizado, facilitando a análise e a tomada de decisões. Foi também durante este período que as ferramentas de BI começaram a se tornar mais user-friendly, com interfaces gráficas e funcionalidades drag-and-drop.

Anos 2000 e além: Com a chegada do século XXI, a BI passou por uma série de inovações significativas. O aumento da capacidade de armazenamento de dados e o avanço das tecnologias de processamento de dados permitiram o surgimento do Big Data e da análise preditiva. Além disso, o surgimento de soluções de BI baseadas em nuvem tornou estas ferramentas acessíveis a um público muito mais amplo.

A BI também começou a se integrar com outras tecnologias emergentes, como inteligência artificial (AI) e aprendizado de máquina (ML), permitindo análises de dados mais avançadas e a capacidade de prever tendências futuras.

Hoje, a Business Intelligence está mais avançada e acessível do que nunca, permitindo que as empresas de todos os tamanhos e setores utilizem os dados para informar a tomada de decisões e impulsionar o sucesso dos negócios.

Componentes Relacionados e Termos Comuns

Sistemas de Business Intelligence (BI) são compostos por vários componentes, cada um contribuindo para a eficácia geral do sistema em coletar, armazenar, analisar e apresentar dados. Os componentes chave de um sistema de BI incluem:

Fontes de Dados: Essas são as várias fontes das quais os dados são coletados para análise. As fontes de dados podem ser internas (por exemplo, bancos de dados da empresa, planilhas, sistemas ERP, CRM) ou externas (por exemplo, mídias sociais, bases de dados públicas, feeds de notícias).

ETL (Extract, Transform, Load): ETL é o processo de extrair dados de várias fontes, transformá-los para um formato adequado para análise e, em seguida, carregá-los em um local centralizado, como um Data Warehouse. A transformação pode incluir a limpeza de dados, a integração de várias fontes de dados e a conversão de dados em um formato padronizado.



Data Warehouse: É o repositório central onde os dados transformados são armazenados para análise. Ele organiza os dados em um formato que facilita a consulta e a análise. Data Warehouses podem armazenar grandes volumes de dados históricos, facilitando a análise de tendências ao longo do tempo.

Data Mart: Data Mart é um subconjunto do Data Warehouse que é geralmente orientado para uma linha de negócios específica ou um departamento. Eles permitem que os usuários acessem e analisem dados relevantes para suas necessidades específicas.

Ferramentas de Análise e Consulta: Estas são as ferramentas que os usuários finais usam para interagir com os dados. Elas permitem que os usuários consultem o Data Warehouse para recuperar os dados de que precisam e realizem análises usando várias técnicas (por exemplo, análise de tendências, análise de segmento, análise preditiva).

Dashboards e Relatórios: Dashboards e relatórios são os mecanismos pelos quais os resultados das análises de dados são apresentados aos usuários. Eles fornecem visualizações de dados (por exemplo, gráficos, tabelas) e métricas de desempenho chave (KPIs) para ajudar os usuários a entender os dados e tomar decisões informadas.

Data Mining: Data Mining é o processo de descobrir padrões e correlações nos dados que podem não ser imediatamente óbvios. Ele usa uma variedade de técnicas, incluindo estatísticas, inteligência artificial e aprendizado de máquina.

OLAP (Online Analytical Processing): OLAP é uma abordagem para responder a consultas multidimensionais de maneira rápida. Ele permite que os usuários visualizem os dados de várias perspectivas.

Em conjunto, esses componentes funcionam juntos para criar um sistema de Business Intelligence que pode transformar dados brutos em insights úteis para informar a tomada de decisões e a estratégia de negócios.

Dashboards

Dashboards, no contexto de Data Warehousing (DW) e Business Intelligence (BI), são interfaces visuais que apresentam métricas e indicadores de desempenho chave (KPIs) de maneira concisa e fácil de entender. Eles são projetados para fornecer aos usuários uma visão rápida e atualizada do desempenho dos negócios ou de um aspecto específico do negócio.

Os dashboards podem incluir uma variedade de elementos gráficos, incluindo gráficos de barras, gráficos de linhas, gráficos de pizza, mapas de calor, tabelas, entre outros, para apresentar dados



de maneira visual. Esses elementos são normalmente interativos, permitindo aos usuários filtrar dados, perfurar para obter mais detalhes ou explorar diferentes visões dos dados.

Eles são alimentados por dados que foram coletados, transformados e armazenados em um DW. Isso permite que o dashboard atualize regularmente com dados novos e precisos.

Dashboards são ferramentas úteis para os tomadores de decisão, pois permitem monitorar o desempenho dos negócios em tempo real, identificar tendências e padrões, detectar problemas ou oportunidades e tomar decisões informadas com base em dados.

Por exemplo, um dashboard de vendas pode mostrar informações sobre vendas totais, vendas por região, vendas por vendedor, vendas por categoria de produto e assim por diante. As informações apresentadas no dashboard podem ser atualizadas em tempo real ou em intervalos regulares, dependendo das necessidades do negócio e da arquitetura do DW.

A figura abaixo mostra um exemplo de Dashboard de dados sobre visitas em uma página web:



Dashboard

Visits Left 30.000 of 60.000

Last 30 Days

02/09/2022 → 03/10/2022

Overall Page Visits

41.580

+ 12.45%

Overall Sessions

29.430

+ 36.36%

New Visitors

29.430

+ 36.36%

Returning Visitors

0

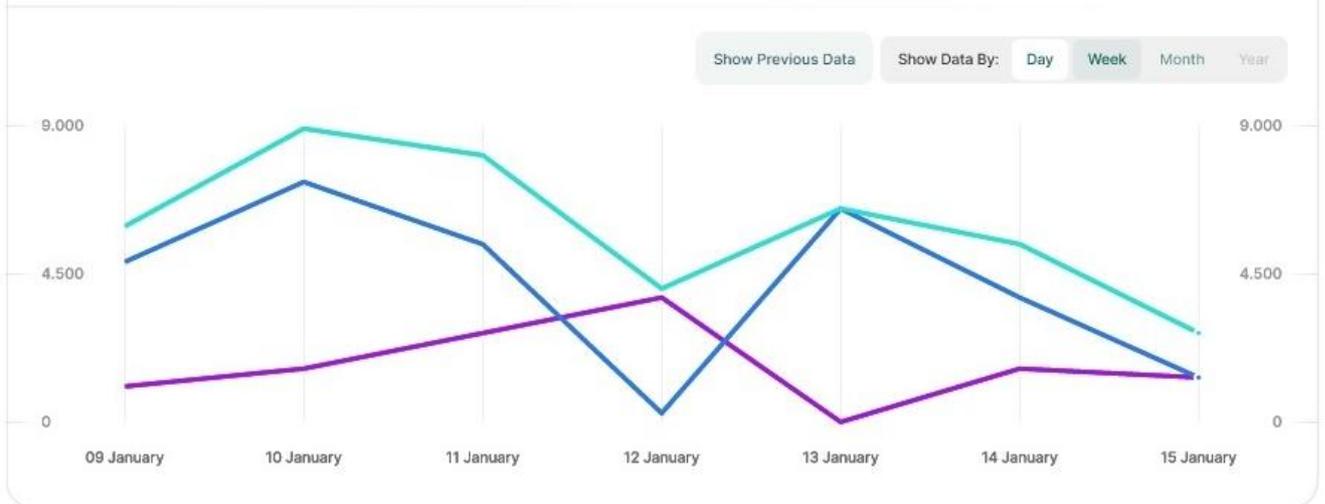
0.00%

Converting Visitors

0

0.00%

8 Conversion Pages



Page Visits

91.350

+ 12.06%

Page	Visits	Change	Bounce Rate
Home Page medium.com	54.810 Visits + 1.50%	60.0%	
Registration Page register.medium.com	13.704 Visits + 6.30%	15.0%	
Blog Page blog.medium.com	9.135 Visits - 2.40%	10.0%	
Features Page features.medium.com	4.566 Visits + 1.08%	5.0%	
Support Page support.medium.com	4.566 Visits + 0.90%	5.0%	
Search page search.medium.com	4.566 Visits	5.0%	

View All Pages

Bounce Rate

45.0%

+ 9.0%

D W M Y



View More

Page Conversions

13.500

0.00%

D W M Y



View More

Responsive Access to Your Data

Direct Access to Analytics, Responsive Design & Allow Others to View Statistics

Request Direct Access



OLTP versus OLAP

No contexto de Business Intelligence, **OLTP (Online Transaction Processing)** e **OLAP (Online Analytical Processing)** são duas estratégias fundamentais para o gerenciamento e uso de dados em um ambiente de negócios. Embora ambos sejam críticos para as operações de negócios, eles têm funções distintas e são otimizados para diferentes tipos de tarefas.

OLTP:

- OLTP é a estratégia que gerencia e manipula dados de transações em tempo real (dia a dia). É caracterizado por um grande número de transações curtas que inserem, modificam e recuperam dados de várias linhas na base de dados. Um exemplo de sistema OLTP seria um sistema de processamento de pedidos online, em que cada pedido seria uma transação.
- Sistemas OLTP são orientados para a velocidade e a confiabilidade, pois precisam processar transações de maneira rápida e precisa. Eles também são projetados para evitar conflitos de dados - por exemplo, dois pedidos simultâneos que tentam comprar o último item em estoque.
- No contexto de um Data Warehouse, os dados OLTP são tipicamente a fonte que será carregada no warehouse para análise. Porém, os dados precisam passar por um processo ETL (Extração, Transformação, Carga) para serem otimizados para análise.

OLAP:

- OLAP, por outro lado, é a estratégia usada para analisar dados em um ambiente de Business Intelligence. Sistemas OLAP são projetados para lidar com consultas complexas e permitir uma análise aprofundada dos dados.
- Em vez de processar transações, sistemas OLAP são usados para explorar tendências, realizar análises comparativas e outras análises de alto nível que podem envolver grandes volumes de dados. Por exemplo, um sistema OLAP pode ser usado para analisar as vendas de um produto ao longo do tempo, comparando diferentes regiões ou segmentos de clientes.
- Os sistemas OLAP são projetados para um desempenho rápido de consulta e análise, geralmente pré-agregando dados e usando técnicas como cubos multidimensionais e indexação bitmap.

Sistemas OLAP e Data Warehouse têm alguns objetivos básicos que os diferem de sistemas transacionais convencionais, quais sejam:

Manter a Informação Acessível:

Um dos principais objetivos de um Data Warehouse (DW) é garantir que os dados estejam facilmente acessíveis para os usuários finais. Isso é particularmente importante em grandes



organizações, onde os dados podem ser gerados por muitos departamentos diferentes e armazenados em muitos sistemas diferentes. Um DW coleta esses dados e os armazena em um local centralizado, permitindo que os usuários acessem os dados de maneira rápida e fácil, independentemente de onde os dados foram originalmente gerados.

Manter a Informação Consistente (Credibilidade):

A consistência dos dados é crucial para garantir a credibilidade do DW. Isso significa que os dados devem ser precisos, completos e confiáveis. Para garantir a consistência, os dados normalmente passam por um processo de ETL (Extração, Transformação, Carga) antes de serem carregados no DW. Durante esse processo, os dados são limpos, verificados para garantir a qualidade e transformados em um formato consistente. Isso garante que, independentemente de onde os dados vêm, eles são apresentados da mesma maneira no DW, tornando-os mais fáceis de usar e entender.

Ser Adaptável a Mudanças:

As necessidades de negócios e os ambientes de dados estão sempre mudando, e um DW eficaz deve ser capaz de se adaptar a essas mudanças. Isso pode incluir a capacidade de integrar novas fontes de dados, mudar a maneira como os dados são armazenados e apresentados, ou ajustar a arquitetura do DW para acomodar mudanças na escala de dados. Uma arquitetura de DW flexível é crucial para garantir que o DW possa continuar atendendo às necessidades da empresa à medida que ela cresce e muda.

Disponibilizar a Informação de Forma Temporal:

Um DW deve ser capaz de armazenar dados históricos e fornecer uma visão temporal dos dados. Isso permite que os usuários realizem análises de tendências ao longo do tempo, o que pode ser essencial para entender padrões e prever o futuro. Para isso, o DW deve ser projetado com um componente de tempo, de modo que cada peça de dado esteja associada a um ponto específico no tempo. Isso pode ser conseguido através do uso de marcas de tempo ou versões de dados.

Self-Service BI

Self-Service Business Intelligence (BI) refere-se a ferramentas e estratégias de BI que permitem aos usuários finais acessar, analisar e visualizar dados **sem a necessidade de envolvimento direto de TI ou especialistas** em análise de dados. Em outras palavras, o Self-Service BI capacita usuários individuais a gerar relatórios, visualizações de dados e insights sem ter que depender de especialistas em análise de dados.



Self-Service BI serve para democratizar o acesso aos dados e a análise de dados em uma organização. Ele permite que os usuários de negócios – que são frequentemente os que mais entendem suas próprias necessidades de dados – acessem e analisem dados de forma independente, sem ter que passar por um intermediário. Isso torna o processo de análise de dados mais rápido e mais ágil, pois os usuários não precisam esperar pela assistência de especialistas em análise de dados.

Vantagens do Self-Service BI:

Agilidade e Velocidade: Com o Self-Service BI, os usuários podem acessar e analisar dados em tempo real, sem atrasos causados pela solicitação de relatórios ou análises a um departamento de TI ou análise de dados.

Democratização dos Dados: Self-Service BI capacita todos os usuários, independentemente de suas habilidades técnicas, a fazer perguntas aos dados e obter respostas. Isso pode levar a um melhor entendimento dos negócios e a melhores decisões.

Alívio para o Departamento de TI: Ao permitir que os usuários finais gerem seus próprios relatórios e análises, o Self-Service BI pode aliviar a carga sobre os recursos de TI ou de análise de dados, permitindo que se concentrem em outras tarefas importantes.

Melhoria na Tomada de Decisões: Com acesso direto aos dados, os tomadores de decisão podem explorar os dados conforme necessário para apoiar suas decisões, levando a decisões de negócios mais informadas.

Várias ferramentas de BI no mercado permitem a análise de self-service, incluindo:

Microsoft Power BI: Esta é uma suíte de ferramentas de análise de negócios que oferece recursos de visualização de dados interativos, relatórios empresariais e dashboards.

Tableau: Tableau é outra poderosa ferramenta de BI que permite que os usuários criem visualizações de dados interativas e dashboards. Ele tem uma interface de arrastar e soltar que facilita a criação de visualizações.

QlikView: QlikView é uma plataforma de BI que permite aos usuários criar e implantar análises interativas e dashboards. Ele usa uma interface de arrastar e soltar e permite aos usuários explorar os dados de várias perspectivas.

A tabela a seguir resume e faz um paralelo entre BI Tradicional e Self-Service BI:



	BI Tradicional	Self-Service BI
Usuários	Principalmente analistas de dados e TI	Usuários de negócios, analistas de dados, equipes de TI, basicamente qualquer pessoa na organização
Dependência de TI	Alta: TI está envolvida em todos os passos	Baixa: os usuários podem acessar e analisar dados sem assistência direta de TI
Velocidade	Pode ser lenta devido ao envolvimento de TI	Rápida, pois os usuários podem gerar relatórios e análises por conta própria
Flexibilidade	Menor devido à estrutura fixa	Alta: os usuários podem adaptar análises e relatórios para atender às suas necessidades específicas
Escalabilidade	Alta, pois é gerenciada por profissionais de TI	Pode ser limitada pela capacidade do usuário de gerenciar e analisar grandes volumes de dados
Capacidade de lidar com a complexidade dos dados	Alta: profissionais de TI podem lidar com dados complexos	Média a alta, dependendo da habilidade do usuário e da ferramenta de BI utilizada
Governança de Dados	Alta: controlada pela TI	Pode ser um desafio, pois a responsabilidade é mais dispersa. As organizações precisam implementar políticas adequadas de governança de dados

Governança de Dados

A **Governança de Dados** e a **Gestão de Dados** são dois conceitos fundamentais na área de Business Intelligence e Data Warehousing, e embora estejam relacionados, eles têm funções distintas.

De acordo com o DAMA International's Data Management Body of Knowledge (DAMA-DMBOK), a Governança de Dados é "**o exercício de autoridade, controle e tomada de decisões compartilhada (planejamento, monitoramento e execução) sobre o gerenciamento de dados e**



recursos de informação". Em outras palavras, Governança de Dados se refere ao conjunto de políticas, procedimentos, funções, estruturas, padrões e medidas que garantem o uso eficaz, eficiente, e seguro dos dados dentro de uma organização.

A Governança é responsável por definir quem dentro de uma organização tem autoridade e controle sobre os dados e recursos de informação, e como esses dados são usados e protegidos. Isso pode incluir questões como privacidade e proteção de dados, qualidade de dados, conformidade com regulamentos e normas, e integração e compatibilidade de dados.

Por outro lado, a **Gestão de Dados** é a prática de planejar, controlar e entregar os dados e recursos de informação de uma organização. Inclui uma série de atividades, como coleta de dados, armazenamento de dados, processamento de dados, análise de dados, distribuição de dados, etc.

A Gestão de Dados é mais focada no aspecto operacional, enquanto a Governança de Dados é mais focada no aspecto estratégico. A Governança de Dados fornece as regras e estruturas que orientam a Gestão de Dados, enquanto a Gestão de Dados implementa essas regras e estruturas no dia a dia da organização.

Objetivos do DAMA-DMBOK2

- Criar consenso para uma visão geralmente aplicável das áreas de conhecimento de gerenciamento de dados.
- Fornecer definições padrão para áreas de conhecimento de gerenciamento de dados comumente usadas, entregáveis, funções e outras terminologias, em conjunto com o DAMA Dictionary of Data Management e, assim, promover uma padronização de conceitos e atividades.
- Identificar princípios orientadores para o gerenciamento de dados.
- Esclarecer o escopo e os limites das atividades de gerenciamento de dados.
- Fornecer uma visão geral das boas práticas comumente aceitas, técnicas amplamente adotadas e abordagens alternativas significativas, sem referência a fornecedores de tecnologia específicos ou seus produtos.
- Apresentar questões organizacionais e culturais comuns.
- Identificar estratégias para análise de maturidade de gerenciamento de dados.
- Fornecer recursos adicionais e material de referência para melhor entendimento do gerenciamento de dados

Princípios de Governança de Dados



Os princípios da Governança de Dados são um conjunto de diretrizes que orientam o gerenciamento e o uso de dados em uma organização. Eles formam a base para a estrutura de governança de dados e as políticas, procedimentos e práticas que são implementados. Vejamos o significado de cada um dos princípios mencionados no DAMA-DMBOK:

Princípio	Descrição
Regra de Ouro	Este princípio estabelece que os dados são um ativo valioso e devem ser tratados como tal. Isso significa proteger, manter e usar os dados de maneira que maximizem seu valor para a organização.
Federação	Este princípio refere-se ao equilíbrio entre o controle centralizado e descentralizado dos dados. O objetivo é permitir que cada unidade ou departamento da organização gerencie seus próprios dados, mas dentro de um conjunto comum de políticas e padrões definidos em nível de organização.
Eficiência	Este princípio foca na otimização do uso dos dados para alcançar os objetivos da organização de forma mais eficaz e eficiente. Isso pode envolver a eliminação de redundâncias, melhorando a precisão dos dados, ou simplificando processos de gestão de dados.
Qualidade	Este princípio enfatiza a importância de manter a qualidade dos dados. Isso envolve garantir que os dados sejam precisos, completos, atualizados e relevantes para as necessidades da organização.
Gestão de Risco	Este princípio envolve a identificação e gestão de riscos associados aos dados, incluindo problemas de segurança, privacidade e conformidade. O objetivo é minimizar os potenciais danos que podem resultar do uso inadequado dos dados.
Colaboração	Este princípio enfatiza a importância da colaboração entre diferentes partes da organização na gestão de dados. Isso pode envolver o compartilhamento de dados entre departamentos, a colaboração na definição de políticas e padrões, ou a cooperação na resolução de problemas de dados.
Contextualização	Este princípio implica que os dados devem ser usados e interpretados no contexto adequado. Isso significa entender as circunstâncias que cercam os dados, como eles foram coletados e para que propósito, para garantir que eles sejam usados corretamente.



Inovação	Este princípio reconhece que a governança de dados não deve restringir a inovação, mas sim apoiá-la. Isso significa criar um ambiente que incentive a exploração e o uso inovador dos dados, ao mesmo tempo em que se gerencia riscos e se protege o valor dos dados.
-----------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Curadoria de Dados

A **Curadoria de Dados** é um conjunto de processos e atividades usados para gerenciar, proteger e aumentar o valor dos dados ao longo do tempo. Isso inclui atividades como coleta de dados, validação de dados, anotação de dados, melhoria de qualidade, controle de versão, preservação de dados e disponibilização de dados para reuso. A curadoria é fundamental para garantir que os dados sejam confiáveis, acessíveis e úteis para a organização.

Um **Curador de Dados** ou **Data Steward** é um indivíduo responsável pela gestão e governança de dados em uma organização. Os Data Stewards desempenham um papel crítico na implementação da curadoria de dados, garantindo que os dados sejam bem gerenciados e usados de acordo com as políticas e princípios de governança de dados da organização.

As responsabilidades de um Data Steward abrangem:

Garantia de Qualidade de Dados: O Data Steward é responsável por garantir a precisão, completude e confiabilidade dos dados. Isso envolve a limpeza de dados, a validação de dados e a resolução de problemas de qualidade de dados.

Implementação de Políticas de Governança de Dados: O Data Steward ajuda a implementar e aplicar as políticas e princípios de governança de dados da organização. Isso inclui garantir a conformidade com regulamentos de privacidade e segurança, gerenciar o acesso a dados e assegurar o uso adequado dos dados.

Melhoria de Processos de Dados: O Data Steward pode trabalhar para melhorar os processos de gestão de dados da organização, como processos de coleta de dados, processos de integração de dados e processos de distribuição de dados.

Apoio à Tomada de Decisões Baseada em Dados: O Data Steward apoia a tomada de decisões baseada em dados, fornecendo informações e insights a partir dos dados e ajudando a interpretar e aplicar os dados de maneira eficaz.

Promoção da Alfabetização de Dados: O Data Steward pode desempenhar um papel na promoção da alfabetização de dados dentro da organização, educando os outros sobre a importância dos dados e como usá-los efetivamente.



Data Warehouse

Inicialmente, vamos ver a definição de Data Warehouse dos três principais autores sobre o assunto:

Kimball: *Um data warehouse é uma cópia de dados transacionais especificamente estruturada para consulta e análise. Kimball também define o processo de data warehousing que 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.*

Características

Um Data Warehouse (DW) reúne dados de várias fontes dentro de uma organização para fornecer uma visão consolidada e unificada da informação e são caracterizados pelos seguintes atributos-chave:

Orientação por Assunto: O DW organiza os dados em torno de assuntos específicos (como clientes, produtos, vendas, etc.) que são relevantes para a organização. Isso permite que os usuários realizem análises e relatórios baseados em áreas de interesse específicas.

Integração: DW integra dados de várias fontes e sistemas diferentes, proporcionando uma visão unificada e consistente dos dados. Os dados de diferentes fontes são limpos, transformados e integrados para garantir que sejam compatíveis e consistentes entre si.

Variação no Tempo: Os dados em um Data Warehouse são armazenados ao longo do tempo, permitindo análises históricas e de tendências. Isso significa que os usuários podem ver como os dados mudaram ao longo do tempo e analisar tendências e padrões.



Não Volatilidade: Uma vez que os dados são inseridos em um Data Warehouse, eles não são modificados ou excluídos. Isso garante que os dados permaneçam consistentes e confiáveis para análise ao longo do tempo.

Granularidade: A granularidade refere-se ao nível de detalhe ou resolução dos dados. Os Data Warehouses podem armazenar dados em vários níveis de granularidade, permitindo análises em um nível muito detalhado ou em um nível mais agregado.

Credibilidade dos Dados: Os Data Warehouses garantem a credibilidade dos dados através de processos rigorosos de qualidade e governança de dados. Isso garante que os dados sejam precisos, completos, atualizados e confiáveis.

Tipos de Data Warehouse

A indústria atualmente reconhece pelo menos três tipos diferentes de Data Warehouses: Data Mart (DM), Data Warehouse Empresarial (EDW) e Armazenamento de Dados Operacionais (ODS).

Data Mart:

Um **Data Mart** é um subconjunto de um Data Warehouse que é **orientado para um assunto ou área de negócio específica**. Enquanto um Data Warehouse reúne todos os dados corporativos em um local central, um Data Mart se concentra em uma **área específica** de negócios, como vendas, marketing, finanças, etc.

Data Marts são utilizados para atender às necessidades de uma equipe ou grupo específico dentro de uma organização. Eles permitem que os usuários acessem e analisem dados relevantes para sua área específica de negócios, sem ter que lidar com a complexidade e a vastidão de um Data Warehouse completo.

Por exemplo, um Data Mart de vendas inclui dados sobre clientes, vendas, produtos e vendas por região. Um Data Mart de marketing pode incluir dados sobre campanhas de marketing, respostas do cliente, análises de mercado e análises de segmento de clientes.

Data Marts têm várias vantagens:

Desempenho: Como os Data Marts contêm menos dados do que um Data Warehouse completo, as consultas geralmente são mais rápidas.

Relevância: Os Data Marts fornecem dados que são relevantes para um grupo ou equipe específica, facilitando a análise e a tomada de decisões.



Facilidade de Uso: Como os Data Marts são menores e mais focados, eles são geralmente mais fáceis de usar do que um Data Warehouse completo.

Implantação Rápida: Data Marts geralmente podem ser criados e implantados mais rapidamente do que um Data Warehouse completo.

É importante notar que a utilização de Data Marts deve ser feita de forma estratégica e cuidadosa, para evitar a criação de "silos de dados" onde os dados são isolados em diferentes partes da organização. Idealmente, Data Marts devem ser usados como complemento a um Data Warehouse, permitindo análises específicas de departamentos sem comprometer uma visão unificada e integrada dos dados em toda a organização.

Data Warehouse Empresarial:

Data Warehouse Empresarial (Enterprise Data Warehouse - EDW) é um sistema centralizado de armazenamento de dados que é utilizado em todo o âmbito de uma organização. O EDW coleta, integra e gerencia grandes volumes de dados de várias fontes e sistemas dentro da empresa, incluindo sistemas de transação, sistemas de relatório, bancos de dados, arquivos planos e até mesmo fontes de dados externas.

O EDW tem um papel crucial no apoio à tomada de decisões estratégicas e operacionais em toda a organização. Ele permite que os tomadores de decisão acessem facilmente os dados necessários para análises de desempenho, identificação de tendências, elaboração de previsões e planejamento estratégico.

Os EDWs também são a espinha dorsal da maioria das iniciativas de Business Intelligence (BI), fornecendo os dados necessários para relatórios, painéis (dashboards), análises e mineração de dados.

Por exemplo, imagine uma grande empresa de varejo. O EDW dessa empresa pode coletar dados de uma variedade de fontes, incluindo sistemas de ponto de venda (POS), sistemas de gestão de relacionamento com o cliente (CRM), sistemas de gerenciamento de estoque e sistemas de e-commerce. Todas essas informações são integradas e armazenadas no EDW.

A alta administração pode então utilizar esses dados para vários propósitos, como:

- Analisar o desempenho de vendas em diferentes lojas ou regiões.
- Identificar tendências de compra dos clientes.
- Monitorar níveis de estoque e otimizar a gestão do estoque.
- Avaliar a eficácia de campanhas de marketing ou promoções.
- Fazer previsões de vendas e crescimento.



Armazenamento de Dados Operacionais:

O **Operational Data Store (ODS)**, ou Armazenamento de Dados Operacionais, é um tipo intermediário de banco de dados projetado para integrar dados de várias fontes para oferecer suporte às operações de negócios em tempo real com um mínimo de latência de dados. Ele combina a natureza operacional dos sistemas OLTP (On-Line Transaction Processing) com as características de relatório dos sistemas OLAP (On-Line Analytical Processing) usados em Data Warehouses.

ODS é um repositório de dados operacionais atuais, mantido a curto prazo, que permite a consolidação de dados de diversas fontes, facilitando o acesso, a análise e a elaboração de relatórios para usuários de negócios.

Um ODS é comumente usado para:

Integrar Dados: O ODS consolida dados de vários sistemas operacionais e os torna acessíveis para uso. Isso é útil em organizações onde diferentes sistemas gerenciam diferentes operações, como vendas, inventário e contabilidade.

Suportar Decisões Operacionais: O ODS permite aos gerentes e outros tomadores de decisões acessar informações atualizadas sobre operações em tempo real, ajudando-os a tomar decisões mais informadas.

Resolver Conflitos de Dados: Em muitas organizações, diferentes sistemas operacionais podem ter informações conflitantes. Um ODS pode ajudar a resolver esses conflitos, fornecendo uma visão única e consistente dos dados.

Apoiar Processos de Negócios: Muitos processos de negócios exigem acesso a vários sistemas operacionais. Um ODS pode facilitar esses processos, fornecendo acesso centralizado a todos os dados necessários.

Por exemplo, em uma empresa de telecomunicações, um ODS pode ser usado para combinar informações de sistemas de faturamento, atendimento ao cliente e provisionamento de rede. Isso permitiria que os gerentes acessassem informações atualizadas sobre a utilização da rede, questões de atendimento ao cliente e dados de faturamento em um único lugar, facilitando a tomada de decisões operacionais e estratégicas.

Vale ressaltar que um ODS é diferente de um Data Warehouse na medida em que é atualizado em tempo real e contém apenas dados atuais e operacionais, enquanto um Data Warehouse contém dados históricos e é atualizado em intervalos regulares. Além disso, um ODS é



normalmente otimizado para leitura e escrita rápidas, enquanto um Data Warehouse é otimizado para leitura e análise de dados.

ODS versus Staging Area:

O ODS (Operational Data Store) e a Staging Area são ambos componentes comuns de um sistema de Business Intelligence/Data Warehouse, mas têm objetivos e funções ligeiramente diferentes.

Operational Data Store (ODS): O ODS é um sistema intermediário entre os sistemas operacionais (OLTP) e o Data Warehouse. É usado para realizar limpeza, integração e alinhamento de dados de diferentes sistemas operacionais antes de serem carregados no Data Warehouse. O ODS é projetado para executar operações de baixa latência, permitindo atualizações e consultas rápidas, e costuma armazenar dados a um nível bastante detalhado.

Staging Area: A Staging Area, por outro lado, é o local onde os dados são preparados para o carregamento no Data Warehouse. Os dados são extraídos dos sistemas de origem, podem ser limpos e transformados para corresponder ao esquema do Data Warehouse e, em seguida, são carregados no Data Warehouse. A Staging Area é temporária e não é projetada para consultas ou atualizações frequentes.

A correlação entre ODS e Staging Area é que ambos são usados como áreas intermediárias para a preparação e limpeza de dados antes do carregamento no Data Warehouse. Ambos são fundamentais para garantir que os dados que chegam ao Data Warehouse sejam consistentes, precisos e de alta qualidade.

Porém, eles não são idênticos. O ODS é um sistema que pode servir a várias finalidades operacionais e analíticas, oferecendo um instantâneo atualizado em tempo real dos sistemas operacionais. A Staging Area, por outro lado, é um componente de uma única direção do processo de ETL, sem a intenção de ser usada para consultas diretas ou funções operacionais.

DW - Arquiteturas em Camadas

A arquitetura de um Data Warehouse é geralmente organizada em camadas para efetivamente lidar com as diferentes necessidades e requisitos de processamento dos dados. Enquanto as especificidades podem variar, uma arquitetura de Data Warehouse típica inclui as seguintes camadas:

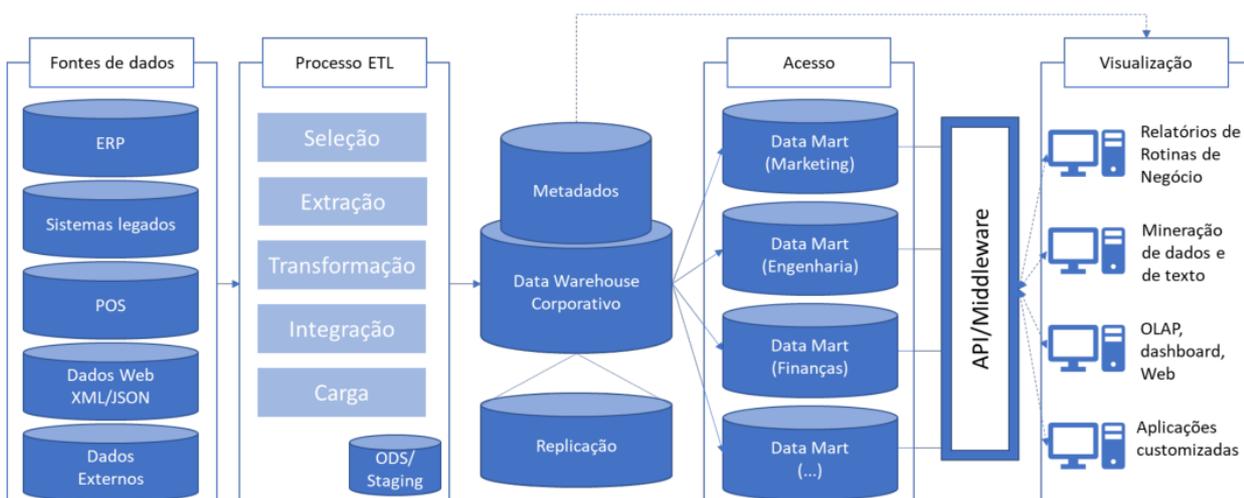
1. **Camada de Fonte de Dados:** Esta é a primeira camada na arquitetura de um Data Warehouse e é composta pelos vários sistemas de origem que fornecem os dados. Inclui sistemas de gerenciamento de banco de dados, aplicativos ERP, arquivos planos, e outras fontes de dados.



Os dados destas fontes podem ser estruturados, semiestruturados ou não estruturados, e podem vir em diferentes formatos.

2. **Camada de Extração, Transformação e Carga (ETL):** Esta camada é responsável por extrair os dados das várias fontes, transformá-los (por exemplo, limpando, validando, agregando, etc.) em um formato consistente e carregá-los no Data Warehouse. As ferramentas de ETL são essenciais nesta camada para garantir que os dados sejam devidamente preparados e integrados.
3. **Camada de Armazenamento de Dados:** Esta é a camada onde os dados são armazenados após passarem pelo processo de ETL. Pode incluir o Data Warehouse em si (que armazena dados históricos para análise) e, em algumas arquiteturas, um Operational Data Store (ODS) que armazena dados mais recentes para relatórios e consultas de curto prazo.
4. **Camada de Data Marts:** Data Marts são subconjuntos do Data Warehouse orientados para um assunto específico ou uma área de negócios específica, como vendas, marketing, finanças, etc. Esta camada permite que diferentes departamentos ou grupos dentro da organização acessem e trabalhem com um subconjunto mais administrável dos dados que são mais relevantes para eles.
5. **Camada de Acesso ou Apresentação:** Esta é a camada onde os usuários finais interagem com os dados. Inclui as ferramentas de Business Intelligence (BI) que os usuários usam para consultar, analisar e visualizar os dados, como relatórios, dashboards, ferramentas de análise e mineração de dados.

Veja a figura a seguir que ilustra as camadas:



DW - Outras Arquiteturas



A seguir descrevemos outras arquiteturas citadas pela literatura.

Data Marts Independentes:

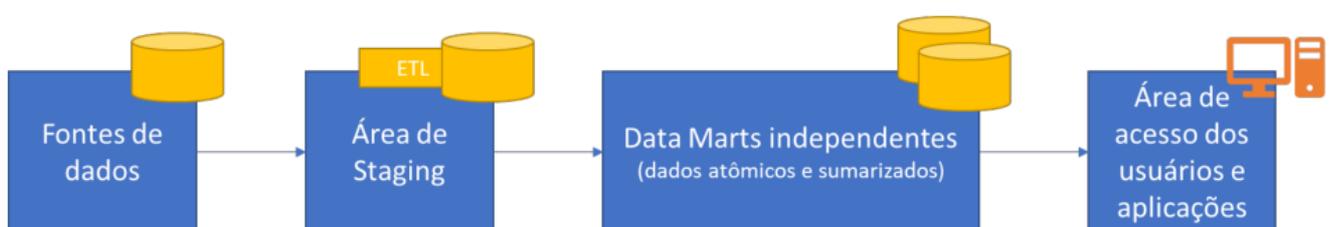
A arquitetura de Data Marts independentes é um modelo no qual vários Data Marts são criados de maneira autônoma para atender a necessidades específicas de um departamento ou área de negócios, sem a presença de um Data Warehouse centralizado.

Os principais aspectos dessa arquitetura incluem:

- **Desenvolvimento Descentralizado:** Cada Data Mart é criado de maneira independente pelo respectivo departamento ou unidade de negócios, sem coordenação central.
- **Fontes de Dados Próprias:** Cada Data Mart extrai, transforma e carrega (ETL) seus dados diretamente das fontes operacionais.
- **Flexibilidade:** Esta arquitetura oferece maior flexibilidade e agilidade, pois cada Data Mart pode ser adaptado às necessidades específicas de cada departamento.

No entanto, a arquitetura de Data Marts independentes pode enfrentar problemas, como redundância de dados, dificuldade em garantir consistência de dados e a falta de uma visão centralizada e unificada dos dados da organização. Essa arquitetura é geralmente mais adequada para organizações menores ou situações em que as necessidades de análise de dados são mais simples e específicas para departamentos individuais, para consultar, analisar e visualizar os dados, como relatórios, dashboards, ferramentas de análise e mineração de dados.

Exemplo:



Barramento de Data Marts (Kimball):

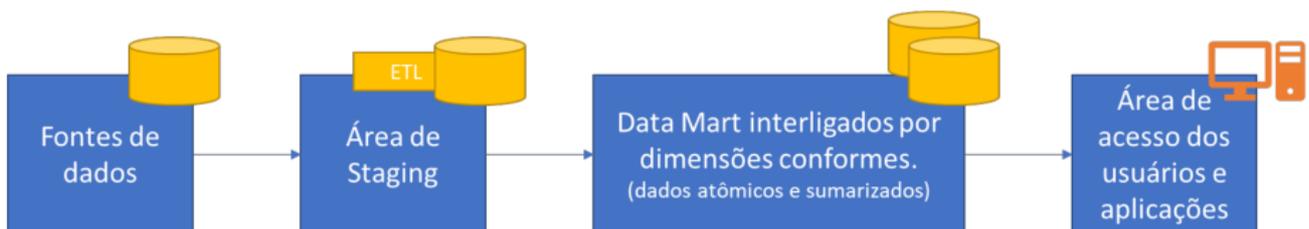
A arquitetura de Barramento de Data Marts, também conhecida como Arquitetura de Data Warehouse baseada em barramento, é um modelo de projeto que busca equilibrar a flexibilidade dos Data Marts independentes com a consistência de um Data Warehouse centralizado.

Aqui estão os principais aspectos desta arquitetura:



- **Barramento Comum:** A ideia central dessa arquitetura é a existência de um barramento de dados comum, que é um conjunto padronizado de definições de dados que são compartilhadas entre todos os Data Marts na organização. Este barramento comum ajuda a garantir a consistência dos dados em toda a organização.
- **Data Marts:** Em vez de ter um Data Warehouse centralizado, a arquitetura de barramento de Data Marts consiste em uma série de Data Marts conectados por meio do barramento de dados comum. Cada Data Mart é projetado para atender a uma área de negócios específica, como vendas, finanças ou marketing.
- **ETL:** Os dados são extraídos das fontes operacionais, transformados para se conformar às definições do barramento de dados comum e, em seguida, carregados em um ou mais Data Marts.
- **Flexibilidade e Consistência:** Esta arquitetura combina a flexibilidade dos Data Marts independentes (permitindo que cada departamento ou unidade de negócios personalize seu próprio Data Mart) com a consistência de um Data Warehouse centralizado (graças ao barramento de dados comum).

Exemplo:



Hub-and-spoke (Inmon):

A arquitetura hub-and-spoke é um modelo de Data Warehouse que visa integrar dados de diferentes fontes para fornecer uma visão unificada e consistente dos dados. A metáfora "hub-and-spoke" (ou "hub e raios" em português) refere-se à maneira como os dados são organizados e gerenciados nesta arquitetura.

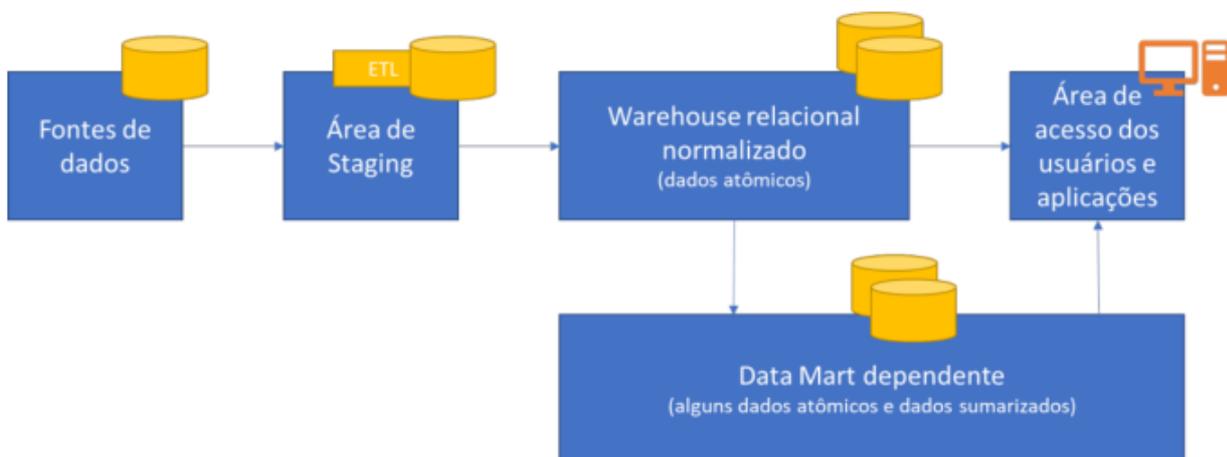
Os aspectos principais desta arquitetura incluem:

- **Hub Central (Núcleo):** O hub, ou núcleo, é o repositório central de dados, geralmente um Data Warehouse, onde os dados de diferentes fontes são integrados e consolidados. Este hub central garante uma única versão da verdade e consistência nos dados.
- **Spokes (Raios):** Os "spokes" ou "raios" são os Data Marts, que são subconjuntos especializados do Data Warehouse principal. Cada Data Mart é projetado para atender às necessidades de uma área de negócios específica, como vendas, marketing ou finanças. Os dados são alimentados a partir do hub central para os respectivos Data Marts.



- **Processo ETL:** Os dados são extraídos das fontes operacionais, transformados e, em seguida, carregados no hub central. De lá, os dados são novamente extraídos, possivelmente transformados para atender a requisitos específicos do Data Mart e, em seguida, carregados nos Data Marts.
- **Consistência e Integração:** Esta arquitetura permite a consistência dos dados (graças ao hub central) enquanto ainda permite a personalização dos Data Marts para atender às necessidades específicas de diferentes departamentos ou áreas de negócios.

Exemplo:



DW Centralizado:

A arquitetura de Data Warehouse (DW) centralizado é um modelo em que todos os dados de uma organização são coletados e integrados em um único local central. Essa abordagem simplifica a gestão de dados e fornece uma visão consistente e unificada dos dados em toda a organização.

Os aspectos principais dessa arquitetura incluem:

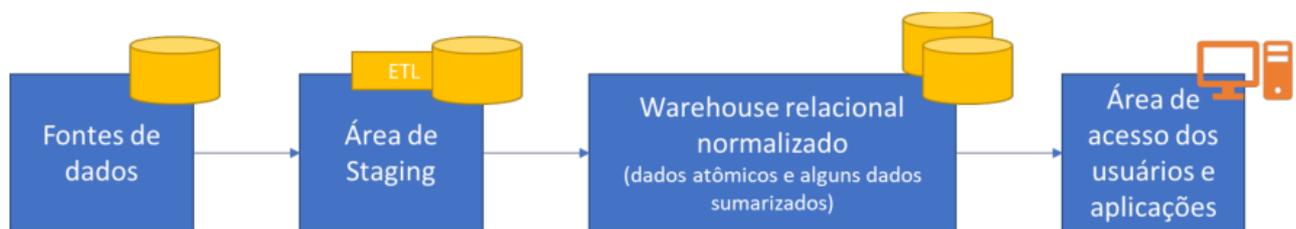
- **Armazenamento Centralizado:** Em um DW centralizado, os dados de várias fontes são extraídos, transformados e carregados (ETL) em um único repositório central. Este Data Warehouse armazena todos os dados relevantes para a organização, muitas vezes mantendo-os em um formato orientado por assunto e integrado para facilitar a análise.
- **Consistência e Integridade de Dados:** O uso de um DW centralizado ajuda a garantir que todos os dados estejam consistentes e sejam tratados da mesma maneira, independentemente de onde se originaram. Isso ajuda a manter a integridade dos dados e oferece uma única versão da verdade.



- **Acesso a Dados:** Os usuários podem acessar e consultar dados do DW centralizado usando várias ferramentas de Business Intelligence (BI). Como todos os dados estão armazenados em um único local, é mais fácil para os usuários encontrar as informações de que precisam e realizar análises complexas.

No entanto, uma desvantagem potencial de um DW centralizado é que ele pode ser mais difícil de escalar à medida que a quantidade de dados aumenta. Além disso, pode não oferecer a mesma flexibilidade que as arquiteturas baseadas em Data Marts para atender às necessidades específicas de diferentes departamentos ou grupos de usuários.

Exemplo:



DW Federado:

A arquitetura de Data Warehouse (DW) federado é um modelo que combina vários Data Warehouses ou Data Marts, cada um possuindo seu próprio banco de dados e regras de negócio, em uma estrutura unificada para proporcionar uma visão integrada dos dados.

Os principais aspectos dessa arquitetura incluem:

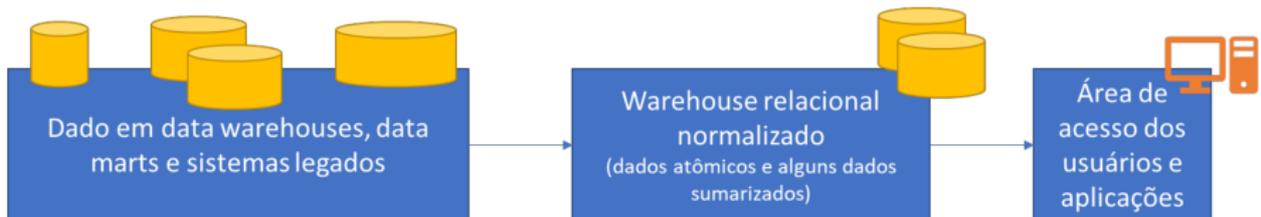
- **Diversos Data Warehouses ou Data Marts:** Em uma arquitetura federada, existem vários Data Warehouses ou Data Marts que são desenvolvidos e gerenciados de forma independente, mas conectados através de uma camada de integração.
- **Camada de Integração:** A camada de integração é responsável por unificar os diferentes Data Warehouses ou Data Marts, facilitando consultas e análises que necessitam de dados de várias fontes. Geralmente, isso é feito através de uma interface comum que traduz as consultas em instruções apropriadas para cada sistema de banco de dados subjacente.
- **Flexibilidade e Escalabilidade:** A arquitetura federada permite que a organização integre dados de várias fontes sem a necessidade de criar um único Data Warehouse centralizado. Isso oferece maior flexibilidade e escalabilidade, pois novos Data Warehouses ou Data Marts podem ser adicionados à federação conforme necessário.

Porém, a arquitetura de DW federado também tem seus problemas. Garantir a consistência e qualidade dos dados entre os diferentes sistemas pode ser complexo. Além disso, a performance



das consultas pode ser um desafio, especialmente quando os dados precisam ser coletados e integrados de vários sistemas em tempo real.

Exemplo:



Abordagens de DW: Top-Down e Bottom-Up

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

Abordagem Top-Down (Bill Inmon):

A abordagem top-down para o design do Data Warehouse foi popularizada por Bill Inmon e envolve a construção de um Data Warehouse empresarial centralizado antes de criar Data Marts departamentais.

Características:

- **Data Warehouse Centralizado:** Inmon defende a criação de um Data Warehouse centralizado que integra todos os dados da organização. Esse Data Warehouse serve como a única fonte de verdade para toda a organização.
- **Data Marts Departamentais:** Depois que o Data Warehouse centralizado é criado, os Data Marts departamentais são construídos a partir dele. Esses Data Marts contêm subconjuntos do Data Warehouse principal e são projetados para atender às necessidades específicas de diferentes departamentos.
- **Foco na Consistência:** A abordagem top-down enfatiza a consistência dos dados em toda a organização. A ideia é que, ao ter um único Data Warehouse centralizado, a organização possa garantir que todos estejam trabalhando com os mesmos dados.

Abordagem Bottom-Up (Ralph Kimball):

A abordagem bottom-up, por outro lado, foi popularizada por Ralph Kimball e envolve a construção de Data Marts departamentais antes de integrá-los em um Data Warehouse empresarial.



Características:

- **Data Marts Departamentais:** Na abordagem de Kimball, os Data Marts departamentais são construídos primeiro. Cada Data Mart é projetado para atender às necessidades específicas de um departamento ou área de negócios.
- **Data Warehouse Empresarial:** Uma vez que vários Data Marts foram construídos, eles são então integrados em um Data Warehouse empresarial. O Data Warehouse atua como um consolidador dos diferentes Data Marts.
- **Foco na Agilidade:** A abordagem bottom-up é focada na agilidade e na entrega rápida de valor para os usuários finais. Ao construir Data Marts departamentais primeiro, a organização pode fornecer rapidamente aos usuários finais acesso aos dados de que precisam.

Comparação:

Embora ambas as abordagens tenham suas vantagens, elas também têm suas desvantagens. A abordagem top-down, enquanto proporciona consistência, pode ser mais lenta e mais cara de implementar, pois requer a construção de um Data Warehouse centralizado antes que os usuários finais possam acessar os dados.

Por outro lado, a abordagem bottom-up pode entregar valor mais rapidamente, pois os Data Marts departamentais são construídos primeiro. No entanto, esta abordagem pode levar a problemas de consistência e integração de dados, especialmente se os diferentes Data Marts não forem projetados com uma visão integrada em mente.

A tabela a seguir resume as diferenças entre as duas abordagens:

Características	Abordagem Top-Down (Bill Inmon)	Abordagem Bottom-Up (Ralph Kimball)
Ordem de Construção	Começa com a construção do Data Warehouse centralizado, seguido pelos Data Marts departamentais	Começa com a construção dos Data Marts departamentais, seguido pela integração desses para formar o Data Warehouse empresarial
Foco	Focado na consistência e integração dos dados em toda a organização	Focado na agilidade e entrega rápida de valor para os usuários finais



Fonte de Verdade	O Data Warehouse centralizado atua como a única fonte de verdade	Cada Data Mart pode atuar como uma fonte de verdade para seu respectivo departamento
Complexidade	Pode ser mais complexo e demorado para implementar, devido à necessidade de construir o Data Warehouse centralizado antes dos Data Marts	Pode ser mais rápido para implementar inicialmente, mas pode aumentar a complexidade a longo prazo devido à necessidade de integrar vários Data Marts
Custo	Pode ser mais caro inicialmente, devido à necessidade de construir o Data Warehouse centralizado	Pode ser menos caro inicialmente, mas os custos podem aumentar a longo prazo devido à necessidade de integrar vários Data Marts
Gestão de Dados	É mais fácil garantir a consistência e qualidade dos dados, pois há um único Data Warehouse centralizado	Pode haver desafios para garantir a consistência e qualidade dos dados, pois há vários Data Marts independentes
Adaptabilidade às Necessidades de Negócios	Pode ser menos flexível para atender às necessidades específicas de diferentes departamentos ou usuários	Pode ser mais flexível para atender às necessidades específicas de diferentes departamentos ou usuários

Modelagem Multidimensional

A modelagem multidimensional é uma técnica de design comumente usada em sistemas de Business Intelligence, como Data Warehouses, para estruturar dados de uma maneira que seja intuitiva e de fácil compreensão para os usuários finais.

Em vez de organizar os dados em tabelas com linhas e colunas, como na modelagem relacional, a modelagem multidimensional organiza os dados em uma estrutura cúbica, onde os dados são acessados por várias dimensões.

Aqui estão os principais componentes da modelagem multidimensional:



Fatos: Os fatos representam dados quantitativos ou mensuráveis que uma empresa pode querer analisar. Por exemplo, as vendas totais, o número de produtos vendidos ou a quantidade de horas trabalhadas.

Dimensões: Dimensões são as várias perspectivas ou ângulos a partir dos quais uma organização pode querer analisar os dados. Exemplos comuns de dimensões incluem tempo, geografia, produtos e departamentos.

Hierarquias: As hierarquias representam os níveis dentro de uma dimensão. Por exemplo, a dimensão tempo pode ter uma hierarquia que vai de dias para meses, trimestres e anos.

Com a modelagem multidimensional, os usuários finais podem realizar análises complexas e consultas ad-hoc simplesmente navegando pelo cubo de dados. Eles podem "fatiar e cortar" os dados para ver diferentes níveis de detalhe, e podem "perfurar" os dados para ver informações mais específicas.

Modelagem multidimensional também é altamente otimizada para consultas de agregação, o que é comum em aplicações de BI. Por exemplo, um usuário pode querer ver o total de vendas por região, por produto ou por trimestre. Essas consultas podem ser realizadas muito mais rapidamente em um modelo multidimensional do que em um modelo relacional.

Por fim, a modelagem multidimensional é intuitiva para os usuários de negócios porque representa os dados de uma maneira que reflete como eles pensam sobre os dados. Isso facilita a adoção e o uso efetivo de ferramentas de BI pelos usuários de negócios.

Dimensões

Em um Data Warehouse (DW), as dimensões representam as várias **perspectivas ou categorias a partir das quais os dados são analisados**. Elas são um componente fundamental da modelagem multidimensional, que é uma técnica de design comumente usada em sistemas de Business Intelligence (BI).

Uma dimensão, em essência, é um ângulo pelo qual os dados de negócios podem ser visualizados ou analisados. Cada dimensão é geralmente composta de uma lista de valores discretos, chamados de membros, e frequentemente tem uma estrutura hierárquica que permite a análise em diferentes níveis de granularidade.

Veja alguns exemplos de dimensões comumente usadas em um DW:



Dimensão de Tempo: Esta é talvez a dimensão mais comum. A dimensão de tempo pode incluir elementos como dia, semana, mês, trimestre e ano, permitindo análises ao longo do tempo. Por exemplo, um analista de negócios pode querer analisar as vendas trimestrais ou o crescimento ano a ano.

Dimensão Geográfica: Esta dimensão permite a análise por localização. Pode incluir elementos como país, estado, cidade, região, etc. Por exemplo, um gerente de vendas pode querer analisar as vendas por região ou cidade.

Dimensão de Produto: Esta dimensão permite a análise por produto. Pode incluir elementos como SKU do produto, nome do produto, categoria de produto, etc. Por exemplo, um gerente de produto pode querer analisar as vendas por produto ou categoria de produto.

Dimensão de Cliente: Esta dimensão permite a análise por cliente. Pode incluir elementos como ID do cliente, nome do cliente, segmento do cliente, etc. Por exemplo, um gerente de marketing pode querer analisar o comportamento de compra por segmento de cliente.

Dimensão de Funcionários: Esta dimensão permite a análise por funcionário. Pode incluir elementos como ID do funcionário, nome do funcionário, departamento, etc. Por exemplo, um gerente de RH pode querer analisar a performance dos funcionários por departamento.

Essas dimensões ajudam a fornecer contexto aos fatos mensuráveis no DW, como vendas, custos ou horas trabalhadas, permitindo que os usuários de negócios realizem análises complexas e tomem decisões informadas.

Fatos

Os fatos são elementos centrais de um Data Warehouse (DW) que são quantificáveis e mensuráveis, e que uma empresa pode querer analisar. Eles geralmente representam o desempenho ou o comportamento dos negócios e são normalmente numéricos. Os fatos são armazenados em tabelas de fatos que estão no centro de um esquema de estrela ou floco de neve (descreveremos os esquemas mais à frente) na modelagem de dados multidimensional.

Existem três tipos principais de fatos:

Fatos Aditivos: São os mais comuns e podem ser somados em qualquer dimensão. Por exemplo, a quantidade total de vendas, a quantidade total de produtos vendidos, etc.



Fatos Semi-Aditivos: Podem ser somados em algumas dimensões, mas não em todas. Um exemplo seria o saldo bancário - podemos somar saldos ao longo do tempo, mas não faria sentido somar saldos entre contas.

Fatos Não Aditivos: Não podem ser somados de forma significativa em nenhuma dimensão. Por exemplo, a temperatura média ou a porcentagem de lucro.

Veja alguns exemplos de fatos comumente usados em um DW:

- **Vendas Totais:** Este é um fato aditivo que representa a quantidade total de vendas feitas por uma empresa. Pode ser analisado por várias dimensões, como tempo, geografia, produto, etc.
- **Número de Produtos Vendidos:** Este é outro fato aditivo que representa o número total de produtos vendidos. Pode ser analisado por várias dimensões, como tempo, geografia, produto, etc.
- **Horas Trabalhadas:** Este é um fato aditivo que representa o número total de horas trabalhadas pelos funcionários. Pode ser analisado por várias dimensões, como tempo, departamento, cargo, etc.
- **Saldo Bancário:** Este é um fato semi-aditivo que representa o saldo atual da conta bancária de um cliente. Pode ser analisado por várias dimensões, como tempo, cliente, tipo de conta, etc.
- **Porcentagem de Lucro:** Este é um fato não aditivo que representa a margem de lucro de um produto ou serviço. Pode ser analisado por várias dimensões, como tempo, produto, categoria de produto, etc.

Fatos são o que os usuários finais estão mais interessados em analisar, pois representam o desempenho real do negócio. Dimensões, por outro lado, fornecem o contexto para esses fatos, permitindo que os usuários entendam por que o desempenho é o que é e como ele pode ser melhorado.

Tipos de Tabela Fato:

A literatura destaca seis tipos de tabelas fato:

Fato Transacional: É a forma mais comum de tabela de fatos e representa eventos individuais que ocorrem na organização. Cada linha na tabela de fatos transacionais corresponde a um evento ou transação específica.

Fato Agregada: Uma tabela de fatos agregados armazena resultados pré-calculados de medidas agregadas, tais como somas e médias, a partir de uma tabela de fatos transacionais. Isso é feito para melhorar a performance das consultas.



Fato Consolidada: A tabela de fatos consolidada é uma versão simplificada da tabela de fatos que consolida dados de diferentes tabelas de fatos para facilitar a análise.

Fato de Snapshot Periódico: Esta tabela de fatos fornece uma imagem dos dados em intervalos regulares, como semanal, mensal, trimestral, etc. Cada registro representa um resumo de um período de tempo específico.

Fato de Snapshot Acumulado: Este tipo de tabela de fatos é útil para processos de negócios que têm claramente um início, meio e fim, onde se quer analisar a duração do ciclo. Cada registro rastreia o estado do evento em intervalos de tempo regulares entre o início e o fim.

Fato Sem Fato: Uma tabela de fatos sem fatos (ou tabela de fatos de eventos) é uma tabela de fatos que não possui medidas numéricas. Ela é usada para registrar a ocorrência de eventos. Por exemplo, em um sistema de atendimento ao cliente, cada chamada de um cliente pode ser registrada sem qualquer medida quantitativa.

O quadro a seguir resume as características de Tabelas Fato e Tabelas de Dimensão:

	Tabelas Fato	Tabelas de Dimensão
Definição	Tabelas de fatos armazenam medidas quantitativas e informações que podem ser analisadas. Essas medidas são geralmente resultados numéricos de um evento de negócio.	As tabelas de dimensões fornecem o contexto para os fatos e são usadas para categorizar os dados de fatos. Elas contêm atributos que são normalmente descrições textuais.
Conteúdo	Contêm chaves primárias que se referem a chaves estrangeiras nas tabelas de dimensões e dados quantitativos associados a essas chaves.	Contêm chaves primárias que são referenciadas por chaves estrangeiras nas tabelas de fatos e atributos associados a essas chaves.
Estrutura	Normalmente contêm menos colunas, mas muito mais linhas do que tabelas de dimensões.	Normalmente contêm muitas colunas e menos linhas do que as tabelas de fatos.
Exemplos de Dados	Total de vendas, quantidade de produtos vendidos, horas trabalhadas, etc.	Detalhes do produto, detalhes do cliente, detalhes de tempo, detalhes da loja, etc.
Utilização	Usada para responder questões quantitativas sobre o negócio, como	Usada para fornecer contexto para as medidas, como "Quais produtos



	"Qual foi o total de vendas no último trimestre?"	foram vendidos no último trimestre?" ou "Quais foram as vendas por região no último trimestre?"
--	---------------------------------------------------	-------------------------------------------------------------------------------------------------

Esquemas Multidimensionais

Esquemas multidimensionais são um tipo de estrutura de banco de dados projetada para melhorar a performance de sistemas de Data Warehouse e Business Intelligence (BI). Em vez de organizar dados de uma forma relacional linear (como em bancos de dados tradicionais), os esquemas multidimensionais organizam os dados em um formato que reflete melhor a maneira como os usuários finais realmente interagem e usam os dados.

Eles são chamados de "multidimensionais" porque permitem aos usuários visualizar e analisar dados de múltiplas dimensões ao mesmo tempo. Relembrando, uma "dimensão" é qualquer categoria que os usuários finais possam querer analisar. Por exemplo, uma loja de varejo pode querer analisar as vendas pela dimensão de "produto", "localização da loja", "tempo" e "cliente". Cada uma dessas dimensões fornece uma perspectiva única sobre os dados.

Existem vários tipos de esquemas multidimensionais, sendo os mais comuns os esquemas de Estrela, Floco de Neve e Constelação.

Esquema em Estrela:

O Esquema em Estrela (Star Schema) é um modelo de dados que organiza os dados em um formato mais simplificado, tornando-se um padrão popular para o design de Data Warehouses e aplicações de Business Intelligence. Ele é chamado de "esquema em estrela" por causa de sua aparência, onde uma tabela de fatos central está cercada por tabelas de dimensão, as quais estão visualmente dispostas de forma a se assemelhar a uma estrela.

Tabela de Fatos: Esta é a tabela central no esquema em estrela, que contém as medidas, métricas ou fatos de um negócio. A tabela de fatos armazena dados quantitativos que podem ser analisados e contém chaves estrangeiras para as tabelas de dimensão. Por exemplo, em um DW de varejo, a tabela de fatos pode conter dados como número de unidades vendidas, total de vendas, lucro, etc.

Tabelas de Dimensão: Estas são as tabelas que circundam a tabela de fatos e fornecem contexto para os dados na tabela de fatos. Elas contêm detalhes descritivos ou atributos de várias dimensões de um negócio. Cada registro na tabela de fatos geralmente tem uma



correspondência em cada uma das tabelas de dimensão. No mesmo DW de varejo, as tabelas de dimensão podem ser Cliente, Produto, Tempo, Loja, etc.

Um exemplo simplificado de um esquema em estrela para um DW de varejo poderia se parecer com isto:

Tabela de Fato: Vendas

ID_Venda (PK)	ID_Cliente (FK)	ID_Produto (FK)	ID_Tempo (FK)	ID_Loja (FK)	Unidades Vendidas	Total de Vendas
1	101	1001	20230101	10	3	1500
2	102	1002	20230102	11	2	2000

Tabela de Dimensão: Cliente

ID_Cliente (PK)	Nome do Cliente	Localização do Cliente	Segmento do Cliente
101	Cliente A	São Paulo	Segmento X
102	Cliente B	Rio de Janeiro	Segmento Y

Tabela de Dimensão: Produto

ID_Produto (PK)	Nome do Produto	Categoria do Produto	Preço do Produto
1001	Produto A	Categoria X	500
1002	Produto B	Categoria Y	1000

Tabela de Dimensão: Tempo

ID_Tempo (PK)	Dia	Mês	Ano
20230101	1	Janeiro	2023
20230102	2	Janeiro	2023

Tabela de Dimensão: Loja

ID_Loja (PK)	Localização da Loja	Tamanho da Loja
10	São Paulo	Grande



11	Rio de Janeiro	Pequena
----	----------------	---------

Neste esquema, a tabela de fatos (Vendas) contém dados mensuráveis e chaves estrangeiras para as tabelas de dimensão. As tabelas de dimensão fornecem o contexto para os dados na tabela de fatos. Este arranjo facilita a análise dos dados de várias perspectivas, como analisar as vendas por produto, por cliente, por localização da loja ou ao longo do tempo.

Floco de Neve:

O Esquema Floco de Neve (Snowflake Schema) é uma extensão do esquema em estrela onde as tabelas de dimensão são normalizadas, ou seja, a redundância de dados é eliminada. Essa normalização quebra a tabela de dimensão em várias tabelas relacionadas, formando uma estrutura que se assemelha a um floco de neve.

O uso do esquema floco de neve leva a economias de armazenamento e pode garantir a consistência dos dados através da redução de redundâncias. Por outro lado, consultas podem se tornar mais complexas devido ao maior número de tabelas envolvidas.

Aqui está um exemplo de Esquema Floco de Neve (Snowflake Schema) para um DW de varejo:

Tabela de Fato: Vendas

ID_Venda	ID_Cliente	ID_Produto	ID_Tempo	ID_Loja	Unidades Vendidas	Total de Vendas
1	101	1001	20230101	10	3	1500
2	102	1002	20230102	11	2	2000

Tabela de Dimensão: Cliente

ID_Cliente	Nome do Cliente	ID_Localização	ID_Segmento
101	Cliente A	1	1
102	Cliente B	2	2

Tabela de Dimensão: Produto

ID_Produto	Nome do Produto	ID_Categoria	Preço do Produto
1001	Produto A	1	500
1002	Produto B	2	1000



Tabela de Dimensão: Tempo

ID_Tempo	Dia	Mês	Ano
20230101	1	Janeiro	2023
20230102	2	Janeiro	2023

Tabela de Dimensão: Loja

ID_Loja	ID_Localização	Tamanho da Loja
10	1	Grande
11	2	Pequena

Tabela de Dimensão: Localização

ID_Localização	Cidade	Estado
1	São Paulo	SP
2	Rio de Janeiro	RJ

Tabela de Dimensão: Categoria de Produto

ID_Categoria	Nome da Categoria
1	Categoria X
2	Categoria Y

Tabela de Dimensão: Segmento de cliente

ID_Segmento	Nome do Segmento
1	Segmento X
2	Segmento Y

Neste esquema, as tabelas de dimensão foram normalizadas em tabelas de dimensão adicionais (Localização, Categoria de Produto, Segmento do Cliente). Isso pode melhorar a consistência e economizar espaço de armazenamento, embora possa tornar as consultas mais complexas devido ao maior número de tabelas envolvidas.

Constelação:



O Esquema de Constelação, também conhecido como Esquema de Galáxia ou Esquema de Fatos Múltiplos, é uma forma avançada de modelagem multidimensional. Nele, há múltiplas tabelas de fatos que compartilham tabelas de dimensão. Esta estrutura é útil quando diferentes processos de negócios têm tabelas de fatos separadas, mas compartilham algumas dimensões comuns.

Vamos criar um exemplo com duas tabelas de fatos: Vendas e Devoluções, e compartilhando dimensões comuns, como Produto, Tempo e Loja.

Tabela de Fato: Vendas

ID_Venda	ID_Produto	ID_Tempo	ID_Loja	Unidades Vendidas	Total de Vendas
1	1001	20230101	10	3	1500
2	1002	20230102	11	2	2000

Tabela de Fato: Devoluções

ID_Devolução	ID_Produto	ID_Tempo	ID_Loja	Unidades Devolvidas	Total de Devoluções
1	1002	20230103	10	1	1000
2	1001	20230104	11	2	1000

Tabela de Dimensão: Produto

ID_Produto	Nome do Produto	Categoria do Produto	Preço do Produto
1001	Produto A	Categoria X	500
1002	Produto B	Categoria Y	1000

Tabela de Dimensão: Tempo

ID_Tempo	Dia	Mês	Ano
20230101	1	Janeiro	2023
20230102	2	Janeiro	2023
20230103	3	Janeiro	2023
20230104	4	Janeiro	2023

Tabela de Dimensão: Loja



ID_Loja	Localização da Loja	Tamanho da Loja
10	São Paulo	Grande
11	Rio de Janeiro	Pequena

Neste esquema, a tabela de Vendas e a tabela de Devoluções são independentes uma da outra, mas compartilham tabelas de dimensão comuns. Isso permite análises complexas e abrangentes. Por exemplo, podemos comparar o total de vendas e devoluções por produto, por loja ou ao longo do tempo.

A tabela a seguir resume as diferenças entre os esquemas *Star Schema*, *Snowflake* e *Constelação*:

	Estrela	Snowflake	Constelação
Descrição	Simple e mais comum. Uma tabela de fatos ligada a várias tabelas de dimensão.	Extensão do esquema estrela com tabelas de dimensão normalizadas.	Modelo mais complexo com várias tabelas de fatos compartilhando tabelas de dimensão.
Simplicidade	Alta. Fácil de entender e implementar.	Média. A normalização adiciona complexidade.	Baixa. Pode ser difícil de entender devido à presença de várias tabelas de fatos.
Performance	Alta. Menor número de joins.	Média. Mais joins devido à normalização.	Média. Pode haver mais joins dependendo do número de tabelas de fatos.
Flexibilidade	Média. Suporta a maioria das necessidades de negócios.	Alta. Suporta necessidades mais complexas.	Alta. Suporta diferentes processos de negócios.
Uso de espaço	Maior devido à redundância.	Menor devido à normalização.	Varia. Dependente do número de tabelas de fatos e dimensões.
Exemplo	Vendas ligadas a Produto, Tempo, Loja, etc.	Vendas ligadas a Produto, Categoria de Produto, Tempo, etc.	Vendas e Devoluções ligadas a Produto, Tempo, Loja, etc.



Processo de Design Dimensional

O design dimensional, também conhecido como modelagem dimensional, é uma técnica de design usada para construir data warehouses, data marts e outros repositórios de dados. Este processo é estruturado em etapas que permitem criar um modelo que seja compreensível e eficiente para as necessidades de análise de dados. As etapas principais são:

1. **Selecionar o processo de negócio:** Esta é a primeira e uma das mais importantes etapas do processo de design dimensional. O primeiro passo é selecionar o processo de negócio que será representado no modelo dimensional. Este processo de negócio pode ser algo como vendas, compras, produção, etc. A seleção do processo de negócio determinará o escopo do modelo dimensional e influenciará todas as decisões subsequentes.
2. **Definir a granularidade:** A granularidade se refere ao nível de detalhe dos dados que serão armazenados na tabela de fatos. A granularidade pode ser alta (detalhada) ou baixa (resumida), dependendo das necessidades de análise de dados. Por exemplo, no caso do processo de vendas, a granularidade pode ser ao nível de cada transação individual (alta granularidade) ou ao nível de vendas diárias por loja (baixa granularidade).
3. **Identificar as dimensões:** As dimensões são as categorias de análise que dão contexto aos fatos. Elas são geralmente aspectos do negócio que são de interesse para a análise, como tempo, localização, produto, cliente, etc. As dimensões são identificadas com base no processo de negócio selecionado e nas perguntas de análise que se quer responder.
4. **Identificar as medidas:** As medidas são os dados quantitativos que são analisados no contexto das dimensões. No caso do processo de vendas, as medidas poderiam ser a quantidade de produtos vendidos, o valor total das vendas, o lucro, etc. As medidas são armazenadas na tabela de fatos e geralmente são numéricas e aditivas.

Estas etapas do design dimensional são repetidas para cada processo de negócio que se deseja representar no data warehouse ou data mart. O resultado é um conjunto de tabelas de fatos e dimensões que representam os processos de negócio e suportam a análise de dados.

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. (FGV / MPE-AL – 2018) No âmbito da utilização de bancos de dados no papel de “data warehouse”, é usualmente aceitável que as tabelas não estejam plenamente normalizadas. Assinale a opção que mostra porque essa característica, embora indesejada, é admitida:

- a) As dependências funcionais tornam-se obsoletas.
- b) As operações de insert/update não estão mais sujeitas ao controle de concorrência.
- c) Facilita os processos de ETL.
- d) Não é possível criar bancos de dados históricos normalizados.
- e) Simplifica as consultas e melhora o tempo de resposta.

Comentários:

(a) Errado, as dependências funcionais não se tornam obsoletas por conta de normalização; (b) Errado, não há operações de update e operações de insert não necessitam de controle de concorrência; (c) Errado, isso é uma consequência, mas não é por essa razão que tabelas desnormalizadas são admitidas; (d) Errado, isso é completamente independente da normalização ou não das tabelas; (e) Correto, a desnormalização simplifica consultas e melhora o tempo de resposta. Ela prejudica o desempenho de inserções, exclusões e atualizações – mas essas transações não devem ocorrer em um Data Warehouse.

Gabarito: E

2. (FGV / IBGE – 2017) Pedro foi contratado para desenvolver uma solução de integração de dados a partir de diversas fontes heterogêneas com o objetivo de apoiar a análise de informações. A solução deve ter as seguintes características: Não-volátil, Histórico, Orientado a Assunto e Variante no Tempo. Para isso, Pedro deve desenvolver um:

- a) Sistema de Informação Transacional;
- b) Data Warehouse;
- c) Otimizador de Consultas;
- d) Sistema Gerenciador de Banco de Dados;
- e) Banco de Dados Distribuído.



Comentários:

Definição de Bill Inmon: Data Warehouse 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.

Gabarito: B

3. (FGV / AL-MA – 2013) Bill Inmon é considerado um dos “pais” da tecnologia de data warehouse. O autor define o termo como sendo um conjunto de _____ integrado, orientado _____, _____ no tempo e _____ que fornece suporte ao processo de tomada de decisão”.

Assinale a alternativa cujos itens completam corretamente as lacunas do fragmento acima.

- a) dados– por assunto – variável – não volátil.
- b) dados – por objetos – não variável – granular.
- c) processos– por assunto – variável – volátil.
- d) processos– por relacionamentos – muito variável–granular.
- e) processos – por relacionamentos – pouco variável –volátil.

Comentários:

De acordo com Bill Inmon, um Data Warehouse 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.

Gabarito: A

4. (FGV / Senado Federal – 2012) Assinale a alternativa que apresenta características dos Datawarehouses.



- a) Processamento transacional, dados históricos e consultas pré-definidas simples.
- b) Processamento analítico, manipulação de pequenos volumes de dados e consultas simples.
- c) Processamento transacional, dados desnormalizados e atualizações online.
- d) Processamento analítico, dados históricos e dados normalizados.
- e) Processamento analítico, dados desnormalizados e consultas ad hoc complexas.

Comentários:

(a) Errado, o processamento é analítico e as consultas são ad hoc complexas; (b) Errado, ele manipula grandes volumes de dados e as consultas são ad hoc complexas; (c) Errado, o processamento é analítico e não há atualizações online; (d) Errado, os dados geralmente são desnormalizados; (e) Correto, o processamento é realmente analítico, os dados geralmente são desnormalizados e as consultas são ad hoc e complexas.

O que é uma consulta ad hoc? A expressão Ad hoc é uma expressão em latim que significa "para este propósito". Ou seja, a consulta é criada apenas para satisfazer aquela necessidade específica, aquele propósito, em um momento específico. É diferente das consultas de bancos de dados tradicionais, que geralmente são consultas pré-determinadas e genéricas.

Gabarito: E

5. (FGV / MEC – 2009) Um termo está associado a uma coleção de dados orientada por assuntos, integrada, variante no tempo, que tem por objetivo dar suporte aos processos de tomada de decisão. É implementado por meio de um banco de dados contendo dados extraídos do ambiente de produção da empresa, que foram selecionados e depurados, tendo sido otimizados para processamento de consulta e não para processamento de transações. Em geral, requer a consolidação de outros recursos de dados além dos armazenados em base de dados relacionais, incluindo informações provenientes de planilhas eletrônicas e documentos textuais. Seu objetivo é fornecer uma "imagem única da realidade do negócio". De uma forma geral, são sistemas que compreendem um conjunto de programas que extraem dados do ambiente de dados operacionais da empresa, um banco de dados que os mantém, e



sistemas que fornecem estes dados aos seus usuários. O termo aqui tratado define o conceito de:

- a) DataMining
- b) DataSystems
- c) DataBusiness
- d) DataProcessing
- e) DataWarehouse

Comentários:

O termo associado a uma coleção de dados orientada por assuntos, integrada, variante no tempo, que tem por objetivo dar suporte aos processos de tomada de decisão é o Data Warehouse.

Gabarito: E

6. (FADESP / SEFA-PA – 2022) Considerando os conceitos e características próprios de um Data Warehouse, julgue verdadeira (V) ou falsa (F) cada uma das afirmativas a seguir.

I. Os Data Warehouses existem como armazenamento persistente em vez de serem materializados por demandas.

II. Os Data Warehouses podem ser indexados para otimizar o desempenho e, caracteristicamente, oferecem suporte específico de funcionalidades.

III. Os Data Warehouses oferecem uma grande quantidade de dados integrados e são livres das restrições do ambiente transacional, permitindo uma eficiência aumentada no processamento da consulta.

A sequência correta é:

- a) I - F; II - F; III - F.
- b) I - V; II - F; III - V.



- c) I - V; II - V; III – F.
- d) I - F; II - V; III - V.
- e) I - V; II - V; III - V.

Comentários:

(I) Correto. Data Warehouses existem como armazenamento persistente (em que os dados são preservados) em vez de serem materializados por demandas (como ocorrem com as views); (II) Correto. Data Warehouses realmente podem ser indexados para otimizar seu desempenho e melhorar a eficiência das consultas, suportando funcionalidades específicas; (III) Correto. Uma das características do Data Warehouse são os dados integrados, onde os dados não têm diversas limitações que existem no ambiente transacional – isso, claro, permite o aumento de eficiência da consulta.

Gabarito: E

7. (CESGRANRIO / BB – 2021) Os sistemas interativos que provêm inteligência de negócio, BI ou business intelligence, em uma organização, são utilizados por seus gestores para:

- a) configuração do controle de acesso aos dados de cada transação da organização.
- b) encadeamento das atividades de um processo de trabalho da organização.
- c) exploração de dados sumarizados para compreensão e inspiração na solução de problemas.
- d) coordenação da execução de transações distribuídas.
- e) correção de dados diretamente em sistemas transacionais.

Comentários:

Sistemas interativos para inteligência de negócio são utilizados por gestores para exploração de dados sumarizados para compreensão e inspiração na solução de problemas. Nenhuma das outras opções faz qualquer sentido!

Gabarito: C



8. (VUNESP / Prefeitura de Campinas - SP – 2019) No contexto de armazéns de dados (data warehouse), a área intermediária na qual os dados coletados pelo processo de ETL são armazenados antes de serem processados e transportados para o seu destino é chamada de:

- a) cubo OLAP.
- b) dicionário de dados.
- c) staging.
- d) data vault.
- e) data mart.

Comentários:

A área intermediária na qual os dados coletados são armazenados antes de serem processados e transportados para o seu destino é chamada de stage area ou staging.

Gabarito: C

9. (IADES / APEX BRASIL - 2018) No jargão empresarial moderno, *business intelligence* é o (a):

- a) mesmo que *data warehouse*, já que ambos têm a finalidade de armazenar dados e criar relatórios gerenciais.
- b) inteligência artificial dos computadores modernos.
- c) inteligência da empresa que tem por base os sistemas modernos de informatização.
- d) processo de recolhimento e tratamento de informações que apoiarão a gestão de um negócio.
- e) conjunto de relatórios preparados pelos executivos de uma empresa.

Comentários:



(a) Errado. DW é uma coleção de dados orientados por assunto, integrados, variáveis com o tempo e não-voláteis – já o BI é um processo que auxilia na tomada de decisões;

(b) Errado. Inteligência artificial se refere à capacidade de simular a inteligência humana e Business Intelligence se refere à capacidade de auxiliar na tomada de decisões;

(c) Errado. Não se trata de inteligência da empresa, mas de processos que ajudam na tomada de decisões empresariais estratégicas;

(d) Correto. Pode ser definido como um processo de recolhimento e tratamento de informações que auxiliarão a gestão de um negócio;

(e) Errado. Esse item não faz o menor sentido. Na verdade, pode-se gerar relatórios para os executivos de uma empresa, mas isso não é a definição de Business Intelligence.

Gabarito: D

10. (GESTÃO CONCURSO / EMATER/MG – 2018) Sobre os dados do Data Warehouse, avalie as afirmações a seguir.

I. São integrados.

II. São orientados por assunto.

III. Contêm poucos níveis de detalhes.

IV. Contêm dados históricos de um período curto.

V. São dados capturados em um determinado momento.

a) I, II e III.

b) I, II e V.

c) I, III e V.

d) II, III e IV.

Comentários:



De acordo com Bill Inmon, um Data Warehouse 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. Em geral, ele contém dados históricos de um período longo – dados de um período curto geralmente estão em bancos de dados transacionais. Por fim, os dados realmente são capturados em um determinado momento pelo Processo ETL.

Gabarito: B

11. (CESGRANRIO / CEF – 2021) As resoluções assumidas por um gestor dependem, fundamentalmente, da consolidação de dados e informações que sustentam o processo de tomada de decisão. Assim sendo, ferramentas que têm por objetivo organizar e apresentar dados e informações relevantes ao processo de tomada de decisão são denominadas:

- a) Codecs.
- b) Dashboards.
- c) Hardening.
- d) Weblogs.
- e) LMS (Learning Management Systems).

Comentários:

Os dashboards são painéis com informações de vários dados relevantes para um negócio – eles normalmente são apresentados por meio de gráficos, facilitando a análise. Esses dados podem, ainda, serem customizados de acordo com a necessidade do gestor. Além disso, dashboards permitem aos profissionais a capacidade de monitorar o desempenho, criar relatórios e definir estimativas e metas para trabalhos futuros. Em suma, os dashboards são uma ótima ferramenta, pois conseguem agregar dados de várias fontes, permitindo uma análise mais rápida pelos funcionários.

Gabarito: B

12. (VUNESP / Prefeitura de Campinas - SP – 2019) No contexto de armazéns de dados (data warehouse), a área intermediária na qual os dados coletados pelo processo de ETL são



armazenados antes de serem processados e transportados para o seu destino é chamada de:

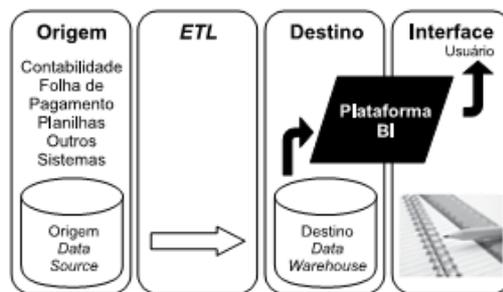
- a) cubo OLAP.
- b) dicionário de dados.
- c) staging.
- d) data vault.
- e) data mart.

Comentários:

A área intermediária na qual os dados coletados são armazenados antes de serem processados e transportados para o seu destino é chamada de *stage area* ou *staging*.

Gabarito: C

13. (VUNESP / MPE-SP - 2016) Observe o esquema a seguir, que representa a estrutura geral e os componentes de uma solução de Business Intelligence – BI (Inteligência Empresarial ou Inteligência de Negócios).



Sobre esse esquema, é correto afirmar que:

- a) a camada “origem de dados” é o ambiente operacional onde se encontram os aplicativos de gestão, com alto grau de interação sistêmica entre eles (os aplicativos), e é onde os consumidores e fornecedores interagem diretamente com os sistemas administrativos.
- b) na camada ETL (do inglês *Extract Transform Load* – Extração, Transformação e Carga), os dados extraídos dos aplicativos de gestão são inseridos diretamente no banco de dados.



c) no centro dessa arquitetura, está o *data warehouse* (banco de dados), que além de conter dados granulares integrados, é considerado o “coração” de informações da fábrica.

d) na última camada, à direita, está a plataforma de BI, representando um conjunto de ferramentas que apenas disponibiliza dados para consulta e processamento pelo usuário, com alta disponibilidade e grande número de acessos.

e) a estrutura geral e os componentes que compõem uma solução de BI não necessariamente demandam que cada um de seus componentes e o fluxo de informações corporativas sejam delineados de forma lógica. Basta apenas as informações estarem organizadas no *data warehouse* (banco de dados).

Comentários:

(a) Errado. Na origem dos dados estão os aplicativos operacionais/transacional e, não, de gestão. Esses dados servem de insumo para popular o Data Warehouse após o Processo de ETL;

(b) Errado. Os dados extraídos dos aplicativos operacionais/transacionais passam por uma Stage Area, são limpos, transformados e só então são carregados no banco de dados dimensional;

(c) Correto. Apesar de não saber o que o examinador quis dizer com “informações da fábrica”, o DW realmente é o centro da arquitetura e contém dados granulares;

(d) Errado. As informações de BI não apenas disponibilizam dados para consultas e processamento pelo usuário – elas possuem diversas outras funcionalidades;

(e) Errado. Por conta do Processo de ETL, os componentes devem necessariamente ser delineados de forma lógica para que os dados sejam armazenados de forma organizada no DW

Gabarito: C



14. (VUNESP / PREF SÃO JOSÉ DOS CAMPOS - 2015) Na modelagem multidimensional, um de seus componentes é a tabela fato, sobre a qual é correto afirmar que, em sua composição típica:

- a) contém um mínimo de 10% dos valores de seus atributos preenchidos com o valor nulo.
- b) contém valores numéricos que representam o desempenho de um negócio.
- c) não admite atributos do tipo literal ou numérico.
- d) não contém chaves primárias nem chaves estrangeiras.
- e) não pode conter mais de 10 atributos do tipo numérico.

Comentários:

Na modelagem multidimensional, a tabela fato é responsável por armazenar os valores numéricos que representam o desempenho de um negócio ou medida que está sendo analisada. A tabela fato contém as métricas ou medidas que são objeto de análise, como vendas, lucro, quantidade, tempo, entre outros.

Gabarito: B

15. (VUNESP / TCE SP - 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

Comentários:

Na modelagem multidimensional, as medições numéricas que representam o desempenho do negócio são armazenadas na tabela fato. A tabela fato contém os dados detalhados e granulares que são analisados, como vendas, lucro, quantidade, tempo, entre outros. Essas medições são geralmente valores numéricos que descrevem o desempenho do negócio em relação às dimensões.



As tabelas dimensão, por sua vez, contêm os atributos que descrevem as características das dimensões, como o tempo, o produto, a localização, entre outros. Essas tabelas são relacionadas à tabela fato por meio de chaves estrangeiras.

Gabarito: C

16. (VUNESP / TJ-PA – 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.

Comentários:

No modelo dimensional, as tabelas dimensão contêm atributos que descrevem as características de cada dimensão. As dimensões representam os aspectos que desejamos analisar, como tempo, produto, localização, cliente, entre outros. Os atributos presentes nas tabelas dimensão ajudam a descrever e classificar os dados na tabela fato.

As tabelas fato, por sua vez, armazenam as medições numéricas que representam o desempenho do negócio. Essas medições estão relacionadas às dimensões por meio de chaves estrangeiras.

Gabarito: C

17. (VUNESP / TJ-SP - 2012) Na especificação de uma tabela fato de um modelo multidimensional de um data warehouse, é importante definir o grau de detalhamento de seus valores. A denominação utilizada para tal detalhamento é:

- a) backroom.
- b) cesta de mercado.
- c) granularidade.
- d) parcelamento.



e) snowflake.

Comentários:

Na especificação de uma tabela fato em um modelo multidimensional de um data warehouse, é importante definir o nível de detalhamento dos valores. A granularidade refere-se ao nível de detalhe ou precisão dos dados armazenados na tabela fato. Ela determina o quão agregados ou detalhados são os dados na tabela fato.

Por exemplo, em uma tabela fato de vendas, a granularidade pode ser definida para registrar as vendas em nível diário, semanal, mensal ou anual. Isso influencia a quantidade de registros e o nível de detalhe dos dados armazenados.

Gabarito: C

18. (VUNESP - 2014 - TJ-PA - Analista Judiciário - Análise de Sistema - Desenvolvimento)

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.

Comentários:

No processo de extração, transformação e carga (ETL) utilizado na montagem de um data warehouse, a transformação de dados é uma etapa importante. Um exemplo de transformação de dados é a conversão de diversas unidades de medida para uma única

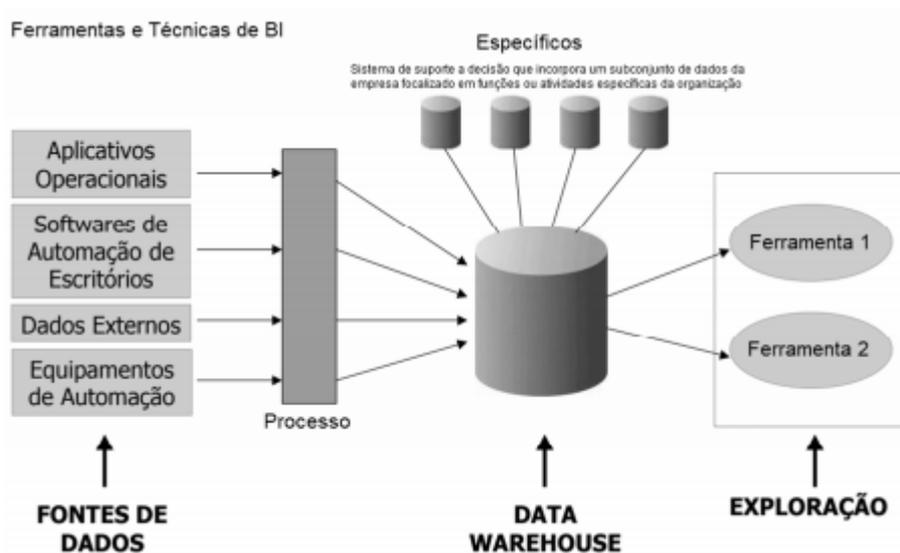


unidade, como converter centímetros (cm), milímetros (mm) e metros (m) para uma única unidade, como centímetros.

As transformações de dados podem incluir limpeza, padronização, conversão de formatos, agregação, cálculos derivados e muito mais. O objetivo é preparar e organizar os dados do ambiente operacional para serem carregados no data warehouse de forma coerente e útil para análise.

Gabarito: B

19. (FCC / SANASA – 2019) Atenção: Para responder à questão, considere a imagem abaixo.



O sistema de suporte a decisão representado em cada um dos cilindros do conjunto denominado Específicos, na imagem, é um:

- Catálogo de Metadados.
- Schema.
- Drill.
- OLTP.
- Data Mart.

Comentários:

Cada um dos cilindros pequenos são Data Marts e o cilindro grande é um Data Warehouse.



Gabarito: Letra E

20. (FCC / SEFAZ-BA – 2019) Nos sistemas transacionais, os dados sofrem diversas alterações como inclusão, alteração e exclusão. Antes de serem carregados no ambiente de um Data Warehouse, os dados são filtrados e limpos, de forma a gerarem informação útil. Após esta etapa, esses dados:

- a) ficam disponíveis para a mineração em tempo real, pois tais dados são constantemente atualizados a partir da chave de tempo que indica o dia em que foram extraídos dos sistemas transacionais.
- b) podem sofrer operações de consulta, mas, devido a sua não volatilidade, não podem ser alterados, não havendo necessidade de bloqueio por concorrência de usuários ao seu acesso.
- c) são reunidos a partir de diversas fontes de dados, o que facilita muito o trabalho do analista, embora este tenha que lidar com a grande redundância das informações.
- d) ficam ordenados pela data da extração do sistema transacional, sendo necessárias técnicas de data mining para fazer a sua recuperação orientada por assunto.
- e) são classificados somente pelo assunto principal de interesse da organização. Por exemplo, em uma organização de arrecadação de impostos, os dados são organizados pelo cadastro de contribuintes que possuem impostos a recolher.

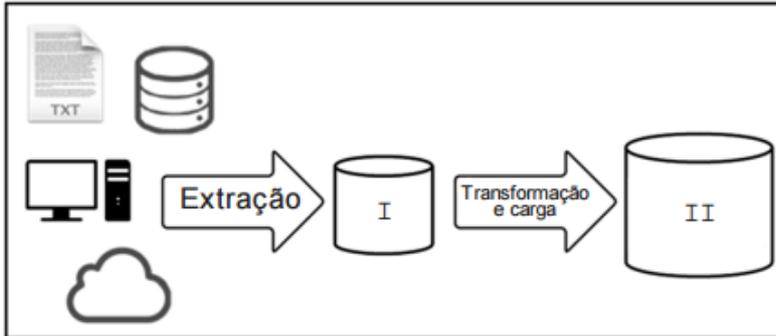
Comentários:

(a) Errado, os dados não são constantemente atualizados; (b) Correto; (c) Errado, a redundância é tratada durante o processo de transformação; (d) Errado, eles ficam ordenados pela data do conteúdo do dado em si e, não, pela data de extração; (e) Errado, eles podem ser organizados por diversos tipos de assuntos diferentes de acordo com o desejo do usuário.

Gabarito: Letra B



21. (FCC / AL - MS – 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.

Comentários:

Após a Extração, os dados ficam em uma **Staging Area** aguardando para serem transformados e carregados em um **Data Warehouse**.

Gabarito: Letra E

22. (CESPE / CNMP - 2023) Em data warehouse, o conceito de granularidade refere-se ao nível de detalhe ou resumo existente em uma unidade de dados, de forma que, quanto mais detalhes, mais alto o nível de granularidade.

Comentários:

É o exato oposto! Quanto maior o nível de detalhe dos dados, menor será a granularidade, e quanto menor o nível de detalhe, maior será a granularidade. Isso significa que, com uma granularidade mais alta, os dados são mais detalhados e específicos, enquanto que com uma granularidade mais baixa, os dados são agregados e resumidos.



Gabarito: Errado

23. (CESPE / CNMP - 2023) Fatos, dimensões e medidas são elementos essenciais de um data warehouse.

Comentários:

Perfeito! Fatos representam os dados numéricos e quantificáveis que são o foco da análise em um data Warehouse; Dimensões são as características ou contextos que fornecem uma estrutura para analisar os fatos; e Medidas são os atributos numéricos associados aos fatos. Elas representam os valores específicos que estão sendo analisados ou calculados.

Gabarito: Correto

24. (CESPE / DPDF – 2022) Um ambiente de data warehouse permite acessos simultâneos ao mesmo registro, para consulta e(ou) atualização, sem a preocupação com bloqueios de deadlock.

Comentários:

Deadlock é uma situação em que ocorre um impasse, e dois ou mais processos ficam impedidos de continuar suas execuções. Essa situação não ocorre em data warehouses, pois eles suportam o acesso simultâneo. Os dados vêm de um ambiente operacional e, depois de carregados no DW, podem ser consultados sem necessidade de nenhum tipo de bloqueio por concorrência de usuários no seu acesso.

Gabarito: Correto

25. (CESPE / Petrobrás - 2022) No processo de preparação de dados para BI, um dado com incompletude é normalmente aquele cujo valor está fora do domínio do atributo.

Comentários:



Um dado que está fora do domínio do atributo (ou que é discrepante em relação aos outros dados) trata-se de um dado inconsistente, e não de um dado incompleto. Um dado incompleto é aquele em que faltam valores de atributos.

Gabarito: Errado

26. (CESPE / SEFAZ-AL – 2021) Um armazém de dados contém um subconjunto de informações obtidas de um repositório de dados, para atender às necessidades de uma unidade de negócios em uma empresa.

Comentários:

Aqui precisamos falar um pouco sobre tradução:

- Armazém de Dados = Data Warehouse;
- Repositório de Dados = Data Mart;

Infelizmente, as bibliografias não são rigorosas em relação a essas traduções e poderemos vê-las sendo utilizadas de forma indiscriminada. De todo modo, a redação correta da questão seria: um repositório de dados contém um subconjunto de informações obtidas de um armazém de dados, para atender às necessidades de uma unidade de negócios em uma empresa. Lembrem-se que os Data Marts são especializados em algum assunto ou área específicos.

Gabarito: Errado

QUESTIONÁRIO DE REVISÃO E APERFEIÇOAMENTO

A ideia do questionário é elevar o nível da sua compreensão no assunto e, ao mesmo tempo, proporcionar uma outra forma de revisão de pontos importantes do conteúdo, a partir de perguntas que exigem respostas subjetivas.

São questões um pouco mais desafiadoras, porque a redação de seu enunciado não ajuda na sua resolução, como ocorre nas clássicas questões objetivas.



O objetivo é que você realize uma auto explicação mental de alguns pontos do conteúdo, para consolidar melhor o que aprendeu ;)

Além disso, as questões objetivas, em regra, abordam pontos isolados de um dado assunto. Assim, ao resolver várias questões objetivas, o candidato acaba memorizando pontos isolados do conteúdo, mas muitas vezes acaba não entendendo como esses pontos se conectam.

Assim, no questionário, buscaremos trazer também situações que ajudem você a conectar melhor os diversos pontos do conteúdo, na medida do possível.

É importante frisar que não estamos adentrando em um nível de profundidade maior que o exigido na sua prova, mas apenas permitindo que você compreenda melhor o assunto de modo a facilitar a resolução de questões objetivas típicas de concursos, ok?

Nosso compromisso é proporcionar a você uma revisão de alto nível!

Vamos ao nosso questionário:

Perguntas

1. O que é Business Intelligence?
2. O que é um Data Warehouse (DW)?
3. O que é um Data Mart?
4. O que é OLTP (Online Transaction Processing)?
5. O que é OLAP (Online Analytical Processing)?
6. O que é Self-Service BI?
7. O que é Governança de Dados?
8. O que é Curadoria de Dados?
9. O que é a orientação por assunto em um DW?
10. O que é a integração em um DW?



11. O que significa que um DW tem uma variação no tempo?
12. O que significa a não volatilidade em um DW?
13. O que é a granularidade em um DW?
14. O que significa a credibilidade dos dados em um DW?
15. O que é um Data Warehouse Empresarial?
16. O que é Armazenamento de Dados Operacionais (ODS - Operational Data Storage)?
17. O que é a Arquitetura em Camadas de DW?
18. O que é a Arquitetura de Data Marts Independentes?
19. O que é a Arquitetura de Barramento de Data Marts?
20. O que é a Arquitetura Hub-and-spoke?
21. O que é a Arquitetura de DW Centralizado?
22. O que é a Arquitetura de DW Federado?
23. O que é a abordagem top-down para DW?
24. O que é a abordagem bottom-up para DW?
25. O que é Modelagem Dimensional?
26. O que é uma Dimensão em um DW?
27. O que é um Fato em um DW?
28. O que é o Esquema em Estrela?
29. O que é o Esquema Floco de Neve?
30. O que é o Esquema em Constelação?



Perguntas e Respostas

1. O que é Business Intelligence?

Resposta: Business Intelligence é o campo de análise de dados que visa fornecer suporte à tomada de decisões nos negócios. Ele envolve a coleta, integração, análise e apresentação de dados de negócios.

2. O que é um Data Warehouse (DW)?

Resposta: Um Data Warehouse é um sistema de armazenamento de dados projetado para análise e relatórios, em vez de processamento de transações. Ele armazena dados históricos e consolidados.

3. O que é um Data Mart?

Resposta: Um Data Mart é um subconjunto de um Data Warehouse que é orientado a um tema ou área de negócios específica, como vendas ou marketing.

4. O que é OLTP (Online Transaction Processing)?

Resposta: OLTP é uma categoria de sistemas de processamento de dados que facilitam e gerenciam aplicações transacionais para dados de entrada e recuperação de transações.

5. O que é OLAP (Online Analytical Processing)?

Resposta: OLAP é uma categoria de software que permite aos usuários analisar facilmente informações de várias dimensões do banco de dados.

6. O que é Self-Service BI?

Resposta: Self-Service BI são práticas e ferramentas de BI que permitem aos usuários finais acessar e trabalhar com dados corporativos mesmo sem a ajuda do departamento de TI.

7. O que é Governança de Dados?

Resposta: A Governança de Dados é o conjunto de processos que garantem a qualidade, disponibilidade, integridade, segurança e usabilidade dos dados gerenciados em uma organização.

8. O que é Curadoria de Dados?

Resposta: A Curadoria de Dados é o processo de gestão de dados que inclui a coleta, organização, administração, proteção e manutenção dos dados.

9. O que é a orientação por assunto em um DW?

Resposta: A orientação por assunto é uma característica de um DW que se refere à organização dos dados em torno de temas ou assuntos específicos para análise.

10. O que é a integração em um DW?

Resposta: No contexto de um DW, a integração se refere à consolidação de dados de várias fontes em uma estrutura unificada.



11. O que significa que um DW tem uma variação no tempo?

Resposta: Variação no tempo significa que um DW armazena dados históricos, permitindo análises de tendências ao longo do tempo.

12. O que significa a não volatilidade em um DW?

Resposta: Não volatilidade significa que, uma vez que os dados são inseridos em um DW, eles não são alterados, o que assegura a consistência dos relatórios ao longo do tempo.

13. O que é a granularidade em um DW?

Resposta: A granularidade se refere ao nível de detalhe ou resumo contido nos dados de um DW.

14. O que significa a credibilidade dos dados em um DW?

Resposta: A credibilidade dos dados se refere à qualidade e confiabilidade dos dados em um DW.

15. O que é um Data Warehouse Empresarial?

Resposta: Um Data Warehouse Empresarial é um DW que serve toda a organização com uma estrutura de dados integrada e consolidada.

16. O que é Armazenamento de Dados Operacionais (ODS - Operational Data Storage)?

Resposta: ODS é uma arquitetura de dados que fornece armazenamento de dados de transações operacionais que são integrados e normalizados, geralmente em tempo real.

17. O que é a Arquitetura em Camadas de DW?

Resposta: A Arquitetura em Camadas de DW é uma estrutura que divide o DW em camadas de armazenamento de dados, cada uma com um propósito específico.

18. O que é a Arquitetura de Data Marts Independentes?

Resposta: Nesta arquitetura, os data marts são criados sem um DW central. Cada data mart é autônomo.

19. O que é a Arquitetura de Barramento de Data Marts?

Resposta: Nesta arquitetura, os data marts são ligados por um "barramento" comum, ou conjunto de dados, que contém elementos de dados comuns.

20. O que é a Arquitetura Hub-and-spoke?

Resposta: A arquitetura hub-and-spoke é uma estrutura onde os data marts (os "spokes") são alimentados a partir de um DW central (o "hub").

21. O que é a Arquitetura de DW Centralizado?

Resposta: A Arquitetura de DW Centralizado é uma estrutura onde todos os dados são armazenados em um único DW.

22. O que é a Arquitetura de DW Federado?

Resposta: Na Arquitetura de DW Federado, múltiplos DWs ou data marts são coordenados e parecem um único DW para os usuários.



23. O que é a abordagem top-down para DW?

Resposta: A abordagem top-down começa com uma visão abrangente e detalhada e depois a desagrega em componentes menores.

24. O que é a abordagem bottom-up para DW?

Resposta: A abordagem bottom-up começa com os componentes menores e mais detalhados e os combina em um todo mais abrangente.

25. O que é Modelagem Dimensional?

Resposta: A Modelagem Dimensional é um método de design de banco de dados que visa melhorar a legibilidade e o desempenho de bases de dados orientadas a consultas.

26. O que é uma Dimensão em um DW?

Resposta: Uma Dimensão é uma categoria de informação usada para análise em um DW, como tempo, localização ou produto.

27. O que é um Fato em um DW?

Resposta: Um Fato é uma medida quantitativa ou qualitativa registrada em uma tabela de fatos em um DW.

28. O que é o Esquema em Estrela?

Resposta: O Esquema em Estrela é um modelo de banco de dados onde uma tabela de fatos está no centro, cercada por tabelas de dimensão.

29. O que é o Esquema Floco de Neve?

Resposta: O Esquema Floco de Neve é um modelo de banco de dados onde tabelas de dimensão estão normalizadas, resultando em uma estrutura que se parece com um floco de neve.

30. O que é o Esquema em Constelação?

Resposta: O Esquema em Constelação é um modelo de banco de dados que consiste em várias tabelas de fatos que compartilham tabelas de dimensão.

LISTA DE QUESTÕES ESTRATÉGICAS

- 1. (FGV / MPE-AL – 2018)** No âmbito da utilização de bancos de dados no papel de “data warehouse”, é usualmente aceitável que as tabelas não estejam plenamente normalizadas. Assinale a opção que mostra porque essa característica, embora indesejada, é admitida:



- a) As dependências funcionais tornam-se obsoletas.
- b) As operações de insert/update não estão mais sujeitas ao controle de concorrência.
- c) Facilita os processos de ETL.
- d) Não é possível criar bancos de dados históricos normalizados.
- e) Simplifica as consultas e melhora o tempo de resposta.

2. (FGV / IBGE – 2017) Pedro foi contratado para desenvolver uma solução de integração de dados a partir de diversas fontes heterogêneas com o objetivo de apoiar a análise de informações. A solução deve ter as seguintes características: Não-volátil, Histórico, Orientado a Assunto e Variante no Tempo. Para isso, Pedro deve desenvolver um:

- a) Sistema de Informação Transacional;
- b) Data Warehouse;
- c) Otimizador de Consultas;
- d) Sistema Gerenciador de Banco de Dados;
- e) Banco de Dados Distribuído.

3. (FGV / AL-MA – 2013) Bill Inmom é considerado um dos “pais” da tecnologia de data warehouse. O autor define o termo como sendo um conjunto de _____ integrado, orientado _____, _____ no tempo e _____ que fornece suporte ao processo de tomada de decisão”.

Assinale a alternativa cujos itens completam corretamente as lacunas do fragmento acima.

- a) dados– por assunto – variável – não volátil.
- b) dados – por objetos – não variável – granular.
- c) processos– por assunto – variável – volátil.
- d) processos– por relacionamentos – muito variável–granular.
- e) processos – por relacionamentos – pouco variável –volátil.

4. (FGV / Senado Federal – 2012) Assinale a alternativa que apresenta características dos Datawarehouses.

- a) Processamento transacional, dados históricos e consultas pré-definidas simples.



- b) Processamento analítico, manipulação de pequenos volumes de dados e consultas simples.
- c) Processamento transacional, dados desnormalizados e atualizações online.
- d) Processamento analítico, dados históricos e dados normalizados.
- e) Processamento analítico, dados desnormalizados e consultas ad hoc complexas.

5. **(FGV / MEC – 2009)** Um termo está associado a uma coleção de dados orientada por assuntos, integrada, variante no tempo, que tem por objetivo dar suporte aos processos de tomada de decisão. É implementado por meio de um banco de dados contendo dados extraídos do ambiente de produção da empresa, que foram selecionados e depurados, tendo sido otimizados para processamento de consulta e não para processamento de transações. Em geral, requer a consolidação de outros recursos de dados além dos armazenados em base de dados relacionais, incluindo informações provenientes de planilhas eletrônicas e documentos textuais. Seu objetivo é fornecer uma "imagem única da realidade do negócio". De uma forma geral, são sistemas que compreendem um conjunto de programas que extraem dados do ambiente de dados operacionais da empresa, um banco de dados que os mantém, e sistemas que fornecem estes dados aos seus usuários. O termo aqui tratado define o conceito de:

- a) DataMining
- b) DataSystems
- c) DataBusiness
- d) DataProcessing
- e) DataWarehouse

6. **(FADESP / SEFA-PA – 2022)** Considerando os conceitos e características próprios de um Data Warehouse, julgue verdadeira (V) ou falsa (F) cada uma das afirmativas a seguir.

I. Os Data Warehouses existem como armazenamento persistente em vez de serem materializados por demandas.

II. Os Data Warehouses podem ser indexados para otimizar o desempenho e, caracteristicamente, oferecem suporte específico de funcionalidades.



III. Os Data Warehouses oferecem uma grande quantidade de dados integrados e são livres das restrições do ambiente transacional, permitindo uma eficiência aumentada no processamento da consulta.

A sequência correta é:

- a) I - F; II - F; III - F.
- b) I - V; II - F; III - V.
- c) I - V; II - V; III - F.
- d) I - F; II - V; III - V.
- e) I - V; II - V; III - V.

7. (CESGRANRIO / BB – 2021) Os sistemas interativos que provêm inteligência de negócio, BI ou business intelligence, em uma organização, são utilizados por seus gestores para:

- a) configuração do controle de acesso aos dados de cada transação da organização.
- b) encadeamento das atividades de um processo de trabalho da organização.
- c) exploração de dados sumarizados para compreensão e inspiração na solução de problemas.
- d) coordenação da execução de transações distribuídas.
- e) correção de dados diretamente em sistemas transacionais.

8. (VUNESP / Prefeitura de Campinas - SP – 2019) No contexto de armazéns de dados (data warehouse), a área intermediária na qual os dados coletados pelo processo de ETL são armazenados antes de serem processados e transportados para o seu destino é chamada de:

- a) cubo OLAP.
- b) dicionário de dados.
- c) staging.
- d) data vault.
- e) data mart.

9. (IADES / APEX BRASIL - 2018) No jargão empresarial moderno, *business intelligence* é o (a):



- a) mesmo que *data warehouse*, já que ambos têm a finalidade de armazenar dados e criar relatórios gerenciais.
- b) inteligência artificial dos computadores modernos.
- c) inteligência da empresa que tem por base os sistemas modernos de informatização.
- d) processo de recolhimento e tratamento de informações que apoiarão a gestão de um negócio.
- e) conjunto de relatórios preparados pelos executivos de uma empresa.

10. (GESTÃO CONCURSO / EMATER/MG – 2018) Sobre os dados do Data Warehouse, avalie as afirmações a seguir.

- I. São integrados.
- II. São orientados por assunto.
- III. Contêm poucos níveis de detalhes.
- IV. Contêm dados históricos de um período curto.
- V. São dados capturados em um determinado momento.

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

11. (CESGRANRIO / CEF – 2021) As resoluções assumidas por um gestor dependem, fundamentalmente, da consolidação de dados e informações que sustentam o processo de tomada de decisão. Assim sendo, ferramentas que têm por objetivo organizar e apresentar dados e informações relevantes ao processo de tomada de decisão são denominadas:

- a) Codecs.
- b) Dashboards.
- c) Hardening.

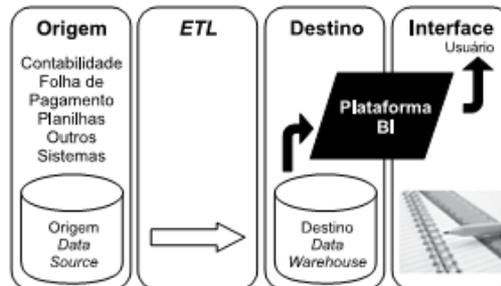


- d) Weblogs.
- e) LMS (Learning Management Systems).

12. (VUNESP / Prefeitura de Campinas - SP – 2019) No contexto de armazéns de dados (data warehouse), a área intermediária na qual os dados coletados pelo processo de ETL são armazenados antes de serem processados e transportados para o seu destino é chamada de:

- a) cubo OLAP.
- b) dicionário de dados.
- c) staging.
- d) data vault.
- e) data mart.

13. (VUNESP / MPE-SP - 2016) Observe o esquema a seguir, que representa a estrutura geral e os componentes de uma solução de Business Intelligence – BI (Inteligência Empresarial ou Inteligência de Negócios).



Sobre esse esquema, é correto afirmar que:

- a) a camada “origem de dados” é o ambiente operacional onde se encontram os aplicativos de gestão, com alto grau de interação sistêmica entre eles (os aplicativos), e é onde os consumidores e fornecedores interagem diretamente com os sistemas administrativos.
- b) na camada ETL (do inglês *Extract Transform Load* – Extração, Transformação e Carga), os dados extraídos dos aplicativos de gestão são inseridos diretamente no banco de dados.
- c) no centro dessa arquitetura, está o *data warehouse* (banco de dados), que além de conter dados granulares integrados, é considerado o “coração” de informações da fábrica.



d) na última camada, à direita, está a plataforma de BI, representando um conjunto de ferramentas que apenas disponibiliza dados para consulta e processamento pelo usuário, com alta disponibilidade e grande número de acessos.

e) a estrutura geral e os componentes que compõem uma solução de BI não necessariamente demandam que cada um de seus componentes e o fluxo de informações corporativas sejam delineados de forma lógica. Basta apenas as informações estarem organizadas no *data warehouse* (banco de dados).

14. (VUNESP / PREF SÃO JOSÉ DOS CAMPOS - 2015) Na modelagem multidimensional, um de seus componentes é a tabela fato, sobre a qual é correto afirmar que, em sua composição típica:

- a) contém um mínimo de 10% dos valores de seus atributos preenchidos com o valor nulo.
- b) contém valores numéricos que representam o desempenho de um negócio.
- c) não admite atributos do tipo literal ou numérico.
- d) não contém chaves primárias nem chaves estrangeiras.
- e) não pode conter mais de 10 atributos do tipo numérico.

15. (VUNESP / TCE SP - 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

16. (VUNESP / TJ-PA – 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.



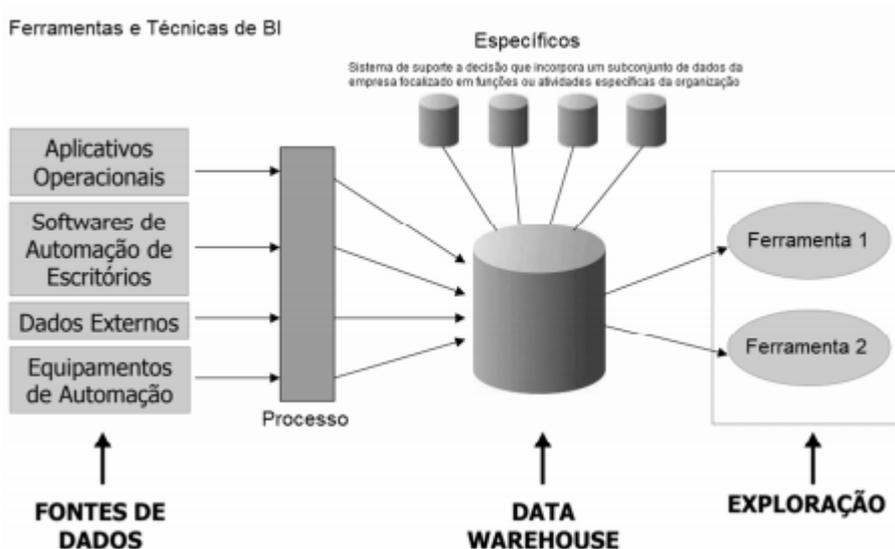
17.(VUNESP / TJ-SP - 2012) Na especificação de uma tabela fato de um modelo multidimensional de um data warehouse, é importante definir o grau de detalhamento de seus valores. A denominação utilizada para tal detalhamento é:

- a) backroom.
- b) cesta de mercado.
- c) granularidade.
- d) parcelamento.
- e) snowflake.

18.(VUNESP - 2014 - TJ-PA - Analista Judiciário - Análise de Sistema - Desenvolvimento) 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.

19.(FCC / SANASA – 2019) Atenção: Para responder à questão, considere a imagem abaixo.



O sistema de suporte a decisão representado em cada um dos cilindros do conjunto denominado Específicos, na imagem, é um:

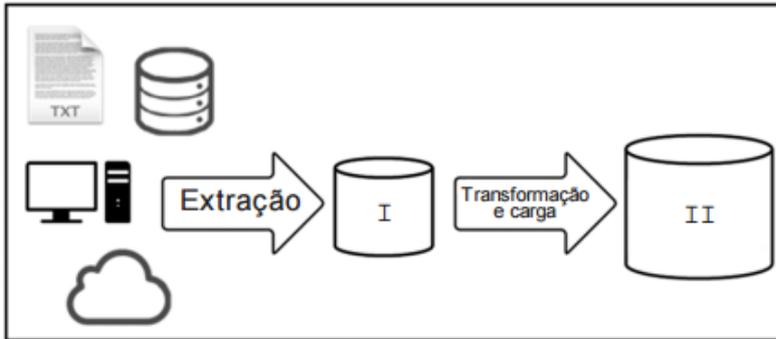
- a) Catálogo de Metadados.
- b) Schema.
- c) Drill.
- d) OLTP.
- e) Data Mart.

20. (FCC / SEFAZ-BA – 2019) Nos sistemas transacionais, os dados sofrem diversas alterações como inclusão, alteração e exclusão. Antes de serem carregados no ambiente de um Data Warehouse, os dados são filtrados e limpos, de forma a gerarem informação útil. Após esta etapa, esses dados:

- a) ficam disponíveis para a mineração em tempo real, pois tais dados são constantemente atualizados a partir da chave de tempo que indica o dia em que foram extraídos dos sistemas transacionais.
- b) podem sofrer operações de consulta, mas, devido a sua não volatilidade, não podem ser alterados, não havendo necessidade de bloqueio por concorrência de usuários ao seu acesso.
- c) são reunidos a partir de diversas fontes de dados, o que facilita muito o trabalho do analista, embora este tenha que lidar com a grande redundância das informações.
- d) ficam ordenados pela data da extração do sistema transacional, sendo necessárias técnicas de data mining para fazer a sua recuperação orientada por assunto.
- e) são classificados somente pelo assunto principal de interesse da organização. Por exemplo, em uma organização de arrecadação de impostos, os dados são organizados pelo cadastro de contribuintes que possuem impostos a recolher.

21. (FCC / AL - MS – 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.

22. (CESPE / CNMP - 2023) Em data warehouse, o conceito de granularidade refere-se ao nível de detalhe ou resumo existente em uma unidade de dados, de forma que, quanto mais detalhes, mais alto o nível de granularidade.

23. (CESPE / CNMP - 2023) Fatos, dimensões e medidas são elementos essenciais de um data warehouse.

24. (CESPE / DPDF – 2022) Um ambiente de data warehouse permite acessos simultâneos ao mesmo registro, para consulta e(ou) atualização, sem a preocupação com bloqueios de deadlock.

25. (CESPE / Petrobrás - 2022) No processo de preparação de dados para BI, um dado com incompletude é normalmente aquele cujo valor está fora do domínio do atributo.

26. (CESPE / SEFAZ-AL – 2021) Um armazém de dados contém um subconjunto de informações obtidas de um repositório de dados, para atender às necessidades de uma unidade de negócios em uma empresa.



Gabaritos

1. E
2. B
3. A
4. E
5. E
6. E
7. C
8. C
9. D
10. B
11. B
12. C
13. C
14. B
15. C
16. C
17. C
18. B
19. E
20. B
21. E
22. E
23. C
24. C
25. E
26. E



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.