

Aula 05 - Profs Diego Carvalho e Emanuele Gouveia

Banco do Brasil (*Escriturário - Agente de
Tecnologia*) Banco de Dados - 2023

(Pós-Edital)
Autor:

Thiago Rodrigues Cavalcanti,
Erick Muzart Fonseca dos Santos,
Diego Carvalho

10 de Fevereiro de 2023

Índice

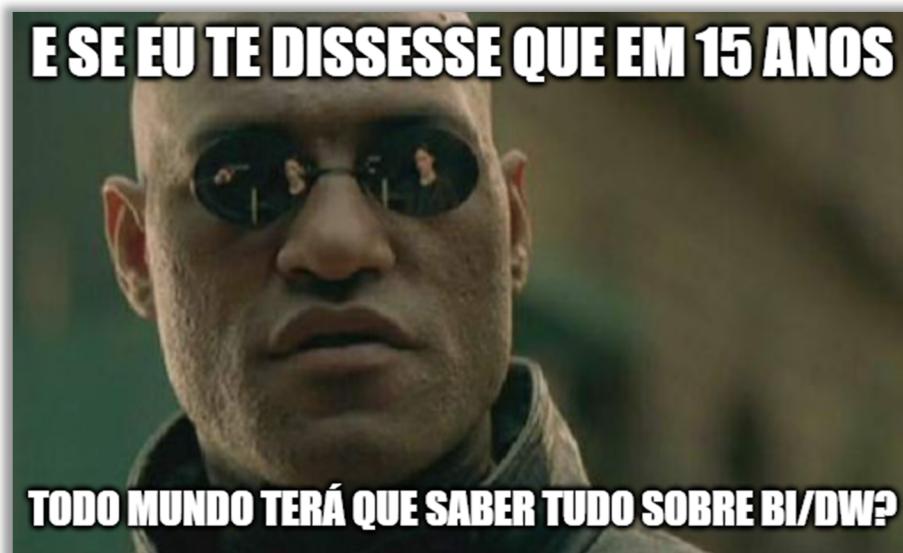
1) Análise de Informações - BI e Data Warehouse - Sistemas de Suporte à Decisão	3
2) Análise de Informações - BI e Data Warehouse - Data Warehouse (DW)	16
3) Resumo - Análise de Informações - BI e Data Warehouse	42
4) Mapa Mental - Análise de Informações - BI e Data Warehouse	45
5) Questões Comentadas - Análise de Informações - BI e Data Warehouse - Multibancas	48
6) Lista de Questões - Análise de Informações - BI e Data Warehouse - Multibancas	112



APRESENTAÇÃO DA AULA

Fala, galera! O assunto da nossa aula de hoje **Business Intelligence e Data Warehouse**. Essa aula é um pouco pesada, mas é necessário estudá-la porque ela é a base para a Mineração de Dados. Aqui veremos um novo tipo de banco de dados, como ele é construído e porque ele é utilizado. Eu juro que tentei ao máximo deixar o mais palatável possível para vocês, mas essa é uma aula que exigirá bastante atenção! Levantem, tomem um energético e venham comigo...

 PROFESSOR DIEGO CARVALHO - [WWW.INSTAGRAM.COM/PROFESSORDIEGOCARVALHO](https://www.instagram.com/professordiegocarvalho)



Galera, todos os tópicos da aula possuem Faixas de Incidência, que indicam se o assunto cai muito ou pouco em prova. Diego, se cai pouco para que colocar em aula? Cair pouco não significa que não cairá justamente na sua prova! A ideia aqui é: se você está com pouco tempo e precisa ver somente aquilo que cai mais, você pode filtrar pelas incidências média, alta e altíssima; se você tem tempo sobrando e quer ver tudo, vejam também as incidências baixas e baixíssimas. Fechado?

INCIDÊNCIA EM PROVA: BAIXÍSSIMA

INCIDÊNCIA EM PROVA: BAIXA

INCIDÊNCIA EM PROVA: MÉDIA

INCIDÊNCIA EM PROVA: ALTA

INCIDÊNCIA EM PROVA: ALTÍSSIMA

Além disso, essas faixas não são por banca – é baseado tanto na quantidade de vezes que caiu em prova independentemente da banca e também em minhas avaliações sobre cada assunto...



#ATENÇÃO

Avisos Importantes



O curso abrange todos os níveis de conhecimento...

Esse curso foi desenvolvido para ser acessível a **alunos com diversos níveis de conhecimento diferentes**. Temos alunos mais avançados que têm conhecimento prévio ou têm facilidade com o assunto. Por outro lado, temos alunos iniciantes, que nunca tiveram contato com a matéria ou até mesmo que têm trauma dessa disciplina. A ideia aqui é tentar atingir ambos os públicos - iniciantes e avançados - da melhor maneira possível..



Por que estou enfatizando isso?

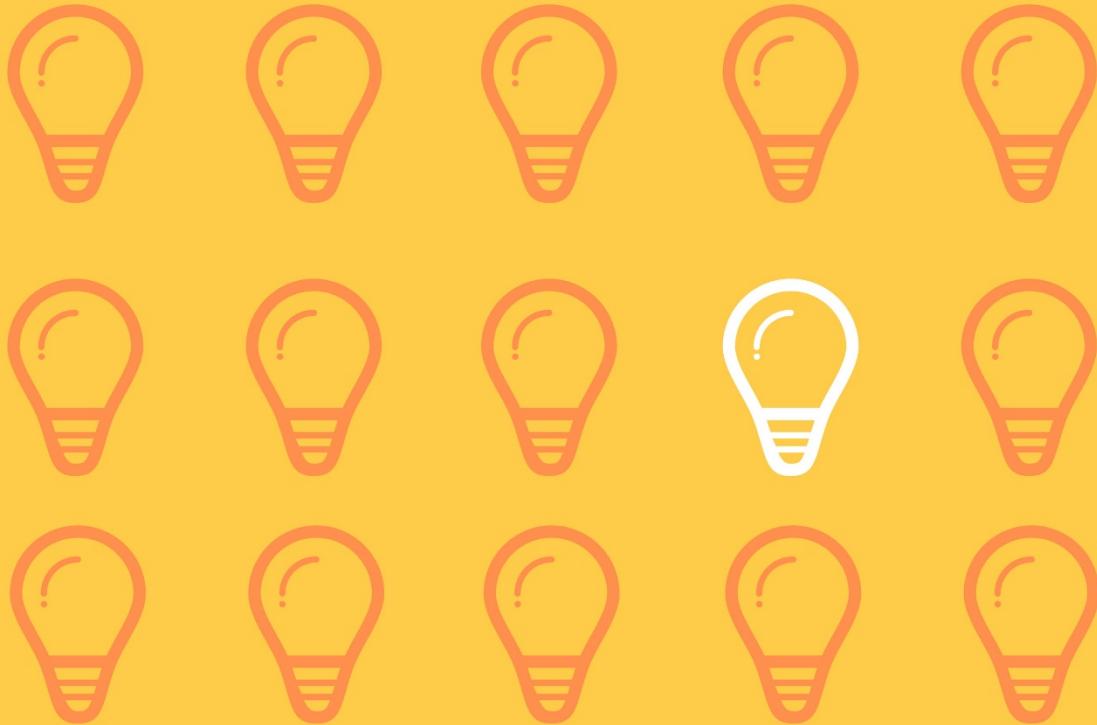
O **material completo** é composto de muitas histórias pessoais, exemplos, metáforas, piadas, memes, questões, desafios, esquemas, diagramas, imagens, entre outros. Já o **material simplificado** possui exatamente o mesmo núcleo do material completo, mas ele é menor e mais objetivo. *Professor, eu devo estudar por qual material?* Se você quiser se aprofundar nos assuntos ou tem dificuldade com a matéria, necessitando de um material mais passo-a-passo, utilize o material completo. Se você não quer se aprofundar nos assuntos ou tem facilidade com a matéria, necessitando de um material mais direto ao ponto, utilize o material simplificado.



Por fim...

O curso contém diversas questões espalhadas em meio à teoria. Essas questões possuem um comentário mais simplificado porque **têm o único objetivo de apresentar ao aluno como bancas de concurso cobram o assunto previamente administrado**. A imensa maioria das questões para que o aluno avalie seus conhecimentos sobre a matéria estão dispostas ao final da aula na lista de exercícios e **possuem comentários bem mais abrangentes**.





• ATENÇÃO •

Existem muitos exercícios sobre esse tema em sites de questões, no entanto a imensa maioria foi aplicada em provas para cargos específicos de Tecnologia da Informação (TI), os quais podem demandar um conhecimento muito mais aprofundado da matéria.

Dessa forma, recomendo que vocês tenham muita atenção na seleção das questões realizadas para que não extrapolhem o nível cobrado na sua prova.

Qualquer dúvida, estou à disposição para maiores esclarecimentos!



SISTEMAS DE SUPORTE À DECISÃO

Conceitos Básicos

INCIDÊNCIA EM PROVA: BAIXA

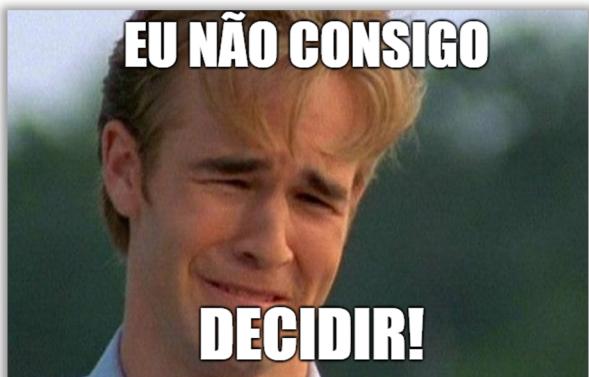
Vamos, galera... gás totaaaaaal! Hoje eu estou animado, vamos matar mais um assunto do seu edital! **Por falar nisso, o assunto da aula de hoje são os Sistemas de Suporte à Decisão!** Galera, decidir sobre algo nem sempre é fácil! Eu vou dar um exemplo: é terça-feira à noite, minha noiva diz que está morrendo de fome. Segue o diálogo apresentado ao lado...

Quem nunca ficou horas vendo as opções de restaurante no iFood? Quem nunca viu todos os cartazes de filme da Netflix tentando escolher algum para assistir com a namorada/namorado e acabou não assistindo nada? Pois é, quando se tem muitas opções, a escolha acaba se tornando mais difícil. Se isso vale para escolhas simples, imagine para grandes decisões!



Pessoal, essas decisões que eu mencionei ocorrem várias vezes durante nossa vida e não têm nenhuma repercussão mais grave. Assistir um filme chato não vai me fazer parar de ver filmes no futuro. Comer algo que eu não quero também não me causará nenhum dano (e ainda ajuda nos relacionamentos). No entanto, **há algumas decisões que ocorrem basicamente uma ou, no máximo, duas vezes na vida e geralmente podem causar grandes impactos em sua vida.**

Dá um exemplo, professor? **Um exemplo clássico é sua carreira! Quando você tem apenas dezessete anos, você tem que decidir qual será sua carreira** (por vezes, isso valerá para o resto da vida). Quando eu fiz vestibular, eu fiquei dias e dias pensando em qual curso escolher: Ciência da Computação ou Economia. *Eu poderia desistir no meio do curso e fazer um novo vestibular?* Sim, mas – em muitos casos – isso pode não ser viável. Concordam?

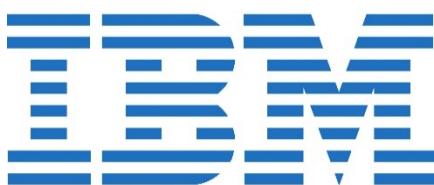


Outro exemplo: eu sou um cara meio enrolado! Eu precisei namorar oito anos para finalmente tomar coragem e pedir minha namorada em casamento. **Essa é uma decisão que se faz uma ou duas vezes na vida** – exceto se você for a Gretchen ou o Vinicius de Moraes! *E se vocês passarem para um concurso em um local muito distante de onde vocês moram? Vão se mudar e deixar amigos e parentes? E se dois concursos importantes terão provas no mesmo dia?*



Uma particularidade de todos esses exemplos é que eles trataram de decisões individuais! No entanto, o que aconteceria se estivéssemos falando de uma organização – seja ela pública ou privada? Pois é, decisões individuais afetam apenas o indivíduo, mas decisões organizacionais podem afetar centenas, milhares ou até bilhões de pessoas. Vocês querem saber de um ótimo exemplo? A assinatura ilimitada!

Vocês já pararam para pensar quanto os sócios-fundadores do Estratégia Concursos devem ter pensado antes de decidirem migrar para um plano de negócio baseado em assinatura ilimitada? Pois é, esse é um tipo de decisão importantíssima para uma organização que pode levá-la a outro patamar ou pode levá-la à ruína. Aliás, o mundo está repleto de exemplos de como decisões erradas podem levar a prejuízos colossais. Vejam só...



- Decisão:
Comprar um sistema operacional da Microsoft para seus computadores pessoais.
- Erro:
A IBM não previu que se tornaria refém da Microsoft.
- Prejuízo:
US\$ 75.000.000.000,00

- Decisão:
Aquisição da concorrente Compaq
- Erro:
A união das empresas não funcionou e os lucros permaneceram inalterados.
- Prejuízo:
US\$ 19.000.000.000,00



- Decisão:
Investir em linhas internacionais, como a rota São Paulo a Frankfurt (Alemanha).
- Erro:
Superestimou a demanda dos usuários.
- Prejuízo:
R\$ 40.000.000,00

Pessoal, para evitar esses prejuízos surreais e ajudar a tomar melhores decisões, foram criados os Sistemas de Suporte/Apoio à Decisão. Eles tratam basicamente de sistemas que suportam o processo de tomada de decisão de organizações. Professor, ele toma decisões sozinho? Não, ele apoia, auxilia, guia um gestor – mas não substitui a intervenção humana! Na verdade, você utiliza ferramentas ou recursos que te auxiliam na tomada de decisões o tempo todo. Quer ver?

Quando você precisa decidir o que comer, você olha no relógio para verificar as horas! Se for de manhã, você provavelmente não vai comer uma feijoada. Se for no almoço, você provavelmente não vai comer apenas uma sopa. Você também verifica sua carteira ou seu saldo bancário para verificar se você tem dinheiro suficiente ou não. **Se você tem apenas cinco reais, o máximo que dá para comer é um churrasquinho na esquina ou um cachorro quente.**

Quando eu vou comer, eu olho sempre para minha barriga. Se ela estiver imensa, eu decido comer uma saladinha leve; mas se estiver tranquila, eu mando aquela rabada com agrião sem peso na



consciência. **O que eu quero mostrar é que nós já utilizamos ferramentas ou recursos para nos auxiliar na tomada de decisões.** Já organizações precisam evidentemente de ferramentas mais parrudas. Por que? Porque em geral decisões necessitam de cálculos, muitos cálculos!

Quando você vê em seu smartphone a previsão do tempo para amanhã, por trás dos panos há um supercomputador com altíssima capacidade de processamento realizando trilhões de cálculos para chegar em um resultado. **Ele analisa umidade, pressão, histórico e centenas de outras variáveis para alcançar uma previsão mais acurada.** Decisões em geral também necessitam de muitos dados – geralmente extraídos de um Data Warehouse (veremos à frente).

Querem ver um exemplo? Em 2015, a Apple lançou seu Apple Watch! Vocês sabem que há pessoas que são muito fãs de produtos da Apple e fazem filas enormes para comprá-los – **algumas vezes sem sequer saber como eles serão.** Por outro lado, a Apple também não sabe exatamente qual será a demanda pelos seus novos produtos antes de lançá-los, logo é complicado estimar quantos devem ser produzidos de antemão para que as lojas tenham estoque para todos os compradores.

Vocês já imaginaram pagar caro por um Apple Watch e só receber-lo após dois meses? Isso seria péssimo para a imagem da empresa. Além disso, deve-se levar em consideração que o lançamento ocorre por vezes em escala mundial, logo a empresa tem que estimar quantos devem ser produzidos considerando onde o produto será lançado, o poder de compra da comunidade local, o histórico de compras de outros produtos da empresa, entre outros.

Em suma, a empresa tem que decidir quantos serão fabricados antes do lançamento e qual a proporção de distribuição em cada país. Exemplo: 40% do estoque ficará nos EUA porque o poder de compra dos americanos é alto e o histórico é favorável; somente 5% será enviado ao Brasil porque o poder de compra dos brasileiros é baixo e as pessoas em geral não têm o costume de investir esse valor em um relógio. Para tomar essas decisões, são necessários muitos dados!

Ainda não acabou! Decisões necessitam de alta consistência. *Como assim, Diego? Dados os mesmos conjuntos de informações de entrada, um sistema deve obter resultados semelhantes.* Vamos imaginar – por exemplo – que você está precisando de uma grana e decide ir a um banco pedir um empréstimo. *Vocês já imaginaram se a quantidade de crédito oferecida dependesse por exemplo do humor do gerente naquele dia específico?* Seria complicado...

Sistemas são mais consistentes que humanos, portanto nós podemos programá-los para seguir diversas regras e sempre aplicar os mesmos critérios para atingir resultados consistentes. Logo, se dez pessoas com remunerações parecidas, quantidade de poupança semelhantes, idades iguais e nenhum histórico de dívidas forem a um banco pedir um empréstimo, o sistema de suporte a decisão deve sugerir valores de crédito semelhantes para as dez pessoas.

Por fim, decisões frequentemente devem ser tempestivas! Em outras palavras, quando você precisa tomar uma decisão, você tem que tomá-la rapidamente. Estou eu aqui pensando em um bom exemplo para vocês quando vejo uma notícia: **Capitania dos Portos alerta para possível ciclone**



marítimo com ventos de até 89 km/h no sul e baixo sul da Bahia. Você é o governador da Bahia agora, o que você faria?

The screenshot shows a news article from G1 BA. At the top, there's a red header bar with 'MENU', the 'G1' logo, the word 'BAHIA' in the center, and a search bar on the right. The main title of the article is 'Capitania dos Portos alerta para possível ciclone marítimo com ventos de até 89 km/h no sul e baixo sul da Bahia'. Below the title, a subtitle reads: 'Fenômeno deve atingir cidades como Itamaraju e Valença. Outras regiões do estado têm previsão de chuvas e rajadas de ventos, mas não é em decorrência do ciclone.' The author is listed as 'Por Thaic Carvalho, G1 BA' with the date '22/03/2019 14h42 · Atualizado há 3 dias'. To the right of the author's name are social media sharing icons for Facebook, Twitter, WhatsApp, LinkedIn, and Pinterest.

Sistemas de Suporte à Decisão podem auxiliar a decidir qual meio deve ser utilizado para alertar os moradores (mensagens, ligações telefônicas, avisos sonoros?); podem auxiliar a decidir para onde mandar ajuda (para cidades onde efetivamente passará o ciclone ou para cidades ao lado que possuam maior infraestrutura?); qual ajuda mandar (comida, água, remédios?). **Se você demorar alguns minutos, pessoas podem morrer! Logo, decisões devem ser rápidas!**

Enfim, galera... **as decisões que passam por esse tipo de sistema geralmente são resultado de muitos cálculos, necessitam de um volume altíssimo de dados, são altamente consistentes e devem ser tempestivas.** Legal? Bem... feita essa introdução, é importante ver uma definição formal: trata-se de um conjunto de procedimentos baseados em modelos para processamento de dados e julgamentos para ajudar o seu utilizador a tomar decisões.

(SERPRO – 2013) Sistemas de apoio à decisão são sistemas que, substituindo a intervenção humana, tomam decisão em situações críticas que envolvem o processamento de grande volume de informações em paralelo.

Comentários: na verdade, ainda é necessária a intervenção humana (Errado).

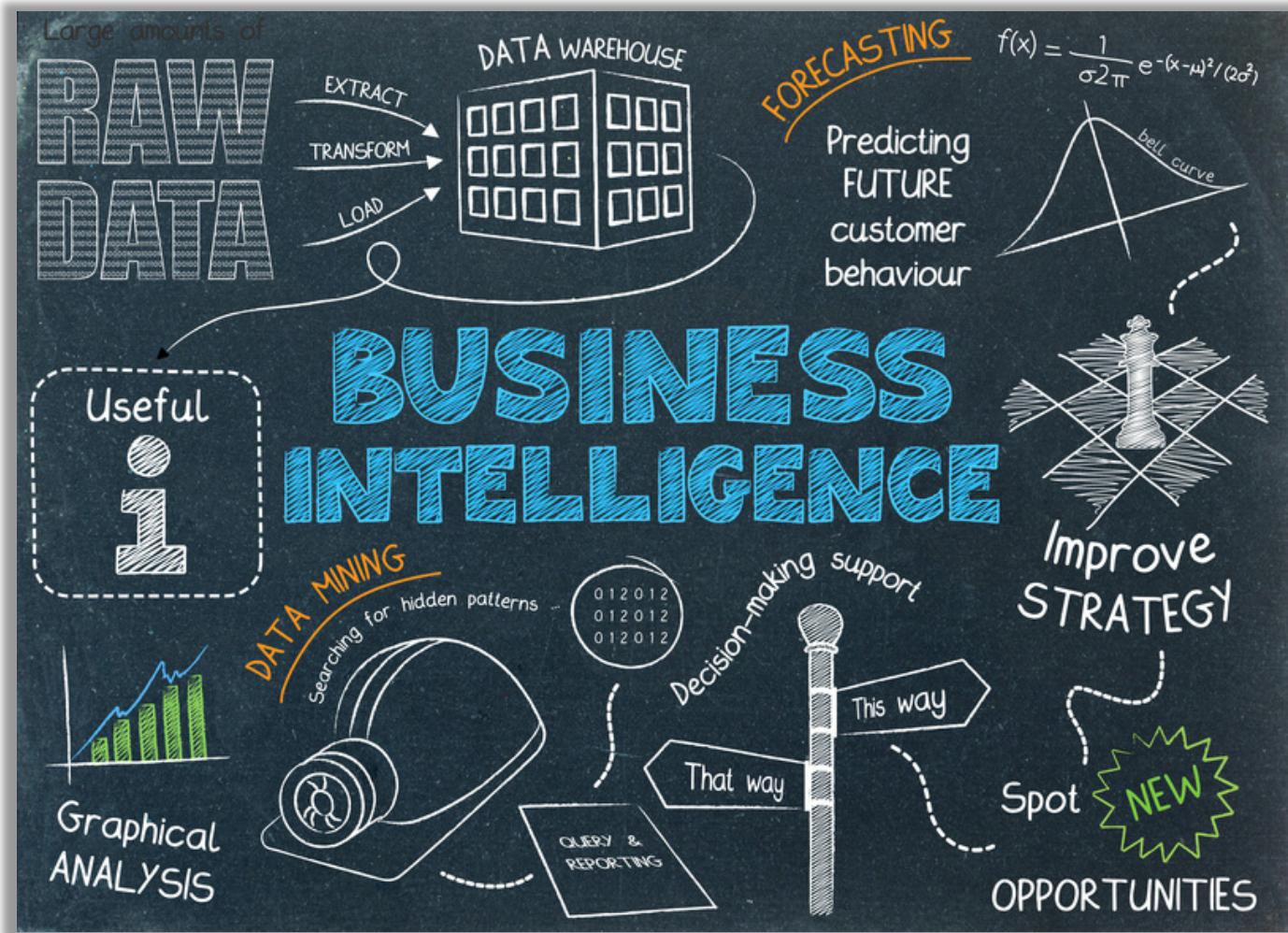
(CFO/DF – 2017) Os sistemas de suporte à decisão utilizam uma coleção de dados relativos a uma empresa. A essa coleção dá-se o nome de Data Warehouse.

Comentários: a questão está perfeita – apesar de essa coleção poder ser de outros tipos (Correto).



Business Intelligence (BI)

INCIDÊNCIA EM PROVA: ALTA



Os tempos foram passando, o mundo foi mudando e o termo Sistema de Suporte à Decisão se tornou o que chamamos hoje de Business Intelligence (BI)! O princípio é o mesmo, porém com alguns detalhes que veremos ao longo da aula. **Vamos ver na tabela a seguir diversas definições formais bastante comuns em prova – percebam como há algumas palavras-chave em comum em quase todas essas definições.**

DEFINIÇÕES

Business Intelligence é o processo de recolhimento e tratamento de informações que apoiarão a gestão de um negócio.

Palavras-chave: processo; tratamento de informações; apoio à gestão de negócios.



Business Intelligence é o processo inteligente de coleta, organização, análise, compartilhamento e monitoração de dados que, depois de processados, geram informações para o suporte e para a tomada de decisões no ambiente de negócios.

Palavras-chave: processo inteligente; coleta, organização, análise, compartilhamento e monitoração de dados; tomada de decisões.

Business Intelligence trata da capacidade de cruzar informações de diferentes bancos de dados, gerando relatórios analíticos diversos.

Palavras-chave: cruzamento de informações; diferentes bancos de dados; relatórios analíticos.

Business Intelligence se refere à aplicação de técnicas analíticas para informações sobre condições de negócio no sentido de melhorá-las, de uma maneira automatizada, mas com a interpretação e respostas humanas, de forma a melhorar a tomada de decisões.

Palavras-chave: técnica analíticas; automatizada; tomada de decisões;

Business Intelligence trata da descrição de habilidades das organizações para acessar dados e explorar informações, analisando e desenvolvendo percepções e entendimentos a seu respeito.

Palavras-chave: habilidade; explorar informações; análise de percepções e entendimentos.

Business Intelligence são técnicas, métodos e ferramentas de análise de dados para subsidiar processos de decisão de uma empresa.

Palavras-chave: técnicos, métodos e ferramentas; análise de dados; processo de decisão.

Business Intelligence nada mais é que uma série de conceitos e metodologias para auxiliar na tomada de decisões estratégicas nas empresas através principalmente de geração de relatórios gerenciais analíticos.

Palavras-chave: conceitos e metodologias; tomada de decisões estratégicas; geração de relatórios analíticos.

Business Intelligence é um termo utilizado para descrever as habilidades das corporações para coletar dados e explorar informações, analisá-las e desenvolver entendimentos para tomada de melhores decisões.

Palavras-chave: habilidades; coleta, exploração e análise de dados; tomada de decisão.

Agora que nós já sabemos tudo isso, nós podemos criar a nossa própria definição:

Business Intelligence (Inteligência de Negócio) é um termo criado pelo Gartner Group utilizado para representar um conjunto de processos, técnicas, metodologias, habilidades, ferramentas e capacidades utilizadas para acessar, coletar, organizar, tratar, analisar, cruzar, processar, compartilhar e monitorar dados de diversas fontes com o intuito de gerar informações e relatórios analíticos que suportem a gestão corporativa, a definição de estratégias e a tomada de decisão em ambientes de negócio.

O Business Intelligence é realizado sobre dados históricos consolidados, provenientes de bases multidimensionais, tais como Data Warehouses. *O que é isso, professor?* Calma, pequeno gafanhoto! Nós veremos isso tudo ainda na aula de hoje. **O que vocês precisam guardar é que o Business Intelligence é capaz de fornecer uma visão dos dados facilitando análises diagnósticas, descritivas e até preditivas para suportar a tomada de decisão¹.** Vejamos...

¹ Não há consenso na literatura sobre os tipos de análise suportadas pelo BI, mas algumas questões não consideram que análises prescritivas são suportadas por essa tecnologia.



- **Análise descritiva:** é o exame de dados ou conteúdos para responder à pergunta: “*O que aconteceu?*” ou “*o que está acontecendo?*”. É caracterizado pelo Business Intelligence tradicional e visualizações como gráficos de pizza, gráficos de barras, etc que servem para que gestores tomem decisões;
- **Análise Diagnóstica:** é uma forma de análise avançada que examina dados ou conteúdos para responder à pergunta: “*Por que isso aconteceu?*”, e é caracterizada por técnicas como detalhamento, descoberta de dados, mineração de dados e correlações. Esse é o tipo de análise empregada em investigações de causa-raiz;
- **Análise Preditiva:** combina técnicas de estatística, mineração de dados e aprendizagem de máquina (*Machine Learning*) para encontrar significado em grandes quantidades de dados, trabalhando com probabilidades, entre outros para antecipar comportamentos futuros com base em eventos passados. Responde à pergunta: “*O que vai acontecer?*”;
- **Análise Prescritiva:** é uma forma de análise avançada que examina os dados ou os conteúdos para responder à pergunta: “*O que deve ser feito?*” ou “*O que podemos fazer caso algo aconteça?*”, e é caracterizada por técnicas como análise de gráficos, simulação, redes neurais, mecanismos de recomendação, heurísticas, aprendizado de máquina, etc.

▪ **(SEFAZ/SC – 2018)** As aplicações de Business Intelligence, nas últimas décadas, se utilizam de dados multidimensionais, armazenados em Data Warehouse, para gerar visões de negócios baseados em análises:

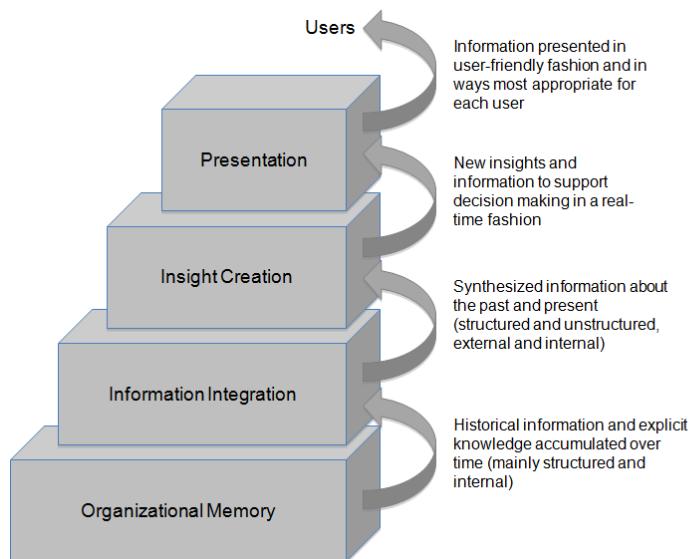
- a) descritivas, diagnósticas, preditivas e prescritivas.
- b) prescritivas, preditivas e diagnósticas, apenas.
- c) preditivas, prescritivas e descritivas, apenas.
- d) diagnósticas, descritivas e preditivas, apenas.
- e) descritivas, diagnósticas e prescritivas, apenas.

Comentários: BI é capaz de fornecer uma visão dos dados facilitando análises diagnósticas, descritivas e até preditivas para suportar a tomada de decisão, mas não prescritiva. *Por que, professor?* As bibliografias não deixam claro e não há um consenso sobre isso, mas eu acho que é porque se trata de um estágio mais avançado em que – além de inferir sobre o que acontecerá – tenta-se inferir as consequências do acontecimento - estando mais relacionado ao Data Science ou Data Mining do que com BI. Questão polêmica! (Letra D).

É interessante mencionar que um usuário que desejar utilizar os conceitos de BI precisará de uma infraestrutura arquitetural específica capaz de extrair, limpar, formatar, transformar e carregar dados estruturados ou não estruturados de diversas fontes em depósitos de informações que possam ser acessados por sistemas analíticos. **Além disso, BI possui quatro habilidades: memória organizacional, informação integrada, criação de conhecimento (*insight*) e apresentação.**



O principal recurso de Business Intelligence é a **Memória Organizacional**. A informação e o conhecimento são armazenados no sistema de BI de forma que possam ser acessados e observados posteriormente. Isso ajuda na próxima habilidade de Business Intelligence: Integração de Informações, que é a habilidade de centralizar informações de diversas fontes. Informações centralizadas em um único local ajudam na Criação de Insights.



Criar insights significa desenvolver a capacidade de ter intuições sobre o negócio para ajudar a melhorar a tomada de decisões. Por fim, a criação de insights fornece contribuições para o recurso de Apresentação. **Por meio dela, o sistema de Business Intelligence tem a habilidade de gerar relatórios e ferramentas adequados, legíveis e inteligíveis para a maioria dos usuários das ferramentas.**

O próprio Gartner Group menciona essas habilidades quando define BI: trata-se das habilidades das corporações de acessar dados e explorar as informações (normalmente contidas em um Data Warehouse ou Data Mart), analisando-as e desenvolvendo percepções e entendimentos a seu respeito – o que as permite incrementar e tornar a tomada de decisões mais pautada em informações. *Entendido?*

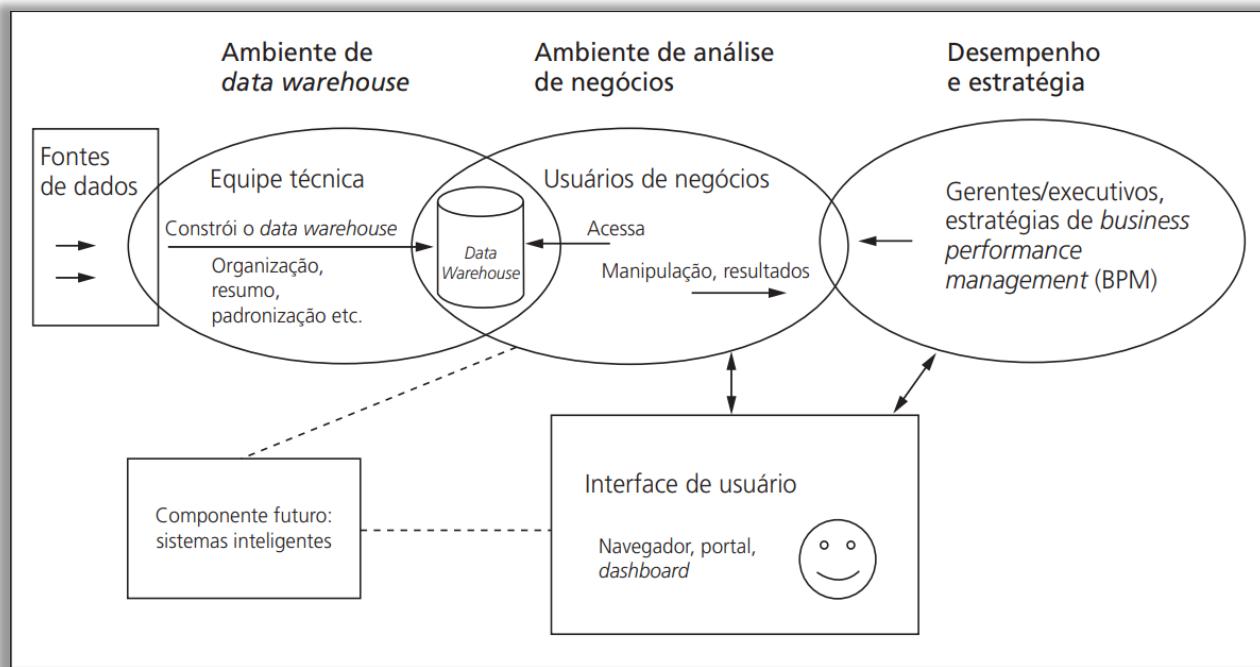
(ANTT – 2013) Uma solução de business intelligence, usualmente, provê as seguintes capacidades: memória da organização, integração da informação, criação de insights e apresentação dos dados.

Comentários: perfeito... todas essas capacidades estão diretamente relacionadas a BI (Correto).

Por fim, vamos falar sobre a Arquitetura de Business Intelligence! **De acordo com E. Turban, BI possui quatro grandes componentes:** (1) Data Warehouse, com seus dados-fonte; (2) Análise de Negócio, uma coleção de ferramentas para manipular e analisar os dados do Data Warehouse, incluindo Data Mining; (3) Business Performance Management (BPM), para monitoria e análise do desempenho; e (4) uma interface de usuário, como um *dashboard*.



Observe que o ambiente de Data Warehousing é sobretudo de responsabilidade de uma equipe técnica, e o ambiente de análise (também conhecido como análise de negócios) está no âmbito dos usuários de negócios. Qualquer usuário pode conectar-se ao sistema por meio de uma interface de usuário, como um navegador, e a alta administração pode usar o componente de BPM e também o dashboard. Vejamos uma imagem dessa arquitetura de alto nível...



COMPONENTES	DESCRÍÇÃO
DATA WAREHOUSE	Partindo do lado esquerdo da figura, vemos um fluxo de dados dos sistemas operacionais (Ex: CRM, ERP, etc) até um Data Warehouse, que é um banco ou repositório de dados de interesse dos gestores preparado para dar suporte a aplicações de tomada de decisão. As aplicações variam de simples gerações de relatórios ou consultas a complexas otimizações.
ANÁLISE DE NEGÓCIOS	Trata-se de um conjunto de ferramentas para manipular, minerar e analisar dados, criar relatórios, realizar consultas sob demanda, entre outros. Dessa forma, usuários de negócio são capazes de identificar com rapidez e facilidade as tendências de desempenho (Ex: usuários podem isolar e identificar produtos, clientes ou regiões que apresentam tendências significativas de alta ou de baixa, ou que constituem fontes de problemas).
BPM	Este componente permite otimizar o desempenho geral de uma organização por meio da conexão de métricas (Ex: informações financeiras) com desempenhos reais da organização. Para tal, utiliza-se da análise, geração de relatórios e consultas para comparar o desempenho de negócios às metas estabelecidas – além de oferecer uma plataforma para compartilhar metas de desempenho e resultados da empresa.
INTERFACE DE USUÁRIO	Esse componente permite uma visão rápida dos dados por meio de ferramentas de visualização, como dashboards, painéis, portais, gráficos, cockpits, etc. Os dashboards fornecem uma visão abrangente das medidas, tendências e exceções do desempenho corporativo provenientes de múltiplas áreas do negócio, propiciando uma visão imediata da saúde da organização.



ARQUITETURA BI

	TRATA-SE DE UMA COLEÇÃO DE DADOS ORIENTADOS POR ASSUNTO, INTEGRADOS, VARIÁVEIS COM O TEMPO E NÃO-VOLÁTEIS, PARA DAR SUPORTE AO PROCESSO DE TOMADA DE DECISÃO.
	TRATA-SE DE UM CONJUNTO DE FERRAMENTAS PARA MANIPULAR, MINERAR E ANALISAR DADOS, CRIAR RELATÓRIOS, REALIZAR CONSULTAS SOB DEMANDA, ENTRE OUTROS.
	TRATA-SE DE UM CONJUNTO DE FERRAMENTAS QUE PERMITEM COMPARAR E COMPARTILHAR O DESEMPENHO DE NEGÓCIOS DE UMA ORGANIZAÇÃO EM RELAÇÃO ÀS METAS ESTABELECIDAS.
	TRATA-SE DE UM CONJUNTO DE FERRAMENTAS QUE PERMITEM A RÁPIDA E ABRANGENTE VISUALIZAÇÃO DE DADOS POR MEIO DE DASHBOARDS E SOFTWARES SIMILARES.

(MPE-AP – 2021) Em um sistema de BI, a coleção de ferramentas utilizada como componente para manipular, minerar e analisar os dados no DW (Data Warehouse) denomina-se:

- a) OLAP (Online Analytical Processing).
- b) BPM (Business Performance Management).
- c) Análise de Negócio.
- d) Dashboard.
- e) Processamento de Transações.

Comentários: a coleção de ferramentas utilizada como componente para manipular, minerar e analisar dados de um Data Warehouse (DW) denomina-se Análise de Negócio (Letra C).

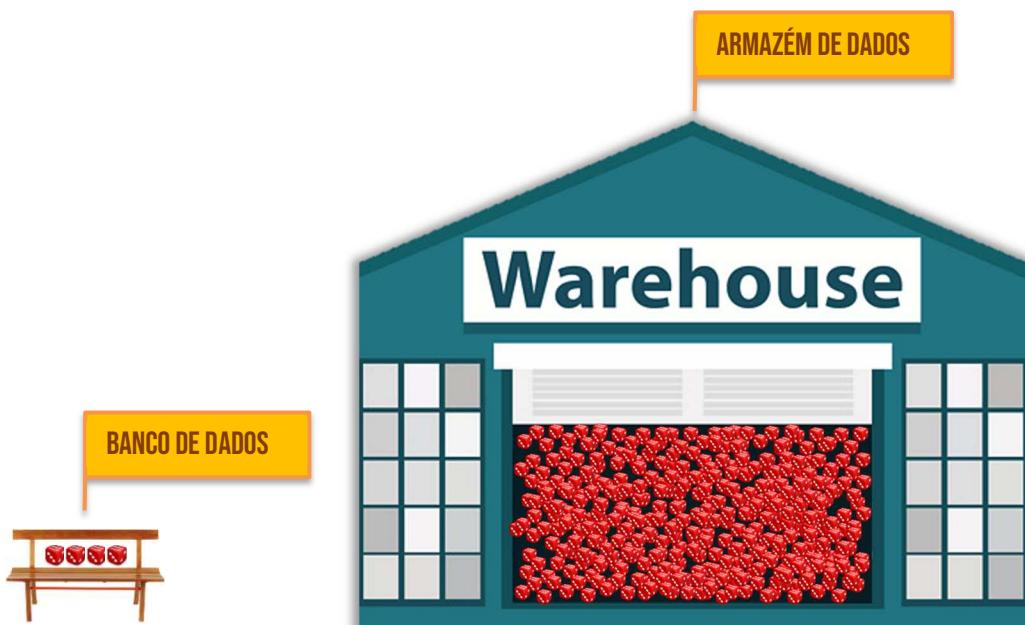


DATA WAREHOUSE (DW)

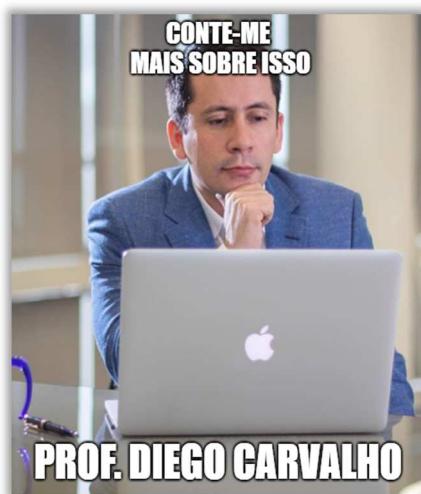
Contexto Geral

INCIDÊNCIA EM PROVA: MÉDIA

Pessoal, nós vamos iniciar agora um estudo um pouco mais aprofundado. Vocês descobrirão várias tecnologias novas, diversos nomes diferentes e – como o assunto vai se tornar bem mais técnico – eu vou ter que utilizar mais analogias e exemplos para que vocês possam absorver com maior facilidade. **Dito isso, eu vou fazer uma grande contextualização inicial para explicar o assunto da nossa aula para que vocês vejam que não é nenhum bicho de sete cabeças...**



Data Warehouse significa Armazém de Dados! Notem que é algo maior que um banco de dados... é um **armazém** de dados! Professor, para que eu vou precisar de um banco de dados enorme?



Vamos imaginar um cenário hipotético: Quatro servidores públicos decidem criar uma empresa de cursos preparatórios para concursos! São poucos funcionários, poucas vendas e todas elas podiam ser registradas em um pequeno banco de dados por meio de um sistema. *E não é que essa empresa começa a dar muito certo?* Começa, então, a surgir alunos de todo canto do país querendo comprar cursos e mais cursos. *Resultado?* O volume de vendas aumentou substancialmente – assim como os dados armazenados em nosso pequeno banco de dados. Além disso, os sócios começaram a ficar estafados de tanto trabalhar e começaram a contratar vários funcionários para trabalhar na área financeira, comercial, recursos humanos, pedagógica, coaching, etc.



Professor, esse tanto de novos funcionários também utilizavam o mesmo sistema e esse mesmo banco de dados? Sim, senhor! No entanto, os bancos de dados tradicionais possuem uma característica específica: eles são extremamente eficientes para trabalhar com transações. Vocês se lembram o que é uma transação? **Basicamente é qualquer leitura, inclusão, alteração ou exclusão de dados.** Como assim, Diego? Vamos abrir um parêntese aqui!



Galera, quando nós lançamos um curso para vocês, nós apresentamos um cronograma de aulas. Quando chega a data específica, nós inserimos a aula no banco de dados (**inclusão**); vocês entram em nosso site e acessam a aula do banco de dados (**leitura**); se eu acabo inserindo a aula errada, eu posso deletá-la do banco de dados (**exclusão**); ou se há algum erro na aula, eu posso editá-la no banco de dados (**alteração**). *Entenderam a transação?* Prosseguindo...

Os bancos de dados tradicionais são bastante eficientes e rápidos para realizar esse tipo de transação. Inclusive, eles são muito recomendados para a galera que trata diretamente com esses dados transacionais. *Adivinha quem?* Professores e Alunos. Nós – professores – que estamos inserindo, excluindo e alterando aulas o tempo todo; e vocês – alunos – que estão acessando aulas diariamente para estudar e passar em um concurso público.



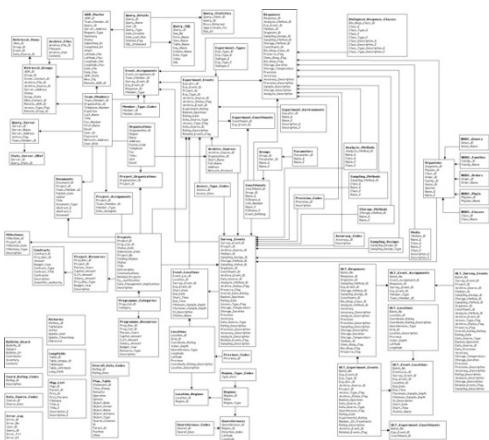
Voltando a nossa história: **vocês devem se lembrar que essa empresa de concursos começou a crescer demais e acabou contratando funcionários para diversas áreas** (Financeira, Comercial, Recursos Humanos, etc). Dito isso, eu tenho uma pergunta: **vocês acham que – para essas áreas – é mais interessante tratar de cada dado individualmente ou é mais interessante tratar mais de relatórios e consultas gerenciais?** Segunda opção!



O pessoal da **área pedagógica** precisa gerar um relatório com os professores que mais atrasam aulas ou perguntas no fórum; o pessoal da **área financeira** precisa gerar um relatório de remuneração mensal dos professores; o pessoal da **área comercial** precisa gerar um relatório de quantidade de vendas e comissões mensais; por fim, o pessoal da **área de recursos humanos** precisa gerar um relatório de cálculos trabalhistas.

Vocês conseguem entender que o intuito de gerar esses relatórios com informações consolidadas é facilitar a tomada de decisão para os gestores? Pois é, se eles possuem esse tipo de informação em mãos, eles podem tomar a decisão de – por exemplo – cobrar o professor que está atrasando as aulas; verificar quem são os professores mais bem remunerados; monitorar porque as vendas subiram ou caíram; entre outros.

No entanto, tudo tem um porém... gerar esses relatórios consolidados requer um alto poder de processamento do banco de dados! Professor, um banco de dados transacional não é capaz de gerar esses relatórios? Em geral, sim! Todavia, eu contei para vocês que essa empresa foi crescendo, crescendo, crescendo e foi se tornando uma gigante com milhões de alunos. O resultado disso foi que o banco de dados passou a ter mais de um bilhão de linhas...

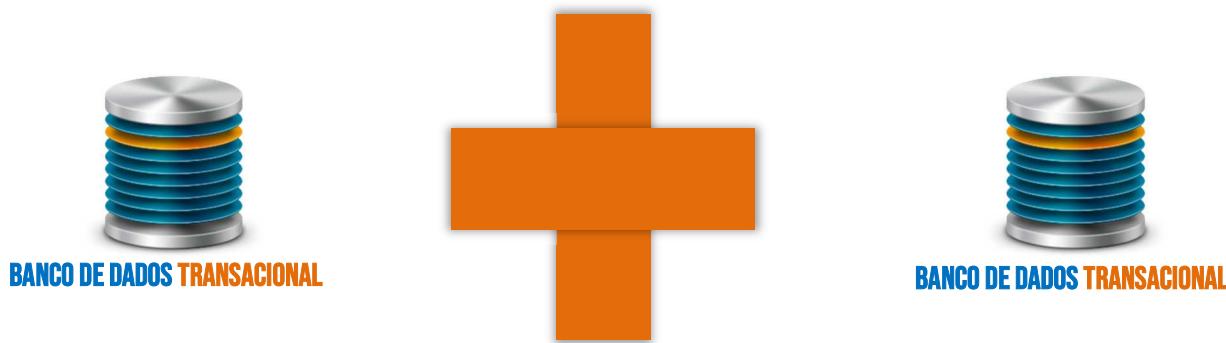


E não é só isso! **Esse banco de dados agora possui um modelo de dados gigantesco com dezenas de relacionamentos complexos entre suas tabelas.** Como esses relatórios gerenciais necessitam de um alto poder de processamento, toda vez que eles são processados, o banco de dados fica lento para todo mundo que estiver acessando. Logo, os alunos ficam irritados porque o download das aulas está mais devagar; os professores ficam furiosos porque o upload das aulas está igual a uma tartaruga; e os funcionários ficam chateados porque todo mundo está dizendo que eles são os culpados por estarem processando esses malditos relatórios gerenciais. *Que pepino!*

Podemos concluir que gerar um relatório consolidado desses não é para qualquer banco de dados tradicional – nós precisamos de mais performance! O que fazer, professor? Bem, você pode adquirir um servidor de banco de dados transacional com o dobro do processamento, porém ele provavelmente será muito caro e, se a empresa crescer mais ainda, ele logo logo também não dará conta do recado. *Concordam comigo?*

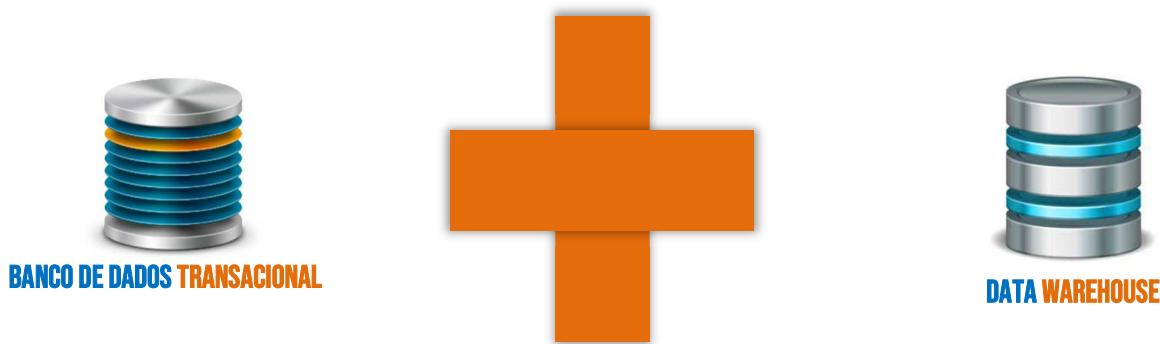


Outra alternativa interessante seria adquirir mais um servidor de banco de dados transacional, copiar os mesmos dados do banco de dados atual, porém utilizá-lo somente para processar relatórios e, não, para processar transações. Opa, essa ideia é interessante porque o custo é menor, no entanto **continuamos tendo que processar relatórios imensos sob demanda toda vez que for necessário – o que ainda pode ser muito lento**. Vejamos o esquema abaixo:



Legal, estamos nos aproximando da solução ideal: *que tal se nós tivéssemos mais um servidor de banco de dados, mas que já guardasse esses relatórios prontos em tabelas?* Ahhhh... que ideia sensacional! **Galera, relatórios são complexos de gerar porque o servidor tem que processar dados de diversos relacionamentos entre dezenas ou até centenas de tabelas diferentes do banco de dados transacional e consolidá-los em um único lugar.**

Nessa solução, essas tabelas continuariam no banco de dados transacional, mas nós teríamos mais um banco de dados um pouquinho diferente chamado: **Data Warehouse (DW)**!



O Data Warehouse deixará esses relatórios já armazenados em tabelas e prontinhos para serem consultados. *Calma, professor... para que ele deixe esses relatórios armazenados em tabelas, ele terá que fazer o mesmo processamento do banco de dados tradicional!* Isso é verdade, porém o **Data Warehouse é um banco de dados especializado em gerar relatórios, logo sua performance para esse tipo de processamento é extremamente eficiente.**

Professor, de todo modo os relatórios não terão que ser alimentados? Sim, senhor! Isso é realizado por meio de um processo chamado **ETL (Extract, Transform, Load)**. Em outras palavras, esse processo realiza três operações: **Extract** – extrai as informações do banco de dados transacional; **Transform**



– transforma as informações para o formato adequado; e **Load** – carrega as informações no Data Warehouse! *Entendido?*

Esse modelo apresenta alta performance, uma vez que os usuários estão distribuídos (alunos e professores acessam o banco de dados transacional; funcionários e gestores acessam o Data Warehouse); é mais fácil realizar consultas e analisar informações, uma vez que os dados já estarão organizados em relatórios pré-definidos; o processamento é reaproveitado, isto é, uma vez que seja realizado o ETL, os relatórios estarão processados e prontos para serem consultados.

Professor, só tem vantagens? Claro que não! **Não é fácil captar e transformar os dados de fontes diversas em algo útil para o Data Warehouse** – isso pode reduzir a relação custo/benefício; ferramentas e infraestrutura extras costumam ser necessárias, nem sempre compensando o custo de aquisição, treinamento e operação; a informação pode ficar obsoleta rapidamente; dados pouco estruturados costumam ser difíceis de se trabalhar; entre outros.

De todo modo, essa empresa era muito sinistra e continuou crescendo exponencialmente! Agora ela não é mais uma empresa – ela é uma *holding* fazendo a gestão de várias outras empresas: uma que cuida somente de concursos; outra que cuida somente de pós-graduação; outra que cuida somente de provas de certificação; outra que cuida somente de vestibulares; outra que cuida somente de um site de questões; e assim por diante.

E o número de usuários explodiu de novo porque os gestores dessas empresas também começaram a utilizar o DW! *Resultado?* Nem ele aguentou tanto acesso e começou a ficar lento de novo. *Oh, céus! De novo isso, professor?* Pois é! Nós até podemos utilizar a mesma estratégia que utilizamos na ocasião anterior, mas existe uma alternativa mais interessante: **nós podemos criar subconjuntos de um Data Warehouse**. *Como é, professor?*



Nós podemos segmentar em subconjuntos de acordo com diversos critérios, por exemplo: assuntos (Concursos, Coaching, Pós-Graduação, Certificação, Vestibulares, Questões, etc); níveis de summarização¹ (Anual, Semestral, Trimestral, Mensal, Semanal, etc); localização (Nordeste, Norte, Centro Oeste, Sudeste, Sul). **Para finalizar, esses subconjuntos de Data Warehouse (DW) são chamados de Data Mart (DM).**

¹ Granularidade é o nome que se dá ao nível de summarização dos elementos de dados de um Data Warehouse. Quanto maior a granularidade, maior será o grão e menor será o nível de detalhes; quanto menor a granularidade, menor será o grão e maior será o nível de detalhes.





Pronto, nós vimos uma visão geral sobre o assunto! Agora seremos mais objetivos, isso foi só para passar uma contextualização uma vez que esse é um assunto extremamente técnico. *Fechou?* 😊

CARACTERÍSTICAS	DADOS OPERACIONAIS	DADOS INFORMACIONAIS
CONTEÚDO	Valores correntes	Valores sumarizados, calculados, integrados de várias fontes
ORGANIZAÇÃO DOS DADOS	Por aplicação/Sistema de informação	Por assuntos/negócios
NATUREZA DOS DADOS	Dinâmica	Estática (até a atualização dos dados, de tempos em tempos)
FORMATO DAS ESTRUTURAS	Relacional, próprio para computação transacional	Dimensional, simplificado, próprio para atividades analíticas
ATUALIZAÇÃO DOS DADOS	Atualização campo a campo	Acesso granular ou agregado, normalmente sem update direto
USO	Altamente estruturado em tabelas, processamento repetitivo	Estruturado em fatos/dimensões, com processamento analítico e preditivo
TEMPO DE RESPOSTA	Otimizado para faixas abaixo de 1 segundo	Análises mais complexas, com tempos de respostas maiores



Definições Básicas

INCIDÊNCIA EM PROVA: MÉDIA

Pessoal, eu dei uma visão geral sobre o tema, mas é interessante ver como são as definições acadêmicas e como isso geralmente cai em prova. Então, vejam a tabela a seguir:

AUTOR	DEFINIÇÕES ACADÊMICAS
RALPH KIMBALL	Conjunto de ferramentas e técnicas de projeto ² que – quando aplicadas às necessidades específicas dos usuários e aos bancos de dados específicos – permitirá que planejem e construam um Data Warehouse.
BILL INMON	Coleção de dados ³ orientados por assunto, integrados, variáveis com o tempo e não-voláteis, para dar suporte ao processo de tomada de decisão.
ARUN SEN	Banco de dados construídos no interesse de suporte à decisão de negócios e contêm dados históricos sumarizados e consolidados provenientes de registros individuais de bancos de dados operacionais.
KENNETH LAUDON	Banco de dados – com ferramentas de consulta e relatório – que armazena dados atuais e históricos extraídos de vários sistemas/ambientes operacionais e consolidados para fins de análises e relatórios administrativos.

Vamos identificar as palavras-chaves que formam essas definições? Coleção de dados; ferramentas de consulta e relatório; orientados por assunto; integrados; variáveis com o tempo; não-voláteis; dados históricos sumarizados e consolidados; suporte à tomada de decisão. **Notem que nós vimos basicamente todos esses conceitos em nossa contextualização.** Fiquem tranquilos porque nós vamos ver em detalhes no tópico a seguir...

(CEGÁS – 2017) Data Warehouse é um depósito de dados digitais, orientado por assunto que transforma dados operacionais em informações voltadas ao suporte à tomada de decisões estratégicas.

Comentários: Depósito de dados? Orientado por assunto? Transforma dados operacionais em informações voltadas ao suporte à tomada de decisões estratégicas? Todas são palavras-chave do conceito de Data Warehouse (Correto).

(MEC – 2009) No que diz respeito ao desenvolvimento de sistemas de informação e da análise de negócios, analise a citação a seguir.

"É um conjunto integrado de dados, extraídos de bancos de dados operacionais, históricos, externos e selecionados, editados e padronizados para recuperação e análise, para fornecer inteligência empresarial e nos negócios no processo de tomada de decisões gerenciais". Essa citação representa o conceito de:

² Note que Ralph Kimball apresenta uma definição bastante abrangente em que trata o Data Warehouse como um conjunto de ferramentas e técnicas de projeto.

³ Esses dados podem ser estruturados ou não-estruturados. No entanto, Data Warehouses geralmente são pouco eficientes ao lidar com dados não-estruturados.



- a) Data Smart.
- b) Data Mining.
- c) Data Business.
- d) Data Warehouse.
- e) Data Information.

Comentários: Conjunto integrado de dados? Extraído de bancos de dados operacionais? Padronizados para recuperação? Fornecer inteligência empresarial? Processo de tomada de decisões gerenciais? Todas são palavras-chave do conceito de Data Warehouse (Letra D).

(MPE/AC – 2013) Considere a seguinte definição: "Uma coleção de dados não-voláteis, integrados, orientados a um determinado tema e utilizados no apoio de processos de tomada de decisão. Geralmente contém componentes complementares como facilidades de extração, mineração e transformação de informação e ferramentas analíticas." Essa é a definição de:

- a) Planilha.
- b) Middleware.
- c) Data Warehouse.
- d) DBMS (Data Base Management System).
- e) Groupware.

Comentários: Coleção de dados? Não-voláteis? Integrados? Orientados a um tema? Utilizados no apoio de processos de tomada de decisão? Facilidade de extração, mineração e transformação? Todas são palavras-chave de DW (Letra C).

(MF – 2013) Um conjunto de dados produzido para oferecer suporte à tomada de decisões, sendo um repositório de dados de interesse aos gerentes, é um:

- a) Management dataset.
- b) Data warehouse.
- c) Data center.
- d) Major warehouse.
- e) Data wareset.

Comentários: Conjunto de dados? Suporte à tomada de decisões? Repositório de dados de interesse aos gerentes? Todas são palavras-chave do conceito de Data Warehouse (Letra B).

(TJ/GO – 2006) O DW pode ser definido como um repositório de informações que armazena de maneira organizada os dados provenientes dos sites de natureza operacional da organização. Ele permite mapear o histórico da organização.

Comentários: Repositório de informações? Armazena dados de forma organizada? Dados provenientes de sites de natureza operacional? Permite mapear o histórico da organização? Todas são palavra-chave do conceito de Data Warehouse (Correto).



Características Essenciais

INCIDÊNCIA EM PROVA: ALTA

Galera, nós vimos um contexto geral e um conjunto de definições básicas. No entanto, nós passamos direto por vários termos importantes que veremos a seguir:



- **Data Warehouse é orientado por assunto:**

Um Data Warehouse é montado por assunto, isto é, ele trata de temas específicos e importantes para o negócio da organização. Por exemplo: uma empresa do ramo de concursos pode ter como assuntos principais: concursos militares, concursos fiscais, concursos policiais, concursos de tribunais, concursos bancários, concursos legislativos, entre outros. Aqui não tem muito o que falar, essa característica é bastante intuitiva.

- **Data Warehouse é não-volátil:**

Galera, vem comigo! Vocês sabem o que significa algo ser volátil? O dicionário nos diz que volátil é aquilo que muda com facilidade. Nós vimos que os dados são carregados a partir do Banco de Dados Transacional para o Data Warehouse. Vocês se lembram? O lance é que o Data Warehouse é não-volátil, logo ele não muda com facilidade. Na verdade, os dados – em tese – jamais devem mudar (serem alterados ou deletados) após inseridos.

Em outras palavras, podemos dizer que os dados ficam disponíveis apenas para que os usuários realizem consultas e façam relatórios que auxiliem a tomada de decisão por parte dos gestores de uma organização. Logo, os dados disponíveis em um Data Warehouse mostram uma



característica de somente leitura para os usuários finais de um banco de dados. *Essa primeira característica foi fácil, concordam?*

■ **Data Warehouse é integrado:**

Professor, o que diabos é isso? Galera, quando eu era desenvolvedor de sistemas, era comum desenvolver um software que – por exemplo – requisitava várias informações dos usuários em um formulário. Dentre essas informações, uma muito comum perguntava em que banco o usuário possuía uma conta corrente. Como funcionava: havia um campo em branco e o usuário preenchia livremente qual era o banco em que ele tinha conta corrente.

Qual é o grande problema disso? O problema é que um usuário escrevia “Caixa Econômica”; outro usuário escrevia “Caixa”; outro usuário escrevia “Caixa Econômica Federal”; e – por fim – outro escrevia “CEF”. **Galera, se os dados representam a mesma entidade, mas estão com nomes diferentes, podemos dizer que os dados não estão integrados.** Em outras palavras, ser integrado significa que os dados estão consistentes e uniformes.

Professor, uma maneira de resolver isso não seria obrigar o usuário a escolher o nome do banco em uma lista em vez de permitir que ele escreva? Booooo... é isso que nós fazemos! *No entanto, o que fazer com os dados que já foram cadastrados?* Além disso, **é comum que vários softwares diferentes, desenvolvidos por empresas diferentes, salvem dados em um mesmo banco de dados.** E se uma empresa escolheu o nome “Caixa Econômica Federal” e outra “CEF”? Já era...

Como vocês fazem para corrigir isso? Galera, nós limpamos a base de dados! *Limpamos, professor?* Sim, esse é o termo utilizado! **Isso significa que as inconsistências são removidas e os conflitos de informações são consertados antes de serem inseridos ou carregados em um Data Warehouse!** Dessa forma, podemos afirmar que os dados de um Data Warehouse são integrados de múltiplas fontes diferentes de forma consistente, padronizada e uniforme.

■ **Data Warehouse é histórico (variável com o tempo):**

Dados históricos são mantidos em um Data Warehouse. Por exemplo: pode-se recuperar dados de três, seis, doze meses ou até mesmo dados mais antigos de um Data Warehouse. Isso contrasta com os bancos de dados transacionais, em que geralmente apenas os dados mais recentes são mantidos. Por exemplo: um banco de dados transacional pode conter o endereço mais recente de um cliente, enquanto um DW pode conter todos os endereços associados a um cliente.

(BAHIAGÁS – 2016) Com relação aos sistemas de Data Warehouse, assinale a alternativa correta:

- a) Um Data Warehouse deve ser orientado a produto, integrado, volátil e invariável com o tempo.



- b) Um Data Warehouse armazena dados históricos usados no processo de tomada de decisão. Este sistema integra os dados corporativos de uma empresa em um único repositório.
- c) Dados de um Data Warehouse provém de uma única fonte, não sendo necessária a padronização para uniformizar nomes ou unidades de medida.
- d) Um Data Warehouse deve ser orientado a departamento, integrado, volátil e invariável com o tempo.
- e) Um Data Warehouse cria uma visão distribuída de um banco de dados centralizado. Este tipo de sistema não permite que usuários finais executem consultas sobre sua base de dados.

Comentários: (a) Errado. DW é orientado a assunto, não-volátil e variável no tempo; (b) Correto; (c) Errado. Dados provém de diversas fontes, sendo necessário a padronização para uniformizar nomes ou unidades de medidas; (d) DW deve ser orientado a assunto, não-volátil e variável no tempo; (e) Errado. Permite, sim, que usuários finais executem consultas sobre sua base de dados (Letra B).

(PETROBRÁS – 2010) Data warehouse tem por característica ser:

- a) orientado a assuntos.
b) temporário.
c) descentralizado.
d) focalizado em aspectos operacionais.
e) normalizado.

Comentários: ele é orientado a assuntos, histórico, centralizado, focalizado em aspectos táticos ou estratégicos e desnormalizado (Letra A).

(INEA/RJ – 2013) Data warehouses são definidos como:

- a) um conjunto de dados integrado, orientados por assunto, variante no tempo e não volátil.
b) um conjunto de dados integrado, orientados por departamentos, variante no tempo e não volátil.
c) um conjunto de regras, orientadas por departamentos, invariante no tempo e não volátil.
d) um conjunto de dados integrado, orientados por assunto, variante no tempo e volátil.
e) um conjunto de regras, orientadas a objeto, invariante no tempo e não volátil.

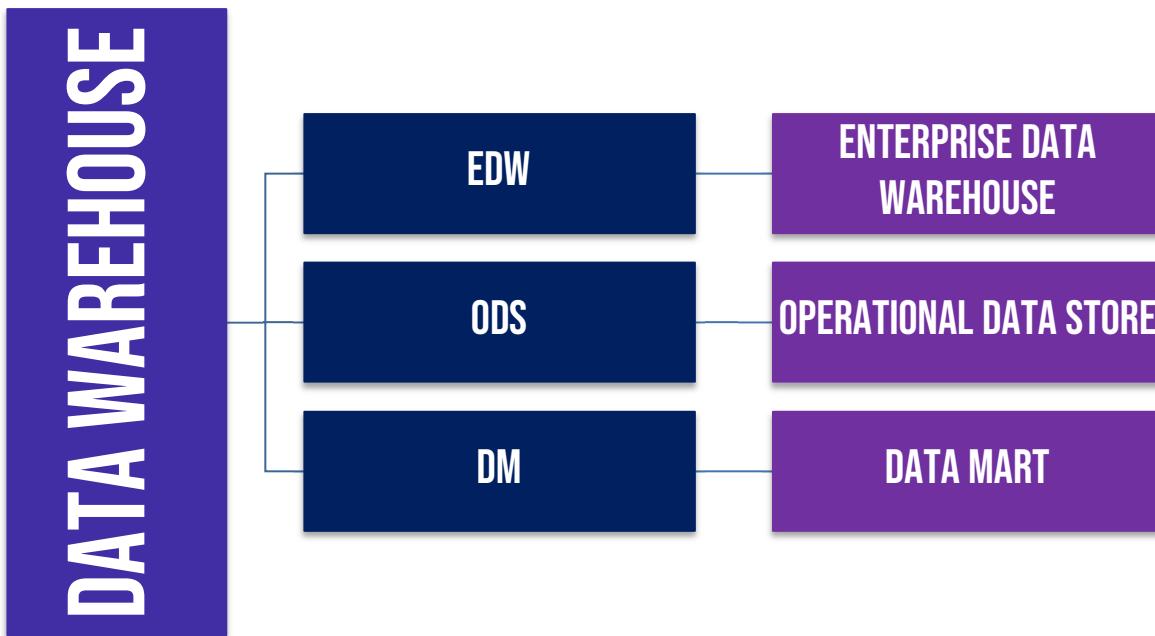
Comentários: trata-se de um conjunto de dados integrado, orientado por assuntos, variante no tempo e não volátil (Letra A).



Tipos de Data Warehouse

INCIDÊNCIA EM PROVA: MÉDIA

Galera, existem basicamente três tipos de Data Warehouse: **Enterprise Data Warehouse (EDW)**, **Operational Data Store (ODS)** e **Data Mart (DM)**! Armaria, professor... que siglas são essas?



Vamos ver a definição dos três tipos, mas não se preocupem tanto com os dois primeiros – **nosso foco aqui é o Data Mart**. Então, vejamos abaixo:

- **Enterprise Data Warehouse (EDW):**

Traduzido como **Armazém de Dados Corporativos/Abrangentes/Empresariais**, é um repositório central tático/estratégico que contém todas as informações de negócio de uma organização de forma acessível para toda empresa com a finalidade de suportar as análises de decisões. Grossso modo, é possível afirmar que um EDW é um DW de larga escala – grande e abrangente – que é utilizado pela organização como um todo e que integra informações de diversas fontes de dados.

- **Operational Data Store (ODS):**

Traduzido como **Armazenamento de Dados Operacionais**, trata-se de um repositório intermediário de dados utilizado para relatórios operacionais e como uma fonte de dados para o **Enterprise Data Warehouse (EDW)**. É um elemento complementar a um EDW em um cenário de suporte a decisões, e é usado para relatórios, controles e tomada de decisões operacionais – em oposição ao EDW, que é utilizado para suporte de decisões táticas e estratégicas.

- **Data Mart (DM):**



Traduzido como Mercado de Dados, trata-se de um subconjunto de dados de um Data Warehouse. Geralmente são dados referentes a um assunto em especial (Ex: Vendas, Estoque, Controladoria, entre outros) ou diferentes níveis de sumarização (Ex: Vendas Anual, Vendas Semestral, Vendas Mensal, entre outros), que focalizam uma ou mais áreas específicas. Seus dados são obtidos do DW de forma indexada para suportar intensas pesquisas.

O Data Mart é considerado um pequeno Data Warehouse – ou um subconjunto/segmentação de um Data Warehouse – que contém dados de uma organização em geral restritos a uma unidade de negócio ou a um processo de negócio específico com a finalidade de dar suporte ao processo decisório da alta direção por meio da criação de uma visão personalizada dos dados. **De acordo com Date, trata-se de um repositório especializado e volátil (mas é o único autor que diz isso)!**

Data Marts geralmente lidam com um único assunto, sendo frequentemente construídos e gerenciados por um único departamento comercial. Eles normalmente obtêm dados de apenas um pequeno número de fontes e – como consequência – são geralmente menos complexos, mais econômicos, mais fáceis de construir e mais fáceis de manter do que Data Warehouses. Vejamos algumas de suas principais características:

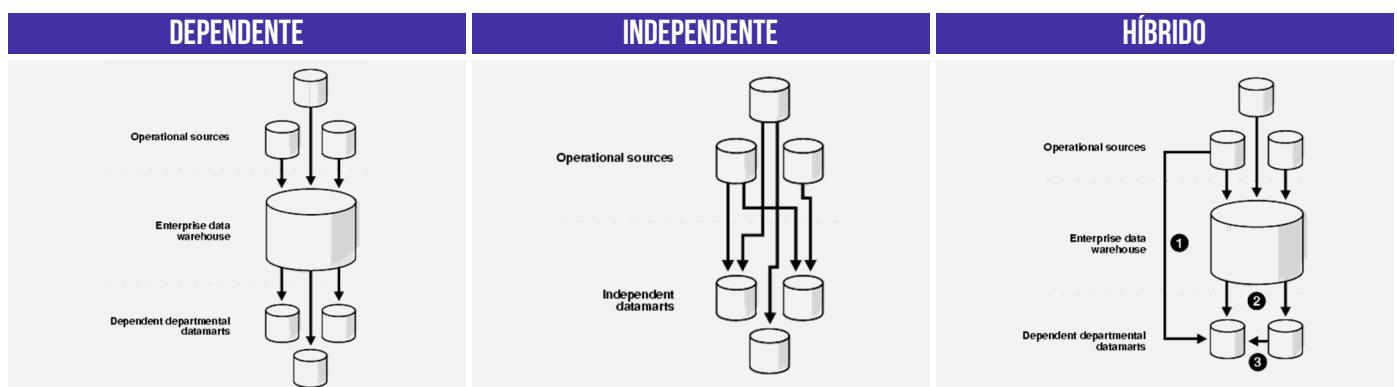
- Um Data Mart permite um acesso mais rápido aos dados, recuperando um conjunto específico de dados para BI e Relatórios. Como resultado, ajuda a acelerar os processos de negócios;
- Sendo focado em um assunto, é mais fácil implementar um Data Mart, sendo mais econômico em comparação com a construção de uma empresa Data Warehouse;
- O uso de um Data Mart é fácil, visto que ele é projetado de acordo com os requisitos de um grupo específico de usuários que trabalham em um departamento específico;
- Um Data Mart é comparativamente mais adaptável que um Data Warehouse. Qualquer alteração no modelo de dados pode ser fácil e rapidamente incorporada no Data Mart;
- Em um Data Mart, os dados são particionados e segmentados, o que permite um controle de acesso em um nível mais granular.

Os Data Marts são utilizados por uma única unidade de negócios para armazenar suas informações. **Digamos que o departamento de vendas de uma empresa tenha que armazenar seus dados de negócios – eles poderão utilizar um Data Mart para armazenar seus dados.** Quando as informações precisam ser visualizadas pelos superiores, elas são carregadas em um Data Warehouse e são, então, transformadas em insights usando software de Business Intelligence.

TIPOS DE DATA MART	DESCRIÇÃO
DEPENDENTE	Nesse caso, um conjunto de fontes de dados operacionais são integradas em um EDW, que – por sua vez – são divididos em diversos Data Marts dependentes.



INDEPENDENTE	Nesse caso, um conjunto de fontes de dados operacionais permitem gerar um ou mais Data Marts independentes, sem a necessidade de criar um EDW.
HÍBRIDO	Nesse caso, um conjunto de fontes de dados operacionais podem ser tanto integradas em um EDW para posteriormente permitir a criação de Data Marts quanto podem gerá-los de maneira independente.



(TRT/11 – 2012) No âmbito dos DWs, representa um armazenamento intermediário que facilita a integração dos dados de ambiente operativo antes da sua atualização no DW. Trata-se de:

- a) ODS b) ETL c) Data Mart d) Star Schema e) Fact Table

Comentários: quem representa o armazenamento intermediário que facilita a integração dos dados de ambiente operativo antes de sua atualização no Data Warehouse é o ODS (Letra A).

(TCE/PA – 2016) Um Data Mart é:

- a) uma técnica de mineração de dados para a extração de conhecimento das bases de dados de uma organização.
- b) um grande data warehouse que consolida todas as informações de uma organização a serem tratadas e organizadas.
- c) um sistema para a monitoração e processamento das funções básicas e rotineiras de uma organização.
- d) um conjunto de metadados para a análise e monitoração de um data warehouse organizacional.
- e) um pequeno data warehouse que dá suporte à decisão para uma específica unidade de negócio.

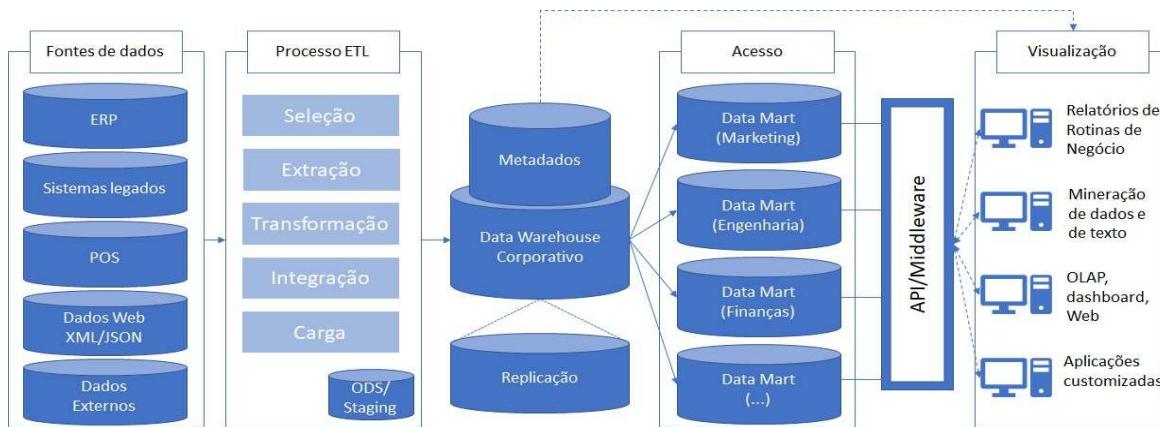
Comentários: é um pequeno Data Warehouse que dá suporte à decisão para uma específica unidade de negócio (Letra E).



Processo de Data Warehouse

INCIDÊNCIA EM PROVA: BAIXA

Galera, nós já vimos componentes suficientes para finalmente entender como funciona o processo de um Data Warehouse. Vejam a figura abaixo:



Vamos por partes: à esquerda, temos os Data Sources! *O que é isso, Diego?* **Cara, essas são as fontes de dados.** Vocês se lembram que nós vimos que um DW possui dados de diversas fontes distintas? Pois é, essa imagem mostra alguns exemplos! Notem que uma fonte de dados pode ser um ERP (Enterprise Resource Planning), que nada mais é que um sistema de informação que integra dados e processos de uma organização. *O que mais?*

Sistemas Legados, que são aqueles sistemas beeeem antigos de uma organização (em geral, com tecnologias ultrapassadas) que ainda fornecem serviços de negócio essenciais. *Eles também servem de fontes de dados para um Data Warehouse?* Claro, uma fonte importantíssima! Point of Sale (POS), que **são máquinas registradoras ou de cartão de crédito que realizam transações de venda – em determinados contextos, também podem ser importantes fontes de dados.**

Outra fonte de dados são aquelas advindas de aplicações web e bancos de dados transacionais – além de quaisquer outras fontes externas de dados. Pessoal, esses são só exemplos típicos de fontes de dados, mas podem existir dezenas de outras. **De todo modo, a imagem quer mostrar que diversas fontes de dados servem de entrada para um Enterprise Data Warehouse (EDW).** No entanto, notem que – antes de serem carregados – eles devem passar pelo processo de ETL.

Vocês se lembram que se trata do processo de ETL – Extract, Transform e Load (Extração, Transformação e Carga)? Essas são as atividades principais, porém existem outras subdivisões: primeiro, selecionam-se os dados que se deseja carregar; segundo, extraí esses dados de suas fontes para uma Stage Area, **que nada mais é que a localização temporária de onde os dados dos sistemas de origem são copiados.**

Em seguida, os dados são transformados, isto é, uma série de regras ou funções são aplicadas aos dados extraídos para prepará-lo para carregar em seu destino final; depois os dados são integrados,



ou seja, são uniformizados em um padrão específico; por fim, os dados são finalmente carregados no nosso armazém de dados empresarial. **Agora notem que existem ainda dois conceitos: metadados e replicação.**

Metadados são dados sobre o próprio Enterprise Data Warehouse, por exemplo: estruturas e atributos de tabelas, especificações do modelo de dados, rotinas de acesso, log de extrações, entre outros. **Notem que os metadados auxiliam na criação da visualização dos dados por meio de diversas ferramentas.** Para finalizar, Replicação é uma cópia que pode ser realizada do próprio Enterprise Data Warehouse para diversas funcionalidades.

Agora notem que ele pode ser dividido em vários Data Marts: Marketing, Engenharia, Finanças, entre outros. Para finalizar, notem que usuários e aplicações utilizam middlewares para acessar Data Marts ou Data Warehouses – eles não acessam os dados diretamente. Fim! **A partir daí, os dados estarão disponíveis para serem analisados e você poderá descobrir informações relevantes para o negócio de uma organização.**

(SUFRAMA – 2014) As estruturas e os atributos das tabelas, a especificação do modelo de dados, as rotinas comuns de acesso a dados e o logging de extrações compõem os metadados de um *datawarehouse*.

Comentários: todos são exemplos de metadados, isto é, dados sobre os dados (Correto).

(Receita Federal – 2012) São componentes principais de um processo de data warehousing:

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

Comentários: basta lembrar da nossa figura – os componentes são as Fontes de Dados (*Data Sources*), Extração de Dados (*Extract*), Carregamento de Dados (*Load*), Banco de Dados Abrangente (*Enterprise Data Warehouse*), metadados (*Metadata*) e Ferramentas de Middleware (Letra A).



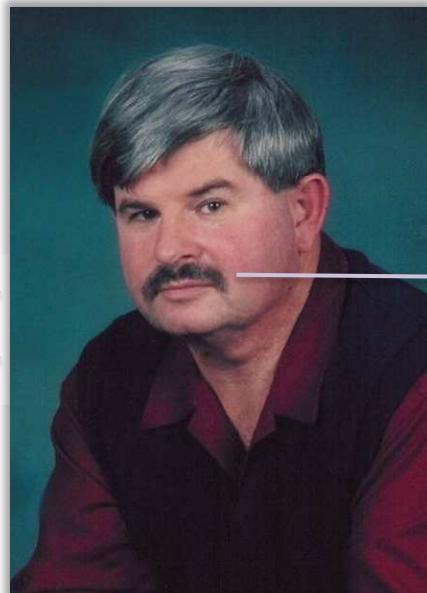
Kimball x Inmon

INCIDÊNCIA EM PROVA: MÉDIA

Pessoal, deixa eu contar uma coisa para vocês: quando o assunto é Data Warehouse, não tem para ninguém – existem dois nomes que comandam a bibliografia! Ei-los...



R. KIMBALL



B. INMON

QUE BIGODE
MARAVILHOSO

Galera... esses caras são tão sinistros que algumas provas cobram o nome deles explicitamente! *Professor, por que tudo isso?* Porque ambos foram pioneiros na teoria sobre Data Warehouse no início da década de noventa, mas não só por isso! **Esses dois caras concordam em váááárias coisas sobre como implementar um Data Warehouse, no entanto há alguns pontos em que eles discordam diametralmente.** Vamos começar pela parte que eles concordam...

A Arquitetura de Inmon é chamada de Enterprise Data Warehouse (sim, aquele que nós vimos lá atrás) e a Arquitetura de Kimball é chamada de Dimensional Data Warehouse. Via de regra, elas são muito parecidas contendo um único repositório integrado de dados corporativos que suporta a análise de informações em toda a organização. Por outro lado, esses discordam veementemente na forma como um Data Warehouse deve ser construído. Vejamos:

De maneira simplista, pode-se afirmar que Inmon defende uma abordagem de cima para baixo (*top-down*) e Kimball defende uma abordagem de baixo para cima (*bottom-up*). Em outras palavras, Inmon acha que a maneira correta é construir o Data Warehouse e depois dividi-lo em Data Marts; já Kimball acha que a maneira correta é construir vários Data Marts e, só então, integrá-los em um Data Warehouse. Em suma:

- **Abordagem de Inmon:**



Recomenda a construção do DW que siga a abordagem de cima para baixo. **Na filosofia de Inmon, é começando com a construção de um grande Enterprise Data Warehouse centralizado**, onde todos os dados disponíveis dos sistemas transacionais são consolidados em uma coleção de dados orientada ao assunto, integrada, variante no tempo e não volátil que suporta a tomada de decisões. Em seguida, os Data Marts são construídos para necessidades analíticas de departamentos.

- **Abordagem de Kimball:**

Recomenda a construção do DW que segue a abordagem de baixo para cima. **Na filosofia de Kimball, inicia-se primeiro com Data Marts de missão crítica que atendem às necessidades analíticas dos departamentos**. Em seguida, integram-se esses Data Marts para consistência de dados por meio do chamado barramento de informações. Kimball faz uso do modelo dimensional para atender às necessidades dos departamentos em várias áreas da empresa.

Vamos detalhar um pouco mais a abordagem de Kimball! Esse autor dividia sua abordagem basicamente em quatro etapas apresentadas a seguir:

ETAPA	NOME	DESCRIÇÃO
1	IDENTIFICAR OS PROCESSOS DE NEGÓCIO	Processos de negócio são atividades de negócio realizadas por uma organização, como receber pedidos, realizar faturamento, receber pagamentos, atender chamadas de serviço, cadastrar alunos, realizar um procedimento médico ou processar sinistros. Ouvir os usuários é o meio mais eficiente para selecionar um processo de negócio. As medições de desempenho que eles analisam no Data Warehouse resultam de processos de medições do negócio.
2	IDENTIFICAR A GRANULARIDADE	Identificar a granularidade significa especificar exatamente o que uma linha específica da tabela de fatos representa. A granularidade exprime o nível de detalhe associado às medições da tabela de fatos. Ela fornece a resposta para a pergunta: "Como você descreve uma única linha na tabela de fatos?". Em suma, identificar a granularidade é identificar o nível de detalhe de um processo de negócio.
3	IDENTIFICAR AS DIMENSÕES	Você precisa rodear a tabelas de fatos com um conjunto robusto de dimensões que representem todas as descrições possíveis que assumem valores únicos no contexto de cada medição. As dimensões representam o "quem, o que, onde, quando, por que e como" associados ao evento. Exemplo: data, produto, cliente, funcionário e local.
4	IDENTIFICAR OS FATOS	Fatos são métricas resultantes do processo de negócio, isto é, quantidades, valores, medidas, entre outros. Os fatos são determinados pela resposta à pergunta: <i>O que estamos medindo?</i> Os fatos também podem ser derivados ou calculados a partir de outras medidas. Os usuários de negócios estão profundamente interessados em analisar essas métricas de desempenho.

CARACTERÍSTICAS	KIMBALL (DM)	INMON (EDW)
ABORDAGEM	Bottom-Up (Data Marts → Data Warehouse)	Top-Down (Data Warehouse → Data Marts)
ORIENTAÇÃO	Orientado a processos de negócio	Orientado a dados
INVESTIMENTO INICIAL	Baixo	Alto
NORMALIZAÇÃO	Desnormalizado	Normalizado (3ª Forma Normal)
DIRECIONAMENTO	Usuários finais	Profissionais de TI
CONSULTAS	Realizadas no Data Warehouse	Realizadas no Data Mart



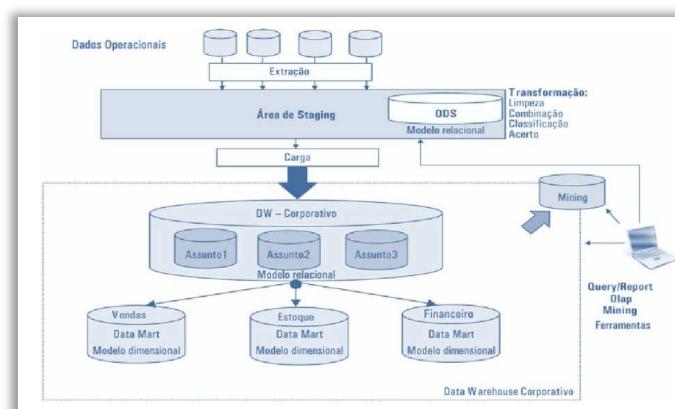
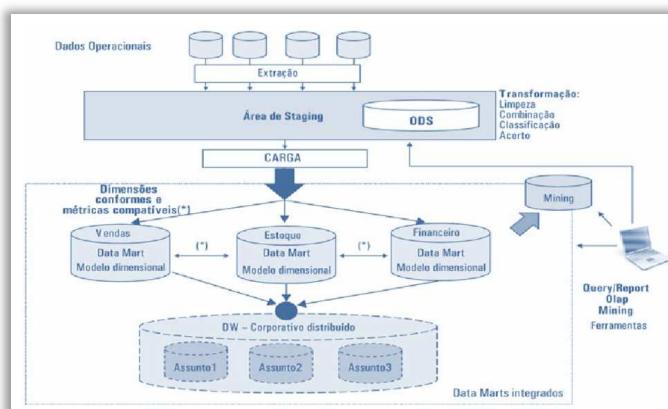
Antes de prosseguir, é importante explicar duas palavrinhas: normalizado e multidimensional!

Explicando bem rapidamente, a normalização é um conjunto de regras aplicadas a um banco de dados para reduzir a redundância, aumentar a integridade e melhorar o desempenho (exceto de consultas). Como mostra a tabela, o banco de dados do Kimball não se importa com dados redundantes, já o banco de dados de Inmon se importa (e é normalizado).

Agora vocês se lembram que temos a modelagem relacional? Temos também a modelagem multidimensional – em geral, utilizada em Data Warehouses! A Modelagem dimensional é uma técnica normalmente usada com Data Warehouses. Segundo Kimball, a modelagem dimensional é a única técnica viável para bancos de dados que devem responder consultas em um DW, uma vez que a modelagem relacional é mais útil para representação de transações.

Não há certo ou errado entre essas duas ideias, uma vez que elas representam apenas diferentes filosofias de armazenamento de dados. Na realidade, a maioria das empresas adota a abordagem de Kimball, porque a maioria dos Data Warehouses começam como um esforço departamental e, portanto, originaram-se como um Data Mart. Somente quando mais Data Marts são construídos posteriormente é que eles evoluem para um Data Warehouse.

Por fim, eu deixo um desafio para vocês: nos diagramas abaixo, um representa a abordagem de Kimball e outro representa a abordagem de Inmon. E aí, qual que é qual?



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

- é conhecida como top down.
- permite um retorno de investimento apenas em longo prazo, ou seja, um slower pay back.



- c) tem como objetivo a construção de um sistema OLAP incremental a partir de DMs independentes.
- d) não garante padronização dos metadados, podendo criar inconsistências de dados entre os DMs.
- e) tem como vantagem a criação de legamarts ou DMs legados que facilitam e agilizam futuras integrações.

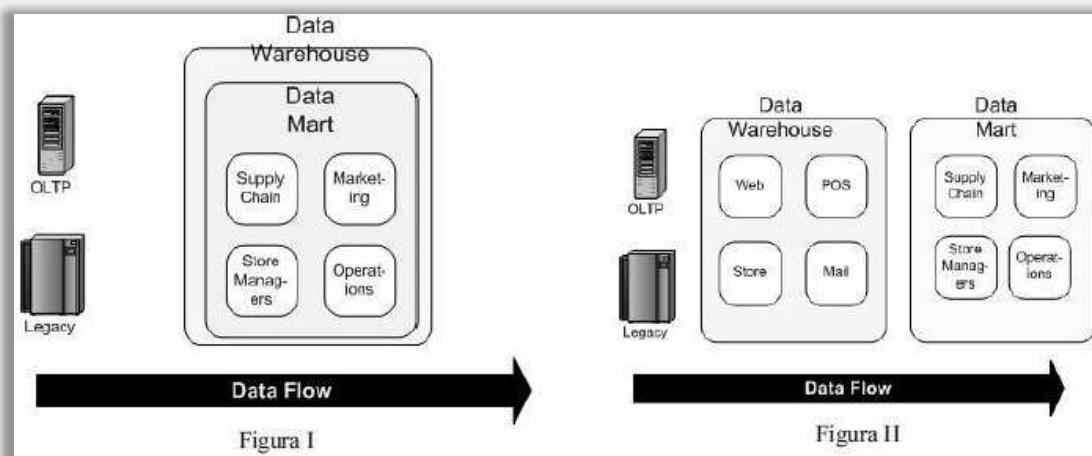
Comentários: (a) Errado, é conhecida como bottom-up; (b) Errado, o investimento inicial é mais baixo, logo permite um retorno em curto prazo; (c) Errado, tem como objetivo a construção de um DW Corporativo a partir de Data Marts independentes; (d) Correto, realmente não garante a padronização dos metadados e permite criar inconsistências de dados entre os Data Marts, uma vez que os eles são criados separadamente; (e) Errado, isso é uma desvantagem – esse tipo de implementação pode gerar uma falta de padronização dos Data Marts que pode acarretar em redundância e dados inconsistentes, uma vez que são construídos geralmente de forma independente que podem se tornar legados e dificultar futuras integrações (Letra D).

(TRE/MT – 2010) Com relação aos conceitos de modelagem dimensional, assinale a opção correta.

- a) Na abordagem utilizada por Kimball, todos os dados de uma empresa são reunidos em um repositório central, caracterizado por um projeto dimensional, e que pode ser consultado diretamente.
- b) A abordagem de Inmon consolida todas as informações de uma empresa em um repositório central chamado data warehouse corporativo. Essa abordagem é caracterizada por um projeto na terceira forma normal, no qual os dados são consultados diretamente por aplicativos de data warehouse.
- c) Na abordagem de Kimball, os data marts são adaptados às necessidades e pontos de vista de um grupo de negócios específico, suportando a modelagem dimensional, e são consultados por aplicativos de data warehouse.
- d) A abordagem de data marts autônomos tem foco empresarial e atende às necessidades dentro de uma área de assunto. Seu desenvolvimento necessita explorar o uso de elementos de dados comuns em toda a empresa.

Comentários: (a) Correto. Kimball parte de DMs para chegar ao DW, que é o repositório central dimensional acessível diretamente aos usuários finais; (b) Errado. Inmon parte do DW para chegar aos DMs, que são consultados diretamente e, não, o DW; (c) Errado. Nessa abordagem, os aplicativos consultam diretamente o DW e, não, os DMs; (d) Errado. Tem foco departamental e, não, empresarial. Ademais, DMs exploram o uso de elementos de dados individuais e, não, comuns (Letra A).





(TCU – 2009) Na figura II, que corresponde à visão de Kimball, o Datawarehousing é visto como uma constituição de Data Marts, cujo foco é entregar objetivos do negócio para departamentos na organização.

Comentários: a descrição está perfeita, mas trata da Figura I (Errado).

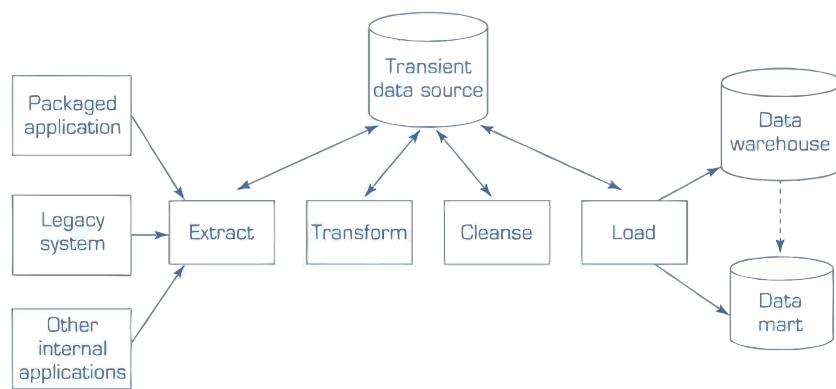


ETL

INCIDÊNCIA EM PROVA: MÉDIA

Galera, você precisa carregar seu Data Warehouse regularmente para que possa servir ao propósito de facilitar a análise de negócios. **Para fazer isso, os dados de um ou mais sistemas/ambientes operacionais precisam ser extraídos e copiados no Data Warehouse.** O desafio nos ambientes de Data Warehouse é integrar, reorganizar e consolidar grandes volumes de dados de muitos sistemas, fornecendo assim uma nova base de informações unificada para análise de dados.

O processo de extrair dados de sistemas de origem e trazê-los para o Data Warehouse é comumente chamado de ETL (*Extract, Transform e Load*). **Galera, é importante entender que o ETL se refere a um processo mais amplo e, não, a três etapas bem definidas.** A verdade é que a sigla ETL é simplista demais, uma vez que omite algumas fases e implica que cada uma das outras fases do processo é distinta. No entanto, o processo como um todo é conhecido como ETL!



Ele é o processo mais crítico e demorado na construção de um Data Warehouse, uma vez que consiste na extração dos dados de fontes homogêneas ou heterogêneas; na transformação e limpeza destes dados; e na carga dos dados no DW. Pessoal, as decisões estratégicas – aquelas mais importantes de uma organização – são tomadas com base nas informações geradas através dos dados armazenados no Data Warehouse.

Se estes dados não forem corretamente trabalhados no processo de extração, as informações geradas através deles farão com que decisões sejam tomadas erroneamente, podendo afetar diretamente os negócios da organização. Já viram a responsabilidade, não é? Galera, hoje em dia todo órgão público de médio/grande porte possui – em sua área de tecnologia da informação – uma gerência responsável por toda essa parte de *Business Intelligence*.

Essa área geralmente é responsável pela criação de Data Warehouses e pelo Processo de ETL! Aliás, não sei se vocês sabem, mas a maior parte do esforço exigido no desenvolvimento de um DW é consumido neste momento – eu diria que cerca de uns 80%! **Logo, ele tem que ser bem feito e refletir a realidade de modo que analistas consigam tirar informações relevantes para o processo de tomada de decisões gerenciais.**



Sabe outra curiosidade? Eu disse que 80% do esforço de desenvolvimento de um DW é ETL! Sabia que 60% do Processo de ETL é só com a Extração? Pois é, essa etapa deve se basear na busca das informações mais importantes em aplicações ou fontes externas. Pessoal, essa busca geralmente não é tão simples! **Dados podem estar distribuídos em dezenas de lugares, tecnologias ou plataformas diferentes necessitando de formas de extração distintas para cada local.**

É, galera – não é tudo mil maravilhas! Você perde muito tempo durante a extração porque frequentemente é necessário implementar diversas formas diferentes de extrair dados de seus locais de origem. E sem carga de dados, não temos Data Warehouse! **No momento da criação de um DW, é comum realizar uma carga de dados inicial, mas com o decorrer do tempo a extração deve estar preparada apenas para fazer cargas incrementais.**

A carga incremental é muito mais eficiente visto que carrega apenas os dados que foram inseridos desde a carga inicial. *Professor, e para onde vão esses dados que foram extraídos, mas ainda não foram transformados e carregados no Data Warehouse?* Excelente pergunta! Resposta: **eles vão para um local chamado Stage Area!** O que é isso, professor? Vocês devem se lembrar que uma das características essenciais de um Data Warehouse é sua integridade.

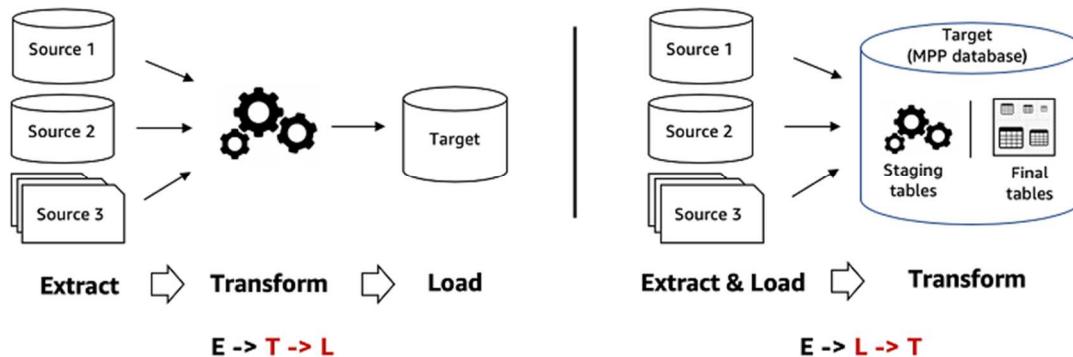
Relembrando: os dados devem estar padronizados, consistentes e uniformizados antes de serem carregados no Data Warehouse! Para tal, eles precisam ser limpos, entre outras atividades. **Isso tudo ocorre na Stage Area, portanto ela é basicamente uma localização temporária de onde os dados de diversas fontes de origem são copiados.** Sabe quando você copia e cola algo em um sistema operacional? O que você copiou vai para um local chamada Área de Transferência!

Adivinhem só: *Stage Area* é similar à Área de Transferência, isto é, uma localização de armazenamento temporário. **Segundo: a transformação dos dados é a fase subsequente à extração. Esta fase não só transforma os dados, mas também realiza a limpeza dos mesmos.** A correção de erros de digitação, a descoberta de violações de integridade, a substituição de caracteres desconhecidos, a padronização de abreviações podem ser exemplos dessa limpeza.

Já a fase de carga consiste no carregamento dos dados para seu destino final. Esse processo pode variar muito, uma vez que alguns Data Warehouses podem adicionar as informações existentes mensalmente (com dados cumulativos e atualizados) ao passo que outros podem adicionar dados a cada hora, dia ou semana. Galera, era isso que era cobrado até pouco tempo atrás, mas as coisas mudaram... agora estão cobrando até o ELT!

Professor, você errou a sigla! Errei, não... é isso mesmo! ELT é uma inversão da ordem das etapas que vimos anteriormente: em vez de fazermos (1) Extração, depois (2) Transformação e só depois a (3) Carga, nós fazemos a (1) Extração, depois a (2) Carga e só depois a (3) Transformação dos dados. **Primeiramente, é importante destacar que os processos de ETL e ELT – apesar de diferentes – solucionam o mesmo problema.**





O ELT é um processo mais ágil para o carregamento e o processamento de dados, visto inverte a ordem das etapas de transformação de dados da abordagem tradicional de ETL. Nesse cenário, a transformação de dados ocorre logo após a coleta e o carregamento das informações em um repositório de dados centralizado e, não, antes. Com isso, é possível transformar dados brutos em dados modelados dentro de um Data Warehouse.

Essa simples inversão de etapas **reduz consideravelmente o tempo de carregamento de dados**, permitindo que a transformação de dados seja feita por analistas de dados comuns, sem a dependência de profissionais altamente técnicos como desenvolvedores e engenheiros de dados. Na prática, os engenheiros de dados se focam apenas nas etapas de extração e carregamento, atribuições características de sua área de atuação.

Dessa forma, a **responsabilidade da transformação de dados fica nas mãos de profissionais próximos à empresa que conhecem as regras de negócio**, como analistas e cientistas de dados, trazendo benefícios como: menor tempo gasto entre extração e transformação de dados; maior eficiência na implementação de projetos; menor dependência de profissionais altamente técnicos; e independência dos analistas para o manuseio de dados brutos.

Quando a ETL foi criada, o armazenamento de dados em repositórios era extremamente caro. Além disso, o processamento dessas informações era extremamente lento e, por isso, pouquíssimas organizações podiam investir nessas tecnologias. **Ocorre que essa realidade mudou com a computação na nuvem (cloud computing)**. Hoje, o que era lento tornou-se rápido; e o que era caro tornou-se acessível.

Assim, considerando o volume massivo de dados para processamento nas operações modernas, muitas empresas têm optado pelo ELT em vez do tradicional ETL, buscando mais flexibilidade e agilidade nos seus pipelines de dados. **O ELT permite analisar grandes volumes de dados de forma rápida, com menos manutenção e gastos de recursos – além de possibilitar uma maior colaboração entre equipes de negócio e profissionais técnicos**.

Por essa razão, ainda que o ELT seja um processo mais complexo, exigindo mais esforços nos quesitos de design e infraestrutura do que o ETL, a primeira oferece mais vantagens no longo prazo. **Além disso, o ELT leva menos tempo no carregamento e na transformação de dados e, por isso,**



é mais econômico. No entanto, caso o repositório de dados de destino não seja robusto para comportar um volume massivo de dados, o processo de ETL pode ser uma opção mais viável.

Bem... agora que as provas começaram a cobrar o ELT, logo 99% das questões ainda serão sobre ETL! De toda forma, vamos resumir as diferenças na tabelinha a seguir:

CRITÉRIOS	ETL	ELT
TEMPO DE CARGA	Uso de sistemas distintos implica uma maior demora/latência para o carregamento de dados.	Uso de um sistema de carregamento integrado implica uma menor demora/latência para o carregamento de dados
TEMPO DE TRANSFORMAÇÃO	Demora considerável – particularmente na transformação de grandes volumes de dados. Além disso, à medida que os volumes crescem, o tempo de transformação aumenta.	Transformação de dados feita de maneira integrada, visto que sua velocidade independe do tamanho ou da complexidade dos dados – graças às tecnologias de infraestrutura na nuvem.
CUSTO DE MANUTENÇÃO	Altos custos de manutenção, dado que a atualização do repositório de dados requer o trabalho recorrente de profissionais caros e escassos no mercado, como engenheiros de dados e desenvolvedores.	Baixos custos de manutenção, pois os dados estão sempre disponíveis e acessíveis para todos sem a necessidade de profissionais extremamente especializados.
COMPLEXIDADE DE IMPLEMENTAÇÃO	Requer menos conhecimento aprofundado de projeto e arquitetura do repositório de dados.	Requer conhecimento profundo de ferramentas avançadas de análise e uma arquitetura bem estruturada do repositório de dados.
LIMITAÇÃO DE DADOS	Dados devem ser selecionados a priori, ou seja, o que não for selecionado é perdido.	Dados brutos ficam disponíveis, por isso a retenção de dados depende apenas da política de retenção eleita pela operação de dados.
SUPORTE PARA BANCOS DE DADOS	ETL foi desenvolvido para dar suporte a banco de dados relacionais, locais e sistemas legados.	ELT foi construído para suportar grandes volumes e diversas fontes de dados, estruturadas e não estruturadas, de forma escalável em infraestruturas de nuvem.
USABILIDADE	Tabelas fixas, cronogramas fixos, predominantemente utilizados por profissionais técnicos de TI, desenvolvedores e engenheiros de dados.	Ad hoc, abordagem escalável, flexível e colaborativa, utilizável tanto por profissionais técnicos quanto pelo usuário final de negócios.
CUSTO/BENEFÍCIO	Abordagem não econômica, especialmente para empresas de pequeno e médio portes.	Por ser escalável, adaptável e acessível para negócios de todos os tamanhos, é uma solução muito mais viável, econômica e moderna.
DADOS SENSÍVEIS	É mais adequado para o armazenamento de dados sensíveis e confidenciais, dado que	É menos adequado para o armazenamento de dados sensíveis e confidenciais, dado que a transformação ocorre somente após a carga.



	permite a transformação antes da carga (Ex: anonimização dos dados).
REPOSITÓRIO	Comumente utilizado em Armazém de Dados (Data Warehouses). Comumente utilizado em Lagos de Dados (Data Lakes).

É importante mencionar que esses sistemas são complementares: não necessariamente é preciso fazer a substituição de um por outro. Pode-se afirmar que o Data Lake é ótimo para realizar análises de Machine Learning/Deep Learning, Data Discovery e Análises Preditivas; já o Data Warehouse é bastante útil para análises de Business Intelligence, Visualização de Dados, Criação de Relatórios e Análises Transacionais.

(Prefeitura de Teresina – 2016) O processo ETL é uma etapa importante no projeto de um Data Warehouse. No processo ETL,

- a) somente é possível realizar a carga periódica de dados no data warehouse uma vez a cada semana.
- b) a fase de transformação de dados envolve apenas alterações nos tipos vindos do ambiente operacional.
- c) a fase de extração envolve a leitura de dados do data warehouse e sua carga posterior nos bancos de dados operacionais.
- d) na fase de transferência de dados são desprezados os puramente literais.
- e) um intervalo possível para a carga periódica de dados no data warehouse é de 24 horas.

Comentários: (a) Errado. Isso não existe – a organização pode determinar a periodicidade da carga de dados; (b) Errado. A fase de transformação de dados envolver alterações nos tipos vindos de diversas fontes de dados e não apenas aquelas vindas do ambiente operacional; (c) Errado. A fase de extração envolve a leitura de dados de diversas fontes diferentes (inclusive dos bancos de dados operacionais) e sua carga posterior no Data Warehouse; (d) Errado. Isso simplesmente não existe; (e) Correto. O intervalo pode ser o que a organização achar adequado e 24 horas é um período possível (Letra E).

(SEFAZ/CE – 2021) Comparado ao ETL, o ELT apresenta vantagens como tempos menores de carregamento e de transformação de dados, e, consequentemente, menor custo de manutenção.

Comentários: o tempo de carregamento e transformação realmente são menores – assim como o custo de manutenção (Correto).



RESUMO

BUSINESS INTELLIGENCE

Business Intelligence (Inteligência de Negócio) é um termo criado pelo Gartner Group utilizado para representar um conjunto de processos, técnicas, metodologias, habilidades, ferramentas e capacidades utilizadas para acessar, coletar, organizar, tratar, analisar, cruzar, processar, compartilhar e monitorar dados de diversas fontes com o intuito de gerar informações e relatórios analíticos que suportem a gestão corporativa, a definição de estratégias e a tomada de decisão em ambientes de negócios.

TIPOS DE ANÁLISES	DESCRIÇÃO
ANÁLISE DESCRIPTIVA	É o exame de dados ou conteúdos para responder à pergunta: "O que aconteceu?" ou "o que está acontecendo?". É caracterizado pelo Business Intelligence tradicional e visualizações como gráficos de pizza, gráficos de barras, etc que servem para que gestores tomem decisões.
ANÁLISE DIAGNÓSTICA	É uma forma de análise avançada que examina dados ou conteúdos para responder à pergunta: "Por que isso aconteceu?", e é caracterizada por técnicas como detalhamento, descoberta de dados, mineração de dados e correlações. Esse é o tipo de análise empregada em investigações de causa-raiz.
ANÁLISE PREDITIVA	Combina técnicas de estatística, mineração de dados e aprendizagem de máquina para encontrar significado em grandes quantidades de dados, trabalhando com probabilidades, previsões, entre outros para antecipar comportamentos futuros com base em eventos passados. Responde à pergunta: "O que vai acontecer?";
ANÁLISE PRESCRITIVA	É uma forma de análise avançada que examina os dados ou os conteúdos para responder à pergunta: "O que deve ser feito?" ou "O que podemos fazer caso algo aconteça?", e é caracterizada por técnicas como análise de gráficos, simulação, redes neurais, mecanismos de recomendação, heurísticas, aprendizado de máquina, etc.

ARQUITETURA BI

	TRATA-SE DE UMA COLEÇÃO DE DADOS ORIENTADOS POR ASSUNTO, INTEGRADOS, VARIÁVEIS COM O TEMPO E NÃO-VOLÁTEIS, PARA DAR SUPORTE AO PROCESSO DE TOMADA DE DECISÃO.
	TRATA-SE DE UM CONJUNTO DE FERRAMENTAS PARA MANIPULAR, MINERAR E ANALISAR DADOS, CRIAR RELATÓRIOS, REALIZAR CONSULTAS SOB DEMANDA, ENTRE OUTROS.
	TRATA-SE DE UM CONJUNTO DE FERRAMENTAS QUE PERMITEM COMPARAR E COMPARTILHAR O DESEMPENHO DE NEGÓCIOS DE UMA ORGANIZAÇÃO EM RELAÇÃO ÀS METAS ESTABELECIDAS.
	TRATA-SE DE UM CONJUNTO DE FERRAMENTAS QUE PERMITEM A RÁPIDA E ABRANGENTE VISUALIZAÇÃO DE DADOS POR MEIO DE DASHBOARDS E SOFTWARES SIMILARES.

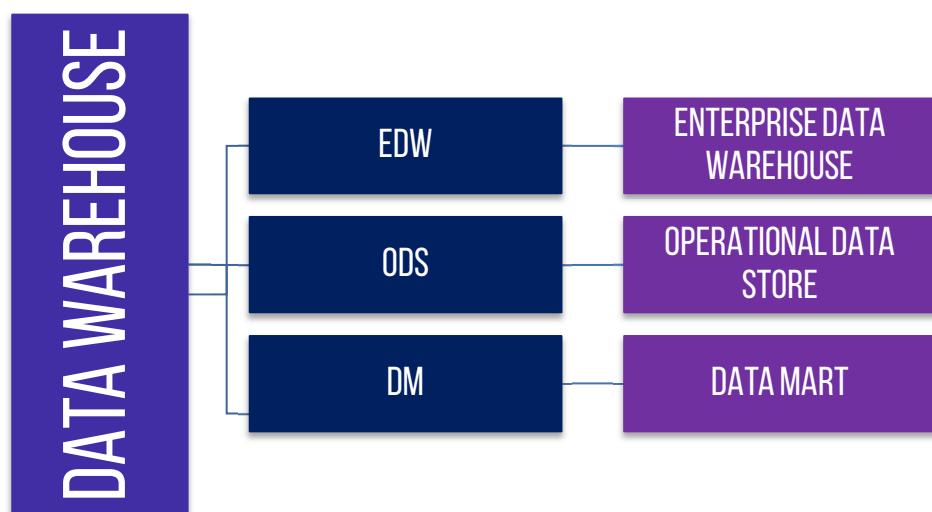
AUTOR	DEFINIÇÕES ACADÊMICAS DE DATA WAREHOUSE
RALPH KIMBALL	Conjunto de ferramentas e técnicas de projeto que – quando aplicadas às necessidades específicas dos usuários e aos bancos de dados específicos – permitirá que planejem e construam um Data Warehouse.

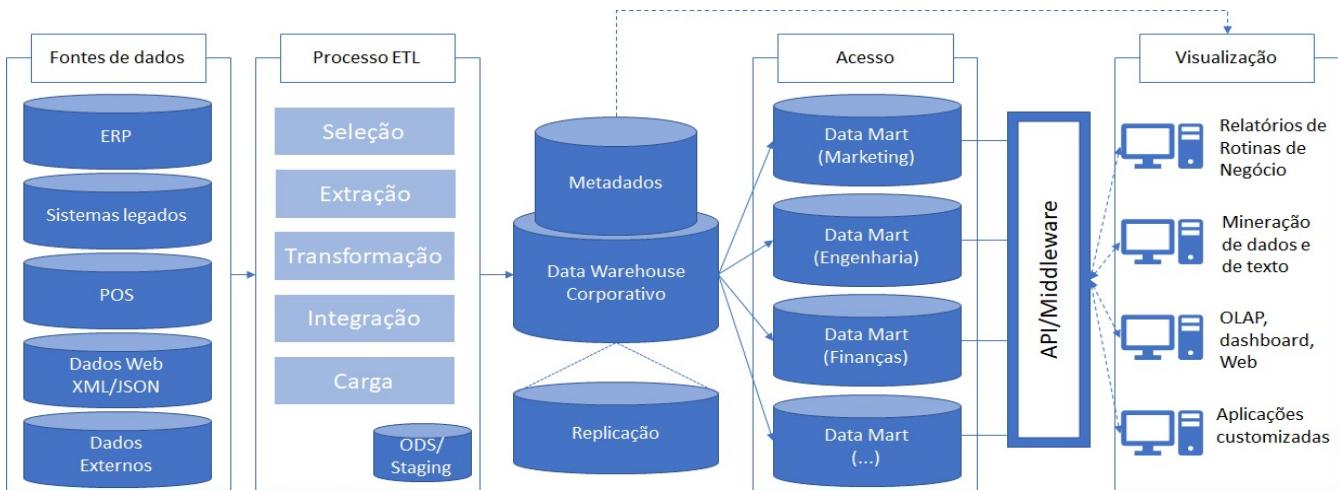


BILL INMON	Coleção de dados orientados por assunto, integrados, variáveis com o tempo e não-voláteis , para dar suporte ao processo de tomada de decisão.
ARUN SEN	Banco de dados construídos no interesse de suporte à decisão de negócios e contêm dados históricos sumarizados e consolidados provenientes de registros individuais de bancos de dados operacionais.
KENNETH LAUDON	Banco de dados – com ferramentas de consulta e relatório – que armazena dados atuais e históricos extraídos de vários sistemas/ambientes operacionais e consolidados para fins de análises e relatórios administrativos .

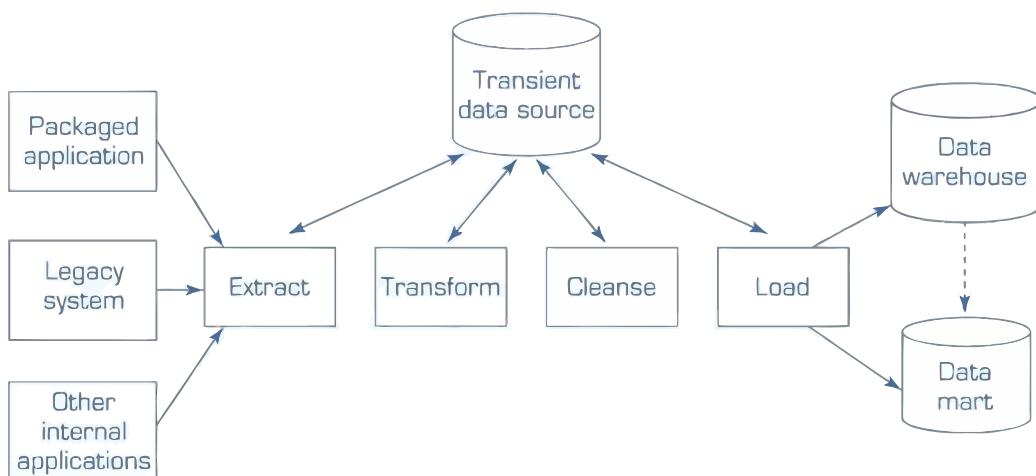
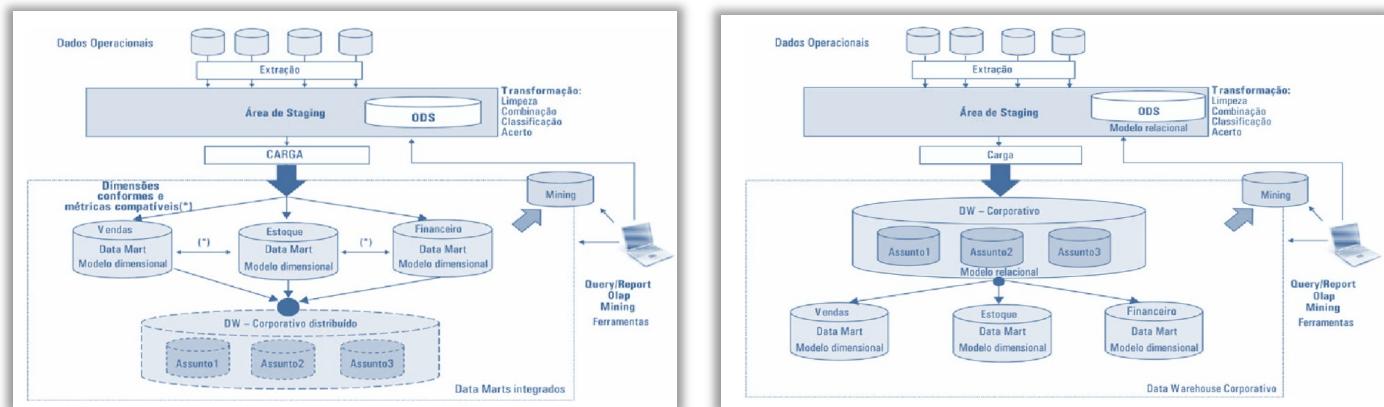


ENTERPRISE DATA WAREHOUSE (EDW)	Traduzido como Armazém de Dados Corporativos ou Abrangentes, é um repositório central tático/estratégico que contém todas as informações de negócio de uma organização de forma acessível para toda empresa com a finalidade de suportar as análises de decisões. Grosso modo é possível afirmar que um EDW é um DW de larga escala – grande e abrangente – que é utilizado pela organização como um todo e que congrega informações de diversas fontes de dados.
OPERATIONAL DATA STORE (ODS)	Traduzido como Armazenamento de Dados Operacionais, trata-se de um repositório intermediário de dados utilizado para relatórios operacionais e como uma fonte de dados para o Enterprise Data Warehouse (EDW). É um elemento complementar a um EDW em um cenário de suporte a decisões, e é usado para relatórios, controles e tomada de decisões operacionais – em oposição ao EDW, que é utilizado para suporte de decisões táticas e estratégicas.
DATA MART (DM)	Traduzido como Mercado de Dados, trata-se de um subconjunto de dados de um Data Warehouse. Geralmente são dados referentes a um assunto em especial (Ex: Vendas, Estoque, Controladoria, entre outros) ou diferentes níveis de summarização (Ex: Vendas Anual, Vendas Semestral, Vendas Mensal, entre outros), que focalizam uma ou mais áreas específicas. Seus dados são obtidos do DW – indexados para suportar intensas pesquisas.

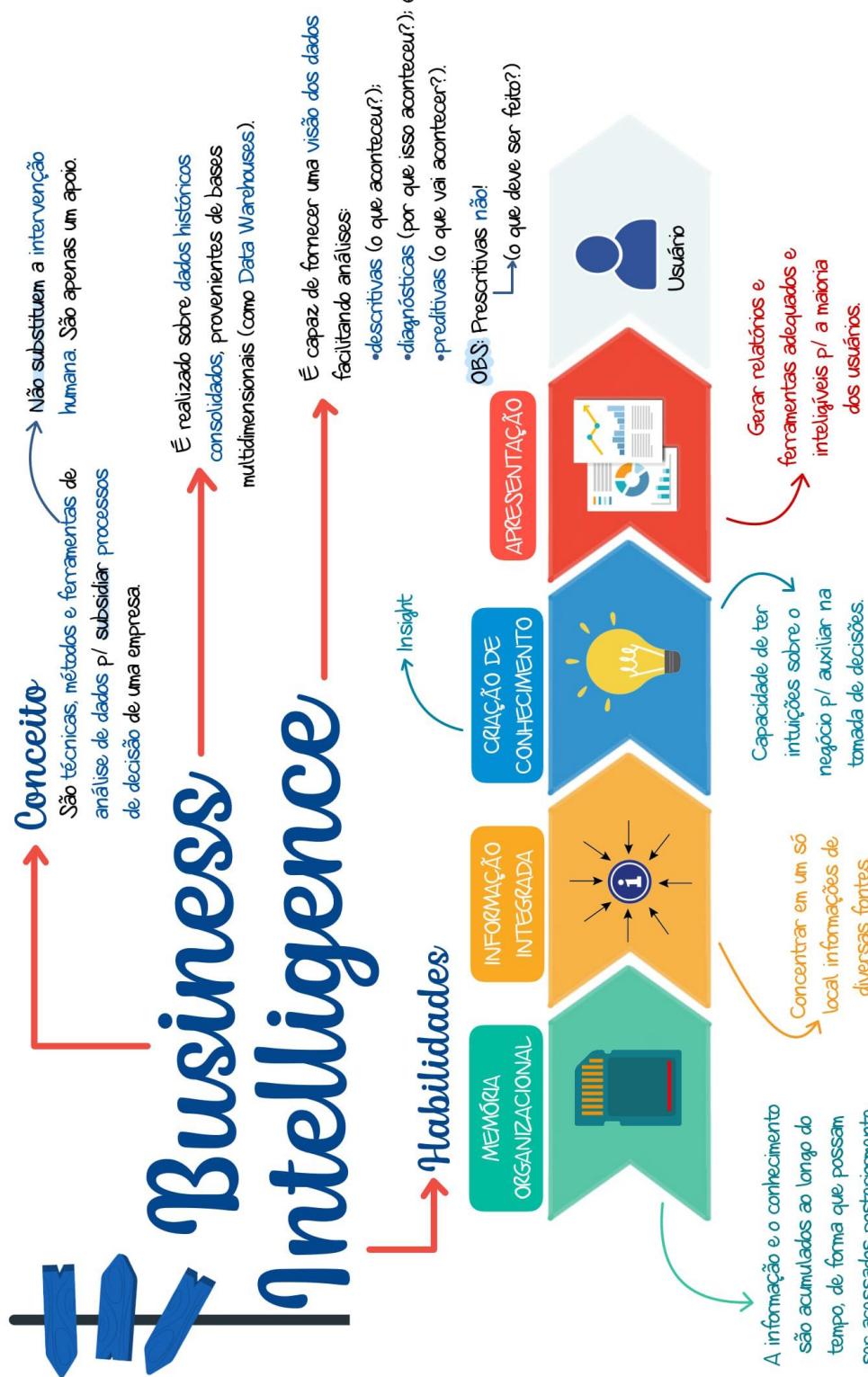




CARACTERÍSTICAS	KIMBALL (DM)	INMON (EDW)
ABORDAGEM	Bottom-Up (Data Marts → Data Warehouse)	Top-Down (Data Warehouse → Data Marts)
ORIENTAÇÃO	Orientado a processos de negócio	Orientado a dados
INVESTIMENTO INICIAL	Baixo	Alto
NORMALIZAÇÃO	Desnormalizado	Normalizado (3ª Forma Normal)
DIRECIONAMENTO	Usuários finais	Profissionais de TI
CONSULTAS	Realizadas no Data Warehouse	Realizadas no Data Mart



MAPA MENTAL



@mapasdathai



Conceito

- Data Warehouse = **armazém de dados**.
- Coleção de dados orientados por assunto, **integrados, variáveis com o tempo e não voláteis**, consolidados para dar suporte ao processo de tomada de decisão.



Orientado por assunto

- Trata de **temas específicos e importantes p/ o negócio da organização**.

Espécies

Enterprise Data Warehouse (EDW)

- **Armazém de Dados Corporativos ou Abrangentes.**
- **Repositório central tático/estratégico.**
- Contém **todas as informações** de negócios de uma organização.
- Acessível p/ toda a empresa.

Operational Data Store (ODS)

- **Armazém de Dados Operacionais.**
- **Repositório intermediário** de dados p/ suporte às **decisões operacionais**.
- Fonte de dados p/ o EDW.

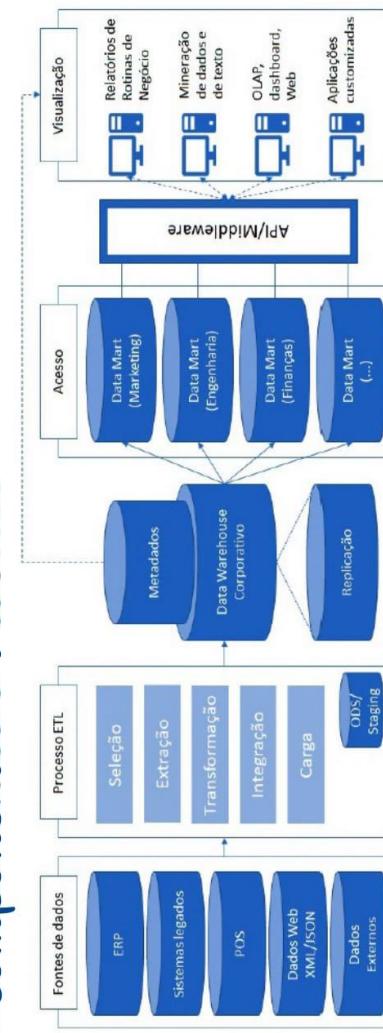
Data Mart (DM)

- **Mercado de Dados**.
- **Subconjunto** de dados de um Data Warehouse.
- Contém **dados restritos a uma unidade** de negócio.
- **Repositório especializado**.

- **Não-volátil**
- Os dados, uma vez inseridos no DW, não podem ser alterados.
- **Somente leitura** p/ usuários finais (**consultas e relatórios**).

Integrado

- Os dados do DW são **integrados de múltiplas fontes** de forma consistente, padronizada e uniforme.
- **Histórico - variável com o tempo**
- **Dados históricos** são mantidos em um DW.
- Em um banco de dados transacional, geralmente só os dados mais recentes são mantidos (ex: o último endereço de um cliente). Já o DW **pode conter todos os endereços já associados a um cliente**.



Componentes e processo



Abordagem de Inmon

- Enterprise Data Warehouse.
- Abordagem top-down (de cima p/ baixo).
- **Construir o DW e dividi-lo em Data Marts.**
- Orientação: **Orientado a dados.**
- Investimento Inicial Alto.
- **Normalizado** (3^a Forma Normal).
- Direcionado a **profissionais de TI**.
- Consultas realizadas no Data Mart



Ambas possuem um único repositório integrado de dados corporativos que suporta a análise de informações em toda a organização.

- Dimensional Data Warehouse.
- Abordagem bottom-up (de baixo p/ cima).
- **Construir vários Data Marts e depois integrá-los em um DW.**
- Orientação: **Orientado a processos de negócio.**
- Investimento Inicial Baixo.
- **Desnormalizado** (não se importa com dados redundantes).
- Direcionado a **usuários finais**.
- Consultas realizadas no Data Warehouse.



Abordagem de Kimball

- P/ Kimball, a modelagem dimensional é a única técnica viável p/ bancos de dados que devem responder consultas em um DW.
- A etapa de extração é a mais demorada.
- Por isso é comum, quando um DW é criado, realizar uma **carga de dados inicial** e com o decorrer do tempo realizar **cargas incrementais**.
- Stage Area é uma localização de **armazenamento temporário**, onde os dados permanecem até estarem prontos p/ serem carregados no DW.
- Dados trabalhados de forma errada podem gerar tomadas de decisões equivocadas na organização.



- **ETL**
- Processo de extrair dados de fontes homogêneas e heterogêneas, transformar e limpar estes dados e **carregá-los** no DW.
- A sigla ETL significa **Extract, Transform e Load**, mas o processo em si é mais complexo do que essas 3 etapas.
- É o processo mais crítico e demorado na construção de um DW.

- Banco do Brasil (Escriturário - Agente de Tecnologia) Banco de Dados - 2023 (Pós-Edital)
- www.estategiaconcursos.com.br

QUESTÕES COMENTADAS – CESPE

1. (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

2. (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

3. (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

Após ter sido privatizada, determinada empresa passou a utilizar novos sistemas administrativos, tais como: folha de pagamento, sistema de pessoal, sistemas de benefícios etc.; porém, constatou-se que os sistemas antigos não poderiam ser descontinuados em um pequeno espaço de tempo. Tais sistemas usavam tecnologias de armazenamento de dados antiquados e dados de baixa qualidade.

Acerca dessa situação hipotética e considerando aspectos diversos pertinentes à inteligência de negócios - business intelligence (BI) -, julgue os seguintes itens.

4. (CESPE / TELEBRÁS - 2021) Na situação hipotética em questão, para transformar os dados brutos em informações estratégicas para a tomada de decisão, deveriam ser utilizadas ferramentas de BI. Tais ferramentas podem ser divididas nas seguintes etapas: coleta, análise, visualização de dados e tomada de decisão. Na etapa da coleta, normalmente, usa-se o método de extração, transformação e carregamento.

Comentários:

O Business Intelligence é o processo de recolhimento e tratamento de informações que apoiarão a gestão de um negócio. As ferramentas de BI permitem a rápida e abrangente visualização de dados por meio de dashboards e softwares similares. Desse modo, os dados devem ser coletados, analisados, cria-se uma visualização para os dados e, por fim, há a tomada de decisão. Ademais, na coleta dos dados é aplicado o processo de ETL, que é o de processo extração dos dados de fontes homogêneas ou heterogêneas; da transformação e limpeza destes dados; e da carga dos dados em um armazém de dados.

Gabarito: Correto

5. (CESPE / TELEBRÁS - 2021) Em situações semelhantes àquela em apreço, as técnicas de um sistema de BI objetivam aumentar a agilidade e assertividade na tomada de decisões, com a definição de regras para a formatação adequada dos dados, visando a transformação desses dados em armazéns estruturados de informações, considerando a origem deles.

Comentários:

Todo o item está correto, com exceção do trecho “considerando a origem deles”. Na verdade, não se considera a origem dos dados, ou seja, os dados podem ter origem de diversas fontes.

Gabarito: Errado



6. (CESPE / TELEBRÁS - 2021) Na situação hipotética considerada, como os dados dos sistemas antigo e novo são heterogêneos, o ETL procurará limpar e preparar os dados para o OLAP.

Comentários:

Pessoal, essa é a função do ETL, ele faz a extração dos dados de fontes homogêneas ou heterogêneas; faz a transformação e limpeza destes dados; e, por fim, faz a carga dos dados no DW. Ademais, com as ferramentas OLAP, é possível acessar, visualizar e analisar grandes quantidades de informações em um DW.

Gabarito: Correto

7. (CESPE / TELEBRÁS - 2021) Na situação hipotética em tela, considerando as bases de dados antigas e novas, seria uma ótima prática a utilização dos conceitos de data warehouse e data mining para, por exemplo, automatizar a análise de currículos de acordo com as competências dos empregados.

Comentários:

Perfeito! O Data Warehouse seria utilizado para armazenar os dados e a mineração de dados para explorar, organizar e analisar de forma automática os dados dos empregados com o intuito de identificar, descobrir, extrair, classificar e agrupar informações implícitas desconhecidas.

Gabarito: Correto

8. (CESPE / TELEBRÁS - 2021) Na situação hipotética em apreço, seria correto optar pela utilização de bancos de dados relacionais, os quais são bancos especializados que foram desenvolvidos para ajudar a extrair dos dados as informações de BI.

Comentários:

Opa, os bancos multidimensionais é que são utilizados para extrair dos dados as informações de BI. De acordo com Kimball, a modelagem dimensional é a única técnica viável para bancos de dados que devem responder consultas em um DW. Os bancos relacionais são mais úteis para representação de transações.

Gabarito: Errado

9. (CESPE / MPE-AP – 2021) Em um sistema de BI, a coleção de ferramentas utilizada como componente para manipular, minerar e analisar os dados no DW (data warehouse) denomina-se:



- a) OLAP (Online Analytical Processing).
- b) BPM (Business Performance Management).
- c) Análise de Negócio.
- d) Dashboard.
- e) Processamento de Transações.

Comentários:

Essa questão foi retirada do livro de E. Turban. De acordo com o autor:

"... o BI tem quatro grandes componentes: um **Data Warehouse (DW)** com seus dados-fonte; a **análise de negócios**, uma coleção de ferramentas para manipular e analisar os dados no data warehouse, incluindo data mining; **Business Performance Management (BPM)** para monitoria e análise do desempenho e uma interface de usuário (como o **Dashboard**)".

Gabarito: Letra C

- 10. (CESPE / SEFAZ-CE – 2021)** Comparado ao ETL, o ELT apresenta vantagens como tempos menores de carregamento e de transformação de dados, e, consequentemente, menor custo de manutenção.

Comentários:

ETL extrai, transforma e carrega; ELT extrai, carrega e transforma. Dito isso, ELT realmente tem a vantagem de ter tempos menores de carregamento e transformação de dados e, logo, menor custo de manutenção. O ETL utiliza sistemas distintos que implicam demora para o carregamento de dados; já o ELT utiliza um sistema de carregamento integrando – realizando-o apenas uma vez; ETL demora a transformar os dados à medida que o volume cresce; já o ELT possui um tempo menor dado que a transformação de dados é feita de maneira integrada; ETL possui altas taxas de manutenção, visto que a atualização dos repositórios requer o trabalho recorrente de profissionais caros e escassos no mercado, como engenheiros de dados e desenvolvedores; já o ELT possui baixos custos de manutenção, pois os dados estão sempre disponíveis e acessíveis para todos.

Gabarito: Correto

- 11. (CESPE / SEFAZ-CE – 2021)** Faz parte do rol de objetivos do BI permitir acesso interativo aos dados e fornecer aos analistas de negócios, por meio da manipulação desses dados, a capacidade de realizar a análise adequada.

Comentários:



Business Intelligence (BI) realmente permite que analistas de negócio possam manipular (aqui no sentido de realizar diferentes tipos de consultas) dados de forma interativa para auxiliar em sua capacidade de analisar dados de uma organização para tomada de decisão.

Gabarito: Correto

12. (CESPE / SEFAZ-CE – 2021) Um Data Warehouse (DW), ainda que seja não volátil - ou seja, após os dados serem inseridos nele os usuários não podem alterá-los - é variável no tempo, pois mantém um conjunto de dados históricos que oferecem suporte à tomada de decisões.

Comentários:

Perfeito, perfeito, perfeito! Essa é uma dúvida muito comum e essa questão ajuda a diferenciar essas duas características. O DW é não-volátil, logo – após serem carregados – seus dados não podem ser alterados. No entanto, ele é variável no tempo, porque ele mantém dados históricos em relação a diferentes períodos de tempo de uma organização – isso ajuda bastante os gestores no suporte à tomada de decisões.

Gabarito: Correto

13. (CESPE / APEX-BRASIL – 2021) Em DW (data warehouse), conjunto de dados voltado para oferecer suporte à tomada de decisões, os dados:

- a) devem ser inseridos em seu formato original, cada um com sua unidade de medida, quando forem provenientes de diferentes fontes.
- b) utilizam, ao mesmo tempo, em um DW em tempo real, as estruturas relacional e multidimensional para proporcionar aos usuários acesso fácil e ativo aos dados.
- c) devem ser alterados de modo a permanecerem atualizados à medida que as fontes de dados sejam registradas e modificadas.
- d) detectam tendências e relações de longo prazo para previsão, com uma qualidade temporal para cada DW, sem necessariamente mostrar o status atual.

Comentários:

(a) Errado, devem ser transformados antes de serem carregados; (b) Errado, utiliza – em regra – estruturas multidimensionais; (c) Errado, não deve haver modificação a fim de atualizar dados; (d) Correto, eles realmente ajudam a detectar tendências e relações de longo prazo para previsão, com uma qualidade temporal para cada DW, sem necessariamente mostrar o status atual.

Gabarito: Letra D



14.(CESPE / APEX-BRASIL – 2021) Os objetivos característicos do BI (Business Intelligence) incluem:

- a) possibilitar acessos interativos a dados, a sua manipulação e a realização de análises apropriadas pelos gestores e analistas, com base na transformação desses dados em informações, decisões e ações.
- b) criar um conjunto de processos para gerenciar relacionamentos com clientes efetivos e potenciais e com parceiros de negócios, por meio de marketing, vendas e serviços, independentemente do canal de comunicação.
- c) gerenciar os produtos e os serviços direcionados aos consumidores, por meio de uma rede de organizações interligadas, com base no gerenciamento de uma cadeia de suprimentos.
- d) estabelecer relação entre a eficácia na gestão e a estratégia de negócio, a partir de uma composição balanceada de medidas, de metas, de objetivos e de requisitos para essa finalidade.

Comentários:

(a) Correto, lembrando que a manipulação aqui é no sentido de consulta apenas; (b) Errado, isso estaria mais voltado para CRM (Customer Relationship Management); (c) Errado, isso estaria mais voltado para SCM (Supply Chain Management); (d) Errado, isso estaria mais voltado para BSC (Balance ScoreCard).

Gabarito: Letra A

15. (CESPE / TCE-RJ – 2021) O Data Warehouse empresarial, ou EDW (Enterprise Data Warehouse), é uma das aplicações do data warehouse que permite a integração em larga escala de dados oriundos de diversas fontes em formato padronizado, para subsidiar a inteligência de negócios.

Comentários:

O EDW é uma das aplicações do Data Warehouse? Sim, junto com o ODS e o Data Mart. Ele permite a integração em larga escala de dados? Sim, ele integra uma grande quantidade de dados. Esses dados são oriundos de diversas fontes? Sim, os dados são obtidos de várias fontes. Esses dados vêm em formato padronizado? Sim, eles ficam em formato padronizado após passar pelo processo de transformação do ETL. Para subsidiar a inteligência de negócio? Sim, para auxiliar o Business Intelligence.

Gabarito: Correto

16.(CESPE / TCE-RJ – 2021) A construção de um Data Mart antecede a criação de um Data Warehouse.



Comentários:

Honestamente, eu detesto esse tipo de questão! *Por que, Diego?* Nós sabemos que existem duas abordagens: Kimball afirma que o Data Warehouse é construído a partir dos Data Marts; Inmon afirma que os Data Marts são construídos a partir do Data Warehouse.

Ora, a construção de um Data Mart antecede a criação de um Data Warehouse? Depende da abordagem utilizada! Esse é o tipo de questão que o examinador pode dar qualquer gabarito infelizmente – ainda bem que ele considerou o item como incorreto.

Gabarito: Errado

17. (CESPE / TCE-RJ – 2021) Enquanto os depósitos de dados e seus subconjuntos Data Marts são bancos estáticos, ou seja, não mais sofrem mudanças depois de agrupados, formatados e armazenados seus dados em um repositório, os bancos de dados típicos mudam constantemente, sendo dinâmicos por natureza.

Comentários:

Os bancos de dados transacionais (chamados de típicos) possuem uma natureza dinâmica ou volátil, no sentido de realizar grandes quantidades de transações, enquanto bancos de dados multidimensional (DW/DM) possuem uma natureza estática ou não-volátil, no sentido de realizar apenas consultas.

Professor, não se modifica, mas se inserem novos dados de tempos em tempos – isso não caracterizaria uma mudança? Concordo com você, no entanto a questão se focou apenas na característica da não-volatilidade dos dados e ignorou a característica histórica dos depósitos de dados. Enfim... péssima questão e eu discordo do gabarito!

Gabarito: Correto

18. (CESPE / TCE-RJ – 2021) Para o processamento de grandes quantidades de transações, de modo rápido e eficaz, é mais indicado o modelo relacional de bancos de dados do que os modelos orientados a objetos ou multidimensional.

Comentários:

Questão polêmica! O modelo relacional é mais adequado para processar grandes quantidades de transações do que os modelos orientados a objetos ou multidimensional. Tanto o modelo multidimensional quanto o modelo orientado a objetos são mais indicados para contextos de consultas ou manipulações complexas de dados. A banca não especificou o tipo de transação, mas o modelo relacional é mais adequado para transações de inclusão, exclusão e atualização, enquanto



o modelo multidimensional é mais adequado para transações de consulta. *E aí, candidato?* Pois é... eu jamais vou concordar com esse tipo de questão!

Gabarito: Errado

19.(CESPE / TCE-RJ - 2021) ETL (Extract Transform Load) é uma ferramenta utilizada para extrair informações e realizar análise multidimensional no Data Warehouse.

Comentários:

O Processo de ETL não é utilizado para realizar análise multidimensional – quem é responsável por essa atividade são as Ferramentas OLAP.

Gabarito: Errado

20.(CESPE / CODEVASF – 2021) Os sistemas de suporte à decisão (DSS, do inglês Decision Support System) são programados para substituir os tomadores de decisão das organizações e podem ser usados rotineiramente em tarefas de tomada de decisão comuns, devendo ser evitados para tarefas de decisões pontuais de alto nível relacionadas às estratégias de negócios.

Comentários:

A ideia não é substituir os tomadores de decisão das organizações e, sim, auxiliá-los com informações estratégicas.

Gabarito: Errado

21.(CESPE / CODEVASF – 2021) A inteligência de negócios lida com as atividades operacionais cotidianas da organização, sendo as suas duas principais ferramentas de implementação o processamento de transações e a integridade transacional com atualização das bases de dados operacionais.

Comentários:

A inteligência de negócio lida com atividades estratégicas e, não, cotidianas. Para lidar com atividades cotidianas por meio de ferramentas de processamento de transações e integridade transacional de bases de dados operacionais, utilizam-se Ferramentas OLTP (Online Transaction Processing).

Gabarito: Errado

22.(CESPE / ME – 2020) Modelagem preditiva é utilizada para antecipar comportamentos futuros, por meio do estudo da relação entre duas ou mais variáveis.



Comentários:

Perfeito! Modelagem Preditiva é caracterizada por tentar estimar eventos futuros com base em eventos passados a partir de um conjunto de dados e ferramentas estatísticas e matemáticas.

Gabarito: Correto

23. (CESPE / ME – 2020) Uma tabela de fatos registra dados dimensionais que explicam os fatos registrados.

Comentários:

Na verdade, são as tabelas dimensões que guardam informações de contexto – a tabela de fatos registra os eventos e as medidas a eles associadas.

Gabarito: Errado

24. (CESPE / TJ-PA – 2020) Assinale a opção que indica um processo de extração e transformação de dados em um data Warehouse:

- a) Big Data
- b) OLAP
- c) OLTP
- d) ETL
- e) machine learning

Comentários:

O processo de extração e transformação de dados para um data Warehouse é o ETL.

Gabarito: Letra D

25. (CESPE / CGM JOÃO PESSOA - 2018) *Business intelligence* pode ser definido como um processo inteligente de coleta, organização, análise, compartilhamento e monitoração de dados que, depois de processados, geram informações para o suporte e para a tomada de decisões no ambiente de negócios.

Comentários:

Perfeito! Business Intelligence (Inteligência de Negócio) é um termo criado pelo Gartner Group utilizado para representar um conjunto de processos, técnicas, metodologias, habilidades, ferramentas e capacidades utilizadas para acessar, coletar, organizar, tratar, analisar, cruzar,



processar, compartilhar e monitorar dados de diversas fontes com o intuito de gerar informações e relatórios analíticos que suportem a gestão corporativa, a definição de estratégias e a tomada de decisão em ambientes de negócio.

Gabarito: Correto

26.(CESPE / SECONT/ES - 2018) Data Warehouse é uma coleção de dados orientados ao assunto, que tem como características visão conceitual multidimensional, operações interdimensionais irrestritas, dimensões e níveis de agregação ilimitados, sendo que os dados são não voláteis e variantes no tempo.

Comentários:

Data Warehouse é uma coleção de dados orientados por assunto? Sim, de acordo com Bill Inmon. *Ele tem como característica visão conceitual multidimensional?* Sim, Data Warehouse utiliza uma modelagem multidimensional e, não, modelo entidade-relacionamento. *Realiza operações interdimensionais irrestritas?* Sim, há infinitas possibilidades. *Dimensões e níveis de agregação são ilimitados?* Sim, você possui diversos níveis de granularidade. Por fim, os dados são não-voláteis e variantes no tempo, de acordo com Bill Inmon.

Gabarito: Correto

27. (CESPE / TJ/SE - 2018) Os dados armazenados em um *DataWarehouse* devem estar integrados, temporalmente identificados, orientados ao assunto e devem ser protegidos, de modo a se evitar o acesso do usuário.

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. No entanto, **os dados devem estar protegidos para evitar o acesso do usuário não autorizado**, mas devem estar acessíveis aos usuários autorizados.

Gabarito: Errado

28.(CESPE / IPHAN – 2018) O *data warehouse* é integrado, não volátil e orientado a assuntos, contudo, embora lide com dados e os armazene para a tomada de decisões gerenciadas, não é considerado um banco de dados, pois é variável em relação ao tempo.

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. No entanto, ele é – sim – considerado um banco de dados.

Arun Sen afirma que: **Data Warehouse é um banco de dados** construídos no interesse de suporte à decisão de negócios e contêm dados históricos sumarizados e consolidados provenientes de registros individuais de bancos de dados operacionais.

Gabarito: Errado

29. (CESPE / STM – 2018) Em processos ligados a ETL, os dados são identificados e extraídos de diferentes fontes, incluindo sistemas de banco de dados e aplicações.

Comentários:

Impecável! Dados realmente são identificados e extraídos de diversas fontes diferentes de forma consistente, padronizada e uniforme no Processo de ETL.

Gabarito: Correto

30. (CESPE / CGM JOÃO PESSOA – 2018) Uma ferramenta de ETL deve ser capaz de extrair dados de fontes heterogêneas, copiá-los para uma área de transição, onde possam ser convertidos conforme o padrão estabelecido, e, ainda, realizar limpeza e correção nesses dados, de acordo com regras preestabelecidas.

Comentários:

Dados heterogêneos são aqueles dados de tipos diferentes, em contraste com os dados homogêneos, que são aqueles dados de tipos iguais. Uma Ferramenta ETL realmente é capaz de extrair dados de fontes heterogêneas (ou homogêneas), copiá-los para uma área de transição (conhecida como *Stage Area*), onde possam ser convertidos (ou transformados) conforme o padrão estabelecido e, ainda, realizar limpeza e correção nos dados, de acordo com regras preestabelecidas.

Gabarito: Correto

31. (CESPE / STJ – 2018) O data warehouse tem como finalidade a apresentação das informações necessárias para a identificação de indicadores e da evolução de valores ao longo de uma grande janela de tempo.

Comentários:



Muitas vezes, a banca não tem rigor no enunciado. Data Warehouse não tem a finalidade de apresentar nada – ele apenas armazena grandes quantidades de dados em um repositório. Por meio de Ferramentas OLAP, é possível acessar, visualizar e analisar grandes quantidades de informações necessárias para a identificação de indicadores e da evolução de valores ao longo de uma grande janela de tempo, uma vez que o Data Warehouse é um depósito de dados variável com o tempo.

Agora, adivinhem só: o gabarito definitivo da banca foi verdadeiro, no entanto eu vou discordar desse gabarito até o fim...

Gabarito: Correto

32. (CESPE / STM – 2018) O ambiente ETL permite a visualização dos dados para análise de maneira multidimensional.

Comentários:

Opa... ele permite extrair, transformar e carregar dados de diversas fontes. A ferramenta que permite acessar, visualizar e analisar dados para análise de maneira multidimensional é o OLAP.

Gabarito: Errado

33. (CESPE / SEDF – 2017) Comparados aos bancos de dados transacionais, os *data warehouses* são mais voláteis porque, para que se mantenham consistentes, são atualizados em tempo real a cada atualização que ocorrer em qualquer uma das bases originais de dados que o componham.

Comentários:

Comparados aos bancos de dados transacionais, os data warehouses são mais voláteis? Não, essa é uma das características básicas de um Data Warehouse. *Para que se mantenham consistentes, são atualizados em tempo real a cada atualização que ocorrer em qualquer uma das bases originais de dados que o componham?* Não, bancos de dados transacionais são atualizados em tempo real a cada atualização de suas bases de dados originais. Data Warehouses são atualizados de tempos em tempos, sempre com uma defasagem em relação aos bancos de dados transacionais e são não-voláteis, isto é, permanecem estáveis ao longo do tempo.

Gabarito: Errado

34. (CESPE / FUB - 2015) Em um ambiente de *Business Intelligence*, o usuário pode deparar com diferentes requisitos de infraestrutura para atender às necessidades de cada uma das instâncias dos sistemas analíticos.

Comentários:



Um usuário que desejar utilizar os conceitos de BI precisará de uma infraestrutura arquitetural específica capaz de extrair, limpar, formatar, transformar e carregar dados estruturados ou não estruturados de diversas fontes em depósitos de informações que possam ser acessados por sistemas analíticos. Logo, ele acabará se deparando com diferentes requisitos de infraestrutura para atender suas necessidades (Ex: Ferramentas de ETL, Data Warehouses; Ferramentas de Análises de Dados, etc).

Gabarito: Correto

35. (CESPE / ANATEL – 2014) Em um Data Warehouse (armazém de dados) que apresente a característica de ser não volátil, os dados não são atualizados após a inserção inicial.

Comentários:

Nos sistemas transacionais, os dados sofrem diversas alterações como, por exemplo, a inclusão, alteração e exclusão de dados. No ambiente do Data Warehouse, os dados – antes de serem carregados – são transformados e limpos. Após essa etapa, esses dados sofrem somente operações de consulta, sem que possam ser atualizados, sendo essa característica conhecida como não-volatilidade.

Gabarito: Correto

36. (CESPE / TJ-SE – 2014) O processo de extração, transformação e carga, comumente referenciado como ETL (*Extract-Transform-Load*), é um processo usado na criação e na atualização de um Datawarehouse.

Comentários:

A questão está correta, visto que o Processo de ETL é realmente utilizado na criação e inserção de um Data Warehouse. No entanto, a parte de sua redação que menciona “*atualização*” gera uma certa ambiguidade e poderia ser retirada para evitar confusão.

Gabarito: Correto

37. (CESPE / TJ-SE – 2014) O processo de carga de um DataWarehouse consiste exclusivamente em adicionar novos dados históricos regularmente.

Comentários:

A questão dá a entender que a carga deve ocorrer regularmente, o que é falso! A carga pode ocorrer também sob demanda com frequências diversas – sem qualquer regularidade.

Gabarito: Errado



38.(CESPE / ANATEL – 2014) As ferramentas para extração, transformação e carga de dados (ETL) copiam todos os dados dos sistemas operacionais e os transferem para o *Data Warehouse* de forma a apoiar a análise corporativa das tendências e a realização de previsões empresariais.

Comentários:

Não são copiados todos os dados – apenas aqueles que sejam relevantes para a tomada de decisões.

Gabarito: Errado

39.(CESPE / ANTAQ – 2014) O Módulo de ETL (*extract transform load*), dedicado à extração, carga e transformação de dados, coleta informações em fontes como sistemas ERP, arquivos com extensão TXT e planilhas Excel.

Comentários:

Perfeito! A extração pode ocorrer a partir de fontes diversas com Sistemas ERP, Arquivos Texto, Planilhas Excel, entre outros.

Gabarito: Correto

40.(CESPE / ANATEL – 2014) As ferramentas de business intelligence, encarregadas de extrair e carregar dados na base de dados de BI, são denominadas ferramentas de ETL.

Comentários:

Perfeito! ETL é a ferramenta de BI responsável por extrair, transformar e carregar dados em uma base de dados.

Gabarito: Correto

41.(CESPE / TJ-SE – 2014) O processo de transformação de dados pode exigir que dados logicamente relacionados, mas fisicamente separados, sejam recompostos, ainda que envolvam registros distintos ou até mesmo estejam em bancos de dados operacionais distintos.

Comentários:

Aqui cabe uma pequena contextualização: o armazenamento físico de dados em um banco de dados é transparente para o usuário. *Como assim, Diego?* Isso significa que o usuário se importa apenas que seus dados sejam armazenados – ele não se preocupa como será o armazenamento. Dito isso, quando dados são armazenados, mesmo que eles pareçam ser logicamente relacionados



(Ex: você armazena nome, endereço e idade de um aluno em uma tabela), eles podem ser fisicamente armazenados de forma completamente separada.

Logo, o processo de transformação de dados – que é a segunda etapa do processo de ETL – realmente pode exigir que dados logicamente relacionados, mas fisicamente separados, sejam recompostos, ainda que envolvam registros distintos ou até mesmo estejam em bancos de dados operacionais distintos. Lembrando que o processo de transformação aplica uma série de regras ou funções aos dados extraídos antes de carregá-los no Data Warehouse.

Gabarito: Correto

42.(CESPE / TJ-SE – 2014) Em corporações que possuem diversos departamentos, a implementação de DataMarts facilita a disseminação das informações por todos os departamentos.

Comentários:

Data Marts são departamentais, logo eles não facilitam a disseminação das informações por todos os departamentos. O ideal para facilitar a disseminação das informações por todos os departamentos é o Data Warehouse.

Gabarito: Errado

43.(CESPE / SERPRO – 2013) Os datamarts são criados a partir de sistemas transacionais, buscando informações relevantes para o público que irá utilizá-lo.

Comentários:

Péssima questão! Os Data Marts podem ser dependentes ou independentes. A questão trata apenas dos independentes, isto é, aqueles que são criados a partir de fontes transacionais sem a necessidade de um Data Warehouse. Logo, ela restringiu a definição e, para mim, deveria ser anulada. No entanto, a banca considerou no gabarito oficial como verdadeira.

Gabarito: Correto

44.(CESPE / INPI – 2013) Um Data Mart pode ser considerado, na visão bottom-up, uma reunião de vários Data Mining, desde que estes sejam desenvolvidos em modelos multidimensionais (MDM).

Comentários:

Um Data Warehouse ~~Data Mart~~ pode ser considerado, na visão bottom-up, uma reunião de vários Data-Mining Data Marts, desde que estes sejam desenvolvidos em modelos multidimensionais.



Vamos relembrar alguns conceitos: grosso modo, a abordagem de cima para baixo (top-down) trata da construção do Data Warehouse e depois da sua divisão em Data Marts; já a abordagem de baixo para cima (bottom-up) trata da construção de vários Data Marts e da posterior integração em um Data Warehouse.

Gabarito: Errado

45. (CESPE / MPU – 2013) Uma característica distinta dos *data warehouses* é o seu direcionamento para aplicações de apoio às decisões. Eles são otimizados para a recuperação de dados, não para o processamento rotineiro de transações.

Comentários:

Perfeito! Data Warehouse busca dar suporte ao processo de tomada de decisão dos gestores de uma organização. Ademais, eles são extremamente eficientes para recuperação/consulta de dados e geração de relatórios, não para o processamento rotineiro de transações.

Gabarito: Correto

46.(CESPE / MPOG – 2013) O processo de extração, transformação e carga visa trabalhar com a manipulação de dados de fontes externas e sua respectiva transformação, para atender às necessidades de negócios e carga dos dados dentro de uma estrutura de data warehouse.

Comentários:

Que questão linda! Definiu perfeitamente o Processo de ETL. Apenas para martelar mais ainda na cabeça de vocês: o processo de extrair dados de sistemas de origem e trazê-los para o Data Warehouse é comumente chamado de ETL (*Extract, Transform e Load*). Extract: extrai as informações do banco de dados transacional; Transform: transforma as informações para o formato adequado; e Load: carrega as informações no Data Warehouse.

Gabarito: Correto

47.(CESPE / MPOG – 2013) Em ETL, frequentemente, é necessário limpar, ajustar e consolidar os dados antes de realizar sua carga. Nesse estágio de transformação, aplicam-se regras ou funções aos dados extraídos para ajustar os dados a serem carregados. A limpeza trata de vários tipos de erros, como, por exemplo, valores ilegais, ou que não obedecem às regras de integridade da base, e erros de ortografia.

Comentários:



Nem há muito para comentar de tão perfeita que está essa questão. A transformação realmente aplica regras aos dados extraídos de forma a ajustá-los antes da carga. Já a limpeza corrige erros de diversos tipos.

Gabarito: Correto

48.(CESPE / ANTT – 2013) Ferramentas ETL são utilizadas na extração, transformação e remoção de dados.

Comentários:

Ferramentas ETL são utilizadas na Extração, Transformação e Carga – em regra, dados não são excluídos de um Data Warehouse.

Gabarito: Errado

49.(CESPE / Banco da Amazônia – 2012) O processo de extração, transformação e carga (ETL) ilustrado permite extrair dados de diversas fontes de dados e migrá-los para o data warehouse, mantendo sua estrutura e normalização originais.

Comentários:

Durante a etapa de transformação do Processo de ETL, os dados geralmente são desnormalizados antes de serem migrados para o Data Warehouse. Logo, não se mantém a estrutura e normalização originais.

Gabarito: Errado

50.(CESPE / TJ - AC – 2012) Por meio de um data mart, que é um subconjunto do data warehouse, é possível criar uma visão personalizada dos dados.

Comentários:

O Data Mart é considerado um pequeno Data Warehouse – ou um subconjunto/segmentação de um Data Warehouse – que contém dados de uma organização em geral restritos a uma unidade de negócio ou a um processo de negócio específico com a finalidade de dar suporte ao processo decisório da alta direção por meio da criação de uma visão personalizada dos dados. De acordo com Date, trata-se de um repositório especializado e volátil!

Gabarito: Correto

51. (CESPE / MPE-PI – 2012) Na construção de um data warehouse, o processo extract, transform and load (ETL), normalmente, é o menos crítico e o mais rápido. Esse processo envolve a



extração dos dados, na sua transformação, e, eventualmente, influencia na limpeza desses dados.

Comentários:

Na verdade, o Processo de ETL é o **processo mais crítico** e demorado na construção de um Data Warehouse, uma vez que consiste na extração dos dados de fontes homogêneas ou heterogêneas; na transformação e limpeza destes dados; e na carga dos dados no DW. Pessoal, as decisões estratégicas – aquelas mais importantes de uma organização – são tomadas com base nas informações geradas através dos dados armazenados no Data Warehouse.

Gabarito: Errado

52. (CESPE / CORREIOS - 2011) Inteligência empresarial, ou *business intelligence*, é um termo utilizado para descrever as habilidades das corporações para coletar dados e explorar informações, analisá-las e desenvolver entendimentos para tomada de melhores decisões.

Comentários:

Definição retirada *ipsis litteris* do próprio Gartner Group – que foi quem cunhou esse termo.

Gabarito: Correto

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

Comentários:

De fato, é possível a análise de grandes volumes de dados, no entanto os dados ficam disponíveis apenas para serem consultados ou eliminados e, não, alterados e manipulados pelo usuário.

Gabarito: Errado

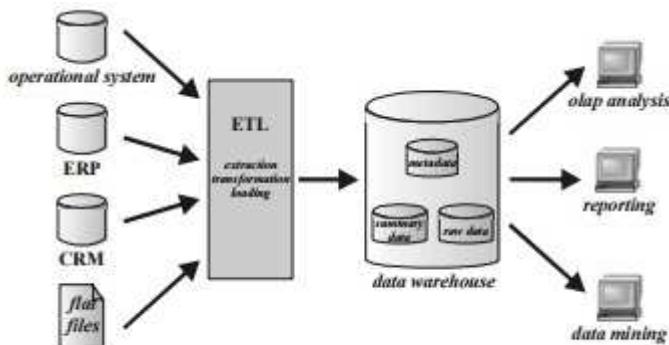
54. (CESPE / CORREIOS – 2011) As ferramentas de software ETL (*extract transform load*) têm como função a extração de dados de diversos sistemas, a transformação desses dados de acordo com as regras de negócio e a carga dos dados em um *data mart* ou um DW.

Comentários:

Item impecável! Não há mais nada a acrescentar – excelente definição de ETL!

Gabarito: Correto





Internet: <www.infoescola.com/informatica/data-warehouse>.

55. (CESPE / Correios – 2011) Considere que, em uma organização que mantenha uma solução de DW, haja grande quantidade de dados para serem manipulados e que uma nova área criada dentro dessa organização necessite acessar parte dos dados desse DW.

Nessa situação, seria adequada a implementação de um *datamart* específico para essa nova área.

Comentários:

O julgamento dessa questão é bastante sutil! Notem que ela trata de um contexto em que já existe um DW com uma grande quantidade de dados de uma organização e que uma nova área dessa organização deseja acessar parte dos dados do DW. Um Data Mart é um subconjunto de um Data Warehouse com dados referentes a uma área específica. *Logo, seria adequada a implementação de um Data Mart específico para essa nova área?* Sim, seria adequado!

Alguns alunos já argumentaram que a questão trata apenas do acesso a parte dos dados do DW! Sim, é possível solucionar esse problema por meio de controles de acesso para os usuários dessa área da organização sem a necessidade de criar um novo Data Mart. No entanto, isso não invalida a questão: a implementação de um Data Mart específico para essa nova área continua sendo uma implementação adequada para o problema (adequada e, não, obrigatória).

Gabarito: Correto

56. (CESPE / SERPRO – 2010) No ambiente de data warehouse, os dados consistem de um snapshot (retrato no tempo) e abrangem períodos de até 10 anos, sendo atualizados periodicamente, e não em tempo real.

Comentários:

Galera, essa questão está desatualizada! À época, o snapshot de um data warehouse realmente abrangia períodos de 10 anos. *Professor, por que você está colocando a questão aqui se ela está desatualizada?* Porque já recebi diversas perguntas no fórum e toda vez tenho que explicar que isso



não faz sentido atualmente. Logo, se vocês virem essa questão em algum site de questões, saibam que esse período de 10 anos já faz mais sentido atualmente.

Gabarito: Correto

57. (CESPE / MPE - PI – 2010) Um *data warehouse* é um repositório de informações de uma fonte única, armazenadas sob diversos esquemas, em um só local, e que fornecem dados consolidados.

Comentários:

Data Warehouse é um repositório de informações de uma fonte única? Não, é um repositório de informações de fontes diversas – homogêneas ou heterogêneas.

Gabarito: Errado

58. (CESPE / INMETRO – 2010) Assinale a opção correta com relação aos conceitos de *data mart*.

- a) O data warehouse é constituído de apenas um data mart.
- b) O data mart pode ser visto como a restrição do data warehouse a um único processo de negócios.
- c) Todo data mart deve ser representado por um modelo orientado a objetos.
- d) O data mart possui uma estrutura diferente do data warehouse.
- e) O que diferencia um data mart de um data warehouse é a forma de acesso aos dados.

Comentários:

(a) Errado. Um Data Warehouse pode ser constituído de diversos Data Marts; (b) Correto. Pode ser uma restrição a uma unidade, assunto ou processo de negócio específico; (c) Errado. Em geral, eles utilizam modelos multidimensionais; (d) Errado. Data Mart e Data Warehouse possuem estruturas semelhantes; (e) Errado. A forma de acesso aos dados também é semelhante.

Gabarito: Letra B

59. (CESPE / BANCO DA AMAZÔNIA – 2010) Por fazerem uso de um processo de modelagem dimensional, os *data warehouses* não podem ser instalados em sistemas de banco de dados relacionais.

Comentários:

Data Warehouses podem ser implementados tanto em bancos de dados multidimensionais quanto em bancos de dados relacionais.



Gabarito: Errado

60.(CESPE / BANCO DA AMAZÔNIA – 2010) Um data mart é uma reunião de vários data warehouses, a fim de fornecer visão mais ampla dos dados.

Comentários:

Na verdade, um Data Mart é um subconjunto de um Data Warehouse, fornecendo uma visão mais restrita e departamental dos dados.

Gabarito: Errado

61.(CESPE / INMETRO - 2009) Entre as operações típicas realizadas em uma organização que adota a abordagem de *business intelligence*, destaca-se o uso de mineração de dados e a construção de *data warehouses*, ambas empregando bancos de dados relacionais com elevado grau de normalização.

Comentários:

Opa! Ambas empregam – em geral – bancos de dados multidimensionais com baixo grau de normalização (alta redundância de dados).

Gabarito: Errado

62.(CESPE / ANTAQ - 2009) Aplicações de *business intelligence* (BI) oferecem visões históricas e atuais de operações de negócios empregando unicamente dados operacionais.

Comentários:

Na verdade, emprega dados gerenciais consolidados a partir de dados operacionais. Logo, de forma genérica, emprega tanto dados operacionais quanto dados gerenciais.

Gabarito: Errado

63.(CESPE / SERPRO – 2008) Os *data warehouses* são depósitos de dados tipicamente direcionados para aplicações de apoio às decisões administrativas. As informações em *data warehouses* mudam mais freqüentemente do que em bancos de dados convencionais. Além disso, os *data warehouse*, em vez de otimizar a recuperação de dados, otimizam o processamento de transações.

Comentários:



As informações em *data warehouses* mudam mais freqüentemente do que em bancos de dados convencionais? Claro que não! Data Warehouses são não voláteis, enquanto bancos de dados convencionais (transacionais) mudam frequentemente suas informações.

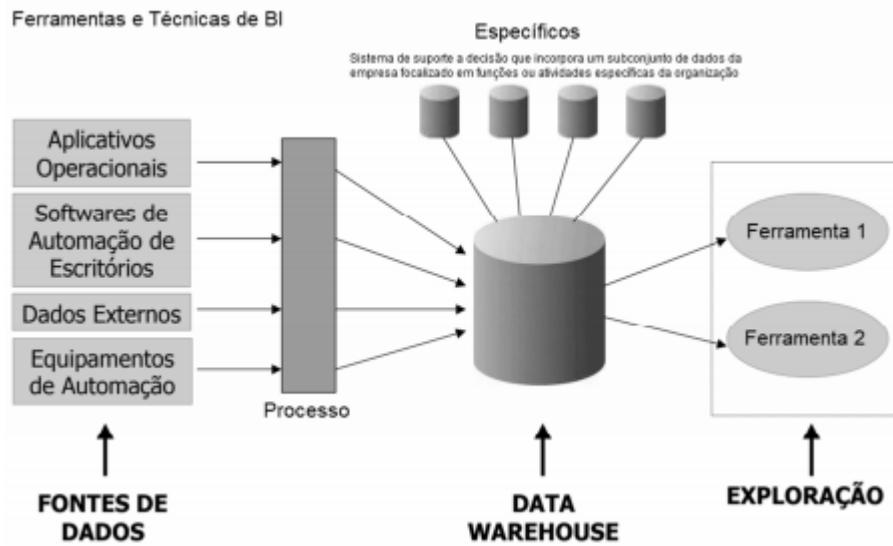
Os data warehouse, em vez de otimizar a recuperação de dados, otimizam o processamento de transações? Pelo contrário, bancos de dados transacionais são otimizados para o processamento de transações, enquanto Data Warehouses são otimizados para recuperação de dados.

Gabarito: Errado



QUESTÕES COMENTADAS – FCC

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



O Processo, representado na imagem por um retângulo vertical, é um método de alimentação do Data Warehouse a partir de diversos dados da organização. Trata-se de:

- a) ODS.
- b) ERP.
- c) ETL.
- d) CRM.
- e) EIS.

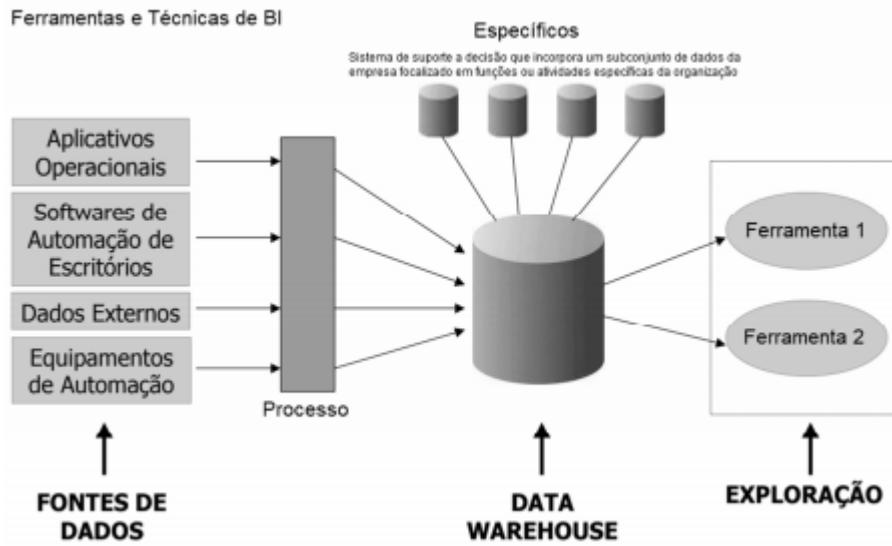
Comentários:

O processo que extrai dados de fontes de dados diferentes, transforma-os e os carrega em um Data Warehouse é o Processo de ETL.

Gabarito: Letra C

65.(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.

Comentários:

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

Gabarito: Letra E

66. (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

67. (FCC / DPE - AM – 2018) Sobre o processo de ETL, aplicado a data warehouse, é correto afirmar que:

- a) a fase de extração de dados consiste em obter os dados do servidor do data warehouse.
- b) a fase de transformação consiste em realizar modificações nos dados carregados, adequando seus valores ao modelo definido para o data warehouse.
- c) as fases de extração e carga de dados são realizadas de forma simultânea.
- d) a fase de carga de dados visa eliminar valores nulos contidos nos bancos de dados transacionais da empresa.
- e) a fase de carga de dados consiste em inserir os dados transformados nos bancos de dados transacionais da empresa.

Comentários:

(a) Errado. Dados são obtidos de fontes diversas, tratados e inseridos no Data Warehouse; (b) Correto. É a fase de transformação que é realizada a conversão e padronização dos dados carregados na área de *stage* conforme o Data Warehouse; (c) Errado. Primeiro é realizada a extração, depois a transformação e, em seguida, a carga; (d) Errado. A fase de transformação é responsável por eliminar valores nulos de um banco de dados transacional; (e) Errado. São extraídos de bancos de dados transacionais e inseridos no Data Warehouse.

Gabarito: Letra B



68. (FCC / PREF SÃO LUÍS – 2018) Para extrair dados de fontes de dados heterogêneas que irão alimentar um *Data Warehouse* de forma homogênea e concisa, servindo de base para gerar relatórios e gráficos para apoiar as decisões da gerência da organização, deve-se utilizar um processo conhecido como:

- a) OLAP.
- b) Data Mart.
- c) ETL.
- d) OLTP.
- e) Data Mining.

Comentários:

Vamos analisar as palavras-chave: extrair dados; fontes heterogêneas; alimentar um DW; gerar relatórios e gráficos; apoiar as decisões. Todas elas nos remetem ao Processo de ETL.

Gabarito: Letra C

69. (FCC / TRT 14ª REGIÃO – 2016) Quando uma empresa utiliza *Data Warehouse* (DW) é necessário carregá-lo para permitir a análise comercial. Para isso, os dados de um ou mais sistemas devem ser extraídos e copiados para o DW em um processo conhecido como:

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

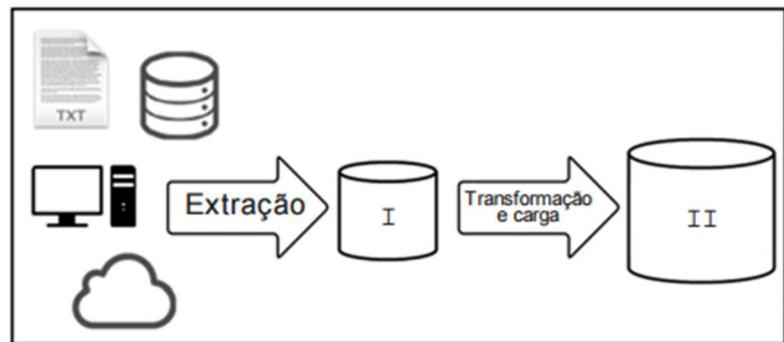
Comentários:

O processo de extrair dados de sistemas de origem e trazê-los para o *Data Warehouse* é comumente chamado de ETL (*Extract, Transform e Load*). Extract: extrai as informações do banco de dados transacional; Transform: transforma as informações para o formato adequado; e Load: carrega as informações no *Data Warehouse*. Observem que a questão nos oferece duas dicas: primeiro, quando fala em “carregá-lo”; e segundo, quando fala em “extraídos”.

Gabarito: Letra D

70. (FCC / AL - MS – 2016) Considere a figura abaixo.





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

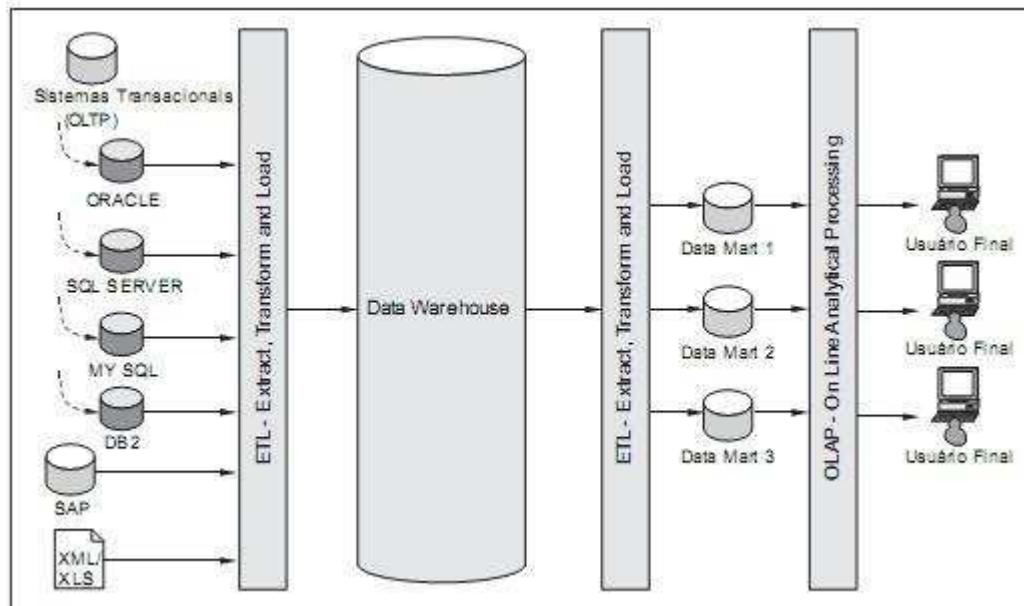
- a) OLTP e Data Warehouse.
- b) OLTP e Staging Area.
- c) Data Mart e Staging Area.
- d) Staging Area e OLTP.
- e) Staging Area e Data Warehouse.

Comentários:

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

Gabarito: Letra E

71. (FCC / AL - MS – 2016) Considere que a equipe de Analistas de Desenvolvimento de Sistemas do CNMP está projetando a arquitetura para o Data Warehouse (DW) da instituição, conforme mostra a figura abaixo:



É correto afirmar que esta arquitetura:

- a) é *bottom-up*. Garante a existência de um único conjunto de aplicações para ETL, ou seja, extração, limpeza e integração dos dados, embora os processos de manutenção e monitoração fiquem descentralizados.
- b) é *bottom-up*, pois primeiro a equipe cria um DW e depois parte para a segmentação, ou seja, divide o DW em áreas menores gerando pequenos bancos orientados por assuntos aos departamentos.
- c) é *bottom-up*. Permite um rápido desenvolvimento, pois a construção dos *Data Marts* é altamente direcionada. Normalmente um *Data Mart* pode ser colocado em produção em um período de 2 a 3 meses.
- d) é *top-down*. A partir do DW são extraídos os dados e metadados para os *Data Marts*. Nos *Data Marts* as informações estão em maior nível de sumarização e, normalmente, não apresentam o nível histórico encontrado no DW.
- e) é *top-down*, pois possui um retorno de investimento muito rápido ou um *faster pay back*. O propósito desta arquitetura é a construção de um DW incremental a partir de *Data Marts* independentes.

Comentários:

Acompanhando a figura da esquerda para a direita, é possível visualizar que os dados são extraídos de diversas fontes (OLTP, ORACLE, SQL SERVER, MYSQL, DB2, SAP, etc) e são carregados em um Data Warehouse. Em seguida, eles são extraídos, transformados e carregados em vários Data Marts. Logo, trata-se de uma abordagem Top-Down (DW > DMs)! A partir do DW são extraídos os dados e metadados para os Data Marts. Nos Data Marts as informações estão em maior nível de sumarização e, normalmente, não apresentam o nível histórico encontrado no DW. Trata-se de abordagem de Inmon ou Kimball? Trata-se da abordagem de Inmon!

Gabarito: Letra D

72. (FCC / DPE - RS – 2017) O processo de ETL é composto por algumas fases, identificadas como entregar, limpar, extraír e preparar. A ordem correta para a execução dessas fases é:

- a) limpar, preparar, extraír e entregar.
- b) preparar, limpar, entregar e extraír.
- c) entregar, preparar, limpar e extraír.
- d) entregar, extraír, limpar e preparar.
- e) extraír, limpar, preparar e entregar.

Comentários:



(a) Errado. Como você vai limpar antes de extrair? (b) Errado. Como você vai preparar antes de extrair? (c) Errado. Como você vai entregar antes de extrair? (d) Errado. Como você vai entregar antes de extrair? (e) Correto. Extração, Limpeza, Preparação e Entrega.

Gabarito: Letra E

73. (FCC / TRE - SP – 2017) No processo ETL de um *data warehouse*:

(I) a tradução de valores codificados conhecida como limpeza de dados (por exemplo, supondo que o sistema de origem armazena 1 para sexo masculino e 2 para feminino, mas o *data warehouse* armazena M para masculino e F para feminino) e (II) a geração de valores de chaves substitutas (*surrogate keys*) são:

- a) integrantes da etapa de transformação.
- b) integrantes das etapas de transformação e carga, respectivamente.
- c) manipulações de dados, aplicáveis no processo *pivoting*.
- d) elementos que possibilitam a conversão de *drill-down* para *roll up*.
- e) integrantes da etapa de carga.

Comentários:

A tradução de valores codificados conhecida como limpeza de dados é parte da etapa de Transformação do Processo ETL – assim como a geração de chaves substitutas. Pessoal, uma chave substituta é uma chave artificial com características semelhantes às chaves primárias que ajuda na modelagem de um Data Warehouse.

Gabarito: Letra A

74. (FCC / TRT 3ª REGIÃO – 2015) Ao contrário do que muitas empresas pensam, um projeto de data warehouses não necessita ser imenso e exigir um grande investimento financeiro de início. Ele pode começar atendendo um subconjunto da organização. Igualmente, isso vale para empresas de médio porte, que podem se beneficiar dessa tecnologia. Isso só é possível porque, quando se conceitua data warehouse, tem-se também o conceito de:

- a) Data Mart.
- b) Data Center.
- c) Data Warehouse Virtual.
- d) Banco de Dados em Nuvem.
- e) Data Warehouse Empresarial.

Comentários:



De fato, é possível começar com Data Marts em vez de Data Warehouses. Como eles são menores, o investimento financeiro inicial também é reduzido, o que é ideal para empresas de médio porte.

Gabarito: Letra A

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

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

Comentários:

Notem as dicas que o enunciado fornece: subconjunto de dados de um Data Warehouse; direcionado à área administrativa; banco de dados multidimensional; capturando dados de sistemas transacionais; informações relevantes para a área administrativa. Todas essas dicas nos remetem ao conceito de Data Mart.

Gabarito: Letra E

76. (FCC / DPE-RS – 2014) O Data Warehouse (DW) é um tipo especial de banco de dados que proporciona uma sólida e concisa integração dos dados de uma empresa para a realização de análises gerenciais estratégicas de seus principais processos de negócio. O DW é um depósito de dados:

- a) especializado, orientado por assunto, integrado, variável no tempo e volátil. Os usuários podem atualizar os dados e até criar novos dados, ou seja, novas tabelas, para algum propósito específico de interesse gerencial.
- b) orientado por assunto, integrado, variável com o tempo e não volátil. Uma vez inseridos, os dados não podem ser alterados, o que implica não ser necessário nenhum tipo de bloqueio por concorrência de usuários no acesso.
- c) orientado por assunto, integrado, variável com o tempo e volátil. Os dados de um DW variam em relação ao tempo, pois representam resultados operacionais que refletem a evolução da empresa. Isso implica que os dados devem ser constantemente atualizados.



d) orientado por processos, integrado, variável com o tempo e volátil. Um DW armazena as informações agrupadas por processos de negócios que são atividades desenvolvidas na organização, como transformação de entrada e saída com agregação de valor, essenciais para a tomada de decisões.

e) especializado, orientado por assunto, integrado, variável com o tempo e não volátil. Em ambientes de múltiplas plataformas sistêmicas, a característica de integração se torna fundamental, pois a organização necessita manter os diferentes padrões existentes entre os diferentes sistemas operacionais.

Comentários:

(a) Errado. Data Warehouses são não-voláteis; (b) Correto. 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. *Por que que implica não ser necessário nenhum tipo de bloqueio por concorrência de usuários no acesso?* Porque o controle de concorrência é utilizado em bancos de dados tradicionais para evitar que usuários diferentes tentem manipular uma mesma informação – como DWs são utilizados basicamente para consultas, não é necessário ter esse tipo de preocupação; (c) Errado. Data Warehouses são não-voláteis e dados não devem ser constantemente atualizados; (d) Errado. Data Warehouses são orientados por assuntos e não-voláteis; (e) Errado. Data Warehouses são genéricos, contendo diversas fontes de dados e diversos assuntos.

Gabarito: Letra B

77. (FCC / MPE-CE - 2013) Em relação ao entendimento do significado do termo Business Intelligence (BI) e da solução que provê, a definição que NÃO é coerente com o termo *Business Intelligence* é a que:

- a) consiste em uma metodologia que fornece objetivos de negócios ligados a objetivos de TI, provendo métricas e modelos de maturidade para medir a sua eficácia e identificando as responsabilidades relacionadas dos donos dos processos de negócios e de TI.
- b) se refere à aplicação de técnicas analíticas para informações sobre condições de negócio no sentido de melhorá-las, de uma maneira automatizada, mas com a interpretação e respostas humanas, de forma a melhorar a tomada de decisões.
- c) reúne recursos que provêm a habilidade para que a pessoa certa receba a informação adequada e no momento correto para tomar a melhor decisão.
- d) consiste em um sistema de negócios que inclui uma estrutura de busca efetiva e acessível, acurada, em tempo real, com informações e relatórios que permitam aos líderes das áreas de negócio se manterem informados para tomar decisões.



e) é uma solução fácil de dizer, mas difícil de fazer corretamente pois envolve mudanças na forma como a organização conduz uma busca efetiva, bem como, a necessidade de se possuir uma base de dados de qualidade para que se possa tomar ações com o objetivo de otimizar a performance corporativa.

Comentários:

Metodologia que fornece objetivos de negócios (?) ligados a objetivos de TI (?), provendo métricas e modelos de maturidade (?) para medir a sua eficácia e identificando as responsabilidades relacionadas dos donos dos processos de negócios e de TI (?). Galera, tudo isso trata de um assunto que não tem absolutamente nada a ver com Business Intelligence – isso é a definição de COBIT (que é um framework de boas práticas em governança de TI).

Gabarito: Letra A

78.(FCC / TST – 2012) O processo de ETL em uma Data Warehouse possui várias fases. Em uma destas fases é efetuada a:

- a) extração dos dados dos sistemas de origem.
- b) introdução de novos produtos no mercado.
- c) validação das interfaces de usuário.
- d) criação de diagramas estáticos e comportamentais das classes e atributos.
- e) definição dos custos e prazos.

Comentários:

O Processo de ETL possui três fases: Extração, Transformação e Carga. A etapa de Extração deve se basear na busca das informações mais importantes em aplicações ou fontes externas. Nenhum dos outros itens faz qualquer sentido lógico.

Gabarito: Letra A

79.(FCC / INFRAERO - 2011) É uma característica de um sistema de *Business Intelligence*:

- a) Capacidade de cruzar informações de diferentes bancos de dados, gerando relatórios analíticos diversos.
- b) Utilização de comandos SQL para a geração de tabelas dinâmicas inteligentes ligadas às regras de negócio.
- c) Utilização de Sistemas Baseados em Conhecimento, onde é possível utilizar técnicas de Inteligência Artificial na geração de relatórios.



d) Análise automática de requisitos funcionais e não funcionais, permitindo a geração de relatórios gerenciais.

e) Análise de informações com o intuito de fornecer subsídio para a criação de sistemas inteligentes.

Comentários:

(a) Correto, ele realmente permite cruzar informações de fontes diversas para gerar relatórios; (b) Errado, BI é um processo ou metodologia e não cabe utilizar comandos SQL; (c) Errado, essa não é uma característica de Business Intelligence; (d) Errado, essa não é uma característica de Business Intelligence; (e) Errado, essa não é uma característica de Business Intelligence;

Gabarito: Letra A

8o.(FCC / TRE-PE – 2011) Um processo importante que ocorre em relação à formação de um data warehouse é a obtenção dos dados de uma ou mais bases de dados da origem. Deve ser rigoroso para evitar a deformação e/ou a perda dos dados quando passados da fonte original para o destino. Trata-se de:

- a) MINING.
- b) DATA MART.
- c) MOLAP.
- d) STAR.
- e) ETL.

Comentários:

A questão trata do Processo de ETL: Extração, Transformação e Carga. No caso, trata com mais foco na fase de extração.

Gabarito: Letra E

81.(FCC / TRT 21ª REGIÃO – 2003) Um banco de dados organizado em estruturas lógicas dimensionais cujas informações são voltadas para o processo decisório e para áreas específicas denomina-se:

- a) data warehouse.
- b) data mining.
- c) data mart.
- d) business intelligence.
- e) competitive intelligence.



Comentários:

Banco de dados? Estruturas dimensionais? Processo decisório? Áreas específicas? São todas palavras-chaves que nos remetem ao Data Mart. O Data Mart é considerado um pequeno Data Warehouse – ou um subconjunto/segmentação de um Data Warehouse – que contém dados de uma organização em geral restritos a uma unidade de negócio ou a um processo de negócio específico com a finalidade de dar suporte ao processo decisório da alta direção por meio da criação de uma visão personalizada dos dados.

Gabarito: Letra C



QUESTÕES COMENTADAS – FGV

82. (FGV / TJDFT – Análise de Sistemas – 2022) Lino precisa preparar um grande volume de dados para minerar realizando operações de limpeza, adequação de formato, exclusão de brancos e inclusão de novos atributos derivados. Para realizar o pré-processamento dos dados, Lino deve usar uma ferramenta do tipo:

- a) ETL;
- b) OLAP;
- c) Apriori;
- d) Data Mart;
- e) Data Lake.

Comentários:

O pré-processamento a ser realizado por Lino é feito por meio de ferramentas de ETL (Extract, Transform, Load) – particularmente na etapa de Transformação. Essa etapa é responsável por compilar, converter, reformatar, limpar os dados, entre outros.

Gabarito: Letra A

83. (FGV / CGU – 2022) Uma organização deseja implementar um pipeline de dados e está avaliando a opção mais adequada para o seu contexto de operação. Em torno de 40% dos dados consumidos pela organização se encontram em planilhas eletrônicas que contêm dados sensíveis, produzidas semanalmente por suas unidades de negócio. Os outros 60% dos dados se encontram em alguns bancos de dados relacionais de sistemas de produção da organização. O tamanho da base é de moderado a pequeno, mas existe a necessidade de conformidade com normas de privacidade e confidencialidade dos dados. O objetivo do pipeline é fornecer insumos para um departamento que realiza análises de dados com métodos não supervisionados de aprendizagem de máquina para elaborar relatórios periódicos mensais. A organização está avaliando a construção de um Armazém de Dados (ETL) ou de um Lago de Dados (ELT).

A proposta de modelo adequada e corretamente justificada é:

- a) Armazém de Dados. Ambos os modelos são adequados, mas Lago de Dados tem maior latência até a carga (L) e custo maior;
- b) Armazém de Dados. Esse modelo possui menor latência até a carga (L) e, ao contrário do Lago de Dados, opera de forma eficiente com dados relacionais;
- c) Armazém de Dados. O processo ETL é mais adequado para o tratamento dos dados sensíveis e os casos de uso são bem conhecidos;



d) Lago de Dados. Esse modelo possui menor latência até a carga (L) e permite a extração (E) de dados semiestruturados e não estruturados;

e) Lago de Dados. Esse modelo não necessita de hardware especializado e, ao contrário do Armazém de Dados, possibilita tarefas de aprendizado de máquina.

Comentários:

(a) Errado. Lago de Dados tem uma menor latência até a carga, além de menor custo; (b) Errado. Armazém de Dados tem uma latência maior até a carga, dado que a transformação ocorre antes; (c) Correto. ETL é mais adequado para o tratamento de dados sensíveis, dado que – antes do carregamento – os dados são transformados (possivelmente anonimizados), o que não ocorre no ELT; (d) Errado. A proposta de modelo adequada e corretamente justificada para armazenamento de dados sensíveis é o Armazém de Dados; (e) Errado. Ambos possuem tarefas de aprendizado de máquina.

Gabarito: Letra C

84.(FGV / SEFAZ-AM – 2022) Os *data warehouses* existem para facilitar que analista e tomadores de decisão executem as consultas complexas e ocasionais, fazendo uso intensivo de dados por meio de processamento OLAP.

A operação analítica que se caracteriza por analisar dados em níveis de agregação progressivamente mais detalhados e de menor granularidade, é denominada:

- a) *pivot*
- b) *roll-up*.
- c) *drill-down*.
- d) *drill-across*
- e) *slice and disse*.

Comentários:

A operação analítica que se caracteriza por analisar dados em níveis de agregação progressivamente mais detalhados e de menor granularidade é o *drill-down*.

(a) Errado, essa operação realiza uma rotação no cubo a fim de alterar a posição das dimensões; (b) Errado, é utilizado para aumentar a granularidade e apresentar menos detalhes; (c) Correto; (d) Errado, essa operação permite navegar com saltos entre os níveis de uma mesma dimensão; (e) Errado, essa operação não existe – o nome correto é *slice-and-dice*, que permite extrair um subcubo.



Gabarito: Letra C

85. (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: Letra E

86. (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:

Acho que já ficou claro como é importante a 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: Letra B



87. (FGV / ALERJ – 2017) Daniel está desenvolvendo um *Data Warehouse* para analisar os dados do Censo Escolar. A fonte de dados está em um arquivo CSV e descrita em um documento, conforme parcialmente ilustrado nas figuras a seguir.

ID_SERIE	CO_TURNO	HR_INICIAL	HR_FINAL	NO_TURMA
3	1	7	11	3º ANO A E...
3	1	7	11	3º ANO A E...
3	1	7	11	3º ANO A E...
3	1	7	11	3º ANO A E...
3	1	7	11	3º ANO A E...

Nome	Descrição	Valor	Descrição do Valor
CO_TURNO	Código do Turno	1	Matutino
		2	Vespertino
		3	Noturno

Para carregar esses dados no *Data Warehouse* com a descrição dos turnos no lugar de seu código, Daniel deve desenvolver um programa para ler os dados do arquivo, realizar transformações e carregar o resultado no banco de dados.

A ferramenta a ser utilizada por Daniel é:

- a) ETL;
- b) OLAP;
- c) Data Mining;
- d) ODBC;
- e) XSLT.

Comentários:

Galera, essa é uma questão bem legal de analisar! Daniel quer carregar esses dados no Data Warehouse, no entanto ele quer carregar a descrição dos turnos ao invés de seu código. Professor, o que isso quer dizer? Vejam a tabela de baixo: trata-se de uma tabela que contém dados sobre o campo Turno. Ela possui o código do turno, a descrição do que é um turno, os valores que um turno pode assumir e a descrição de cada valor que um turno pode assumir.

Dessa forma, o campo Turno tem como código CO_TURNO (Coturno? Hehehe); pode ser descrito como código do turno; pode possuir três valores diferentes (1, 2, 3); e esses valores podem ser descrito respectivamente como (Matutino, Vespertino e Noturno). Agora vejam que a tabela de cima apresenta o valor do código do turno (1, 2, 3), e Daniel deseja mostrar a descrição em vez do valor do código. Convenhamos que é mais fácil entender “Noturno” do que “3”. Concordam?

Para transformar esses valores de código em descrição antes de carregar no Data Warehouse, Daniel deve fazer uma ETL: extrai os dados do Arquivo CSV, realiza algumas transformações (Código > Descrição) e carrega no DW.



Gabarito: Letra A

88. (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: Letra A

89. (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: Letra E

90.(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: Letra E

91.(FGV / MEC – 2009) As alternativas a seguir apresentam maneiras de recuperar informações de um Data Warehouse, à exceção de uma. Assinale-a.

- a) OLAP.
- b) Data Mart.
- c) Data Mining.
- d) EIS (Executive Information Systems).
- e) Ferramentas de consulta e emissão de relatórios.

Comentários:

Todas essas opções permitem recuperar informações de um Data Warehouse, exceto o Data Mart. O Data Mart é um sub-conjunto setorial de um Data Warehouse. Lembrando que EIS (Sistema de Informações Executivas) é um tipo de sistema de suporte à decisão que fornece acesso facilitado a informações relevantes de uma organização.



Gabarito: Letra B

92.(FGV / MEC – 2009) No que diz respeito ao desenvolvimento de sistemas de informação e da análise de negócios, analise a citação a seguir.

"É um conjunto integrado de dados, extraídos de bancos de dados operacionais, históricos, externos e selecionados, editados e padronizados para recuperação e análise, para fornecer inteligência empresarial e nos negócios no processo de tomada de decisões gerenciais."

Essa citação representa o conceito de:

- a) Data Smart.
- b) Data Mining.
- c) Data Business.
- d) Data Warehouse.
- e) Data Information.

Comentários:

Conjunto integrado de dados? Extraídos de bancos de dados operacionais? Históricos? Para fornecer inteligência empresarial? Tomada de decisão gerencial? Todas essas são características típicas de um Data Warehouse (Obs: Data Smart é engraçado...).

Gabarito: Letra D

93.(FGV / SEFAZ-RJ – 2009) O grande desafio do profissional de TI que gerencia qualquer processo é a análise dos fatos relacionados à função que exerce em uma organização. Essa análise deve ser feita com as ferramentas e os dados disponíveis, permitindo aos executivos e gerentes detectar as tendências e tomar as decisões com eficiência e eficácia. Devido a essa necessidade, surgiu o conceito de Business Intelligence - "BI". Assinale a alternativa que indique duas características dos atuais sistemas de Business Intelligence.

- a) procurar relações de causa e efeito / extrair e integrar dados de múltiplas fontes.
- b) evitar a utilização de ferramentas automatizadas / desprezar dados contextualizados.
- c) extrair e integrar dados de múltiplas fontes / evitar a utilização de ferramentas automatizadas.
- d) desprezar dados contextualizados / trabalhar exclusivamente com fatos reais e não hipotéticos.
- e) trabalhar exclusivamente com fatos reais e não hipotéticos / procurar relações de causa e efeito.

Comentários:



(a) Correto; (b) Errado, incentiva a utilização de ferramentas automatizadas e aprecia dados contextualizados; (c) Errado, incentiva a utilização de ferramentas automatizadas; (d) Errado, aprecia dados contextualizados e não trabalha exclusivamente com fatos reais e não hipotéticos; (e) Errado, ele não trabalha exclusivamente com fatos reais e não hipotéticos.

Gabarito: Letra A



QUESTÕES COMENTADAS – DIVERSAS BANCAS

94.(FEPESE / ISS-Criciúma – 2022) No processo de criação de um Data Warehouse, quais etapas devem ser cumpridas, na ordem correta, de modo a se obter os dados a partir dos sistemas transacionais para um melhor resultado?

- a) Definição; planejamento; execução; gestão.
- b) Planejamento; levantamento de requisitos; execução; monitoramento.
- c) Planejamento; extração; gestão e acompanhamento; melhoria contínua.
- d) Extração; Limpeza e transformação; entrega e carga; gestão e acompanhamento.
- e) Limpeza e transformação; gestão e acompanhamento; atuação e correção; medição.

Comentários:

No processo de criação de um Data Warehouse, alguns passos são bem consolidados na bibliografia consagrada: Extração, Limpeza/Transformação, Entrega/Carga. Eu não encontrei nos textos de nenhum autor consagrado as etapas de Gestão e Acompanhamento, mas o item (d) é realmente o mais próximo da realidade – dado que apresenta o processo ETL. Nenhum dos outros itens faz muito sentido, então aqui temos que ir pelo bom senso!

Gabarito: Letra D

95.(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: Letra E

96. (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: Letra B

97. (CESGRANRIO / BB – 2021) Os sistemas interativos que provêm inteligência de negócios, 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: Letra C

98. (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: Letra C

99. (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: Letra D

100. (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: Letra B

101. (CESGRANRIO / TRANSPETRO – 2018) Os sistemas de apoio à decisão devem prover um conjunto de capacidades, tais como:



- a) ser resiliente contra condições temporais mutantes.
- b) prescindir do julgamento humano no processo de análise.
- c) permitir a fácil execução das análises de sensibilidade.
- d) apoiar decisões interdependentes, exceto as sequenciais.
- e) evitar sua utilização em ambientes voltados para internet (web).

Comentários:

(a) Errado, deve ser flexível com relação a condições temporais mutantes (que se modificam); (b) Errado, ele não substitui a intervenção humana; (c) Correto, permite a fácil execução de análises de sensibilidade de forma tempestiva; (d) Errado, apoia decisões interdependentes, inclusive as sequenciais; (e) Errado, é fomentada sua utilização nesses ambientes.

Gabarito: Letra C

102. (IESES / CEGÁS – 2017) Assinale a alternativa que conceitua corretamente *Data Warehouse*:

- a) É um processo analítico em tempo real que transforma dados em informação. Está orientada a assuntos e utiliza tecnologias semânticas.
- b) É um depósito de dados digitais, orientado por assunto que, transforma dados operacionais em informações voltadas ao suporte à tomada de decisões estratégicas.
- c) Trata-se de uma visão explícita, formal, estruturada e compartilhada de um domínio do conhecimento.
- d) É composto por triplas RDF e tem o propósito de formar bancos de conhecimento para suporte a sistemas baseados em conhecimento.

Comentários:

(a) Errado. Data Warehouse não é um processo; (b) Correto. Descrição impecável de Data Warehouse; (c) Errado. Data Warehouse não é uma visão de um domínio do conhecimento; (d) Errado. Isso não tem nenhuma relação com a definição de Data Warehouse.

Gabarito: Letra B

103. (IBFC/TJ - PE – 2017) Ferramentas de software cuja função é a extração de dados de diversos sistemas, transformação desses dados conforme regras de negócios e por fim o carregamento dos dados geralmente para um Data Mart e/ou *Data Warehouse* são chamados pela sigla, em inglês:



- a) DTB - Draw Transform Buren
- b) ETL - Extract Transform Load
- c) ECB - Extract Convert Buren
- d) DCL - Draw Convert Load
- e) ETB - Extract Transform Buren

Comentários:

Vamos analisar as palavras-chave: extração de dados; diversos sistemas (fontes); transformação de dados; carregamento de dados. Todas elas nos remetem ao Processo de ETL (Extract, Transform e Load).

Gabarito: Letra B

104. (IBFC / TJ - PE – 2017) Numa aplicação que necessita de *Data Warehouse* uma das fases mais críticas é a forma pela qual os dados são efetivamente carregados ou introduzidos em um Data Warehouse. As ferramentas de software cuja função é apoiar essa atividade são conhecidas pela sigla em inglês:

- a) OLAP
- b) SQL
- c) ETL
- d) BI
- e) OLTP

Comentários:

A questão menciona uma fase crítica responsável pelo efetivo carregamento dos dados em um Data Warehouse – trata-se do Processo de ETL!

Gabarito: Letra C

105. (QUADRIX / CFO-DF – 2017) Os sistemas de suporte à decisão são sistemas simples que acessam bases de dados específicas de um determinado setor da empresa e auxiliam os desenvolvedores na escolha da tecnologia mais adequada para desenvolver os sistemas.

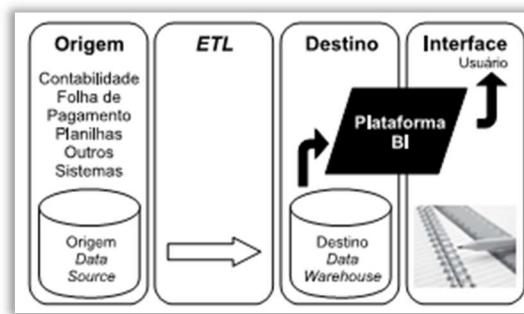
Comentários:

Na verdade, são sistemas extremamente complexos.

Gabarito: Errado



106. (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: Letra C

107. (IESES / BAHIAGÁS – 2016) Com relação aos sistemas de *Data warehouse*, assinale a alternativa correta:

- a) Um Data warehouse deve ser orientado a produto, integrado, volátil e invariável com o tempo.
- b) Um Data warehouse armazena dados históricos usados no processo de tomada de decisão. Este sistema integra os dados corporativos de uma empresa em um único repositório.
- c) Dados de um Data warehouse provém de uma única fonte, não sendo necessária a padronização para uniformizar nomes ou unidades de medida.
- d) Um Data warehouse deve ser orientado a departamento, integrado, volátil e invariável com o tempo.
- e) Um Data warehouse cria uma visão distribuída de um banco de dados centralizado. Este tipo de sistema não permite que usuários finais executem consultas sobre sua base de dados.

Comentários:

(a) Errado. Deve ser orientado a assunto, integrado, não-volátil e variável com o tempo; (b) Correto. Item está perfeito; (c) Errado. Provém de diversas fontes, sendo necessário padronizar/uniformar nomes e unidades; (d) Errado. Deve ser orientado a assunto, integrado, não-volátil e variável com o tempo; (e) Errado. Cria uma visão corporativa e permite, sim, a consulta por usuários finais.

Gabarito: Letra B

108. (COSEAC / UFF – 2015) A capacidade de os dados não permanecerem atualizados nos datawarehouses indica uma característica conhecida como:

- a) atomicidade.
- b) volatilidade.
- c) consistência.
- d) integração.



e) organização.

Comentários:

A capacidade de os dados não permanecerem atualizados nos Data Warehouses indica uma característica conhecida como volatilidade. Lembrando que volátil é aquilo que muda com facilidade. O Data Warehouse é não-volátil, logo ele não muda com facilidade. Uma vez inseridos no Data Warehouse, você não pode mais alterar os dados.

Gabarito: Letra B

109. (FUNCAB / MJ – 2015) O data warehouse é um componente significativo do business intelligence. De acordo com Immon, uma coleção de assuntos organizados para suporte a decisão, com base nas necessidades de um determinado departamento, é conhecido como data:

- a) mart.
- b) slice.
- c) group.
- d) search.
- e) text.

Comentários:

A dica para acertar essa questão é que ela menciona “*com base nas necessidades de um determinado departamento*”. O Data Mart é considerado um pequeno Data Warehouse – ou um subconjunto/segmentação de um Data Warehouse – que contém dados de uma organização em geral restritos a uma unidade de negócio ou a um processo de negócio específico com a finalidade de dar suporte ao processo decisório da alta direção por meio da criação de uma visão personalizada dos dados.

Gabarito: Letra A

110. (FUNCAB / PRODAM/AM – 2014) No contexto da modelagem dimensional, um depósito de dados digitais serve para armazenar informações detalhadas relativamente a uma empresa, criando e organizando relatórios por meio de históricos que são posteriormente utilizados pela empresa para ajudar na tomada de decisões. É orientado por assunto, integrado, não volátil, variável com o tempo, e empregado no apoio às decisões gerenciais. Esse depósito de dados está diretamente relacionado ao seguinte conceito:

- a) OLTP
- b) OLAP
- c) DATAMART



- d) DATAMINING
- e) DATAWAREHOUSE

Comentários:

Notem as palavras-chave: modelagem dimensional; depósito de dados; informações de uma empresa; relatórios por meio de históricos; ajuda na tomada de decisões; orientado por assunto; integrado; não-volátil; variável com o tempo. Todas essas palavras nos remetem à Data Warehouse!

Gabarito: Letra E

111. (CESGRANRIO / LIQUIGÁS – 2014) Data Warehouse pode ser conceituado como uma coleção de dados orientada por assuntos, integrada, assim como:

- a) não variante no tempo e volátil.
- b) não variante no tempo e não volátil.
- c) pouco variante no tempo e volátil
- d) variante no tempo e volátil
- e) variante no tempo e não volátil

Comentários:

De acordo com Bill Inmon, pela milésima vez, 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: Letra E

112. (FUNCAB / MDA – 2014) O principal objetivo de um *Data warehouse* é:

- a) disponibilizar informações para apoio às decisões de uma empresa.
- b) excluir a padronização e integração dos dados existentes nos diversos sistemas.
- c) representar uma tecnologia para a avaliação da qualidade de softwares empresariais.
- d) ser uma metodologia orientada a objetos para desenvolvimento de sistema de informação.
- e) eliminar o uso de dados históricos em operações de consultas e análises de negócio.

Comentários:

(a) Correto. Disponibiliza informações para apoio às decisões de uma empresa é realmente um dos objetivos de um Data Warehouse; (b) Errado. Exclui a padronização e integração dos dados existentes nos diversos sistemas; (c) Errado. Avaliar a qualidade de softwares empresariais definitivamente não é um dos objetivos de um Data Warehouse; (d) Errado. Não se trata de uma



metodologia; é orientada por assuntos; e não serve para desenvolvimento de sistemas de informação; (e) Errado. Eliminar Estimular o uso de dados históricos em operações de consultas e análises de negócio.

Gabarito: Letra A

113. (FUNCAB / MDA – 2014) O principal objetivo de um *Data warehouse* é:

- a) disponibilizar informações para apoio às decisões de uma empresa.
- b) excluir a padronização e integração dos dados existentes nos diversos sistemas.
- c) representar uma tecnologia para a avaliação da qualidade de softwares empresariais.
- d) ser uma metodologia orientada a objetos para desenvolvimento de sistema de informação.
- e) eliminar o uso de dados históricos em operações de consultas e análises de negócio.

Comentários:

(a) Correto. Disponibiliza informações para apoio às decisões de uma empresa é realmente um dos objetivos de um Data Warehouse; (b) Errado. Exclui Inclui a padronização e integração dos dados existentes nos diversos sistemas; (c) Errado. Avaliar a qualidade de softwares empresariais definitivamente não é um dos objetivos de um Data Warehouse; (d) Errado. Não se trata de uma metodologia; é orientada por assuntos; e não serve para desenvolvimento de sistemas de informação; (e) Errado. Eliminar Estimular o uso de dados históricos em operações de consultas e análises de negócio.

Gabarito: Letra A

114. (INSTITUTO AOCP / IFN-MG – 2014) Um Data Warehouse pode ser subdividido em pequenas porções para atender um departamento da empresa, por exemplo. Essa subdivisão do Data Warehouse é conhecida como:

- a) Database.
- b) Data Mart.
- c) Data Center.
- d) Data Mining.
- e) Data Store.

Comentários:

A subdivisão do Data Warehouse é chamada de Data Mart. O Data Mart é considerado um pequeno Data Warehouse – ou um subconjunto/segmentação de um Data Warehouse – que contém dados de uma organização em geral restritos a uma unidade de negócio ou a um processo de negócio específico com a finalidade de dar suporte ao processo decisório da alta direção por meio da criação de uma visão personalizada dos dados.



Gabarito: Letra B

115. (FUNDEP / IFN-MG – 2014) Assinale a alternativa CORRETA que apresenta o nome de como é conhecido em data warehouse o subconjunto de dados da organização como, por exemplo, os dados de um departamento os quais apresentam um foco restrito para análise e descoberta de informações úteis à tomada de decisões.

- a) Data sector.
- b) Data marts.
- c) Data fator.
- d) Data info.

Comentários:

Subconjunto de dados de um departamento com foco na tomada de decisões... só pode ser um Data Mart. O Data Mart é considerado um pequeno Data Warehouse – ou um subconjunto/segmentação de um Data Warehouse – que contém dados de uma organização em geral restritos a uma unidade de negócio ou a um processo de negócio específico com a finalidade de dar suporte ao processo decisório da alta direção por meio da criação de uma visão personalizada dos dados.

Gabarito: Letra B

116. (ESAF / MF - 2013) Um conjunto de dados produzido para oferecer suporte à tomada de decisões, sendo um repositório de dados de interesse aos gerentes, é um:

- a) Management dataset.
- b) Data warehouse.
- c) Data center.
- d) Major warehouse.
- e) Data wareset.

Comentários:

Vamos analisar as palavras-chave: conjunto de dados; suporte à tomada de decisões; repositório de dados de interesse aos gerentes. Todas essas palavras nos remetem ao Data Warehouse.

Gabarito: Letra B

117. (FMP CONCURSOS / MPE - AC – 2013) Considere a seguinte definição: "Uma coleção de dados não-voláteis, integrados, orientados a um determinado tema e utilizados no apoio de processos de tomada de decisão. Geralmente contém componentes complementares como



facilidades de extração, mineração e transformação de informação e ferramentas analíticas." Essa é a definição de:

- a) Planilha.
- b) Middleware.
- c) Data warehouse.
- d) DBMS (Data Base Management System).
- e) Groupware.

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: Letra C

118. (ESPP / MPE-PR – 2013) Como abrangem um grande volume de dados, os data warehouses geralmente são de maior ordem de magnitude que os bancos de dados de origem. Os data warehouses voltados para um subconjunto da organização, como um departamento, são conhecidos por:

- a) data marts.
- b) data sectors.
- c) data virtuais.
- d) data segments.
- e) data sections.

Comentários:

Data Warehouses voltados para um subconjunto da organização? Trata-se de um Data Mart! Lembrando que eles geralmente são dados referentes a um assunto em especial (Ex: Vendas, Estoque, Controladoria, entre outros) ou diferentes níveis de sumarização (Ex: Vendas Anual, Vendas Semestral, Vendas Mensal, entre outros), que focalizam uma ou mais áreas específicas.

Gabarito: Letra A

119. (ESPP / MPE - PR – 2013) Um *Data Warehouse* pode ser descrito como sendo uma coleção de tecnologias de apoio à decisão, visando a habilitar o analista a tomar decisões melhores e mais rápidas. Para a criação de um *Data Warehouse* há uma fase que consiste em limpeza e reformatação dos dados antes que sejam carregados no data warehouse. Esse processo é tratado por ferramentas conhecidas por:



- a) ETL.
- b) DWL.
- c) LDW.
- d) HRD.
- e) DHR.

Comentários:

A fase que consiste em limpeza e reformatação dos dados antes que eles sejam carregados é parte do Processo de ETL.

Gabarito: Letra A

120. (QUADRIX / DATAPREV – 2012) Um ambiente de *Data Warehouse* Oracle inclui várias ferramentas, além de uma base de dados relacional. Um ambiente típico inclui, entre outras ferramentas, um processo de extração de dados dos sistemas de origem, para levar esses dados ao data warehouse. Esse processo corresponde a:

- a) Motor OLAP.
- b) Solução ETL.
- c) Ferramenta de análise de clientes.
- d) Ferramenta Oracle Warehouse Builder.
- e) Aplicação que reúne dados para entregá-los aos usuários.

Comentários:

O processo de extração de dados dos sistemas de origem para levar esses dados ao Data Warehouse corresponde à Solução ETL.

Gabarito: Letra B

121. (IBFC / INEP - 2012) A Inteligência Empresarial, ou *Business Intelligence*, é um termo que foi criado por:

- a) EEE
- b) Gartner Group
- c) Ernst & Young
- d) PricewaterhouseCoopers
- e) ABNT

Comentários:

O termo Business Intelligence foi criado pelo Gartner Group.



Gabarito: Letra B

122. (AOCP / TCE-PA - 2012) Segundo o Gartner Group, é correto afirmar que um BI (*Business Intelligence*) é:

- a) uma política
- b) de informação que especifica regras para compartilhar, disseminar, adquirir, padronizar, classificar e inventariar informações.
- b) um banco de dados que armazena os dados históricos de interesse de toda a empresa para a tomada de decisão, é também conhecido como *Data Warehouse*.
- c) a descrição de habilidades das organizações para acessar dados e explorar informações, analisando e desenvolvendo percepções e entendimentos a seu respeito.
- d) uma tecnologia de software que visa fornecer informações empresariais por meio de uma base de dados.
- e) um conjunto de indicadores de desempenho, conhecidos como KPI, com a finalidade de medir qualquer etapa de um processo ou resultado.

Comentários:

O Gartner Group define BI como as habilidades das corporações de acessar dados e explorar as informações (normalmente contidas em um Data Warehouse ou Data Mart), analisando-as e desenvolvendo percepções e entendimentos a seu respeito – o que as permite incrementar e tornar a tomada de decisões mais pautada em informações. BI não é uma política de informação, nem um banco de dados, muito menos uma tecnologia de software ou um conjunto de indicadores de desempenho.

Gabarito: Letra C

123. (OBJETIVA / EPCT – 2012) Jorge está trabalhando na modelagem de dados de um grande repositório de dados que será usado, entre outras coisas, para armazenamento de dados históricos, geração de relatórios, além de obtenção de informações estratégicas para tomada de decisão e análise de grandes volumes de dados. Conceitualmente, esse repositório de dados é chamado de?

- a) Banco de dados transacional.
- b) Data warehouse.
- c) Banco de dados auxiliar.
- d) SIG (Sistema de Informações Gerenciais).



Comentários:

Vamos analisar as palavras-chave: grande repositório de dados; dados históricos; geração de relatórios; obtenção de informações estratégicas; tomada de decisão. Todas essas palavras-chave nos remetem ao conceito de Data Warehouse.

Gabarito: Letra B

124. (QUADRIX / DATAPREV – 2012) A fim de melhorar o processo de tomada de decisão a partir de dados correntes e históricos da empresa, uma porção resumida e descentralizada dos dados pode ser armazenada em um _____. Essa base menor é destinada a uma área específica de negócio e contém dados altamente focalizados. A alternativa que preenche corretamente a lacuna é:

- a) Data warehouse
- b) Data mart
- c) Data mining
- d) ERP
- e) OLAP

Comentários:

Porção resumida e descentralizada dos dados? Trata-se de um Data Mart! O Data Mart é considerado um pequeno Data Warehouse – ou um subconjunto/segmentação de um Data Warehouse – que contém dados de uma organização em geral restritos a uma unidade de negócio ou a um processo de negócio específico com a finalidade de dar suporte ao processo decisório da alta direção por meio da criação de uma visão personalizada dos dados.

Gabarito: Letra B

125. (IBFC / INEP – 2012) Devido à construção de um projeto de um Data Warehouse (DW) ser dispendioso e demorado, os arquitetos resolveram segmentar a construção do DW através dos:

- a) Data Minings
- b) Minings Data
- c) Data Marts
- d) Data Smarts
- e) Marts Data

Comentários:



O investimento inicial para construção de um Data Warehouse é realmente alto. Logo, algumas empresas optam por construir Data Marts. O Data Mart é considerado um pequeno Data Warehouse – ou um subconjunto/segmentação de um Data Warehouse – que contém dados de uma organização em geral restritos a uma unidade de negócio ou a um processo de negócio específico com a finalidade de dar suporte ao processo decisório da alta direção por meio da criação de uma visão personalizada dos dados.

Gabarito: Letra C

126. (AOCP / TCE-PA – 2012) Extrair os dados de diversos sistemas, transformá-los conforme as regras de negócios e carregá-los em um Data Mart ou em um Data Warehouse é a função das ferramentas de:

- a) SGBD.
- b) BPMS.
- c) XML.
- d) ETL.
- e) UML.

Comentários:

Observem que a questão menciona extrair dados, transformá-los e carrega-los em um DW ou DM. Apenas para solidificar o conhecimento: o processo de extrair dados de sistemas de origem e trazê-los para o Data Warehouse é comumente chamado de ETL (*Extract, Transform e Load*). Extract: extrai as informações do banco de dados transacional; Transform: transforma as informações para o formato adequado; e Load: carrega as informações no Data Warehouse ou em um Data Mart.

Gabarito: Letra D

127. (AOCP / TCE-PA – 2012) Sobre ETL (Extract, Transform and Load), é correto afirmar que:

- a) a extração e carga são opcionais no processo, porém a transformação é obrigatória.
- b) é o processo para tratamento dos dados de uma ou mais bases de dados de origem, para uma ou mais bases de dados de destino.
- c) concentra a menor parte do esforço exigido no desenvolvimento de um Data Warehouse.
- d) não necessariamente os dados necessitam ficar homogêneos para serem carregados no Data Warehouse, pois uma das funções deste último é resolver os conflitos que não foram resolvidos pela ETL.



e) na fase de transformação dos dados não devem ser corrigidos erros de digitação ou descoberta de violações de integridade, por exemplo, para os dados serem mantidos como os originais.

Comentários:

(a) Errado. Extração e Carga são obrigatórios – transformação é opcional; (b) Correto; (c) Errado. Concentra a maior parte do esforço exigido no desenvolvimento de um Data Warehouse; (d) Errado. O DW não é capaz de resolver esse tipo de conflito – essa é uma atribuição do ETL; (e) Errado. Pelo contrário, devem ser corrigidos erros de digitação, violações de integridade, etc.

Gabarito: Letra B

128. (ESAF / MPOG - 2010) BI - Business Intelligence:

- a) é uma técnica de otimização da árvore de decisão.
- b) é um método de formação avançada de gestores.
- c) comprehende ferramentas de análise de dados para otimizar os processos produtivos de uma empresa.
- d) são técnicas, métodos e ferramentas para mineração de dados na área de negócios de uma empresa.
- e) são técnicas, métodos e ferramentas de análise de dados para subsidiar processos de decisão de uma empresa.

Comentários:

BI é um conjunto de técnicas, métodos e ferramentas de análise de dados para subsidiar processos de decisão de uma empresa – nenhum dos outros itens faz qualquer sentido.

Gabarito: Letra E

129. (CETAP / DETRAN RR - 2010) Sobre Sistemas de Apoio à Decisão, leia os enunciados seguintes e marque a alternativa CORRETA:

I - O termo Business Intelligence (BI) nada mais é que uma série de conceitos e metodologias para auxiliar na tomada de decisões estratégicas nas empresas através principalmente de geração de relatórios gerenciais analíticos;



II - Um Data Warehouse é um banco de dados multidimensional grande, de escopo organizacional (ou seja, abrange toda a empresa) e reúne dados de todos os departamentos de forma a permitir a busca rápida de informações para auxiliar a tomada de decisões estratégicas;

III - Um Data Mart é um banco de dados multidimensional de escopo departamental, ou seja, abrange apenas um determinado departamento;

IV - Data Mining (Mineração de dados) é o termo usado para definir uma série de procedimentos, técnicas e ferramentas para recuperar e analisar dados de um Data Warehouse ou Data Mart;

V - Nos Sistemas de Apoio à Decisão é comum desnormalizar os dados para atingir uma melhor performance nos Data Mart e Data Warehouse.

- a) Somente o enunciado III está correto.
- b) Os enunciados I, II e III estão corretos.
- c) Os enunciados I, II e V estão corretos.
- d) Somente o enunciado V está correto.
- e) Todos os enunciados estão corretos.

Comentários:

(I) Correto. São conceitos e metodologias, auxiliam na tomada de decisões estratégicas, ajudam a geração de relatórios gerenciais analíticos;

(II) Correto. DW é realmente um banco de dados multidimensional, possui escopo organizacional e permite a busca rápida de informações para auxiliar a tomada de decisões estratégicas;

(III) Correto. DM também é um banco de dados multidimensional, mas realmente possui um escopo departamental;

(IV) Correto. Essa definição está tão genérica que eu acredito que ela poderia ser tanto para Mineração de Dados quanto para OLAP;

(V) Correto. Desnormalizar é manter dados redundantes no banco de dados. Sistemas de Apoio à Decisão realmente desnormalizam dados para melhorar o desempenho de DW ou DM.

Gabarito: Letra E

130. (CESGRANRIO / PETROBRÁS – 2010) Data Warehouse tem por característica ser:

- a) orientado a assuntos.
- b) temporário.
- c) descentralizado.
- d) focalizado em aspectos operacionais.



e) normalizado.

Comentários:

(a) Correto; (b) Errado, ele é não-volátil; (c) Errado, ele é centralizado; (d) Errado, ele é focado em aspectos táticos e estratégicos; (e) Errado, ele é desnormalizado.

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: Letra A

131. (FUNCAB / PRODAM/AM – 2010) No tocante aos datawarehouses, a característica desses armazéns de dados que está relacionada ao fato de que "uma vez inseridos, os dados não podem ser alterados, embora possam ser eliminados", é:

- a) organização (orientado por assunto).
- b) cliente/servidor.
- c) não volatilidade.
- d) variante de tempo.
- e) relacional.

Comentários:

Nos sistemas transacionais, os dados sofrem diversas alterações como, por exemplo, a inclusão, alteração e exclusão de dados. No ambiente do Data Warehouse, os dados – antes de serem carregados – são transformados e limpos. Após essa etapa, esses dados sofrem somente operações de consulta e exclusão, sem que possam ser alterados, sendo essa característica conhecida como não-volatilidade.

Gabarito: Letra C

132. (CESGRANRIO / CASA DA MOEDA – 2009) Uma empresa deseja criar uma base de dados para apoio à tomada de decisão. As informações corporativas encontram-se em 4 bancos de dados distintos e possuem representações diferentes para atributos de uma pessoa física. Por exemplo, em determinado banco de dados, uma pessoa física do sexo masculino é representada por um atributo com valor M, enquanto que em outro banco de dados, com valor 1.

Em que etapa de um procedimento de ETL para essa base de dados deve ocorrer uma padronização desse tipo de valor?

- a) Extração.



- b) Deleção.
- c) Transformação.
- d) Paralelismo.
- e) Carga.

Comentários:

A etapa de Transformação é utilizada para padronizar dados, corrigir erros de digitação, descobrir violações de integridade e substituir caracteres, entre outros.

Gabarito: Letra C

133. (CESGRANRIO / EPE – 2007) Existem muitos relatórios para diretores em todos os sistemas de informação transacionais de uma empresa, gerando uma sobrecarga significativa no banco de dados. Assinale a providência indicada para essa situação:

- a) Implantar a técnica de COCOMO.
- b) Implantar um Data warehouse.
- c) Aumentar a quantidade de testes de estresse.
- d) Desenvolver os sistemas utilizando modelagem orientada a objeto.
- e) Melhorar o uso das ferramentas CASE.

Comentários:

Quando há uma sobrecarga de dados de bancos de dados transacionais, recomenda-se a implantação de um Data Warehouse – nenhuma das outras alternativas faz sentido.

Gabarito: Letra B

134. (UEG / TJ – GO – 2006) A gestão de uma organização pode ser fortemente amparada por informações gerenciais e de apoio à decisão. Desta forma, o uso do *data warehouse* (DW) surge como um grande aliado. Assinale abaixo a afirmação CORRETA acerca do DW:

- a) O processo de implantação de um DW provoca a alteração de todos os sistemas da organização, pois eles devem deixar de alimentar as bases operacionais, utilizando o próprio DW.
- b) O DW pode ser definido como um repositório de informações que armazena de maneira organizada os dados provenientes dos sites de natureza operacional da organização. Ele permite mapear o histórico da organização.



c) O DW é um repositório que gradualmente substitui as bases de dados dos vários sistemas da organização gerando uma estrutura organizada e integrada para os sistemas informatizados da organização.

d) O processo de implantação de um DW provoca a alteração na estrutura de todos os sistemas da organização, pois eles devem alimentar as bases operacionais, bem como o próprio DW.

Comentários:

(a) Errado. Não há alteração alguma nos sistemas da organização – eles continuam alimentando bases operacionais. O Data Warehouse é alimentado por essas bases transacionais;

(b) Correto. Trata-se de um repositório de informações provenientes de sites de natureza operacional (transacional) e realmente permite mapear o histórico da organização;

(c) Errado. O Data Warehouse não substitui as bases de dados transacionais dos vários sistemas da organização;

(d) Errado. Não há alteração na estrutura dos sistemas da organização – eles continuam alimentando bases operacionais. O Data Warehouse é alimentado por essas bases transacionais.

Gabarito: Letra B



LISTA DE QUESTÕES – CESPE

1. **(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.
2. **(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.
3. **(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.

Após ter sido privatizada, determinada empresa passou a utilizar novos sistemas administrativos, tais como: folha de pagamento, sistema de pessoal, sistemas de benefícios etc.; porém, constatou-se que os sistemas antigos não poderiam ser descontinuados em um pequeno espaço de tempo. Tais sistemas usavam tecnologias de armazenamento de dados antiquados e dados de baixa qualidade.

Acerca dessa situação hipotética e considerando aspectos diversos pertinentes à inteligência de negócios - business intelligence (BI) -, julgue os seguintes itens.

4. **(CESPE / TELEBRÁS - 2021)** Na situação hipotética em questão, para transformar os dados brutos em informações estratégicas para a tomada de decisão, deveriam ser utilizadas ferramentas de BI. Tais ferramentas podem ser divididas nas seguintes etapas: coleta, análise, visualização de dados e tomada de decisão. Na etapa da coleta, normalmente, usa-se o método de extração, transformação e carregamento.
5. **(CESPE / TELEBRÁS - 2021)** Em situações semelhantes àquela em apreço, as técnicas de um sistema de BI objetivam aumentar a agilidade e assertividade na tomada de decisões, com a definição de regras para a formatação adequada dos dados, visando a transformação desses dados em armazéns estruturados de informações, considerando a origem deles.
6. **(CESPE / TELEBRÁS - 2021)** Na situação hipotética considerada, como os dados dos sistemas antigo e novo são heterogêneos, o ETL procurará limpar e preparar os dados para o OLAP.
7. **(CESPE / TELEBRÁS - 2021)** Na situação hipotética em tela, considerando as bases de dados antigas e novas, seria uma ótima prática a utilização dos conceitos de data warehouse e data mining para, por exemplo, automatizar a análise de currículos de acordo com as competências dos empregados.



8. **(CESPE / TELEBRÁS - 2021)** Na situação hipotética em apreço, seria correto optar pela utilização de bancos de dados relacionais, os quais são bancos especializados que foram desenvolvidos para ajudar a extrair dos dados as informações de BI.
9. **(CESPE / MPE-AP – 2021)** Em um sistema de BI, a coleção de ferramentas utilizada como componente para manipular, minerar e analisar os dados no DW (data warehouse) denominase:
 - a) OLAP (Online Analytical Processing).
 - b) BPM (Business Performance Management).
 - c) Análise de Negócio.
 - d) Dashboard.
 - e) Processamento de Transações.
10. **(CESPE / SEFAZ-CE – 2021)** Comparado ao ETL, o ELT apresenta vantagens como tempos menores de carregamento e de transformação de dados, e, consequentemente, menor custo de manutenção.
11. **(CESPE / SEFAZ-CE – 2021)** Faz parte do rol de objetivos do BI permitir acesso interativo aos dados e fornecer aos analistas de negócios, por meio da manipulação desses dados, a capacidade de realizar a análise adequada.
12. **(CESPE / SEFAZ-CE – 2021)** Um Data Warehouse (DW), ainda que seja não volátil - ou seja, após os dados serem inseridos nele os usuários não podem alterá-los - é variável no tempo, pois mantém um conjunto de dados históricos que oferecem suporte à tomada de decisões.
13. **(CESPE / APEX-BRASIL – 2021)** Em DW (data warehouse), conjunto de dados voltado para oferecer suporte à tomada de decisões, os dados:
 - a) devem ser inseridos em seu formato original, cada um com sua unidade de medida, quando forem provenientes de diferentes fontes.
 - b) utilizam, ao mesmo tempo, em um DW em tempo real, as estruturas relacional e multidimensional para proporcionar aos usuários acesso fácil e ativo aos dados.
 - c) devem ser alterados de modo a permanecerem atualizados à medida que as fontes de dados sejam registradas e modificadas.
 - d) detectam tendências e relações de longo prazo para previsão, com uma qualidade temporal para cada DW, sem necessariamente mostrar o status atual.
14. **(CESPE / APEX-BRASIL – 2021)** Os objetivos característicos do BI (Business Intelligence) incluem:



- a) possibilitar acessos interativos a dados, a sua manipulação e a realização de análises apropriadas pelos gestores e analistas, com base na transformação desses dados em informações, decisões e ações.
- b) criar um conjunto de processos para gerenciar relacionamentos com clientes efetivos e potenciais e com parceiros de negócios, por meio de marketing, vendas e serviços, independentemente do canal de comunicação.
- c) gerenciar os produtos e os serviços direcionados aos consumidores, por meio de uma rede de organizações interligadas, com base no gerenciamento de uma cadeia de suprimentos.
- d) estabelecer relação entre a eficácia na gestão e a estratégia de negócio, a partir de uma composição balanceada de medidas, de metas, de objetivos e de requisitos para essa finalidade.

15. (CESPE / TCE-RJ – 2021) O Data Warehouse empresarial, ou EDW (Enterprise Data Warehouse), é uma das aplicações do data warehouse que permite a integração em larga escala de dados oriundos de diversas fontes em formato padronizado, para subsidiar a inteligência de negócios.

16. (CESPE / TCE-RJ – 2021) A construção de um Data Mart antecede a criação de um Data Warehouse.

17. (CESPE / TCE-RJ – 2021) Enquanto os depósitos de dados e seus subconjuntos Data Marts são bancos estáticos, ou seja, não mais sofrem mudanças depois de agrupados, formatados e armazenados seus dados em um repositório, os bancos de dados típicos mudam constantemente, sendo dinâmicos por natureza.

18. (CESPE / TCE-RJ – 2021) Para o processamento de grandes quantidades de transações, de modo rápido e eficaz, é mais indicado o modelo relacional de bancos de dados do que os modelos orientados a objetos ou multidimensional.

19. (CESPE / TCE-RJ - 2021) ETL (Extract Transform Load) é uma ferramenta utilizada para extrair informações e realizar análise multidimensional no Data Warehouse.

20. (CESPE / CODEVASF – 2021) Os sistemas de suporte à decisão (DSS, do inglês Decision Support System) são programados para substituir os tomadores de decisão das organizações e podem ser usados rotineiramente em tarefas de tomada de decisão comuns, devendo ser evitados para tarefas de decisões pontuais de alto nível relacionadas às estratégias de negócios.

21. (CESPE / CODEVASF – 2021) A inteligência de negócios lida com as atividades operacionais cotidianas da organização, sendo as suas duas principais ferramentas de implementação o processamento de transações e a integridade transacional com atualização das bases de dados operacionais.



22. (CESPE / ME – 2020) Modelagem preditiva é utilizada para antecipar comportamentos futuros, por meio do estudo da relação entre duas ou mais variáveis.

23. (CESPE / ME – 2020) Uma tabela de fatos registra dados dimensionais que explicam os fatos registrados.

24. (CESPE / TJ-PA – 2020) Assinale a opção que indica um processo de extração e transformação de dados em um data Warehouse:

- a) Big Data
- b) OLAP
- c) OLTP
- d) ETL
- e) machine learning

25. (CESPE / CGM JOÃO PESSOA - 2018) *Business intelligence* pode ser definido como um processo inteligente de coleta, organização, análise, compartilhamento e monitoração de dados que, depois de processados, geram informações para o suporte e para a tomada de decisões no ambiente de negócios.

26. (CESPE / SECONT/ES - 2018) Data Warehouse é uma coleção de dados orientados ao assunto, que tem como características visão conceitual multidimensional, operações interdimensionais irrestritas, dimensões e níveis de agregação ilimitados, sendo que os dados são não voláteis e variantes no tempo.

27. (CESPE / TJ/SE - 2018) Os dados armazenados em um *DataWarehouse* devem estar integrados, temporalmente identificados, orientados ao assunto e devem ser protegidos, de modo a se evitar o acesso do usuário.

28. (CESPE / IPHAN – 2018) O *data warehouse* é integrado, não volátil e orientado a assuntos, contudo, embora lide com dados e os armazene para a tomada de decisões gerenciadas, não é considerado um banco de dados, pois é variável em relação ao tempo.

29. (CESPE / STM – 2018) Em processos ligados a ETL, os dados são identificados e extraídos de diferentes fontes, incluindo sistemas de banco de dados e aplicações.

30. (CESPE / CGM JOÃO PESSOA – 2018) Uma ferramenta de ETL deve ser capaz de extrair dados de fontes heterogêneas, copiá-los para uma área de transição, onde possam ser convertidos conforme o padrão estabelecido, e, ainda, realizar limpeza e correção nesses dados, de acordo com regras preestabelecidas.

31. (CESPE / STJ – 2018) O data warehouse tem como finalidade a apresentação das informações necessárias para a identificação de indicadores e da evolução de valores ao longo de uma grande janela de tempo.



- 32. (CESPE / STM – 2018)** O ambiente ETL permite a visualização dos dados para análise de maneira multidimensional.
- 33. (CESPE / SEDF – 2017)** Comparados aos bancos de dados transacionais, os *data warehouses* são mais voláteis porque, para que se mantenham consistentes, são atualizados em tempo real a cada atualização que ocorrer em qualquer uma das bases originais de dados que o componham.
- 34. (CESPE / FUB - 2015)** Em um ambiente de *Business Intelligence*, o usuário pode deparar com diferentes requisitos de infraestrutura para atender às necessidades de cada uma das instâncias dos sistemas analíticos.
- 35. (CESPE / ANATEL – 2014)** Em um Data Warehouse (armazém de dados) que apresente a característica de ser não volátil, os dados não são atualizados após a inserção inicial.
- 36. (CESPE / TJ-SE – 2014)** O processo de extração, transformação e carga, comumente referenciado como ETL (*Extract-Transform-Load*), é um processo usado na criação e na atualização de um Datawarehouse.
- 37. (CESPE / TJ-SE – 2014)** O processo de carga de um DataWarehouse consiste exclusivamente em adicionar novos dados históricos regularmente.
- 38. (CESPE / ANATEL – 2014)** As ferramentas para extração, transformação e carga de dados (ETL) copiam todos os dados dos sistemas operacionais e os transferem para o *Data Warehouse* de forma a apoiar a análise corporativa das tendências e a realização de previsões empresariais.
- 39. (CESPE / ANTAQ – 2014)** O Módulo de ETL (*extract transform load*), dedicado à extração, carga e transformação de dados, coleta informações em fontes como sistemas ERP, arquivos com extensão TXT e planilhas Excel.
- 40. (CESPE / ANATEL – 2014)** As ferramentas de business intelligence, encarregadas de extrair e carregar dados na base de dados de BI, são denominadas ferramentas de ETL.
- 41. (CESPE / TJ-SE – 2014)** O processo de transformação de dados pode exigir que dados logicamente relacionados, mas fisicamente separados, sejam recompostos, ainda que envolvam registros distintos ou até mesmo estejam em bancos de dados operacionais distintos.
- 42. (CESPE / TJ-SE – 2014)** Em corporações que possuem diversos departamentos, a implementação de DataMarts facilita a disseminação das informações por todos os departamentos.
- 43. (CESPE / SERPRO – 2013)** Os datamarts são criados a partir de sistemas transacionais, buscando informações relevantes para o público que irá utilizá-lo.



44.(CESPE / INPI – 2013) Um Data Mart pode ser considerado, na visão bottom-up, uma reunião de vários Data Mining, desde que estes sejam desenvolvidos em modelos multidimensionais (MDM).

45.(CESPE / MPU – 2013) Uma característica distinta dos *data warehouses* é o seu direcionamento para aplicações de apoio às decisões. Eles são otimizados para a recuperação de dados, não para o processamento rotineiro de transações.

46.(CESPE / MPOG – 2013) O processo de extração, transformação e carga visa trabalhar com a manipulação de dados de fontes externas e sua respectiva transformação, para atender às necessidades de negócios e carga dos dados dentro de uma estrutura de data warehouse.

47.(CESPE / MPOG – 2013) Em ETL, frequentemente, é necessário limpar, ajustar e consolidar os dados antes de realizar sua carga. Nesse estágio de transformação, aplicam-se regras ou funções aos dados extraídos para ajustar os dados a serem carregados. A limpeza trata de vários tipos de erros, como, por exemplo, valores ilegais, ou que não obedecem às regras de integridade da base, e erros de ortografia.

48.(CESPE / ANTT – 2013) Ferramentas ETL são utilizadas na extração, transformação e remoção de dados.

49.(CESPE / Banco da Amazônia – 2012) O processo de extração, transformação e carga (ETL) ilustrado permite extrair dados de diversas fontes de dados e migrá-los para o data warehouse, mantendo sua estrutura e normalização originais.

50.(CESPE / TJ - AC – 2012) Por meio de um data mart, que é um subconjunto do data warehouse, é possível criar uma visão personalizada dos dados.

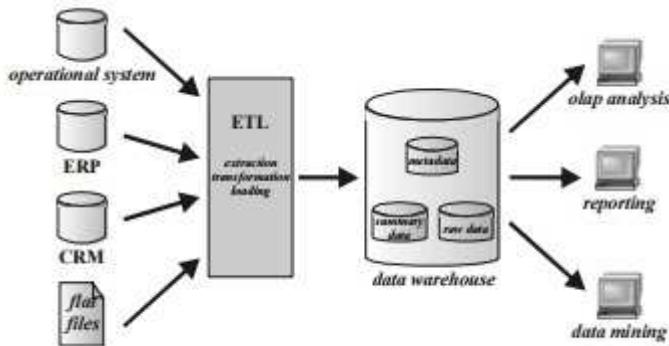
51.(CESPE / MPE-PI – 2012) Na construção de um data warehouse, o processo extract, transform and load (ETL), normalmente, é o menos crítico e o mais rápido. Esse processo envolve a extração dos dados, na sua transformação, e, eventualmente, influencia na limpeza desses dados.

52.(CESPE / CORREIOS - 2011) Inteligência empresarial, ou *business intelligence*, é um termo utilizado para descrever as habilidades das corporações para coletar dados e explorar informações, analisá-las e desenvolver entendimentos para tomada de melhores decisões.

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

54.(CESPE / CORREIOS – 2011) As ferramentas de software ETL (*extract transform load*) têm como função a extração de dados de diversos sistemas, a transformação desses dados de acordo com as regras de negócio e a carga dos dados em um *data mart* ou um DW.





Internet: <www.infoescola.com/informatica/data-warehouse>.

55. (CESPE / Correios – 2011) Considere que, em uma organização que mantenha uma solução de DW, haja grande quantidade de dados para serem manipulados e que uma nova área criada dentro dessa organização necessite acessar parte dos dados desse DW.

Nessa situação, seria adequada a implementação de um *datamart* específico para essa nova área.

56. (CESPE / SERPRO – 2010) No ambiente de data warehouse, os dados consistem de um snapshot (retrato no tempo) e abrangem períodos de até 10 anos, sendo atualizados periodicamente, e não em tempo real.

57. (CESPE / MPE - PI – 2010) Um *data warehouse* é um repositório de informações de uma fonte única, armazenadas sob diversos esquemas, em um só local, e que fornecem dados consolidados.

58. (CESPE / INMETRO – 2010) Assinale a opção correta com relação aos conceitos de *data mart*.

- a) O data warehouse é constituído de apenas um data mart.
- b) O data mart pode ser visto como a restrição do data warehouse a um único processo de negócios.
- c) Todo data mart deve ser representado por um modelo orientado a objetos.
- d) O data mart possui uma estrutura diferente do data warehouse.
- e) O que diferencia um data mart de um data warehouse é a forma de acesso aos dados.

59. (CESPE / BANCO DA AMAZÔNIA – 2010) Por fazerem uso de um processo de modelagem dimensional, os *data warehouses* não podem ser instalados em sistemas de banco de dados relacionais.

60. (CESPE / BANCO DA AMAZÔNIA – 2010) Um data mart é uma reunião de vários data warehouses, a fim de fornecer visão mais ampla dos dados.



61.(CESPE / INMETRO - 2009) Entre as operações típicas realizadas em uma organização que adota a abordagem de *business intelligence*, destaca-se o uso de mineração de dados e a construção de *data warehouses*, ambas empregando bancos de dados relacionais com elevado grau de normalização.

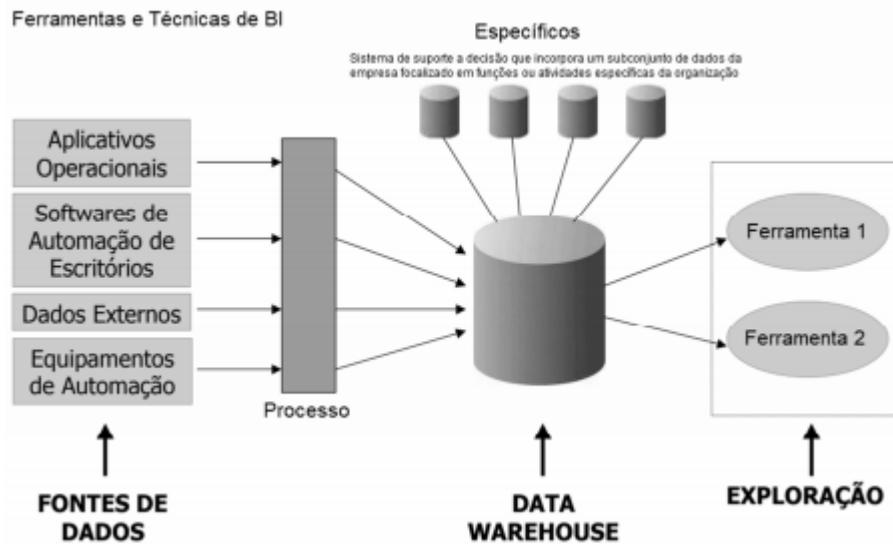
62.(CESPE / ANTAQ - 2009) Aplicações de *business intelligence* (BI) oferecem visões históricas e atuais de operações de negócios empregando unicamente dados operacionais.

63.(CESPE / SERPRO – 2008) Os *data warehouses* são depósitos de dados tipicamente direcionados para aplicações de apoio às decisões administrativas. As informações em *data warehouses* mudam mais freqüentemente do que em bancos de dados convencionais. Além disso, os *data warehouse*, em vez de otimizar a recuperação de dados, otimizam o processamento de transações.



LISTA DE QUESTÕES – FCC

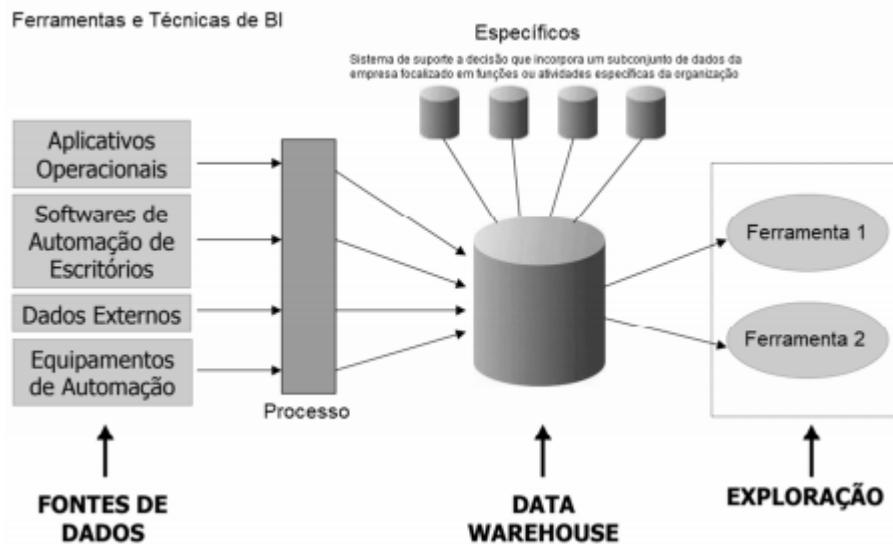
64.(FCC / SANASA – 2019) Atenção: Para responder à questão, considere a imagem abaixo.



O Processo, representado na imagem por um retângulo vertical, é um método de alimentação do Data Warehouse a partir de diversos dados da organização. Trata-se de:

- a) ODS.
- b) ERP.
- c) ETL.
- d) CRM.
- e) EIS.

65.(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.

66. (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.

67. (FCC / DPE - AM – 2018) Sobre o processo de ETL, aplicado a data warehouse, é correto afirmar que:

- a) a fase de extração de dados consiste em obter os dados do servidor do data warehouse.
- b) a fase de transformação consiste em realizar modificações nos dados carregados, adequando seus valores ao modelo definido para o data warehouse.
- c) as fases de extração e carga de dados são realizadas de forma simultânea.
- d) a fase de carga de dados visa eliminar valores nulos contidos nos bancos de dados transacionais da empresa.



e) a fase de carga de dados consiste em inserir os dados transformados nos bancos de dados transacionais da empresa.

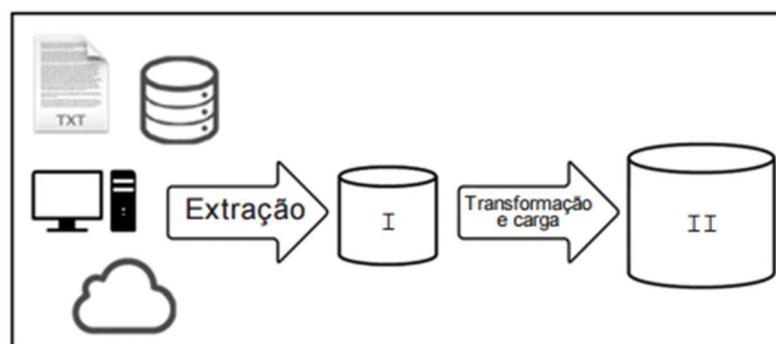
68. (FCC / PREF SÃO LUÍS – 2018) Para extrair dados de fontes de dados heterogêneas que irão alimentar um *Data Warehouse* de forma homogênea e concisa, servindo de base para gerar relatórios e gráficos para apoiar as decisões da gerência da organização, deve-se utilizar um processo conhecido como:

- a) OLAP.
- b) Data Mart.
- c) ETL.
- d) OLTP.
- e) Data Mining.

69. (FCC / TRT 14ª REGIÃO – 2016) Quando uma empresa utiliza Data Warehouse (DW) é necessário carregá-lo para permitir a análise comercial. Para isso, os dados de um ou mais sistemas devem ser extraídos e copiados para o DW em um processo conhecido como:

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

70. (FCC / AL - MS – 2016) Considere a figura abaixo.

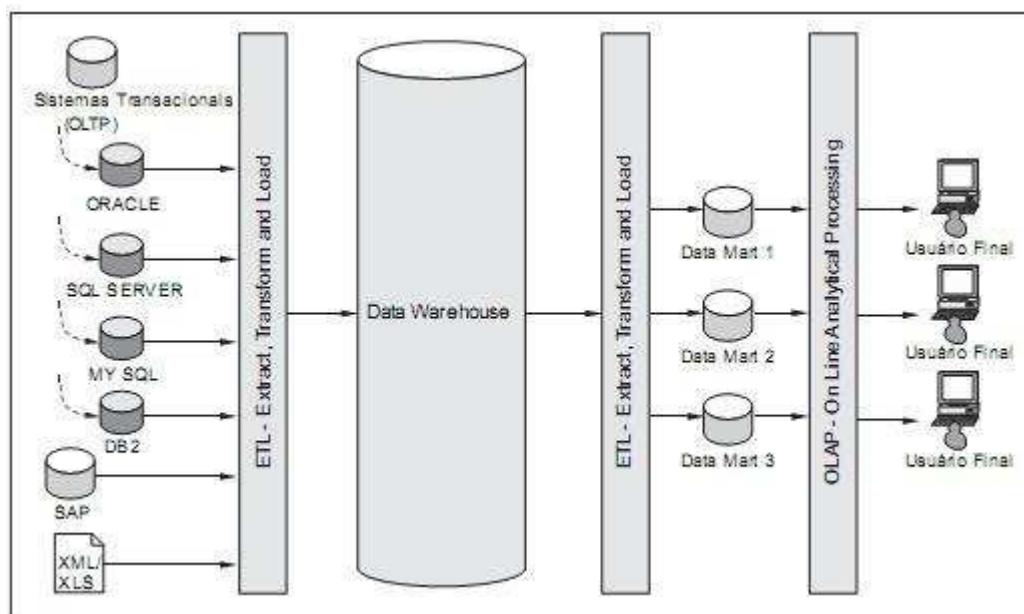


No processo de ETL mostrado na figura, I e II correspondem, respectivamente, a:

- 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.



71. (FCC / AL - MS – 2016) Considere que a equipe de Analistas de Desenvolvimento de Sistemas do CNMP está projetando a arquitetura para o Data Warehouse (DW) da instituição, conforme mostra a figura abaixo:



É correto afirmar que esta arquitetura:

- a) é *bottom-up*. Garante a existência de um único conjunto de aplicações para ETL, ou seja, extração, limpeza e integração dos dados, embora os processos de manutenção e monitoração fiquem descentralizados.
- b) é *bottom-up*, pois primeiro a equipe cria um DW e depois parte para a segmentação, ou seja, divide o DW em áreas menores gerando pequenos bancos orientados por assuntos aos departamentos.
- c) é *bottom-up*. Permite um rápido desenvolvimento, pois a construção dos *Data Marts* é altamente direcionada. Normalmente um *Data Mart* pode ser colocado em produção em um período de 2 a 3 meses.
- d) é *top-down*. A partir do DW são extraídos os dados e metadados para os *Data Marts*. Nos *Data Marts* as informações estão em maior nível de sumarização e, normalmente, não apresentam o nível histórico encontrado no DW.
- e) é *top-down*, pois possui um retorno de investimento muito rápido ou um *faster pay back*. O propósito desta arquitetura é a construção de um DW incremental a partir de *Data Marts* independentes.

72. (FCC / DPE - RS – 2017) O processo de ETL é composto por algumas fases, identificadas como entregar, limpar, extrair e preparar. A ordem correta para a execução dessas fases é:



- a) limpar, preparar, extrair e entregar.
- b) preparar, limpar, entregar e extrair.
- c) entregar, preparar, limpar e extrair.
- d) entregar, extrair, limpar e preparar.
- e) extrair, limpar, preparar e entregar.

73. (FCC / TRE - SP – 2017) No processo ETL de um *data warehouse*:

(I) a tradução de valores codificados conhecida como limpeza de dados (por exemplo, supondo que o sistema de origem armazena 1 para sexo masculino e 2 para feminino, mas o *data warehouse* armazena M para masculino e F para feminino) e (II) a geração de valores de chaves substitutas (*surrogate keys*) são:

- a) integrantes da etapa de transformação.
- b) integrantes das etapas de transformação e carga, respectivamente.
- c) manipulações de dados, aplicáveis no processo *pivoting*.
- d) elementos que possibilitam a conversão de *drill-down* para *roll up*.
- e) integrantes da etapa de carga.

74. (FCC / TRT 3ª REGIÃO – 2015) Ao contrário do que muitas empresas pensam, um projeto de data warehouses não necessita ser imenso e exigir um grande investimento financeiro de início. Ele pode começar atendendo um subconjunto da organização. Igualmente, isso vale para empresas de médio porte, que podem se beneficiar dessa tecnologia. Isso só é possível porque, quando se conceitua data warehouse, tem-se também o conceito de:

- a) Data Mart.
- b) Data Center.
- c) Data Warehouse Virtual.
- d) Banco de Dados em Nuvem.
- e) Data Warehouse Empresarial.

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

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



76. (FCC / DPE-RS – 2014) O Data Warehouse (DW) é um tipo especial de banco de dados que proporciona uma sólida e concisa integração dos dados de uma empresa para a realização de análises gerenciais estratégicas de seus principais processos de negócio. O DW é um depósito de dados:

- a) especializado, orientado por assunto, integrado, variável no tempo e volátil. Os usuários podem atualizar os dados e até criar novos dados, ou seja, novas tabelas, para algum propósito específico de interesse gerencial.
- b) orientado por assunto, integrado, variável com o tempo e não volátil. Uma vez inseridos, os dados não podem ser alterados, o que implica não ser necessário nenhum tipo de bloqueio por concorrência de usuários no acesso.
- c) orientado por assunto, integrado, variável com o tempo e volátil. Os dados de um DW variam em relação ao tempo, pois representam resultados operacionais que refletem a evolução da empresa. Isso implica que os dados devem ser constantemente atualizados.
- d) orientado por processos, integrado, variável com o tempo e volátil. Um DW armazena as informações agrupadas por processos de negócios que são atividades desenvolvidas na organização, como transformação de entrada e saída com agregação de valor, essenciais para a tomada de decisões.
- e) especializado, orientado por assunto, integrado, variável com o tempo e não volátil. Em ambientes de múltiplas plataformas sistêmicas, a característica de integração se torna fundamental, pois a organização necessita manter os diferentes padrões existentes entre os diferentes sistemas operacionais.

77. (FCC / MPE-CE - 2013) Em relação ao entendimento do significado do termo Business Intelligence (BI) e da solução que provê, a definição que NÃO é coerente com o termo *Business Intelligence* é a que:

- a) consiste em uma metodologia que fornece objetivos de negócios ligados a objetivos de TI, provendo métricas e modelos de maturidade para medir a sua eficácia e identificando as responsabilidades relacionadas dos donos dos processos de negócios e de TI.
- b) se refere à aplicação de técnicas analíticas para informações sobre condições de negócio no sentido de melhorá-las, de uma maneira automatizada, mas com a interpretação e respostas humanas, de forma a melhorar a tomada de decisões.
- c) reúne recursos que provêm a habilidade para que a pessoa certa receba a informação adequada e no momento correto para tomar a melhor decisão.



d) consiste em um sistema de negócios que inclui uma estrutura de busca efetiva e acessível, acurada, em tempo real, com informações e relatórios que permitam aos líderes das áreas de negócio se manterem informados para tomar decisões.

e) é uma solução fácil de dizer, mas difícil de fazer corretamente pois envolve mudanças na forma como a organização conduz uma busca efetiva, bem como, a necessidade de se possuir uma base de dados de qualidade para que se possa tomar ações com o objetivo de otimizar a performance corporativa.

78.(FCC / TST – 2012) O processo de ETL em uma Data Warehouse possui várias fases. Em uma destas fases é efetuada a:

- a) extração dos dados dos sistemas de origem.
- b) introdução de novos produtos no mercado.
- c) validação das interfaces de usuário.
- d) criação de diagramas estáticos e comportamentais das classes e atributos.
- e) definição dos custos e prazos.

79.(FCC / INFRAERO - 2011) É uma característica de um sistema de *Business Intelligence*:

- a) Capacidade de cruzar informações de diferentes bancos de dados, gerando relatórios analíticos diversos.
- b) Utilização de comandos SQL para a geração de tabelas dinâmicas inteligentes ligadas às regras de negócio.
- c) Utilização de Sistemas Baseados em Conhecimento, onde é possível utilizar técnicas de Inteligência Artificial na geração de relatórios.
- d) Análise automática de requisitos funcionais e não funcionais, permitindo a geração de relatórios gerenciais.
- e) Análise de informações com o intuito de fornecer subsídio para a criação de sistemas inteligentes.

80.(FCC / TRE-PE – 2011) Um processo importante que ocorre em relação à formação de um data warehouse é a obtenção dos dados de uma ou mais bases de dados da origem. Deve ser rigoroso para evitar a deformação e/ou a perda dos dados quando passados da fonte original para o destino. Trata- se de:

- a) MINING.
- b) DATA MART.
- c) MOLAP.
- d) STAR.



e) ETL.

81. (FCC / TRT 21ª REGIÃO – 2003) Um banco de dados organizado em estruturas lógicas dimensionais cujas informações são voltadas para o processo decisório e para áreas específicas denomina-se:

- a) data warehouse.
- b) data mining.
- c) data mart.
- d) business intelligence.
- e) competitive intelligence.



LISTA DE QUESTÕES – FGV

82. (FGV / TJDFT – Análise de Sistemas – 2022) Lino precisa preparar um grande volume de dados para minerar realizando operações de limpeza, adequação de formato, exclusão de brancos e inclusão de novos atributos derivados. Para realizar o pré-processamento dos dados, Lino deve usar uma ferramenta do tipo:

- a) ETL;
- b) OLAP;
- c) Apriori;
- d) Data Mart;
- e) Data Lake.

83. (FGV / CGU – 2022) Uma organização deseja implementar um pipeline de dados e está avaliando a opção mais adequada para o seu contexto de operação. Em torno de 40% dos dados consumidos pela organização se encontram em planilhas eletrônicas que contêm dados sensíveis, produzidas semanalmente por suas unidades de negócio. Os outros 60% dos dados se encontram em alguns bancos de dados relacionais de sistemas de produção da organização. O tamanho da base é de moderado a pequeno, mas existe a necessidade de conformidade com normas de privacidade e confidencialidade dos dados. O objetivo do pipeline é fornecer insumos para um departamento que realiza análises de dados com métodos não supervisionados de aprendizagem de máquina para elaborar relatórios periódicos mensais. A organização está avaliando a construção de um Armazém de Dados (ETL) ou de um Lago de Dados (ELT).

A proposta de modelo adequada e corretamente justificada é:

- a) Armazém de Dados. Ambos os modelos são adequados, mas Lago de Dados tem maior latência até a carga (L) e custo maior;
- b) Armazém de Dados. Esse modelo possui menor latência até a carga (L) e, ao contrário do Lago de Dados, opera de forma eficiente com dados relacionais;
- c) Armazém de Dados. O processo ETL é mais adequado para o tratamento dos dados sensíveis e os casos de uso são bem conhecidos;
- d) Lago de Dados. Esse modelo possui menor latência até a carga (L) e permite a extração (E) de dados semiestruturados e não estruturados;
- e) Lago de Dados. Esse modelo não necessita de hardware especializado e, ao contrário do Armazém de Dados, possibilita tarefas de aprendizado de máquina.



84.(FGV / SEFAZ-AM – 2022) Os *data warehouses* existem para facilitar que analista e tomadores de decisão executem as consultas complexas e ocasionais, fazendo uso intensivo de dados por meio de processamento OLAP.

A operação analítica que se caracteriza por analisar dados em níveis de agregação progressivamente mais detalhados e de menor granularidade, é denominada:

- a) *pivot*
- b) *roll-up*.
- c) *drill-down*.
- d) *drill-across*
- e) *slice and disse*.

85.(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.

86. (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.

87.(FGV / ALERJ – 2017) Daniel está desenvolvendo um *Data Warehouse* para analisar os dados do Censo Escolar. A fonte de dados está em um arquivo CSV e descrita em um documento, conforme parcialmente ilustrado nas figuras a seguir.



ID_SERIE	CO_TURNO	HR_INICIAL	HR_FINAL	NO_TURMA
3	1	7	11	3º ANO A E...
3	1	7	11	3º ANO A E...
3	1	7	11	3º ANO A E...
3	1	7	11	3º ANO A E...
3	1	7	11	3º ANO A E...

Nome	Descrição	Valor	Descrição do Valor
CO_TURNO	Código do Turno	1	Matutino
		2	Vespertino
		3	Noturno

Para carregar esses dados no *Data Warehouse* com a descrição dos turnos no lugar de seu código, Daniel deve desenvolver um programa para ler os dados do arquivo, realizar transformações e carregar o resultado no banco de dados.

A ferramenta a ser utilizada por Daniel é:

- a) ETL;
- b) OLAP;
- c) Data Mining;
- d) ODBC;
- e) XSLT.

88. (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.

89. (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.



90.(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

91.(FGV / MEC – 2009) As alternativas a seguir apresentam maneiras de recuperar informações de um Data Warehouse, à exceção de uma. Assinale-a.

- a) OLAP.
- b) Data Mart.
- c) Data Mining.
- d) EIS (Executive Information Systems).
- e) Ferramentas de consulta e emissão de relatórios.

92.(FGV / MEC – 2009) No que diz respeito ao desenvolvimento de sistemas de informação e da análise de negócios, analise a citação a seguir.

"É um conjunto integrado de dados, extraídos de bancos de dados operacionais, históricos, externos e selecionados, editados e padronizados para recuperação e análise, para fornecer inteligência empresarial e nos negócios no processo de tomada de decisões gerenciais."

Essa citação representa o conceito de:

- a) Data Smart.
- b) Data Mining.
- c) Data Business.
- d) Data Warehouse.
- e) Data Information.

93.(FGV / SEFAZ-RJ – 2009) O grande desafio do profissional de TI que gerencia qualquer processo é a análise dos fatos relacionados à função que exerce em uma organização. Essa análise deve



ser feita com as ferramentas e os dados disponíveis, permitindo aos executivos e gerentes detectar as tendências e tomar as decisões com eficiência e eficácia. Devido a essa necessidade, surgiu o conceito de Business Intelligence - "BI". Assinale a alternativa que indique duas características dos atuais sistemas de Business Intelligence.

- a) procurar relações de causa e efeito / extrair e integrar dados de múltiplas fontes.
- b) evitar a utilização de ferramentas automatizadas / desprezar dados contextualizados.
- c) extrair e integrar dados de múltiplas fontes / evitar a utilização de ferramentas automatizadas.
- d) desprezar dados contextualizados / trabalhar exclusivamente com fatos reais e não hipotéticos.
- e) trabalhar exclusivamente com fatos reais e não hipotéticos / procurar relações de causa e efeito.



LISTA DE QUESTÕES – DIVERSAS BANCAS

94. (FEPESE / ISS-Criciúma – 2022) No processo de criação de um Data Warehouse, quais etapas devem ser cumpridas, na ordem correta, de modo a se obter os dados a partir dos sistemas transacionais para um melhor resultado?

- a) Definição; planejamento; execução; gestão.
- b) Planejamento; levantamento de requisitos; execução; monitoramento.
- c) Planejamento; extração; gestão e acompanhamento; melhoria contínua.
- d) Extração; Limpeza e transformação; entrega e carga; gestão e acompanhamento.
- e) Limpeza e transformação; gestão e acompanhamento; atuação e correção; medição.

95. (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, characteristicamente, 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.

96. (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).

97. (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.

98. (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.

99. (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.

100. (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.

101. (CESGRANRIO / TRANSPETRO – 2018) Os sistemas de apoio à decisão devem prover um conjunto de capacidades, tais como:

- a) ser resiliente contra condições temporais mutantes.
- b) prescindir do julgamento humano no processo de análise.
- c) permitir a fácil execução das análises de sensibilidade.
- d) apoiar decisões interdependentes, exceto as sequenciais.
- e) evitar sua utilização em ambientes voltados para internet (web).

102. (IESES / CEGÁS – 2017) Assinale a alternativa que conceitua corretamente *Data Warehouse*:

- a) É um processo analítico em tempo real que transforma dados em informação. Está orientada a assuntos e utiliza tecnologias semânticas.
- b) É um depósito de dados digitais, orientado por assunto que, transforma dados operacionais em informações voltadas ao suporte à tomada de decisões estratégicas.
- c) Trata-se de uma visão explícita, formal, estruturada e compartilhada de um domínio do conhecimento.
- d) É composto por triplas RDF e tem o propósito de formar bancos de conhecimento para suporte a sistemas baseados em conhecimento.

103. (IBFC/TJ - PE – 2017) Ferramentas de software cuja função é a extração de dados de diversos sistemas, transformação desses dados conforme regras de negócios e por fim o carregamento dos dados geralmente para um Data Mart e/ou *Data Warehouse* são chamados pela sigla, em inglês:

- a) DTB - Draw Transform Buren
- b) ETL - Extract Transform Load
- c) ECB - Extract Convert Buren
- d) DCL - Draw Convert Load
- e) ETB - Extract Transform Buren

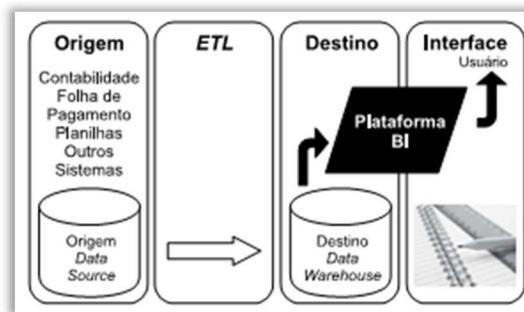
104. (IBFC / TJ - PE – 2017) Numa aplicação que necessita de *Data Warehouse* uma das fases mais críticas é a forma pela qual os dados são efetivamente carregados ou introduzidos em um Data Warehouse. As ferramentas de software cuja função é apoiar essa atividade são conhecidas pela sigla em inglês:



- a) OLAP
- b) SQL
- c) ETL
- d) BI
- e) OLTP

105. (QUADRIX / CFO-DF – 2017) Os sistemas de suporte à decisão são sistemas simples que acessam bases de dados específicas de um determinado setor da empresa e auxiliam os desenvolvedores na escolha da tecnologia mais adequada para desenvolver os sistemas.

106. (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).



107. (IESES / BAHIAGÁS – 2016) Com relação aos sistemas de *Data warehouse*, assinale a alternativa correta:

- a) Um Data warehouse deve ser orientado a produto, integrado, volátil e invariável com o tempo.
- b) Um Data warehouse armazena dados históricos usados no processo de tomada de decisão. Este sistema integra os dados corporativos de uma empresa em um único repositório.
- c) Dados de um Data warehouse provém de uma única fonte, não sendo necessária a padronização para uniformizar nomes ou unidades de medida.
- d) Um Data warehouse deve ser orientado a departamento, integrado, volátil e invariável com o tempo.
- e) Um Data warehouse cria uma visão distribuída de um banco de dados centralizado. Este tipo de sistema não permite que usuários finais executem consultas sobre sua base de dados.

108. (COSEAC / UFF – 2015) A capacidade de os dados não permanecerem atualizados nos datawarehouses indica uma característica conhecida como:

- a) atomicidade.
- b) volatilidade.
- c) consistência.
- d) integração.
- e) organização.

109. (FUNCAB / MJ – 2015) O data warehouse é um componente significativo do business intelligence. De acordo com Immon, uma coleção de assuntos organizados para suporte a decisão, com base nas necessidades de um determinado departamento, é conhecido como data:

- a) mart.
- b) slice.
- c) group.
- d) search.
- e) text.

110. (FUNCAB / PRODAM/AM – 2014) No contexto da modelagem dimensional, um depósito de dados digitais serve para armazenar informações detalhadas relativamente a uma empresa, criando e organizando relatórios por meio de históricos que são posteriormente utilizados pela empresa para ajudar na tomada de decisões. É orientado por assunto, integrado, não volátil, variável com o tempo, e empregado no apoio às decisões gerenciais. Esse depósito de dados está diretamente relacionado ao seguinte conceito:



- a) OLTP
- b) OLAP
- c) DATAMART
- d) DATAMINING
- e) DATAWAREHOUSE

111. (CESGRANRIO / LIQUIGÁS – 2014) Data Warehouse pode ser conceituado como uma coleção de dados orientada por assuntos, integrada, assim como:

- a) não variante no tempo e volátil.
- b) não variante no tempo e não volátil.
- c) pouco variante no tempo e volátil
- d) variante no tempo e volátil
- e) variante no tempo e não volátil

112. (FUNCAB / MDA – 2014) O principal objetivo de um *Data warehouse* é:

- a) disponibilizar informações para apoio às decisões de uma empresa.
- b) excluir a padronização e integração dos dados existentes nos diversos sistemas.
- c) representar uma tecnologia para a avaliação da qualidade de softwares empresariais.
- d) ser uma metodologia orientada a objetos para desenvolvimento de sistema de informação.
- e) eliminar o uso de dados históricos em operações de consultas e análises de negócio.

113. (FUNCAB / MDA – 2014) O principal objetivo de um *Data warehouse* é:

- a) disponibilizar informações para apoio às decisões de uma empresa.
- b) excluir a padronização e integração dos dados existentes nos diversos sistemas.
- c) representar uma tecnologia para a avaliação da qualidade de softwares empresariais.
- d) ser uma metodologia orientada a objetos para desenvolvimento de sistema de informação.
- e) eliminar o uso de dados históricos em operações de consultas e análises de negócio.

114. (INSTITUTO AOCP / IFN-MG – 2014) Um Data Warehouse pode ser subdividido em pequenas porções para atender um departamento da empresa, por exemplo. Essa subdivisão do Data Warehouse é conhecida como:

- a) Database.
- b) Data Mart.
- c) Data Center.
- d) Data Mining.
- e) Data Store.

115. (FUNDEP / IFN-MG – 2014) Assinale a alternativa CORRETA que apresenta o nome de como é conhecido em data warehouse o subconjunto de dados da organização como, por exemplo, os



dados de um departamento os quais apresentam um foco restrito para análise e descoberta de informações úteis à tomada de decisões.

- a) Data sector.
- b) Data marts.
- c) Data fator.
- d) Data info.

116. (ESAF / MF - 2013) Um conjunto de dados produzido para oferecer suporte à tomada de decisões, sendo um repositório de dados de interesse aos gerentes, é um:

- a) Management dataset.
- b) Data warehouse.
- c) Data center.
- d) Major warehouse.
- e) Data wareset.

117. (FMP CONCURSOS / MPE - AC – 2013) Considere a seguinte definição: "Uma coleção de dados não-voláteis, integrados, orientados a um determinado tema e utilizados no apoio de processos de tomada de decisão. Geralmente contém componentes complementares como facilidades de extração, mineração e transformação de informação e ferramentas analíticas." Essa é a definição de:

- a) Planilha.
- b) Middleware.
- c) Data warehouse.
- d) DBMS (Data Base Management System).
- e) Groupware.

118. (ESPP / MPE-PR – 2013) Como abrangem um grande volume de dados, os data warehouses geralmente são de maior ordem de magnitude que os bancos de dados de origem. Os data warehouses voltados para um subconjunto da organização, como um departamento, são conhecidos por:

- a) data marts.
- b) data sectors.
- c) data virtuais.
- d) data segments.
- e) data sections.

119. (ESPP / MPE - PR – 2013) Um *Data Warehouse* pode ser descrito como sendo uma coleção de tecnologias de apoio à decisão, visando a habilitar o analista a tomar decisões melhores e mais rápidas. Para a criação de um *Data Warehouse* há uma fase que consiste em limpeza e



reformatação dos dados antes que sejam carregados no data warehouse. Esse processo é tratado por ferramentas conhecidas por:

- a) ETL.
- b) DWL.
- c) LDW.
- d) HRD.
- e) DHR.

120. (QUADRIX / DATAPREV – 2012) Um ambiente de *Data Warehouse* Oracle inclui várias ferramentas, além de uma base de dados relacional. Um ambiente típico inclui, entre outras ferramentas, um processo de extração de dados dos sistemas de origem, para levar esses dados ao data warehouse. Esse processo corresponde a:

- a) Motor OLAP.
- b) Solução ETL.
- c) Ferramenta de análise de clientes.
- d) Ferramenta Oracle Warehouse Builder.
- e) Aplicação que reúne dados para entregá-los aos usuários.

121. (IBFC / INEP - 2012) A Inteligência Empresarial, ou *Business Intelligence*, é um termo que foi criado por:

- a) EEE
- b) Gartner Group
- c) Ernst & Young
- d) PricewaterhouseCoopers
- e) ABNT

122. (AOCP / TCE-PA - 2012) Segundo o Gartner Group, é correto afirmar que um BI (*Business Intelligence*) é:

- a) uma política
- b) de informação que especifica regras para compartilhar, disseminar, adquirir, padronizar, classificar e inventariar informações.
- b) um banco de dados que armazena os dados históricos de interesse de toda a empresa para a tomada de decisão, é também conhecido como *Data Warehouse*.
- c) a descrição de habilidades das organizações para acessar dados e explorar informações, analisando e desenvolvendo percepções e entendimentos a seu respeito.
- d) uma tecnologia de software que visa fornecer informações empresariais por meio de uma base de dados.



e) um conjunto de indicadores de desempenho, conhecidos como KPI, com a finalidade de medir qualquer etapa de um processo ou resultado.

123. (OBJETIVA / EPCT – 2012) Jorge está trabalhando na modelagem de dados de um grande repositório de dados que será usado, entre outras coisas, para armazenamento de dados históricos, geração de relatórios, além de obtenção de informações estratégicas para tomada de decisão e análise de grandes volumes de dados. Conceitualmente, esse repositório de dados é chamado de?

- a) Banco de dados transacional.
- b) Data warehouse.
- c) Banco de dados auxiliar.
- d) SIG (Sistema de Informações Gerenciais).

124. (QUADRIX / DATAPREV – 2012) A fim de melhorar o processo de tomada de decisão a partir de dados correntes e históricos da empresa, uma porção resumida e descentralizada dos dados pode ser armazenada em um _____. Essa base menor é destinada a uma área específica de negócio e contém dados altamente focalizados. A alternativa que preenche corretamente a lacuna é:

- a) Data warehouse
- b) Data mart
- c) Data mining
- d) ERP
- e) OLAP

125. (IBFC / INEP – 2012) Devido à construção de um projeto de um Data Warehouse (DW) ser dispendioso e demorado, os arquitetos resolveram segmentar a construção do DW através dos:

- a) Data Minings
- b) Minings Data
- c) Data Marts
- d) Data Smarts
- e) Marts Data

126. (AOCP / TCE-PA – 2012) Extrair os dados de diversos sistemas, transformá-los conforme as regras de negócios e carregá-los em um Data Mart ou em um Data Warehouse é a função das ferramentas de:

- a) SGBD.
- b) BPMS.
- c) XML.
- d) ETL.



e) UML.

127. (AOCP / TCE-PA – 2012) Sobre ETL (Extract, Transform and Load), é correto afirmar que:

- a) a extração e carga são opcionais no processo, porém a transformação é obrigatória.
- b) é o processo para tratamento dos dados de uma ou mais bases de dados de origem, para uma ou mais bases de dados de destino.
- c) concentra a menor parte do esforço exigido no desenvolvimento de um Data Warehouse.
- d) não necessariamente os dados necessitam ficar homogêneos para serem carregados no Data Warehouse, pois uma das funções deste último é resolver os conflitos que não foram resolvidos pela ETL.
- e) na fase de transformação dos dados não devem ser corrigidos erros de digitação ou descoberta de violações de integridade, por exemplo, para os dados serem mantidos como os originais.

128. (ESAF / MPOG - 2010) BI - Business Intelligence:

- a) é uma técnica de otimização da árvore de decisão.
- b) é um método de formação avançada de gestores.
- c) comprehende ferramentas de análise de dados para otimizar os processos produtivos de uma empresa.
- d) são técnicas, métodos e ferramentas para mineração de dados na área de negócios de uma empresa.
- e) são técnicas, métodos e ferramentas de análise de dados para subsidiar processos de decisão de uma empresa.

129. (CETAP / DETRAN RR - 2010) Sobre Sistemas de Apoio à Decisão, leia os enunciados seguintes e marque a alternativa CORRETA:

I - O termo Business Intelligence (BI) nada mais é que uma série de conceitos e metodologias para auxiliar na tomada de decisões estratégicas nas empresas através principalmente de geração de relatórios gerenciais analíticos;

II - Um Data Warehouse é um banco de dados multidimensional grande, de escopo organizacional (ou seja, abrange toda a empresa) e reúne dados de todos os departamentos de forma a permitir a busca rápida de informações para auxiliar a tomada de decisões estratégicas;



III - Um Data Mart é um banco de dados multidimensional de escopo departamental, ou seja, abrange apenas um determinado departamento;

IV - Data Mining (Mineração de dados) é o termo usado para definir uma série de procedimentos, técnicas e ferramentas para recuperar e analisar dados de um Data Warehouse ou Data Mart;

V - Nos Sistemas de Apoio à Decisão é comum desnormalizar os dados para atingir uma melhor performance nos Data Mart e Data Warehouse.

- a) Somente o enunciado III está correto.
- b) Os enunciados I, II e III estão corretos.
- c) Os enunciados I, II e V estão corretos.
- d) Somente o enunciado V está correto.
- e) Todos os enunciados estão corretos.

130. (CESGRANRIO / PETROBRÁS – 2010) Data Warehouse tem por característica ser:

- a) orientado a assuntos.
- b) temporário.
- c) descentralizado.
- d) focalizado em aspectos operacionais.
- e) normalizado.

131. (FUNCAB / PRODAM/AM – 2010) No tocante aos datawarehouses, a característica desses armazéns de dados que está relacionada ao fato de que "uma vez inseridos, os dados não podem ser alterados, embora possam ser eliminados", é:

- a) organização (orientado por assunto).
- b) cliente/servidor.
- c) não volatilidade.
- d) variante de tempo.
- e) relacional.

132. (CESGRANRIO / CASA DA MOEDA – 2009) Uma empresa deseja criar uma base de dados para apoio à tomada de decisão. As informações corporativas encontram-se em 4 bancos de dados distintos e possuem representações diferentes para atributos de uma pessoa física. Por exemplo, em determinado banco de dados, uma pessoa física do sexo masculino é representada por um atributo com valor M, enquanto que em outro banco de dados, com valor 1.

Em que etapa de um procedimento de ETL para essa base de dados deve ocorrer uma padronização desse tipo de valor?

- a) Extração.



- b) Deleção.
- c) Transformação.
- d) Paralelismo.
- e) Carga.

133. (CESGRANRIO / EPE – 2007) Existem muitos relatórios para diretores em todos os sistemas de informação transacionais de uma empresa, gerando uma sobrecarga significativa no banco de dados. Assinale a providência indicada para essa situação:

- a) Implantar a técnica de COCOMO.
- b) Implantar um Data warehouse.
- c) Aumentar a quantidade de testes de estresse.
- d) Desenvolver os sistemas utilizando modelagem orientada a objeto.
- e) Melhorar o uso das ferramentas CASE.

134. (UEG / TJ – GO – 2006) A gestão de uma organização pode ser fortemente amparada por informações gerenciais e de apoio à decisão. Desta forma, o uso do *data warehouse* (DW) surge como um grande aliado. Assinale abaixo a afirmação CORRETA acerca do DW:

- a) O processo de implantação de um DW provoca a alteração de todos os sistemas da organização, pois eles devem deixar de alimentar as bases operacionais, utilizando o próprio DW.
- b) O DW pode ser definido como um repositório de informações que armazena de maneira organizada os dados provenientes dos sites de natureza operacional da organização. Ele permite mapear o histórico da organização.
- c) O DW é um repositório que gradualmente substitui as bases de dados dos vários sistemas da organização gerando uma estrutura organizada e integrada para os sistemas informatizados da organização.
- d) O processo de implantação de um DW provoca a alteração na estrutura de todos os sistemas da organização, pois eles devem alimentar as bases operacionais, bem como o próprio DW.



GABARITO

- | | | | | | |
|-----|---------|-----|---------|------|---------|
| 1. | CORRETO | 41. | CORRETO | 81. | LETRA C |
| 2. | ERRADO | 42. | ERRADO | 82. | LETRA A |
| 3. | ERRADO | 43. | CORRETO | 83. | LETRA C |
| 4. | CORRETO | 44. | ERRADO | 84. | LETRA C |
| 5. | ERRADO | 45. | CORRETO | 85. | LETRA E |
| 6. | CORRETO | 46. | CORRETO | 86. | LETRA B |
| 7. | CORRETO | 47. | CORRETO | 87. | LETRA A |
| 8. | ERRADO | 48. | ERRADO | 88. | LETRA A |
| 9. | LETRA C | 49. | ERRADO | 89. | LETRA E |
| 10. | CORRETO | 50. | CORRETO | 90. | LETRA E |
| 11. | CORRETO | 51. | ERRADO | 91. | LETRA B |
| 12. | CORRETO | 52. | CORRETO | 92. | LETRA D |
| 13. | LETRA D | 53. | ERRADO | 93. | LETRA A |
| 14. | LETRA A | 54. | CORRETO | 94. | LETRA D |
| 15. | CORRETO | 55. | CORRETO | 95. | LETRA E |
| 16. | ERRADO | 56. | CORRETO | 96. | LETRA B |
| 17. | CORRETO | 57. | ERRADO | 97. | LETRA C |
| 18. | ERRADO | 58. | LETRA B | 98. | LETRA C |
| 19. | ERRADO | 59. | ERRADO | 99. | LETRA D |
| 20. | ERRADO | 60. | ERRADO | 100. | LETRA B |
| 21. | ERRADO | 61. | ERRADO | 101. | LETRA C |
| 22. | CORRETO | 62. | ERRADO | 102. | LETRA B |
| 23. | ERRADO | 63. | ERRADO | 103. | LETRA B |
| 24. | LETRA D | 64. | LETRA C | 104. | LETRA C |
| 25. | CORRETO | 65. | LETRA E | 105. | ERRADO |
| 26. | CORRETO | 66. | LETRA B | 106. | LETRA C |
| 27. | ERRADO | 67. | LETRA B | 107. | LETRA B |
| 28. | ERRADO | 68. | LETRA C | 108. | LETRA B |
| 29. | CORRETO | 69. | LETRA D | 109. | LETRA A |
| 30. | CORRETO | 70. | LETRA E | 110. | LETRA E |
| 31. | CORRETO | 71. | LETRA D | 111. | LETRA E |
| 32. | ERRADO | 72. | LETRA E | 112. | LETRA A |
| 33. | ERRADO | 73. | LETRA A | 113. | LETRA A |
| 34. | CORRETO | 74. | LETRA A | 114. | LETRA B |
| 35. | CORRETO | 75. | LETRA E | 115. | LETRA B |
| 36. | CORRETO | 76. | LETRA B | 116. | LETRA B |
| 37. | ERRADO | 77. | LETRA A | 117. | LETRA C |
| 38. | ERRADO | 78. | LETRA A | 118. | LETRA A |
| 39. | CORRETO | 79. | LETRA A | 119. | LETRA A |
| 40. | CORRETO | 80. | LETRA E | 120. | LETRA B |



- | | | |
|---------------------|---------------------|---------------------|
| 121. LETRA B | 126. LETRA D | 131. LETRA C |
| 122. LETRA C | 127. LETRA B | 132. LETRA C |
| 123. LETRA B | 128. LETRA E | 133. LETRA B |
| 124. LETRA B | 129. LETRA E | 134. LETRA B |
| 125. LETRA C | 130. LETRA A | |



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

Concursado(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.