Aula 00 - (Prof. Diego Carvalho e Emannuelle Gouveia)

TCDF (Auditor de Controle Externo - Área Auditoria) Noções de Análise de Dados

Autor:
André Castro, Diego Carvalho, Emannuelle Gouveia Rolim, Renato da Costa

29 de Novembro de 2024

Índice

1) Análise de Informações - Teoria da Informação 3
2) Análise de Informações - Dados Abertos 16
3) Análise de Informações - Metadados 23
4) Análise de Informações - Dados Estruturados e Não-Estruturados 30
5) Análise de Informações - Ciclo de Vida de Dados 37
6) Questões Comentadas - Análise de Informações - Teoria da Informação - Multibancas 44
7) Questões Comentadas - Análise de Informações - Dados Abertos - Multibancas 61
8) Questões Comentadas - Análise de Informações - Metadados - Multibancas 70
9) Questões Comentadas - Análise de Informações - Dados Estruturados e Não-Estruturados - Multibancas 77
10) Questões Comentadas - Análise de Informações - Ciclo de Vida de Dados - Multibancas 83
11) Lista de Questões - Análise de Informações - Teoria da Informação - Multibancas 93
12) Lista de Questões - Análise de Informações - Dados Abertos - Multibancas 101
13) Lista de Questões - Análise de Informações - Metadados - Multibancas 106
14) Lista de Questões - Análise de Informações - Dados Estruturados e Não-Estruturados - Multibancas 110
15) Lista de Questões - Análise de Informações - Ciclo de Vida de Dados - Multibancas 114

TEORIA DA INFORMAÇÃO

Dado, Informação, Conhecimento e Inteligência

Pessoal, vamos começar falando de um assunto que não cai tanto em prova, mas que é importante para compreender o que veremos mais à frente – estou falando da Hierarquia DIKW. **Essa é a sigla em inglês para *Data*, *Information*, *Knowledge* e *Wisdom* – Dados, Informação, Conhecimento e Inteligência (ou Sabedoria).** Esses são elementos fundamentais para a comunicação e para a tomada de decisão nas organizações, mas têm significados diferentes.

Essas palavras formam um sistema hierárquico em que um dado para um indivíduo pode ser uma informação e/ou conhecimento para outro. Galera, em nosso dia a dia, nós costumamos usar alguns desses termos indiscriminadamente – como se fossem sinônimos. **No entanto, no contexto de Sistemas de Informações, esses termos são – na verdade – complementares.** Para entender melhor essas diferenças, vamos analisar como são definidos cada um desses conceitos.

Dado

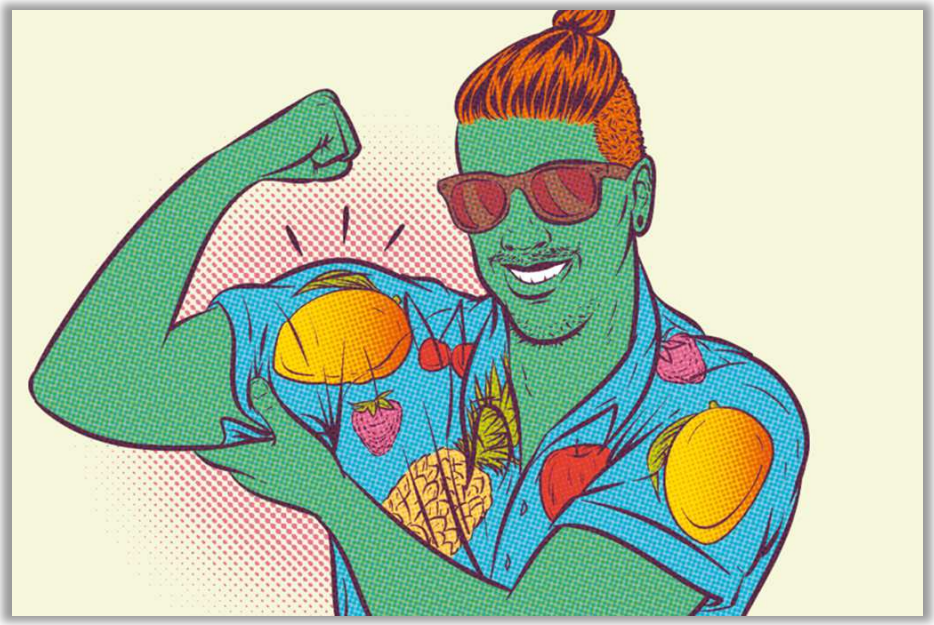
INCIDÊNCIA EM PROVA: ALTÍSSIMA

DADO
Dados são correspondências de um atributo, característica ou propriedade que, sozinho, não tem significado.
Dados são elementos brutos, sem significado, desvinculados da realidade.
Dados são simples observações sobre o estado do mundo.
Dados são um conjunto de fatos objetivos e discretos sobre eventos.
Dados são a menor partícula estruturada que compõe uma informação.

O que podemos concluir? **Podemos concluir que dados são fatos ou estatísticas coletadas para servir de referência ou análise.** O grande lance quando se pensa em dados é que eles não têm significado ou semântica quando vistos isoladamente – eles são brutos, simples e desvinculados de contexto/realidade. Professor, ainda não entendi! Cara... imagine que sem nenhum contexto, eu te mostro um relatório com uma tabela escrita a palavra abaixo em todas as linhas e colunas.

manga

Você consegue tirar alguma informação desse relatório? Não, você pode até inferir que se trata – por exemplo – de algum relatório sobre frutas. No entanto, você não pode afirmar com certeza! *Por quê?* Porque isso pode ser tanto um relatório de um hortifruti quanto um relatório de uma loja de camisetas. *E aí, sem nenhum contexto, tem como saber?* Não, é impossível! *Por quê?* Porque isso é apenas um dado: **bruto, simples, objetivo e desvinculado da realidade.**



Dados são sucessões de fatos brutos, que não foram organizados, processados, relacionados, contextualizados, avaliados ou interpretados, representando apenas partes isoladas de eventos, situações ou ocorrências. Eles se constituem de unidades básicas a partir das quais informações poderão ser elaboradas ou obtidas. **Além disso, eles são frequentemente quantificados, facilmente estruturados, facilmente transferíveis e facilmente obtidos por máquinas.**

Quando os dados são organizados, processados, relacionados, contextualizados, avaliados ou interpretados, ele passa a ser uma informação, conforme veremos no tópico seguinte.

(COREN/PA – 2013) Nos Sistemas de Informação, dados são:

a) informações que possuem significado isoladamente.
b) representações do conhecimento contextualizado.
c) definições do real sentido das coisas.
d) correspondências de um atributo, característica ou propriedade que, sozinho, não tem significado.

_____

Comentários: dados são correspondências de um atributo, característica ou propriedade que, sozinho, não tem significado (Letra D).

(CEFET – 2014) Para os sistemas de informação, definem-se dados como:

a) melhoria contínua de processos organizacionais.
b) resultado da análise do ambiente de negócios.
c) fatos organizados e apresentados de forma útil.
d) sequências de fatos ainda não analisados.
e) decisões tomadas pelos gestores da organização.

_____

Comentários: dados são sequências de fatos ainda não analisados, isto é, brutos (Letra D).

(CGE/PB – 2008 – Letra B) Dados são elementos brutos, sem significado, desvinculados da realidade.

_____

Comentários: eles realmente são elementos brutos, sem significado e desvinculados da realidade (Correto).

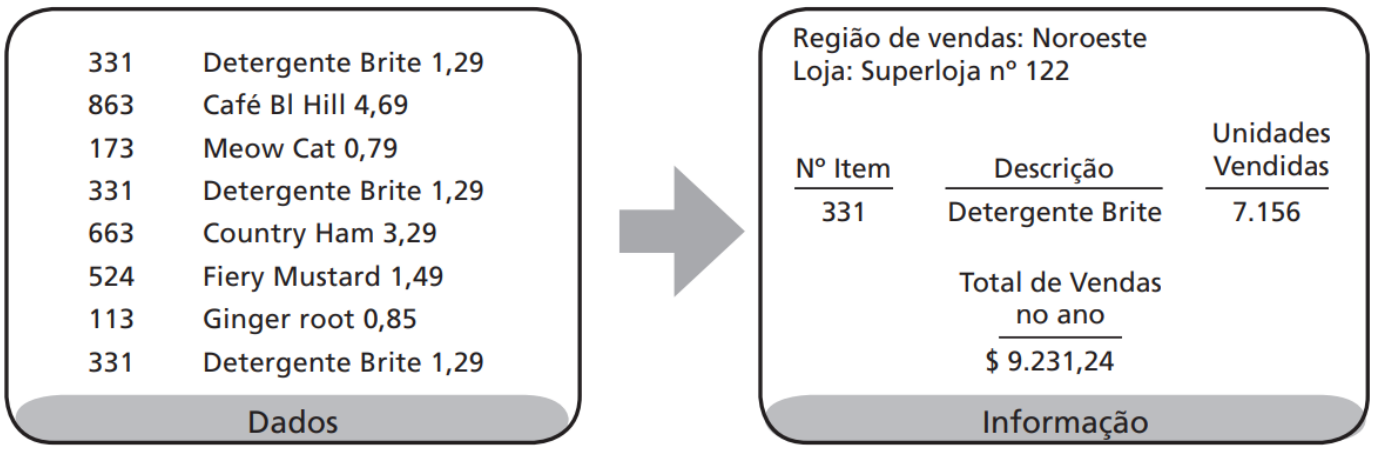
Informação

INCIDÊNCIA EM PROVA: ALTÍSSIMA

INFORMAÇÃO
Conjunto de dados com significado que reduza a incerteza ou que permita o conhecimento a respeito de algo.
Conjunto dos dados presentes em um contexto, carregado de significados e entregue à pessoa adequada.
Conhecimento inscrito (gravado) sob a forma escrita (impressa ou numérica), oral ou audiovisual.
Conjunto de dados contextualizados que visam fornecer uma solução para determinada situação de decisão.
Fatos e/ou dados que encontramos nas publicações, na internet ou mesmo aquilo que as pessoas trocam entre si.
Resultado do processo de acrescentar significado aos dados.
Dados sobre determinado assunto que possam ser interpretados ou tenham significado para o receptor.

Agora ficou fácil saber o que é informação: é simplesmente um dado contextualizado. No momento em que um dado é tratado, ele passa a transmitir uma mensagem e, então, temos uma informação. **A informação é a ordenação e organização dos dados de forma que passa a transmitir uma mensagem compreensiva dentro de um determinado contexto.** Seguindo nosso exemplo anterior, se eu digo: "*Caraca, essa manga está deliciosa!*", nós já conseguimos inferir que se trata, na verdade, da fruta.

Entre os métodos utilizados para transformar dados em informação, temos: **contextualização; categorização; cálculos; correção; e sumarização**. Em tempo, as principais características de uma informação são: precisão; completude; confiabilidade; relevância; verificabilidade; acessibilidade; e segurança. *Galera, preciso dizer que a informação é importantíssima e que pode afetar um comportamento, uma decisão ou um resultado?*



Enfim... informação requer análise, exige consenso em relação ao significado, além de exigir necessariamente a mediação humana. *Fechou?*

(TCM/BA – 2018) O diretor de uma montadora de veículos necessita tomar uma decisão acerca da continuidade ou não de um dos produtos vendidos no Brasil. Para tanto, solicitou um relatório sobre as vendas de carros da marca do último trimestre de 2018, por faixa de preço, região, modelo e cor.

Nessa situação, no contexto de análise da informação, o relatório representa:

a) conhecimento.
b) inteligência.
c) dados.
d) informação.
e) sabedoria.

_____

Comentários: o relatório em si representa a informação. Trata-se de um conjunto de dados contextualizados – veremos conhecimento e inteligência mais à frente (Letra D).

(CRA/SC – 2017) É um conjunto de dados com determinado significado, ou seja, que reduz a incerteza a respeito de algo ou que permite o conhecimento a respeito de algo:

a) Emissor.
b) Dado.
c) Comunicação.
d) Informação.

_____

Comentários: trata-se da definição de informação, isto é, dados com significado (Letra D).

(COREN/PA – 2013) Nos Sistemas de Informação, informação:

a) tem o mesmo significado que dados.

b) não tem conexão com geração de conhecimento.

c) é qualquer dado coletado e mantido por uma pessoa ou organização, de forma não contextualizada.

d) é o conjunto dos dados presentes em um contexto, carregado de significados e entregue à pessoa adequada.

_____

Comentários: (a) Errado, informação é dado processado e com significado; (b) Errado, tem conexão forte com a geração de conhecimento; (c) Errado, a informação é dado contextualizado; (d) Correto, trata-se realmente de um conjunto de dados presentes em um contexto, carregado de significados e entregue à pessoa adequada – excelente definição (Letra D).

Existe tanto um ciclo de vida dos dados quanto um ciclo de vida da informação. Da mesma forma, não há um padrão de classificação – infelizmente cada autor descreve o ciclo de vida com suas fases. **No caso da informação, uma possível classificação compreende as fases de produção, manuseio, armazenamento, transporte e descarte, considerando sua autenticidade, confidencialidade, integridade e disponibilidade.** Agora vejamos os princípios de segurança...

PRINCÍPIOS DE SEGURANÇA	DESCRIÇÃO
CONFIDENCIALIDADE	Capacidade de um sistema de não permitir que informações estejam disponíveis ou sejam reveladas a entidades não autorizadas – incluindo usuários, máquinas, sistemas ou processos.
INTEGRIDADE	Capacidade de garantir que a informação manipulada está correta, fidedigna e que não foi corrompida – trata da salvaguarda da exatidão e completeza da informação.
DISPONIBILIDADE	Propriedade de uma informação estar acessível e utilizável sob demanda por uma entidade autorizada.
AUTENTICIDADE	Propriedade que trata da garantia de que um usuário é de fato quem alega ser. Em outras palavras, ela garante a identidade de quem está enviando uma determinada informação.

(ME – 2020) O ciclo de vida da informação compreende as fases de produção, manuseio, armazenamento, transporte e descarte.

_____

Comentários: essas realmente são as fases do ciclo de vida da informação (Correto).

(ME – 2020) Integridade é a característica que garante o acesso à informação somente para quem estiver autorizado.

_____

Comentários: essa é a definição de confidencialidade e, não, integridade (Errado).

Conhecimento

INCIDÊNCIA EM PROVA: ALTÍSSIMA

Legal, mas como uma informação se torna conhecimento? Essa é uma diferença mais sutil! **Basicamente, o conhecimento acontece quando a informação é aplicada.** *Vamos mudar um pouco o exemplo?* Vejam só: vocês decidem aprender inglês. Palavras em um dicionário de inglês são apenas dados. Quando essas palavras estão em um livro, por exemplo, elas estão processadas e contextualizadas, logo o livro contém informações.

Agora imaginem que vocês fazem cinco anos de escola de inglês – do nível básico ao avançado. Vocês aprendem várias palavras, aprendem a utilizá-las em frases, conseguem entender textos, formar frases e pronunciar tudo perfeitamente. Ora, vocês adquiriram conhecimento de inglês. **Se um curso de inglês ensinar vocabulário, morfologia e sintaxe, mas você não aprender a se comunicar em inglês, você terá informação, mas não terá conhecimento.**

CONHECIMENTO
Propriedade subjetiva, inerente a quem analisa os dados ou as informações.
Conhecer é o processo de compreender e interiorizar as informações recebidas, possivelmente combinando-as de forma a gerar mais conhecimento.
É o que cada indivíduo constrói como produto do processamento, da interpretação, da compreensão dos dados e fatos – é o significado que atribuímos e representamos em nossas mentes sobre a realidade.
O que cada indivíduo constrói como produto do processamento, da interpretação, da compreensão da informação.

Notem as palavras-chave das definições acima: propriedade **subjetiva**; **interiorizar** informações; construção **individual**. É preciso entender que o conhecimento é o recurso de integração da informação. Isso quer dizer que conhecer algo significa saber onde aquilo se integra como parte de algo maior, pois o conhecimento tem a propriedade de um sistema: compõe um todo; tem sua lógica; estabelece um meio. **Segundo Turban, conhecimento é a informação aplicada (em ação).**

O conhecimento vai além de informações, pois ele – além de ter um significado – tem uma aplicação. As informações são valiosas, mas o conhecimento constitui um saber. Produz ideias e experiências que as informações por si só não serão capazes de mostrar. **Se informação é dado processado, então conhecimento é informação processada.** *Bacana?* É importante destacar também que o conhecimento pode ser explícito (tangível) ou tácito (intangível).

No primeiro caso, trata-se da forma mais básica de conhecimento e facilidade de transmissão, uma vez que é geralmente escrito e acessível. Quando os dados são processados, organizados, estruturados e interpretados, o resultado é conhecimento explícito – ele é facilmente articulado, gravado, comunicado e armazenado. Ele inclui livros (inclusive esse que vocês estão lendo), manuais, patentes, base de dados, relatórios, bibliotecas, políticas, procedimentos, entre outros.

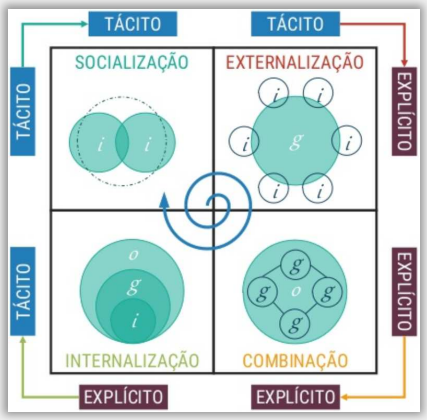
No segundo caso, trata-se do conhecimento que possuímos, obtido da experiência e do contexto pessoal. São as informações que, se solicitadas, seriam as mais difíceis de escrever, articular ou apresentar de forma tangível. *Como assim, Diego?* Galera... minha mãe faz um bolinho de polvilho frito que é de choraaaaaaar! Se ela me der a receita e eu tentar segui-la, vocês podem ter certeza que vai sair uma... porcaria! *Por quê?*

Porque ela tem anos de experiência e aprendeu a sensação exata da massa ou exatamente quanto tempo algo deve ficar no forno e a qual temperatura ideal. Não é algo que ela possa escrever – é *feeling*! No ambiente de trabalho é a mesma coisa... *quem aí já viu um Programa de TV chamado Aeroportos na National Geographic Channel?* **Galera, os auditores fiscais e os policiais federais conseguem identificar facilmente quem está escondendo algo ou cometendo algum crime.**

Como eles conseguem isso? Só com tempo e experiência! Em suma: o conhecimento tácito envolve dimensões técnica e cognitiva. Trata de conceitos como *know-how*, modelos mentais, crenças, percepções cotidianas e práticas adquiridas no dia a dia da relação com outras entidades e indivíduos. **É específico de um contexto, difícil de extrair, codificar, estruturar, capturar por máquinas e transmitir – inclui insights, intuições e sentimentos.**

CONHECIMENTO EXPLÍCITO	CONHECIMENTO TÁCITO
Objetivo, racional e técnico	Subjetivo, cognitivo e experimental
Estruturado e externalizado	Pessoal e internalizado
Fácil de capturar, documentar e compartilhar	Difícil de capturar, documentar e compartilhar
Proveniente da racionalidade	Proveniente da experiência (*expert*)
Conhecimento sequencial	Conhecimento empírico

Existem quatro formas básicas pelas quais o conhecimento pode ser transmitido ou convertido. Elas formam um conceito chamado de **Espiral do Conhecimento (ou Modelo SECI)**:

FORMAS DE TRANSMISSÃO	DESCRIÇÃO	REPRESENTAÇÃO
SOCIALIZAÇÃO	Interação do conhecimento através da troca de ideias e do compartilhamento de experiências, por meio de observação ou/e prática (Tácito → Tácito).	
EXTERNALIZAÇÃO	O conhecimento pessoal passa para o domínio público por meio de uma documentação (Tácito → Explícito).	
COMBINAÇÃO	Agrupamento do conhecimento explícito de diferentes fontes, visando agregar valor para gerar um novo conhecimento (Explícito → Explícito).	
INTERNALIZAÇÃO	O conhecimento é adquirido por meio de alguma publicação escrita que pode estar disponível em livro, revistas, Internet, etc (Explícito → Tácito).	

(CORENPA – 2013) Dado, informação e conhecimento são elementos fundamentais para a comunicação e a tomada de decisão nas organizações, mas seus significados não são tão evidentes. Eles formam um sistema hierárquico de difícil delimitação. O que é um dado para um indivíduo pode ser informação e/ou conhecimento para outro. Davenport corrobora esse ponto de vista colocando resistência em fazer essa distinção e a considera nitidamente imprecisa.

Com isso podemos afirmar que o grande desafio dos tomadores de decisão:

a) é o de minimizar e transformar as interferências individuais em dados e dados em informação, nesse processo de transformação do conhecimento;

b) é o de transformar conhecimento em informação e informação em dados, minimizando as interferências individuais nesse processo de transformação;

c) é o de transformar dados em informação e informação em conhecimento, minimizando as interferências individuais nesse processo de transformação;

d) é o de transformar informação em dados e conhecimento em informação, minimizando as interferências individuais nesse processo de transformação;

e) na transformação da informação e dos dados é interferir o mínimo nesse processo de conhecimento.

_____

Comentários: em nossa hierarquia temos a transformação de dados em informação e informação em conhecimento (Letra C).

Inteligência

INCIDÊNCIA EM PROVA: ALTA

Bacana! Já sabemos que um dado é um fato ou estado não processado; informação é quando você processa um dado; conhecimento é quando você aplica a informação; **e a inteligência (ou sabedoria) é quando você utiliza o seu conhecimento para algum propósito.** *Como assim, professor?* Galera, vejam a tabela a seguir – ela resumirá tudo que nós vimos nos tópicos anteriores. Venham comigo...

NOME	SEXO	IDADE	SALÁRIO	NATURALIDADE
Alice	Feminino	18	R$ 8.500,00	DF
Bernardo	Masculino	26	R$ 2.000,00	GO
Carolina	Feminino	30	R$ 9.500,00	TO
Davi	Masculino	22	R$ 1.500,00	RR
Eduarda	Feminino	32	R$ 7.500,00	SP
Felipe	Masculino	24	R$ 3.000,00	RJ
Gabriela	Feminino	28	R$ 6.500,00	PA
Hugo	Masculino	20	R$ 1.000,00	RS

Se vocês entenderam o que nós vimos anteriormente, vocês serão capazes de me responder algumas perguntas. *Se nós olhássemos isoladamente apenas o número 28, nós teríamos um...?* Dado! *Por quê?* Porque esse número poderia ser o peso de um animal, o número de gols do Gabigol, as polegadas de um monitor, enfim... **sem um contexto, uma análise, um processamento seria impossível identificar do que se trata porque é apenas um dado bruto e simples.**

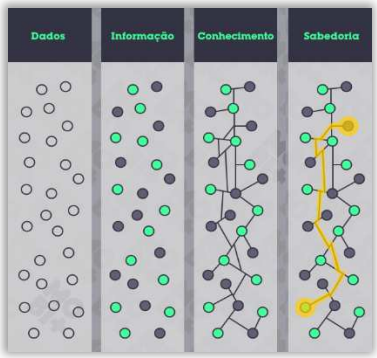
Agora se víssemos os títulos das colunas dessa tabela, poderíamos inferir que 28 é a idade – em anos – de Gabriela. *O que nós acabamos de fazer?* Transformamos dado em informação, isto é, contextualizamos, analisamos, processamos esse dado em uma informação. **Em geral, computadores trabalham com dados, quem é capaz de processá-los somos nós – humanos!** *Tudo legal até aqui?* Então vamos seguir...

Nesse caso, o que seria um exemplo de conhecimento? O salário médio do sexo feminino é mais que duas vezes maior que o salário médio do sexo masculino. *Galera, isso estava explícito na tabela?* Não, eu precisei cruzar informações diferentes da tabela para chegar a essa conclusão. **Logo, eu processei as informações que eu coletei de forma a gerar novos conhecimentos.** *Tudo legal?* Agora vamos para a parte final: inteligência.

A inteligência tem uma premissa interessante: deve ser utilizada com algum propósito, isto é, para resolver algum problema. *Como assim?* Imagine que uma empresa de pesquisa esteja curiosa para entender por que, contrariando estatísticas nacionais, essa amostra de pessoas apresenta uma desigualdade salarial em favor das mulheres. Ela chama todos os seus pesquisadores – todos com formações e níveis de conhecimento semelhantes – e apresenta esse problema a cada um.

Os pesquisadores isoladamente analisam o problema, mas apenas alguns conseguem resolvê-lo. *Por que isso ocorre uma vez que os pesquisadores têm acesso aos mesmos dados, coletam as mesmas informações e possuem os mesmos conhecimentos estatísticos?* Porque é isso que difere a inteligência de conhecimento. **Cada pesquisador utiliza seus conhecimentos de uma forma particular, utilizando conexões cerebrais, memória e experiência para resolver um problema concreto.**

Ora... alguns pesquisadores são mais sábios! *Vocês já viram aqueles problemas de lógica que são utilizados para avaliar o QI (Quociente de Inteligência) de um indivíduo?* Dizem que o Einstein tinha um QI entre 160 e 190! *Por quê?* **Porque dado um conjunto de dados, informações e conhecimentos, ele conseguia resolver problemas que ninguém antes dele sequer chegou perto de resolver.** Einstein era um gênio: sábio e inteligente! *Entendido?*



A inteligência envolve a capacidade de julgamento (Ex: Certo/Errado, Bom/Ruim, etc), isto é, usar o conhecimento para tomar a melhor decisão ou aplicá-lo a alguma situação concreta – trata-se de uma habilidade essencialmente humana e extremamente difícil de ser obtida por meio de máquinas. *Professor, e a inteligência artificial?* Pois é, trata-se de uma linha de pesquisa que busca fazer com que máquinas façam análises inteligentes sobre um conjunto de dados.

(Polícia Federal – 2018) O conhecimento é embasado na inteligência das informações que são coletadas e analisadas para uma organização.

_____

Comentários: a questão inverteu os conceitos – a inteligência é embasada no conhecimento das informações que são coletadas e analisadas para uma organização (Errado).

RESUMO

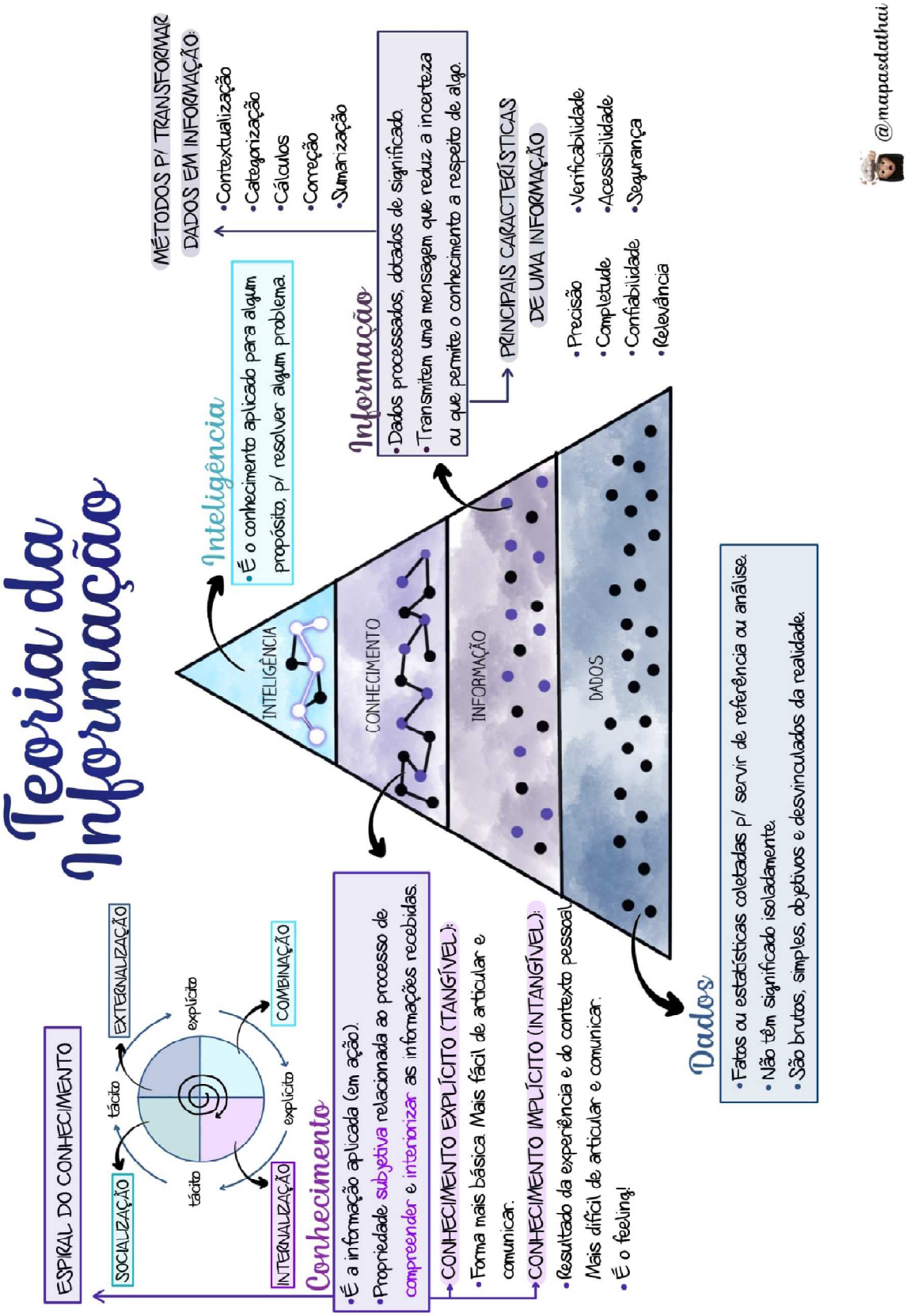
DADO
Dados são correspondências de um atributo, característica ou propriedade que, sozinho, não tem significado.
Dados são elementos brutos, sem significado, desvinculados da realidade.
Dados são simples observações sobre o estado do mundo.
Dados são um conjunto de fatos objetivos e discretos sobre eventos.
Dados são a menor partícula estruturada que compõe uma informação.

INFORMAÇÃO
Conjunto de dados com significado que reduza a incerteza ou que permita o conhecimento a respeito de algo.
Conjunto dos dados presentes em um contexto, carregado de significados e entregue à pessoa adequada.
Conhecimento inscrito (gravado) sob a forma escrita (impressa ou numérica), oral ou audiovisual.
Conjunto de dados contextualizados que visam fornecer uma solução para determinada situação de decisão.
Fatos e/ou dados que encontramos nas publicações, na internet ou mesmo aquilo que as pessoas trocam entre si.
Resultado do processo de acrescentar significado aos dados.
Dados sobre determinado assunto que possam ser interpretados ou tenham significado para o receptor.

CONHECIMENTO
Propriedade subjetiva, inerente a quem analisa os dados ou as informações.
Conhecer é o processo de compreender e interiorizar as informações recebidas, possivelmente combinando-as de forma a gerar mais conhecimento.
É o que cada indivíduo constrói como produto do processamento, da interpretação, da compreensão dos dados e fatos – é o significado que atribuímos e representamos em nossas mentes sobre a realidade.
O que cada indivíduo constrói como produto do processamento, da interpretação, da compreensão da informação.

INTELIGÊNCIA
Trata-se do mais alto nível de abstração, com visão prospectiva e a capacidade de ver além do horizonte.
Trata-se do conhecimento sintetizado e aplicado a determinada situação para ganhar maior profundidade de consciência da mesma.

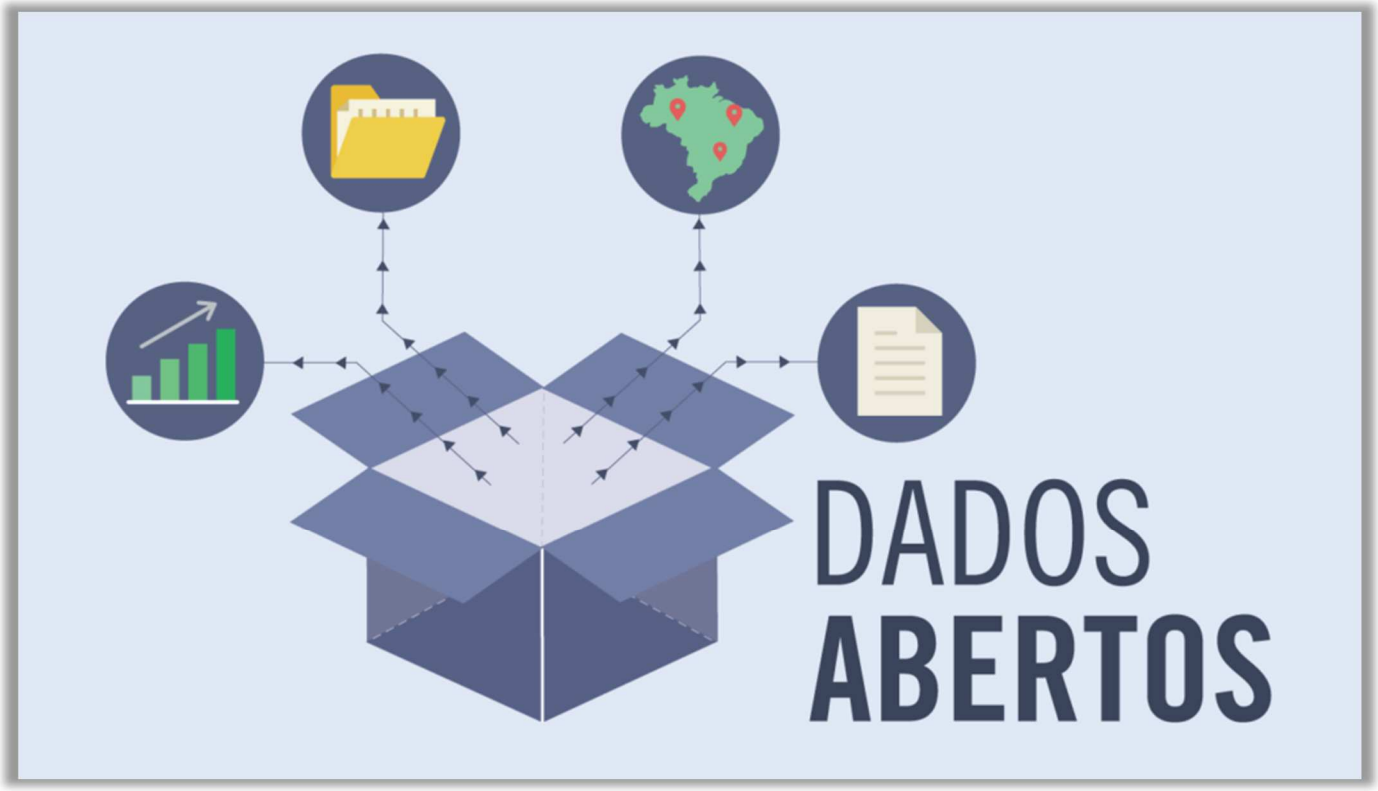
MAPA MENTAL


Teoria da Informação
Dados
•Fatos ou estatísticas coletadas p/ servir de referência ou análise.
•Não têm significado isoladamente.
•São brutos, simples, objetivos e desvinculados da realidade.
DADOS
INFORMAÇÃO
CONHECIMENTO
INTELIGÊNCIA
Informação
•Dados processados, dotados de significado.
•Transmitem uma mensagem que reduz a incerteza ou que permite o conhecimento a respeito de algo.
PRINCIPAIS CARACTERÍSTICAS DE UMA INFORMAÇÃO
•Precisão
•Completude
•Confiabilidade
•Relevância
•Verificabilidade
•Acessibilidade
•Segurança
MÉTODOS P/ TRANSFORMAR DADOS EM INFORMAÇÃO:
•Contextualização
•Categorização
•Cálculos
•Correção
•Sumarização
Inteligência
•É o conhecimento aplicado para algum propósito, p/ resolver algum problema.
Conhecimento
•É a informação aplicada (em ação).
•Propriedade subjetiva relacionada ao processo de compreender e intencionar as informações recebidas.
→ CONHECIMENTO EXPLÍCITO (TANGÍVEL):
•Forma mais básica. Mais fácil de articular e comunicar.
→ CONHECIMENTO IMPLÍCITO (INTANGÍVEL):
•Resultado da experiência e do contexto pessoal. Mais difícil de articular e comunicar.
•É o feeling!
ESPIRAL DO CONHECIMENTO
SOCIALIZAÇÃO
EXTERNALIZAÇÃO
COMBINAÇÃO
INTERNALIZAÇÃO
tácito
explícito
tácito
explícito
@mapasdathai

DADOS ABERTOS

Conceitos Básicos

INCIDÊNCIA EM PROVA: BAIXA



Vamos lá... está acabando! Ânimo e altivez! Segundo a Open Knowledge Foundation – OKFn, "***dados são abertos quando qualquer pessoa pode livremente usá-los, reutilizá-los e redistribuí-los, estando sujeito à, no máximo, exigência de creditar a sua autoria e compartilhar pela mesma licença***". Quando os dados são produzidos, coletados ou custodiados por autoridades públicas e disponibilizados em formato aberto, diz-se que são dados abertos governamentais.

Já segundo o Decreto 8.777/2016, os dados abertos são "*dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte*". Dito isso, para que um dado seja considerado ***aberto***, ele deve reunir – no mínimo – as três características:

DISPONIBILIDADE E ACESSO	Os dados devem estar disponíveis como um todo e sob custo não maior que um custo razoável de reprodução, preferencialmente possíveis de serem baixados pela internet. Os dados devem também estar disponíveis de uma forma conveniente e modificável.

REUTILIZAÇÃO E REDISTRIBUIÇÃO	Os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição – é permitida a combinação com outros conjuntos de dados não abertos (um conjunto de dados sobre um assunto específico também pode ser chamado de *data set*).
PARTICIPAÇÃO UNIVERSAL	Todos devem ser capazes de usar, reutilizar e redistribuir – não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos. Por exemplo, restrições de uso 'não-comercial' que impediriam o uso 'comercial', ou restrições de uso para certos fins (ex: somente educativos) excluem determinados dados do conceito de 'abertos'.

Essas três características podem ser interpretadas como leis ou normas fundamentais dos dados abertos e podem ser detalhadas como apresentado abaixo:

I - Se o dado não pode ser encontrado e indexado na Web, ele não existe;
II - Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e
III - Se algum dispositivo legal não permitir sua replicação, ele não é útil.

As chamadas três "leis" dos dados abertos não são leis no sentido literal, promulgadas por algum Estado. **São, em suma, um conjunto de testes para avaliar se um dado pode, de fato, ser considerado aberto e surgem em consequência das normas estudadas acima.** As leis foram propostas para os dados abertos governamentais, mas pode-se dizer que elas se aplicam aos dados abertos de forma geral, mesmo fora de ambientes governamentais (empresas privadas, ONGs, etc).

O Banco Mundial, por exemplo, disponibiliza dados abertos. Dados também podem ser abertos voluntariamente por organizações privadas, por diversos motivos. **Nos últimos anos, especialistas têm discutido a abertura de dados pelo setor privado para ações que beneficiam o interesse público, os chamados "colaborativos de dados".** Vamos agora conhecer quais são os princípios que norteiam os dados abertos:

PRINCÍPIOS	DESCRIÇÃO
COMPLETOS	Todos os dados públicos são disponibilizados. Dados são informações eletronicamente gravadas, incluindo, mas não se limitando a documentos, bancos de dados, transcrições e gravações audiovisuais. Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou controle de acesso, reguladas por estatutos.
PRIMÁRIOS	Os dados são publicados na forma coletada na fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada.
ATUAIS	Os dados são disponibilizados o quão rapidamente seja necessário para preservar o seu valor.
ACESSÍVEIS	Os dados são disponibilizados para o público mais amplo possível e para os propósitos mais variados possíveis.

PROCESSÁVEIS POR MÁQUINA	Os dados são razoavelmente estruturados para possibilitar o seu processamento automatizado.
ACESSO NÃO DISCRIMINATÓRIO	Os dados estão disponíveis a todos, sem que seja necessária identificação ou registro.
FORMATOS NÃO PROPRIETÁRIOS	Os dados estão disponíveis em um formato sobre o qual nenhum ente tenha controle exclusivo (ex: .csv, .html, .txt, .json, .xml, .ods, .rdf, .svg).
LIVRES DE LICENÇAS	Os dados não estão sujeitos a regulações de direitos autorais, marcas, patentes ou segredo industrial. Restrições razoáveis de privacidade, segurança e controle de acesso podem ser permitidas na forma regulada por estatutos (não é completamente livre de licenças).

Sobre o penúltimo ponto, é importante destacar que a filosofia de dados abertos não define exaustivamente uma lista de formatos permitidos. **Para ser considerado um dado aberto, o conjunto de dados deve estar disponível em um formato de especificação aberta, não proprietário, e estruturado, ou seja, que possibilite seu uso irrestrito e automatizado através da Web.** Além disso, é imprescindível que seja utilizado um formato amplamente conhecido.

MNEMÔNICO DOS PRINCÍPIOS							
CONTA	PARA	AQUELAS	ANTES	PRUDENTES	ACEITAREM	FLAMENGO	LIVREMENTE

Em 2007, um grupo de trabalho de 30 pessoas reuniu-se na Califórnia, Estados Unidos da América, para definir os princípios dos Dados Abertos Governamentais. Além disso, o grupo afirmou que a conformidade com esses princípios precisa ser verificável e uma pessoa deve ser designada como contato responsável pelos dados. Por fim, o TCU fez uma publicação com cinco motivos para a abertura de dados na Administração Pública, que vocês podem ver a seguir:

MOTIVOS PARA ABERTURA DOS DADOS
Transparência na gestão pública;
Contribuição da sociedade com serviços inovadores ao cidadão;
Aprimoramento na qualidade dos dados governamentais;
Viabilização de novos negócios;
Obrigatoriedade por lei.

Para a administração pública, abrir dados é uma obrigação legal. A Lei de Acesso à Informação – LAI (Lei 12.527/2011) se aplica aos órgãos públicos da administração direta e entes da administração indireta dos Poderes Executivo, Legislativo, incluindo as Cortes de Contas, e Judiciário e do Ministério Público, em todas as esferas (art. 1ª, parágrafo único). No que diz respeito à transparência ativa, a LAI traz consigo conceitos de dados abertos, em especial em seu art. 8º:

> *Art. 8º É dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas.*
>
> *(...)*
> *§ 2º Para cumprimento do disposto no caput, os órgãos e entidades públicas deverão utilizar todos os meios e instrumentos legítimos de que dispuserem, sendo obrigatória a divulgação em sítios oficiais da rede mundial de computadores (internet).*
>
> *§ 3º Os sítios de que trata o § 2º deverão, na forma de regulamento, atender, entre outros, aos seguintes requisitos:*
>
> *(...)*
> *II – possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações;*
>
> *III – possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina;*

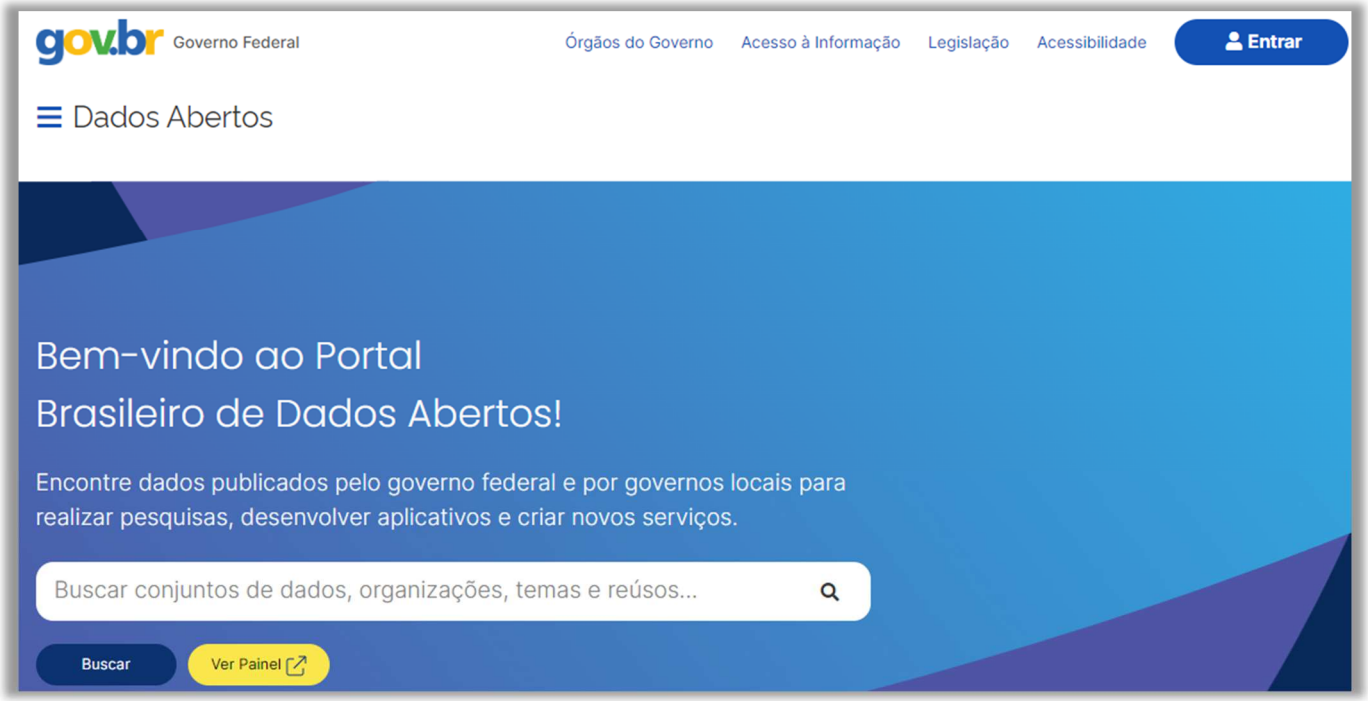
A lei também define as hipóteses de sigilo e de informações pessoais, que são consideradas exceções à regra geral de que os dados devem ser abertos.

O objetivo, além da promoção da transparência governamental, é fornecer insumos aos cidadãos e à sociedade civil para que sejam desenvolvidas ferramentas que proporcionem uma percepção mais efetiva das atividades e dos gastos do governo e contribuam para a melhoria da gestão pública, o estímulo ao controle e participação social, a geração de emprego e renda e o fomento à inovação tecnológica. *Compreendido?*

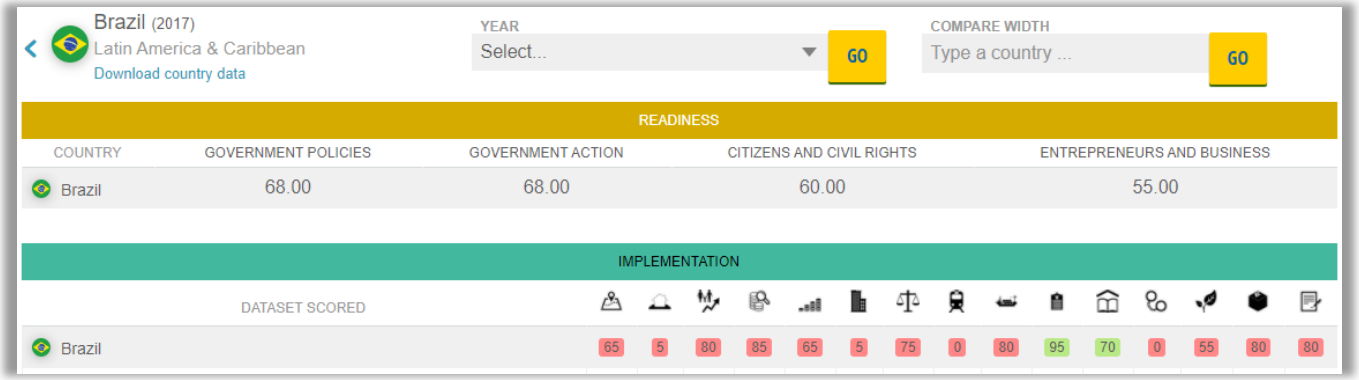
Por fim, é importante falarmos sobra a Infraestrutura Nacional de Dados Abertos (INDA). Trata-se de um conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas no modelo de Dados Abertos, em conformidade com o disposto na e-PING (Padrão de Interoperabilidade de Governo Eletrônico).

O principal componente da INDA é o Portal Brasileiro de Dados Abertos – ferramenta para que todos possam encontrar e utilizar os dados e as informações públicas, prezando pela simplicidade e organização, a fim de disponibilizar, facilmente, os dados e informações que o usuário necessita. **O portal também tem o objetivo de promover a interlocução entre atores da sociedade e governo, para pensar a utilização mais adequada dos dados em prol de uma sociedade melhor.**

HTTP://WWW.DADOS.GOV.BR



Curiosidade: existe um Barômetro de Dados Abertos (Open Data Barometer), que é um indicador estratégico que mede a prevalência e o impacto de iniciativas de dados abertos em todo o mundo. Ele analisa tendências globais e fornece dados comparativos sobre países e regiões para classificar os governos quanto à prontidão, à implementação e ao impacto dos dados abertos. O indicador é uma referência importante para os formuladores de políticas e a sociedade civil.



Atualmente, ele está sendo usado para medir o progresso e mudar as políticas, cobrindo uma coleta de dados de 115 países na edição de 2016. Nessa edição, o Brasil encontrava-se na 14ª posição mundial no indicador de acesso à informação, com score total de 59 pontos. Na América do Sul, o país ocupava a 2ª posição na avaliação, ficando atrás apenas do Uruguai (61 pontos). **É uma curiosidade, mas já caiu em prova...**

RESUMO

DADOS ABERTOS
Dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte.

DISPONIBILIDADE E ACESSO	Os dados devem estar disponíveis como um todo e sob custo não maior que um custo razoável de reprodução, preferencialmente possíveis de serem baixados pela internet. Os dados devem também estar disponíveis de uma forma conveniente e modificável.
REUTILIZAÇÃO E REDISTRIBUIÇÃO	Os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição – é permitida a combinação com outros conjuntos de dados não abertos (um conjunto de dados sobre um assunto específico também pode ser chamado de *data set*).
PARTICIPAÇÃO UNIVERSAL	Todos devem ser capazes de usar, reutilizar e redistribuir - não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos. Por exemplo, restrições de uso 'não-comercial' que impediriam o uso 'comercial', ou restrições de uso para certos fins (ex: somente educativos) excluem determinados dados do conceito de 'abertos'.

LEIS DOS DADOS ABERTOS
I - Se o dado não pode ser encontrado e indexado na Web, ele não existe;
II - Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e
III - Se algum dispositivo legal não permitir sua replicação, ele não é útil.

PRINCÍPIOS	DESCRIÇÃO
COMPLETOS	Todos os dados públicos são disponibilizados. Dados são informações eletronicamente gravadas, incluindo, mas não se limitando a documentos, bancos de dados, transcrições e gravações audiovisuais. Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou controle de acesso, reguladas por estatutos.
PRIMÁRIOS	Os dados são publicados na forma coletada na fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada.
ATUAIS	Os dados são disponibilizados o quão rapidamente seja necessário para preservar o seu valor.
ACESSÍVEIS	Os dados são disponibilizados para o público mais amplo possível e para os propósitos mais variados possíveis.
PROCESSÁVEIS POR MÁQUINA	Os dados são razoavelmente estruturados para possibilitar o seu processamento automatizado.

ACESSO NÃO DISCRIMINATÓRIO	Os dados estão disponíveis a todos, sem que seja necessária identificação ou registro.
FORMATOS NÃO PROPRIETÁRIOS	Os dados estão disponíveis em um formato sobre o qual nenhum ente tenha controle exclusivo.
LIVRES DE LICENÇAS	Os dados não estão sujeitos a regulações de direitos autorais, marcas, patentes ou segredo industrial. Restrições razoáveis de privacidade, segurança e controle de acesso podem ser permitidas na forma regulada por estatutos.

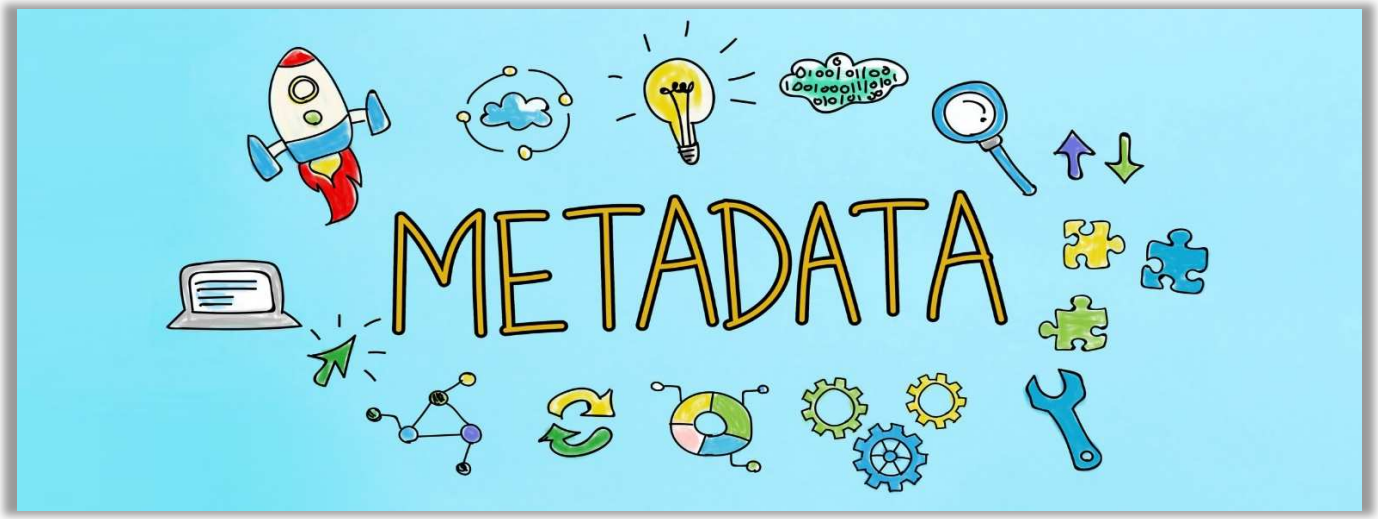
MNEMÔNICO DOS PRINCÍPIOS							
CONTA	PARA	AQUELAS	ANTES	PRUDENTES	ACEITAREM	FLAMENGO	LIVREMENTE

MOTIVOS PARA ABERTURA DOS DADOS
Transparência na gestão pública;
Contribuição da sociedade com serviços inovadores ao cidadão;
Aprimoramento na qualidade dos dados governamentais;
Viabilização de novos negócios;
Obrigatoriedade por lei.

METADADOS

Conceitos Básicos

INCIDÊNCIA EM PROVA: BAIXÍSSIMA

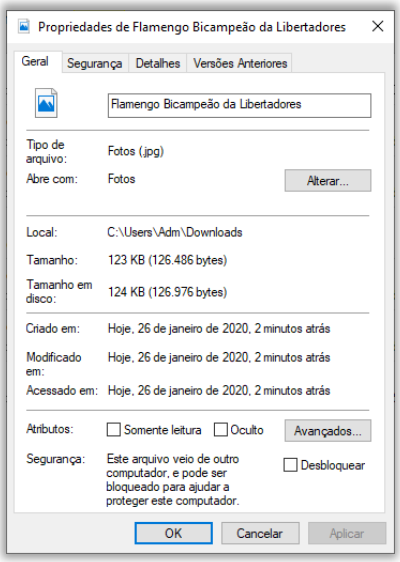


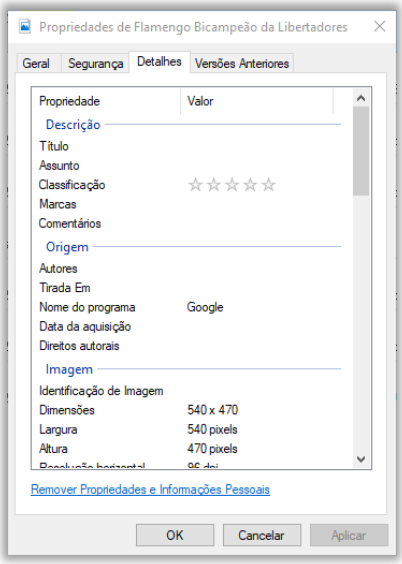
Galera... assunto tranquilaço! *O que seria um metadado?* Vamos pensar na etimologia da palavra: o prefixo "*meta*" vem do grego e significa "*além de*". Dessa forma, metadados são dados que acrescem informações aos próprios dados com o objetivo de fornecer informações sobre eles e de tornar mais fácil a sua compreensão e organização. **Uma definição mais simples é: metadados são dados sobre dados.**

Desde tempos antigos, esse tipo de informação é utilizado para classificar, organizar e pesquisar. Na Suméria (atual Iraque), as placas de argila eram identificadas por fios coloridos conforme seu tipo e arranjadas em prateleiras com indicações escritas ao lado. Os escribas romanos atavam documentos relacionados, etiquetavam-nos e penduravam-nos no teto! **O que agora é diferente é que a informação é eletrônica, dispersa e cresce a uma velocidade exponencial.**

É importante ressaltar que um dado isoladamente não significa nada! **Os metadados são utilizados para ajudar a descrever melhor os dados.** Eles facilitam o entendimento de suas utilidades; ajudam os usuários a encontrar informações relevantes e a descobrir recursos; auxiliam na organização, identificação, arquivamento e preservação de recursos eletrônicos, entre outros. Legal, mas a melhor maneira de entender o que eles são é por meio de exemplos...

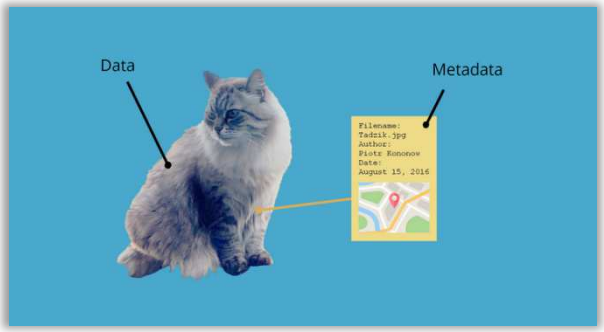
Você está usando seu computador nesse momento? Se sim, procure uma foto qualquer, clique com o botão direito do mouse e acesse suas propriedades. Veja na imagem a seguir que as propriedades de um arquivo trazem diversos dados sobre a imagem (que também é um dado). **Na Aba Geral, ela informa o nome; tipo; software de abertura; local de armazenamento; tamanho; data de criação, modificação e acesso; atributos, segurança, entre outros.**





Na Aba Detalhes, podemos inserir o título, assunto, classificação, marcas, comentários, autores, data em que foi tirada a foto, nome do programa, data de aquisição, direitos autorais, dimensões e diversos outros dados. Essas informações podem ou não ser preenchidas. **Em geral, quando se tira uma foto com uma câmera (não é caso dessa imagem), diversas outras informações são preenchidas sobre a fotografia.**

Galera, é evidente que cada dado terá tipos diferentes de metadados. Um arquivo de vídeo, por exemplo, conterá informações sobre taxa de dados, taxa de bits total, taxa de frames por segundo, taxa de amostragem, canais de áudio, entre outros; um arquivo de música, conterá dados sobre artista, título, álbum, duração, letra, capa, entre outros; e assim por diante para cada tipo de arquivo.



Engana-se quem acha que isso só existe no mundo digital. **Os metadados eram tradicionalmente usados nos catálogos de cartões das bibliotecas até a década de noventa.** *Professor, eu tenho vinte anos – não sei nem o que é isso!* Então, novinho(a)... antigamente os metadados sobre livros (título, autor, quantidade de páginas, nome da editora, etc) eram guardados em pequenos cartões conforme mostra a imagem ao lado.

Com o passar do tempo, os metadados se tornaram primariamente digitais. **Hoje em dia, grande parte dos dados que nós criamos ou manipulamos diariamente contém metadados – por meio deles, recursos digitais podem ser pesquisados por meio de critérios relevantes.** Os metadados das atividades de telecomunicações, incluindo o tráfego da Internet, são amplamente coletados por várias organizações governamentais nacionais. Isso pode ser perigoso...

Esses dados podem ser utilizados para fins de análise de tráfego e para vigilância em massa. *Como assim, Diego?* Galera... tudo que fazemos na internet é criar e manipular dados. Você abre o seu navegador e acessa o Facebook – ele guarda a hora que você entrou, quais páginas você visitou, em que posts você deu like; você acessa o Youtube – ele guarda o que você assistiu, quanto tempo você navegou, em quais vídeos você comentou; entre outros.

Em muitos países, organizações governamentais costumam armazenar metadados sobre e-mails, telefonemas, páginas web, tráfego de vídeo, conexões IP e localizações de telefones. Quando um usuário utiliza seu smartphone para tirar uma foto de algo que deseja vender e publica, por exemplo, a sua imagem em algum site de anúncios ou em redes sociais, **há grande chance de que criminosos extraiam os metadados para fins maliciosos.**

Por vezes, os sites podem ocultar os metadados para outros usuários, mas é sempre bom ter cuidado com esse tipo de exposição de metadados. *Querem outro exemplo?* Em fevereiro de 2003, o departamento pessoal do primeiro-ministro britânico –Tony Blair – publicou um dossiê sobre as organizações de segurança e inteligência do Iraque, além de uma suposta arma de destruição em massa iraquiana.

Um professor de política na Universidade de Cambridge fez o download do arquivo .doc (escrito com o MS-Word) e rapidamente encontrou metadados que indicavam quem eram os autores do documento – mais precisamente, as pessoas que fizeram as dez últimas edições e encontraram os nomes de Paul Hamill, John Pratt, Alison Blackshaw e Murtaza Khan. **Os metadados acabaram expondo os funcionários que criaram o documento.**

Por meio dos metadados, foi possível descobrir também que o arquivo havia sido plagiado. **Essa informação levantou algumas bandeiras sobre qualidade, autenticidade e credibilidade de um relatório que deveria ser ultrassecreto.** Após esse incidente, devido a quantidade de metadados associada ao documento, o governo escolheu o uso da versão .pdf para relatório, uma vez que esse formato contém menos metadados.

Em suma: metadados documentam e organizam os dados das organizações, com objetivo de minimizar duplicações de esforços e facilitar a manutenção de dados. Ademais, são utilizados para criação e uso de documentos em sistema informatizados disponíveis aos usuários; metadados provêm de uma descrição concisa a respeito dos dados. São eles: documentos, tabelas, imagens, vídeos, coleção de documentos, entre outros. Vejamos agora as classificações de metadados...

Poucas organizações entendem os metadados e como podem fazer para implementar uma estratégia de metadados. Nesse contexto, temos a seguinte classificação em termos de uso:

METADADOS	DESCRIÇÃO
TÉCNICOS	Descrição dos dados necessários para as diversas ferramentas que precisem armazenar, manipular ou movimentar dados (Ex: SGBD, Ferramentas CASE, Ferramentas OLAP, etc).

DE NEGÓCIO	Descrição de dados necessários pelos usuários de negócios, para entender o contexto do negócio e o significado dos dados.

Seus padrões também são outra forma de distinguir tipos de metadados. Nesse sentido, eles podem ser classificados em sintático, estruturais ou semânticos:

METADADOS	DESCRIÇÃO
SINTÁTICOS	Dados que descrevem a sintaxe dos dados.
ESTRUTURAIS	Dados que descrevem a estrutura/organização dos dados.
SEMÂNTICOS	Dados que descrevem o significado/conteúdo dos dados em um domínio específico.

Uma terceira classificação que também cai em prova divide os metadados em três tipos de acordo com a sua função: descritivos, estruturais e administrativos.

METADADOS	DESCRIÇÃO
DESCRITIVOS	Estão relacionados à descrição de um objeto para torná-lo mais fácil de ser descoberto e identificado; podem incluir elementos, tais como título, autor, resumo, palavras-chave e identificador – dependem dos metadados estruturais para serem construídos;
ESTRUTURAIS	Estão relacionados à documentação de como um recurso é composto ou organizado – facilitando a navegação e apresentação de recursos. Um livro digitalizado, por exemplo, pode ser publicado como arquivos de imagens individuais das páginas, arquivos PDF ou HTML.
ADMINISTRATIVOS	Estão relacionados ao fornecimento de informações que apoiam os processos de gerenciamento de recursos informacionais. Incluem, por exemplo, informações sobre como e quando o recurso foi criado e a razão da sua criação.

É importante mencionar também que metadados estão bastante relacionados a bancos de dados. Todo banco de dados possui uma completa definição ou descrição da sua estrutura em um lugar chamado catálogo de dados. Ele mantém informações como a estrutura de cada arquivo, o tipo e o formato de armazenamento de cada item de dado. **A informação armazenada no catálogo é chamada de metadados e descreve a estrutura do banco de dados.**

Por fim, é impossível falar de metadados sem mencionar XML (eXtensible Markup Language). *O que diabos é isso, Diego?* **É uma linguagem de marcação que define um conjunto de regras para codificar documentos em um formato que seja legível tanto por humanos quanto por máquinas.** *Caraca, professor... entendi foi nada!* Vamos por partes: assim como humanos falam diversos idiomas, computadores entendem diversas linguagens!

Você tem algum amigo programador? Pois é, esse cara domina algumas linguagens de programação. O XML não é exatamente uma linguagem de programação – trata-se de uma linguagem de marcação. Isso significa que ele possui um conjunto de tags que marcam o início de o fim de informações, além de definir seu significado. **Eu sei que está abstrato, então vamos ver um exemplo para ficar mais claro – vejam o quadro abaixo e tentem entender antes de prosseguir...**

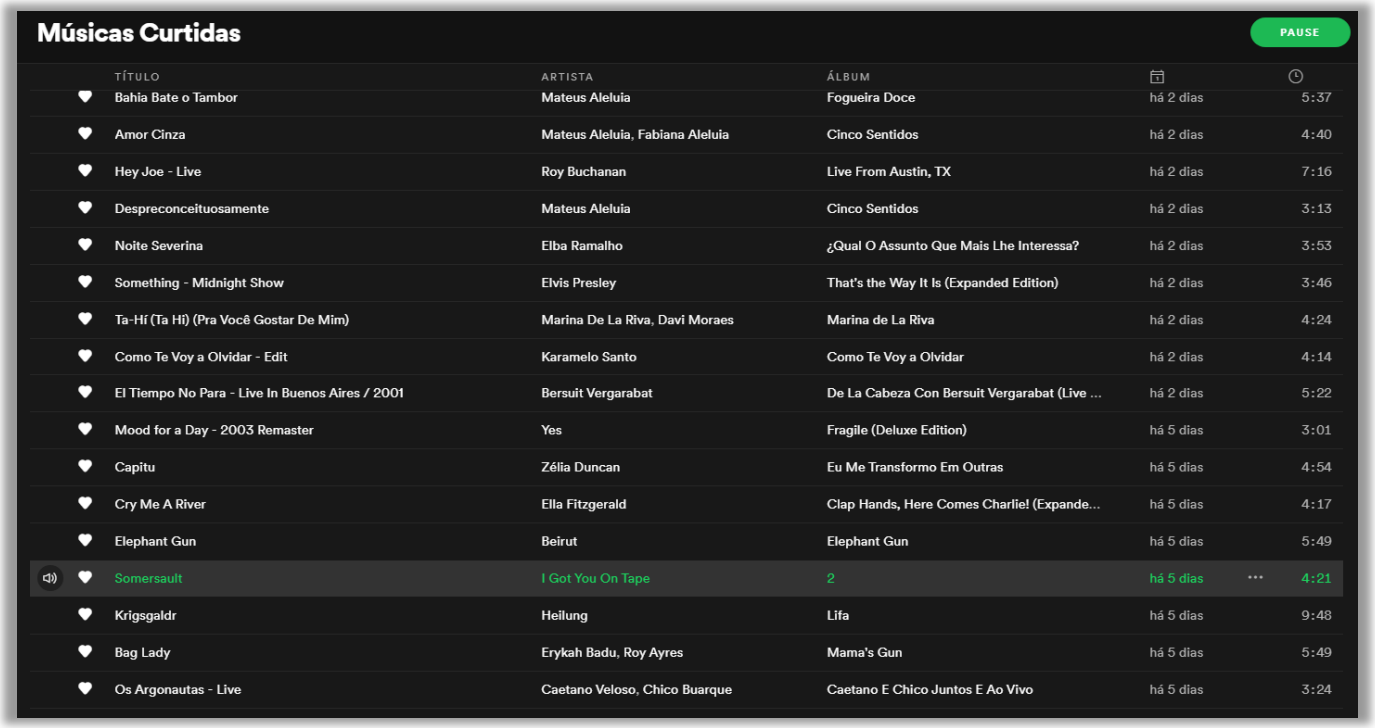
```
<?xml version="1.0" encoding="UTF-8"?>
<DISCO>
   <ARTISTA nome="Molejo">
      <ALBUM titulo="Brincadeira de Criança">
         <FAIXA titulo="Tá Maluca?" tempo="5:15" />
         <FAIXA titulo="Dança da Vassoura" tempo="4:41" />
         <FAIXA titulo="Pula por Cima da Dor" tempo="4:27" />
         <FAIXA titulo="Self-Service" tempo="3:13" />
         <FAIXA titulo="Garoto Zona Sul" tempo="5:01" />
         <FAIXA titulo="Assim Oh!" tempo="4:47" />
         <FAIXA titulo="Lalá" tempo="4:50" />
         <FAIXA titulo="Etiqueta" tempo="5:20" />
         <FAIXA titulo="Menina" tempo="3:21" />
         <FAIXA titulo="Dá Ela Pra Mim?" tempo="2:59" />
         <FAIXA titulo="Amor Estou Sofrendo" tempo="4:32" />
         <FAIXA titulo="Pode Parar" tempo="3:23" />
         <FAIXA titulo="Minha Fortaleza" tempo="4:25" />
      </ALBUM>
      <PAIS> Brasil </PAIS>
      <ANO> 1997 </ANO>
      <FORMATO> CD </FORMATO>
      <GRAVADORA> Continental East West </GRAVADORA>
      <GENERO> Pagode </GRAVADORA>
      <DESCRICAO link=" https://pt.wikipedia.org/wiki/Brincadeira_de_Crian%C3%A7a>
           Discaaaaaaaaaaço da melhor banda de pagode dos anos 90! Quem discorda, é clubista!
      </DESCRICAO>
   </ARTISTA>
</DISCO>
```

Sem eu explicar nada sobre a linguagem, vocês conseguem entender alguma coisa sobre do que se trata esse código acima? Claramente traz dados organizados sobre um disco de um grupo musical! *Estão vendo como XML é uma linguagem compreensível tanto para um humano quanto para o computador?* **Pois é... as palavras em azul são as *tags*!** São simplesmente marcações que indicam o início e o fim de informações que você pode criar para descrever o que quiser. Exemplo:

```
<FORMATO> INFORMAÇÃO </FORMATO>
```

O XML é uma linguagem de marcação utilizada para descrever dados! *Ora, então são dados sobre dados?* Sim, então estamos falando sobre metadados! *Quem aí utiliza Spotify, Deezer ou outro aplicativo de streaming de música?* Galera, esses aplicativos possuem uma base de dados gigantesca de músicas e cada música possui uma infinidade de metadados para que nós – usuários – possamos pesquisar músicas por nome, banda, ano, gênero, país, etc.

As tags (em português, etiquetas) são palavras-chave relevantes utilizadas para agrupar um conjunto de informações. Elas são, usualmente, escolhidas informalmente e como escolha pessoal do autor ou criador do item de conteúdo, isto é, não é parte de um esquema formal de classificação. Elas podem ser utilizadas para descrever qualquer tipo de informação em diferentes contextos. Vejam na imagem seguinte diversos metadados das minhas músicas no Spotify...



Músicas Curtidas — PAUSE

TÍTULO	ARTISTA	ÁLBUM		
Bahia Bate o Tambor	Mateus Aleluia	Fogueira Doce	há 2 dias	5:37
Amor Cinza	Mateus Aleluia, Fabiana Aleluia	Cinco Sentidos	há 2 dias	4:40
Hey Joe - Live	Roy Buchanan	Live From Austin, TX	há 2 dias	7:16
Despreconceituosamente	Mateus Aleluia	Cinco Sentidos	há 2 dias	3:13
Noite Severina	Elba Ramalho	¿Qual O Assunto Que Mais Lhe Interessa?	há 2 dias	3:53
Something - Midnight Show	Elvis Presley	That's the Way It Is (Expanded Edition)	há 2 dias	3:46
Ta-Hí (Ta Hi) (Pra Você Gostar De Mim)	Marina De La Riva, Davi Moraes	Marina de La Riva	há 2 dias	4:24
Como Te Voy a Olvidar - Edit	Karamelo Santo	Como Te Voy a Olvidar	há 2 dias	4:14
El Tiempo No Para - Live In Buenos Aires / 2001	Bersuit Vergarabat	De La Cabeza Con Bersuit Vergarabat (Live ...	há 2 dias	5:22
Mood for a Day - 2003 Remaster	Yes	Fragile (Deluxe Edition)	há 5 dias	3:01
Capitu	Zélia Duncan	Eu Me Transformo Em Outras	há 5 dias	4:54
Cry Me A River	Ella Fitzgerald	Clap Hands, Here Comes Charlie! (Expande...	há 5 dias	4:17
Elephant Gun	Beirut	Elephant Gun	há 5 dias	5:49
Somersault	I Got You On Tape	2	há 5 dias	4:21
Krigsgaldr	Heilung	Lifa	há 5 dias	9:48
Bag Lady	Erykah Badu, Roy Ayres	Mama's Gun	há 5 dias	5:49
Os Argonautas - Live	Caetano Veloso, Chico Buarque	Caetano E Chico Juntos E Ao Vivo	há 5 dias	3:24

Agora eu vou falar uma fanfarronice do CESPE! Na última prova do TCE/RJ (2021), ele cobrou um padrão ou esquema de metadados chamado Dublin Core. **Ele visa descrever objetos digitais, tais como, vídeos, sons, imagens, textos e sites na web.** *Para quê, professor?* O intuito é promover a adoção de padrões de interoperabilidade de metadados e desenvolver vocabulários especializados para descrever fontes e recursos da web.

Dessa forma, sistemas de busca e recuperação de informações podem fazer buscas mais rápidas e flexíveis. **Dentre os metadados, temos: título, criador, assunto, descrição, produtor, colaborador, data, tipo, formato, identificador, fonte, idioma, relação, cobertura e direitos.** Eu não acho bacana cobrar padrões específicos sem deixar explícito no edital, porque – assim como esse – existem diversos outros. No entanto, caiu em prova, então temos que falar em aula...

RESUMO

METADADOS DE ARQUIVOS

Metadados são dados que descrevem, identificam, explicam, localizam e, portanto, facilitam a recuperação, uso e gestão de recursos de informação – uma definição mais simples é: metadados são dados sobre dados.

METADADOS	DESCRIÇÃO
TÉCNICOS	Descrição dos dados necessários para as diversas ferramentas que precisem armazenar, manipular ou movimentar dados (Ex: SGBD, Ferramentas CASE, Ferramentas OLAP, etc).
DE NEGÓCIO	Descrição de dados necessários pelos usuários de negócios, para entender o contexto do negócio e o significado dos dados.

METADADOS	DESCRIÇÃO
SINTÁTICOS	Dados que descrevem a sintaxe dos dados.
ESTRUTURAIS	Dados que descrevem a estrutura/organização dos dados.
SEMÂNTICOS	Dados que descrevem o significado/conteúdo dos dados em um domínio específico.

METADADOS	DESCRIÇÃO
DESCRITIVOS	Estão relacionados à descrição de um objeto para torná-lo mais fácil de ser descoberto e identificado; podem incluir elementos, tais como título, autor, resumo, palavras-chave e identificador – dependem dos metadados estruturais para serem construídos;
ESTRUTURAIS	Estão relacionados à documentação de como um recurso é composto ou organizado – facilitando a navegação e apresentação de recursos. Um livro digitalizado, por exemplo, pode ser publicado como arquivos de imagens individuais das páginas, arquivos PDF ou HTML.
ADMINISTRATIVOS	Estão relacionados ao fornecimento de informações que apoiam os processos de gerenciamento de recursos informacionais. Incluem, por exemplo, informações sobre como e quando o recurso foi criado e a razão da sua criação.

DADOS ESTRUTURADOS E NÃO-ESTRUTURADOS

Conceitos Básicos

INCIDÊNCIA EM PROVA: MÉDIA

Dados são elementos brutos, sem significado, sobre eventos, no entanto nem todo dado é organizado da mesma maneira. **Em outras palavras, os dados gerados por um aplicativo de rede social (Instagram, Twitter, Facebook, etc) é completamente diferente dos dados gerados por um sistema de estoque de mercadorias de um supermercado.** *Como é, Diego?* É isso mesmo, alguns dados são estruturados, mas a maioria é não-estruturado.

E qual é a diferença entre dados estruturados e não-estruturados? **Basicamente, dados estruturados são rigidamente organizados e formatados de uma maneira que seja facilmente pesquisável em bases de dados relacionais.** *Professor, você falou grego agora!* Calma, tudo isso será detalhado futuramente, mas – simplificando – bases de dados relacionais nada mais são que tabelas (linhas e colunas) que armazenam dados.

Logo, se eu consigo organizar um conjunto de dados em um formato que possa ser exibido em uma tabela, então esse é um conjunto de dados estruturados. *Certinho?* **Já os dados não-estruturados não possuem um formato ou uma organização predefinida, tornando muito mais difícil sua coleta, processamento e análise.** Sim, a maneira como os dados são coletados, processados e analisados depende completamente do formato em que eles estão.

Dados Estruturados

INCIDÊNCIA EM PROVA: ALTÍSSIMA

Dados estruturados são frequentemente categorizados como dados quantitativos. No seu trabalho, provavelmente você está acostumado a trabalhar com eles. Tente se lembrar! *Os dados que você trabalha podem ser organizados em linhas e colunas de bases de dados relacionais?* Se sim, você trabalha com dados estruturados! Exemplos: nomes, datas, endereços, números de cartão de crédito, tabelas do Excel, metadados (que são dados sobre dados), entre outros.

Galera, o bacana dos dados estruturados é que eles podem ser facilmente compreendidos e manipulados por linguagens de máquinas – pode-se manipular dados estruturados com relativa rapidez e essa é uma de suas maiores vantagens: **você pode ordenar, agregar, separar, juntar, inserir, atualizar, deletar e consultar dados e tabelas com facilidade.** A tabela que vimos no exemplo do tópico anterior é um exemplo de dado estruturado.

NOME	SEXO	IDADE	SALÁRIO	NATURALIDADE
Alice	Feminino	18	R$ 8.500,00	DF
Bernardo	Masculino	26	R$ 2.000,00	GO
Carolina	Feminino	30	R$ 9.500,00	TO
Davi	Masculino	22	R$ 1.500,00	RR
Eduarda	Feminino	32	R$ 7.500,00	SP
Felipe	Masculino	24	R$ 3.000,00	RJ
Gabriela	Feminino	28	R$ 6.500,00	PA
Hugo	Masculino	20	R$ 1.000,00	RS

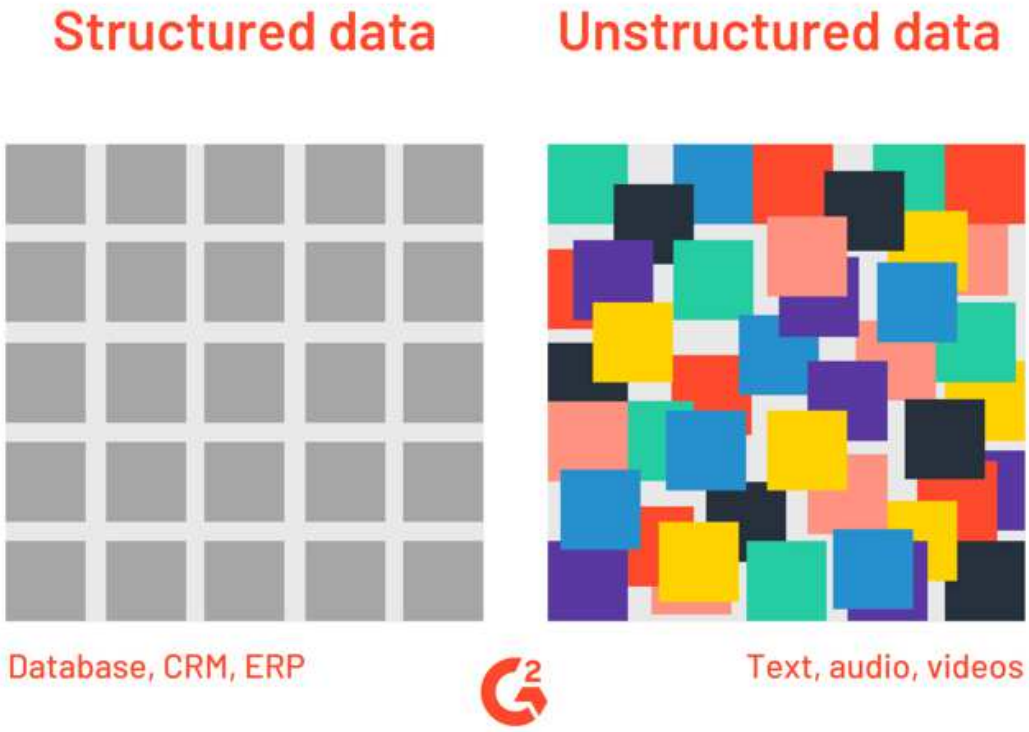
Note que, para cada linha dessa tabela, nós temos sempre as mesmas cinco colunas, com os mesmos atributos e os atributos são sempre do mesmo tipo para cada coluna (Ex: Salário é sempre um Número; Sexo é sempre uma Palavra; Naturalidade são sempre duas letras; entre outras). *Viram como é rígido?* E tem mais: eu não consigo inserir um novo atributo na tabela sem antes modificá-la (Ex: CPF) – caso eu queira inserir essa informação, devo adicionar uma coluna!

Dados Não-Estruturados

INCIDÊNCIA EM PROVA: ALTÍSSIMA

Dados não-estruturados são dados que não possuem um formato ou organização predefinida. Assim como os dados estruturados são frequentemente categorizados como dados quantitativos, os dados não-estruturados são frequentemente categorizados como qualitativos, e não podem ser processados e analisados utilizando ferramentas e métodos convencionais. Exemplos: texto, vídeo, áudio, atividades de mídias sociais, entre outros.

Dados não-estruturados são difíceis de desconstruir porque não têm um formato ou modelo predefinido, significando que não podem ser organizados em uma base de dados relacional. Mais de oitenta por cento de todos os dados gerados atualmente são considerados não-estruturados, e esse número só tende a continuar crescente com o surgimento da Internet das Coisas – tecnologia que permite que vários objetos estejam conectados à internet.



Dados Semiestruturados

INCIDÊNCIA EM PROVA: BAIXA

Dados semiestruturados é uma mistura dos dados estruturados com os dados não-estruturados. Eles não estão de acordo com a estrutura formal dos modelos de dados associados com bancos de dados relacionais ou outras formas de tabelas de dados, mas contêm tags ou outros marcadores para separar elementos semânticos e impor hierarquias de registros e campos dentro dos dados. Temos os dados junto com um esquema de representação parcialmente organizados!

Pensem em um dado que não pode ser perfeitamente encaixado em uma base de dados relacional, mas ele também não é completamente sem formato. Um exemplo clássico é o XML, que é uma linguagem que permite autodescrever um dado! Em outras palavras, ela apresenta o dado e, junto com ele, apresenta uma autodescrição. Eu sei que está difícil de imaginar, mas eu vou dar um exemplo...

```
<xml version="1.0" encoding="UTF-8" standalone="yes" ?>

<MUSICAS xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <MUSICA>
        <NOME> THE SCRIPT FOR MY REQUIEM </NOME>
        <BANDA> BLIND GUARDIAN </BANDA>
        <ALBUM> IMAGINATIONS FROM THE OTHER SIDE </ALBUM>
    </MUSICA>
    <MUSICA>
        <NOME> STARLESS </NOME>
        <BANDA> KING CRIMSON </BANDA>
        <ALBUM> MELTDOWN </ALBUM>
    </MUSICA>
    <MUSICA>
        <NOME> CALADVWCH </NOME>
        <BANDA> DARK AVENGER </BANDA>
        <ALBUM> ALIVE IN THE DARK </ALBUM>
    </MUSICA>
</MUSICAS>
```

Esse é um exemplo de dado escrito em linguagem XML! **Notem que os dados em si são os que estão com a cor preta – todo o resto são dados que auxiliam a descrever os dados.** Apesar de, nesse exemplo, termos estruturas idênticas para cada música/banda/álbum, nada impede que haja estruturas diferentes para cada um. Por essa razão, não é possível dizer que se trata de dados completamente estruturados ou dados completamente não-estruturados. *Então, o que ele é?*

XML descreve dados semiestruturados, porque veja que é possível identificar marcações em seu código, portanto você poderia dizer que são dados estruturados. No entanto, essas tags não

possuem uma ordem ou padrão específico. Você pode organizá-las da maneira que achar adequado. **Se meus dados possuem marcações identificáveis, mas sem ordem e padrão fixo, trata-se de dados semiestruturados – assim como HTML e outras linguagens de marcação!**

Por fim, é possível classificar os tipos de dados quanto à capacidade de terem suas estruturas descritas implicitamente ou explicitamente. Vejamos:

(1) Dados Estruturados: possuem uma estrutura predefinida, logo ela pode ser descrita de forma explícita – você bate o olho e identifica um padrão definido;

(2) Dados Não-Estruturados: não possuem uma estrutura predefinida, logo ela não pode ser descrita (implícita ou explicitamente) – você bate o olho e não identifica um padrão definido;

(3) Dados Semiestruturados: possuem parte de sua estrutura predefinida, logo alguns dados podem ser descritos implicitamente – você bate o olho e consegue inferir um padrão nos dados.

(TCE/PB – 2018) Com relação a dados estruturados e não estruturados, assinale a opção correta.

a) Dados não estruturados podem ser caracterizados por arquivos de diversos tipos — textos, imagens, vídeos, entre outros —, cujas estruturas não são descritas implicitamente.

b) Por padrão, documentos do tipo XML (eXtensible Markup Language) são estruturados.

c) Dados não estruturados de um mesmo grupo possuem as mesmas descrições e, consequentemente, os mesmos atributos.

d) Por padrão, dados não estruturados são organizados em blocos semânticos.

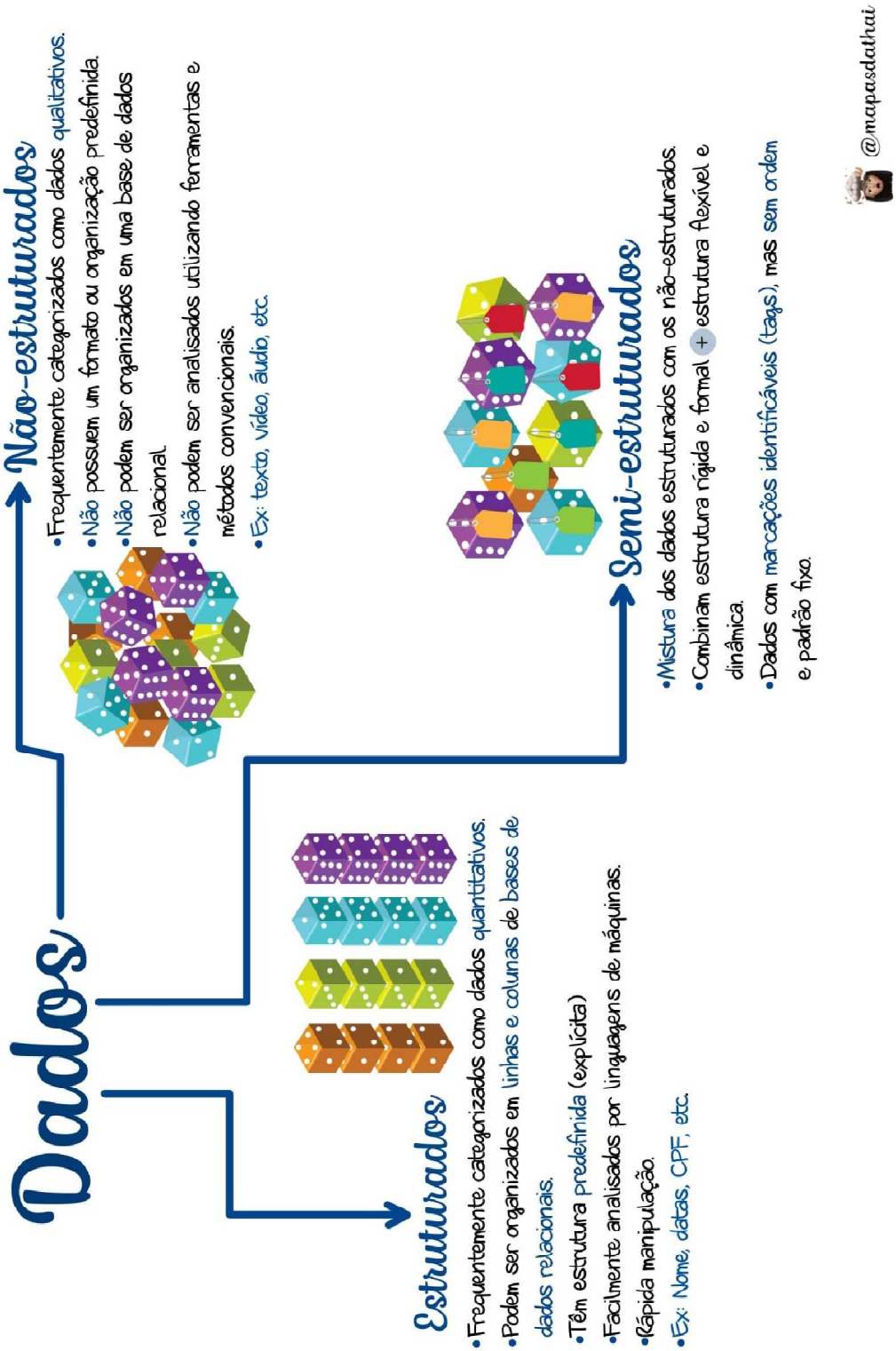
e) A alta heterogeneidade facilita as consultas aos dados não estruturados, desde que estes estejam ligados por ponteiros.

Comentários: (a) Correto, as estruturas não são descritas implicitamente porque sequer há estrutura; (b) Errado, são semiestruturados; (c) Errado, são os dados estruturados de um mesmo grupo que possuem as mesmas descrições e atributos; (d) Errado, dados estruturados são organizados em blocos semânticos iguais; (e) Errado, ela dificulta as consultas aos dados não-estruturados (Letra A).

RESUMO

DADOS	REPRESENTAÇÃO	DESCRIÇÃO
ESTRUTURADOS		Trata-se dos dados que são organizados e representados por meio de uma estrutura rígida e formal.
NÃO-ESTRUTURADOS		Trata-se dos dados que não possuem uma organização rígida e formal, isto é, não seguem uma estrutura padronizada, sendo totalmente flexível e dinâmica.
SEMIESTRUTURADOS		Trata-se dos dados que combinam uma estrutura rígida e formal com uma estrutura flexível e dinâmica.

MAPA MENTAL


Dados
Não-estruturados
•Frequentemente categorizados como dados qualitativos.
•Não possuem um formato ou organização predefinida.
•Não podem ser organizados em uma base de dados relacional.
•Não podem ser analisados utilizando ferramentas e métodos convencionais.
•Ex: texto, vídeo, áudio, etc.
Semi-estruturados
•Mistura dos dados estruturados com os não-estruturados.
•Combinam estrutura rígida e formal + estrutura flexível e dinâmica.
•Dados com marcações identificáveis (tags), mas sem ordem e padrão fixo.
Estruturados
•Frequentemente categorizados como dados quantitativos.
•Podem ser organizados em linhas e colunas de bases de dados relacionais.
•Têm estrutura predefinida (explícita)
•Facilmente analisados por linguagens de máquinas.
•Rápida manipulação.
•Ex: Nome, datas, CPF, etc.
@mapasdathai

Ciclo de Vida de Dados

Conceitos Básicos

INCIDÊNCIA EM PROVA: BAIXA

O papo agora é sobre o ciclo de vida dos dados! **Galera, o ciclo de vida dos dados é basicamente um conjunto de etapas ou fases pelas quais um dado passa desde sua criação até o seu descarte.** *E por que é importante saber isso?* Porque é interessante que saibamos evidenciar os diferentes momentos e fatores envolvidos em cada etapa desse processo para entender mais sobre os dados. Dito isso, nós temos um problema...

Existem dezenas e dezenas e dezenas de ciclos de vida diferentes! *Como assim, Diego?* Ora, isso depende basicamente de perspectiva. Se eu te perguntar quais são as fases pelas quais um dado passa desde a sua criação até o seu descarte, você pode dividir em três, cinco ou dez fases – é uma resposta muito pessoal. **Logo, cada autor tem o seu conjunto de fases de acordo com sua visão particular.** No entanto, há um ciclo muito comum em provas...

Principais Etapas

INCIDÊNCIA EM PROVA: BAIXA

Etapa de Coleta

Em um primeiro momento, é preciso obter os dados que podem ser utilizados para atender uma necessidade específica ou uma demanda prevista de informações sobre um determinado contexto – essa etapa é conhecida como coleta de dados. Nesse contexto, definem-se os dados que serão utilizados para algum propósito, identifica-se uma estrutura compatível, e finalmente os dados são efetivamente coletados.

Entre outras questões desta fase, pode-se destacar: *Qual é o escopo da necessidade da informação? Que tipo de resultado se espera? Com quais características? Quais são os dados necessários? Onde estão as fontes para estes dados? Como os dados podem ser coletados? Em que formato estão? Quais são os tratamentos necessários para que fiquem adequados ao que se precisa? A coleta destes dados não proporciona risco de privacidade para os indivíduos ou entidades referenciadas por eles?*

Elementos, que em alguns casos poderiam ser considerados como secundários, que permitam a integração entre os diversos dados coletados estão sendo obtidos? Como avaliar sua integridade física e lógica, além de outros elementos que garantam sua qualidade? Como identificar sua procedência? Têm-se o direito ou permissão de coletar estes dados? Estão sendo coletados dados que permitam que estes venham a ser identificáveis e recuperáveis em um momento futuro? Entre outras...

Etapa de Tratamento

Dados podem ser coletados a partir de diversas fontes diferentes. Há dados advindos de sistemas e aplicações, dados advindos de planilhas, dados advindos de arquivos de texto, dados advindos de arquivos de apresentação, dados advindos de ferramentas colaborativas, dados advindos de pastas compartilhadas, dados advindos da intranet, dados advindos de portais corporativos, entre outros. Logo, há uma infinidade de fontes de dados diferentes...

Agora vocês já imaginaram quão heterogêneos são esses dados? Eles podem conter tipos ou formatos diferentes, podem conter erros, podem não estar padronizados, podem conter inconsistências, podem conter problemas de qualidade, etc. **Logo, a etapa de tratamento é a responsável justamente por transformar os dados de acordo com as regras de negócio, para que possam ser analisadas pelos usuários responsáveis.**

A etapa de tratamento buscará tratar, transformar, processar, limpar, padronizar e sanitizar os dados obtidos de diferentes fontes de acordo com o padrão de qualidade predefinido. *Diego, ainda não entendi!* Então, deixa eu contar uma história para vocês: há muito tempo atrás, antes da era do streaming, nós tínhamos os arquivos de música em formato .mp3. *Quem aí é dessa época?* Pois é, você carregava o seu player de .mp3 (Ex: iPod) por todo canto.

```
#1 Sepultura - Isolation.mp3
#2 Live Jazz In Rio - O Sapo - João Donato.mp3
#3 Fado Tropical - Chico Buarq.mp3
#4 Coming Home - Leon Bridges - River.mp3
#5 Prowler.mp3
```

Eu fazia o download de várias músicas de diversas fontes diferentes, mas cada uma chegava com o nome em um formato completamente oposto da anterior. Aquilo foi me dando tanta agonia que um dia eu decidi que não podia mais suportar: eu tinha que padronizar o nome das músicas urgentemente. *O que eu fiz?* Eu criei um padrão que todas as músicas deveriam seguir. Esse padrão seguia uma ordem bastante simples:

```
#2 NomeDoAutor - NomeDoAlbum - NomeDaMúsica.mp3
```

A partir desse padrão de qualidade preestabelecido, eu tratei os meus dados (limpei, sanitizei, transformei...) até chegar ao seguinte ponto:

```
#1 Sepultura - Quadra - Isolation.mp3
#2 João Donato - Live Jazz In Rio - O Sapo.mp3
#3 Chico Buarque - Calabar - Fado Tropical.mp3
```

```
#4 Leon Bridges - Coming Home - River.mp3
#5 Iron Maiden - Iron Maiden - Prowler.mp3
```

E só após terminar de padronizar a última música, eu pude colocar a cabeça no travesseiro com o coração leve de quem havia organizado tudo. Ufaaa...

Etapa de Armazenamento

Essa etapa busca processar, transformar, inserir, modificar, carregar, migrar, transmitir ou qualquer ação semelhante que vise persistir os dados em um dispositivo. Persistir – no jargão de tecnologia da informação – é o mesmo que armazenar em algum dispositivo de memória não-volátil. Ora, nós já coletamos os dados e fizemos todo tipo de tratamento... agora é o momento de armazená-lo em alguma base de dados.

Dentre outras questões fundamentais desta fase pode-se destacar: *Quais são os dados disponíveis? Quais destes dados serão armazenados? Qual estrutura (física e lógica) será utilizada para seu armazenamento? Como garantir a permanência dos dados complementares sobre a coleta para que se tenha garantido o contexto de sua obtenção? Estes dados podem representar um risco à privacidade dos indivíduos ou instituições neles referenciados de alguma forma?*

Como as partes de sua estrutura lógica serão interligadas e como serão mantidas as interligações com outros conjuntos de dados? Como garantir que os elementos que sustentam a sua qualidade sejam mantidos? Tem-se o direito de armazenar estes dados? Todos os aspectos que podem contribuir para sua "encontrabilidade" estão sendo armazenados? Todos os fatores para sua utilização ao longo do tempo estão sendo mantidos?

Etapa de Integração

A etapa de integração, na verdade, ocorre desde a etapa de coleta. Ora, nós coletamos dados de diferentes fontes, tratamos e armazenamos em um banco de dados. No entanto, um banco de dados possui uma característica importante: ele é basicamente uma coleção coerente de dados inter-relacionados. Em outras palavras, um banco de dados não é um conjunto de aleatório de dados livres, leves e soltos.

Dito isso, a etapa de integração busca integrar dados de diferentes fontes em uma base de dados unificada. Ao escolher os dados que serão coletados, ao transformá-los de acordo com um padrão de qualidade, ao definir como serão armazenados – tudo isso faz parte da etapa de integração porque todas essas atividades auxiliam, de alguma forma, a alcançar uma base de dados que podem se relacionar de diversas maneiras a fim de auxiliar o usuário.

Etapa de Recuperação

Ora, de que adianta coletar, tratar, armazenar e integrar dados em uma bela base dados se eu não conseguir recuperar esses dados? Nesta fase, preocupa-se com meios que ampliem os níveis de

utilização destes dados, seja por ampliação das possibilidades de acesso via cópia ou obtenção de conjuntos para análise seja por meio da disponibilização de recursos de visualização destes dados (Ex: dashboard, painéis, portais, gráficos, cockpits virtuais, etc).

Essa fase busca encontrar, acessar, consultar e interpretar dados armazenados pelos usuários. **Uma vez que os dados tenham sido coletados, tenham sido efetivamente tratados e já estejam armazenados em um dispositivo de memória não-volátil, pode-se proporcionar uma nova fase (recuperação) que permita tornar estes dados disponíveis para acesso e uso.** Agora vamos para a última etapa...

Etapa de Descarte

Essa etapa busca basicamente descartar os dados armazenados. **Uma vez concluídas as reflexões sobre as fases de coleta, tratamento, armazenamento, integração e recuperação, poderia se supor que o ciclo de vida dos dados estaria completo, principalmente em um momento em que o limite para o volume de dados parece cada vez mais alto.** No entanto, não é o que ocorre – os dados ainda podem ser descartados.

Dentre outras questões fundamentais desta fase, pode-se destacar: *Quais dados já não são mais necessários? Os dados a serem descartados foram persistidos? Em quais suportes? Estes dados estão replicados em outras bases? Como garantir e explicitar que estes dados foram realmente excluídos e não simplesmente ocultos? A eliminação destes dados não prejudicará a integridade ou interligação de outros dados?*

O descarte destes dados não prejudicará a qualidade do conjunto de dados como um todo? Tem-se o direito de excluir este dado? Ao eliminar estes dados qual o impacto em sua encontrabilidade e acesso? Para o descarte foi considerada a necessidade de preservação em seus diversos aspectos? Bem, eu acho que é isso sobre esse tema! Apesar de toda essa teoria, eu nunca encontrei uma única questão que cobre exatamente essas etapas, mas se veio no edital, eventualmente pode cair...

Processo de ETL

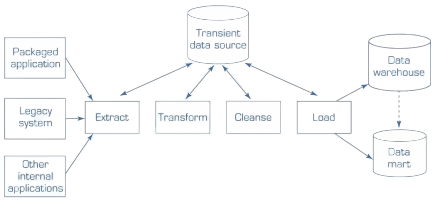
Antes de entrar nesse tópico, é preciso explicar alguns conceitos. **Comecemos com Business Intelligence (BI): trata-se de um conjunto de técnicas, métodos e ferramentas de análise de dados para subsidiar processos de decisão de uma empresa.** Nesse contexto, utiliza-se frequentemente um armazém de dados chamado Data Warehouse. *O que é isso, Diego?* É basicamente um banco de dados gigantesco com algumas características peculiares...

Entre essas características, temos que ele é orientado por assunto, integrado, histórico e não-volátil. Em suma: ele é orientado por assunto, porque ele é projetado a partir de assuntos diferentes – como marketing, financeiro, recursos humanos, etc; ele é não-volátil, porque seus dados não são modificados após inseridos; ele é integrado, porque combina dados de diferentes fontes; e ele é histórico, porque seus dados são mantidos na base sem alterações e só variam com o tempo.

Dito isso, já podemos tratar de ETL...

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, Diego?* Bem, uma das características essenciais de um Data Warehouse é sua integridade.

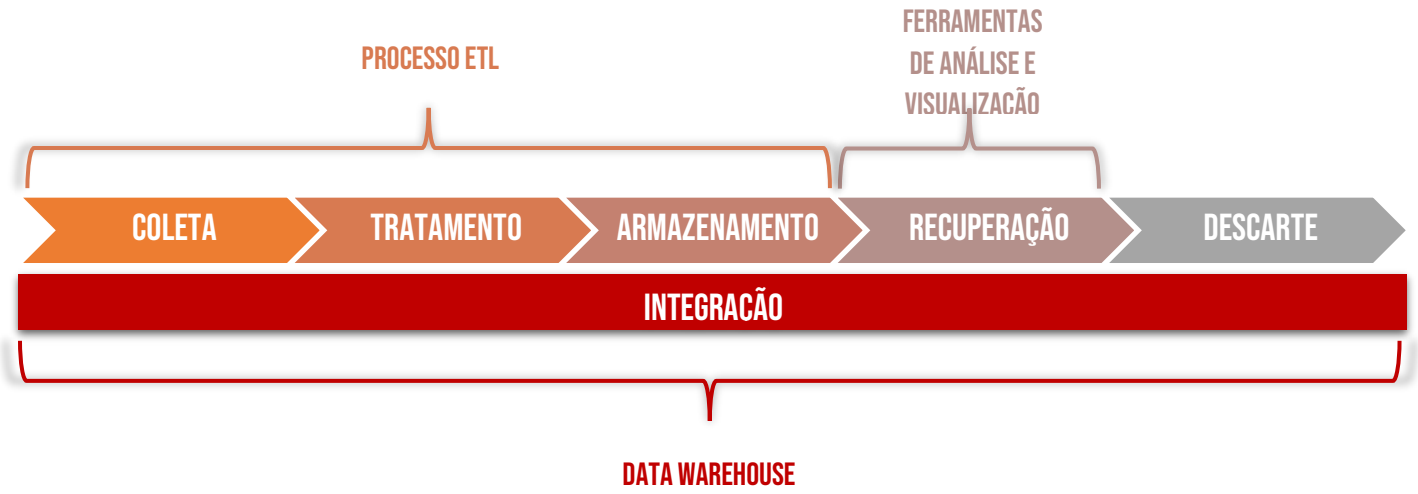
Logo, os dados devem estar padronizados, consistentes e uniformizados antes de serem carregados no Data Warehouse! Para tal, eles precisam ser limpos, entre outras atividades. **Isso tudo ocorre na Stage Area, portanto ela é basicamente uma localização temporária para onde os dados de diversas fontes 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. **Seguindo: 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. *Professor, eu entendi o processo, mas como podemos relacionar isso com as etapas do ciclo de vida de dados?*

Ora, notem que o Processo de ETL é basicamente uma representação da coleta, tratamento e armazenamento. Já a integração é realizada pelo Data Warehouse ao combinar de forma

relacionada dados de diferentes fontes. Por fim, a recuperação de dados pode ser executada por meio de relatórios, dashboards, cockpits virtuais, gráficos, portais, ferramentas de mineração de dados, entre outros.



Veremos mais detalhes sobre Business Intelligence, Data Warehouse e Processo de ETL em outras aulas, essa parte foi apenas uma pequena introdução sobre esses assuntos...

QUESTÕES COMENTADAS – DIVERSAS BANCAS

1. **(CESGRANRIO / BB – 2023)** No processo de aprendizagem de máquina, busca-se identificar padrões compreensíveis em bases de dados. Desse modo, supõe-se a existência de uma hierarquia de três níveis: a base, com os itens elementares, captados e armazenados por recursos de Tecnologia da Informação; o nível intermediário, no qual esses itens são processados, com significados e contextos bem definidos; e o nível mais alto, contendo os padrões ou os conjuntos cuja formulação pode envolver ou relacionar os níveis inferiores. Os três níveis dessa hierarquia, listados, respectivamente, da base para o nível mais alto, são:

   a) conhecimento, informação e dado
   b) dado, conhecimento e informação
   c) dado, informação e conhecimento
   d) informação, conhecimento e dado
   e) informação, dado e conhecimento.

Comentários:

A hierarquia correta começa com dados, que são processados para gerar informações, e, em seguida, a informação é utilizada para adquirir conhecimento.

Gabarito: Letra C

---

2. **(CESPE / SEPLAN-RR - 2023)** Dados de redes sociais são exemplos de dados não estruturados.

Comentários:

Perfeito! Os dados das redes sociais, como posts, comentários, fotos, vídeos e mensagens, são exemplos típicos de dados não estruturados. Esses dados são gerados pelos usuários de forma livre e podem variar muito em termos de formato, conteúdo e contexto.

Gabarito: Correto

---

3. **(CESPE / TCE-SC - 2022)** A gestão do conhecimento está relacionada à ação, à atitude e a uma intenção específica dos seres humanos.

Comentários:

Perfeito! Segundo Turban, conhecimento é a informação aplicada (em ação). O conhecimento vai além de informações, pois ele – além de ter um significado – tem uma aplicação. As informações são valiosas, mas o conhecimento constitui um saber. Produz ideias e experiências que as informações por si só não serão capazes de mostrar.

Gabarito: Correto

---

4. **(CESPE / TCE-SC - 2022)** Informação pode ser definida como um conjunto de fatos diferentes relativos a uma transação e que é, geralmente, desprovida de significado; ela é considerada a matéria-prima dos sistemas de informação.

Comentários:

Enquanto os dados podem ser simples fatos ou observações, a informação resulta da interpretação e contextualização desses dados, tornando-se relevante para tomada de decisões e compreensão de situações. Logo, a informação é o dado acrescido de significado, contexto, interpretação, entre outros.

Gabarito: Errado

---

5. **(CESPE / TCE-SC - 2022)** O conhecimento serve para dar suporte ao processamento decisório; ele representa a informação tratada, confiável e íntegra.

Comentários:

O conhecimento desempenha um papel crucial no processo decisório, pois representa a informação que foi analisada, contextualizada e interpretada, tornando-se um recurso valioso para a tomada de decisões informadas e eficazes. O conhecimento, ao agregar significado à informação, contribui diretamente para uma tomada de decisão mais informada e embasada.

Gabarito: Correto

---

6. **(FEPESE / ISS-Criciúma – 2022)** Analise as afirmativas abaixo com relação a dado, informação e conhecimento no domínio subjetivo.

1. Dados são estímulos sensoriais, que nós percebemos através dos nossos sentidos.
2. Informação é, mais do que ser somente um estágio intermediário entre dado e conhecimento, um conhecimento empírico em si.
3. Conhecimento é o significado desses estímulos sensoriais que nós percebemos através de nossos sentidos.

a) É correta apenas a afirmativa 2.
b) São corretas apenas as afirmativas 1 e 2.
c) São corretas apenas as afirmativas 1 e 3.
d) São corretas apenas as afirmativas 2 e 3.
e) São corretas as afirmativas 1, 2 e 3.

Comentários:

O assunto de Teoria da Informação (Dado, Informação, Conhecimento e Inteligência) é abordado em diversas áreas do conhecimento: Tecnologia da Informação, Comunicação, Filosofia, Arquivologia, entre outros. Por conta disso, autores diferentes oferecem definições de acordo com a sua área (Ex: um autor de livros de arquivologia dará uma definição mais voltada à sua área, assim como ocorrem com outras áreas). *E qual é o problema disso?*

O problema é que examinadores eventualmente trazem definições de outras áreas para provas da área de tecnologia da informação. A prova trouxe definições de Chaim Zins, que é um autor de Ciências da Informação (área mais próxima da filosofia do que da computação). De acordo com ele, "Dados" são os estímulos sensoriais que percebemos por meio de nossos sentidos. "Informação" é o significado destes estímulos sensoriais (por exemplo, a percepção empírica).

Em outras palavras, um ruído que ouçamos é considerado um dado; o significado desse ruído (por exemplo, um motor ligado do carro) é considerado informação. Ainda de acordo com ele, "Informação", no domínio subjetivo, é conhecimento empírico. Sendo assim, no exemplo mencionado, o conhecimento que o motor está em movimento é informação, pois ele está empiricamente baseado.

Ainda: "Conhecimento", no domínio subjetivo, como observado, é um pensamento na mente do indivíduo, caracterizado pela crença justificável do indivíduo de que ele é verdade. Ele pode ser empírico e não empírico, como no caso do conhecimento lógico e matemático (por exemplo, "todo triangulo tem três lados"), conhecimento religioso (por exemplo, "Deus existe"), conhecimento filosófico (por exemplo, "cogito, ergo sum"), e semelhantes.

Viram como é uma definição bastante filosófica dos conceitos? Pois é! Na minha opinião, examinadores não deveriam trazer esse tipo de visão para provas de tecnologia da informação, uma vez que destoam completamente do que estamos acostumados a ver. Enfim... infelizmente não temos o que fazer. Quanto à questão, (1) e (2) estão corretas de acordo com o autor e (3) está errado porque o autor afirma que isso seria a definição de Informação, e não Conhecimento.

Gabarito: Letra B

---

7. **(CESPE / Petrobrás - 2022)** A inteligência é a informação que é expressa na forma de números, letras ou símbolos, os quais são os insumos dos gráficos.

Comentários:

Na verdade, esse é o conceito de um dado bruto. Os dados são coletados e servem de insumos para os gráficos.

Gabarito: Errado

---

8. **(CESPE / PCDF – 2021)** A eficiência na representação de dados relaciona-se à codificação de fonte, cujo objetivo é remover redundâncias no conteúdo da informação e, com isso, reduzir o tamanho da mensagem.

Comentários:

Nada melhor do que a justificativa da própria banca: *"Eficiência de representação: codificação de fonte. Em artigos, discute-se a codificação de fonte, que trata da representação eficiente de dados. Hoje, o termo é sinônimo de compactação de dados. O objetivo básico da codificação de fonte é remover a redundância nas informações para tornar a mensagem menor. Em sua exposição, ele discute um método sem perdas de compactar dados na fonte, usando um código de bloco de taxa variável, mais tarde chamado de código de Shannon-Fano".*

Codificar a fonte de dados busca remover redundâncias, reduzir o tamanho da mensagem e compactar os dados a fim de melhorar a eficiência na representação dos dados.

Gabarito: Correto

---

9. **(CESPE / PF - 2021)** Os dados estruturados diferenciam-se dos dados não estruturados pela rigidez em seu formato e pelo fato de poderem ser armazenados em campos de tabelas de um banco de dados relacional.

Comentários:

Perfeito! Esta questão trata da distinção entre dados estruturados e não estruturados. Os dados estruturados são aqueles que possuem um formato definido e organizado, facilitando seu armazenamento, acesso e análise. Eles são comumente armazenados em bancos de dados relacionais, onde cada elemento de dado está organizado em tabelas, linhas e colunas, permitindo que sejam facilmente acessados e manipulados por meio de consultas estruturadas, como SQL.

Por outro lado, dados não estruturados não possuem um formato ou modelo predefinido, o que os torna mais desafiadores para serem gerenciados e analisados usando métodos tradicionais. Exemplos de dados não estruturados incluem textos, imagens, vídeos, e-mails e postagens em redes sociais. Eles geralmente requerem técnicas mais complexas de processamento e armazenamento, como o uso de sistemas de arquivos ou bancos de dados NoSQL.

Logo, a afirmação da questão é correta ao destacar a rigidez no formato e a capacidade de armazenamento em bancos de dados relacionais como características distintivas dos dados estruturados em comparação com os não estruturados.

Gabarito: Correto

---

10. **(CESPE / PCDF – 2021)**. Embora não haja conceitos únicos para dados e informação, é possível afirmar que os dados são matéria-prima para a informação.

Comentários:

Nada melhor do que a justificativa da própria banca: *"Há muitos conceitos diferentes para dados, informação e conhecimento. Entretanto, é comumente aceita a ideia de que os dados são matéria prima para conhecimento e informação. "Fatos desorganizados e não processados, números brutos, figuras, imagens, palavras, sons, derivados de observações ou medições. Normalmente, os dados são de natureza estática, um conjunto de fatos discretos e objetivos sobre eventos. Observa que dados não são conhecimento. Os dados são a matéria-prima para a criação do conhecimento. Não há significado inerente aos dados. Dados são matéria-prima para informação. Ao agregar valor, os dados são transformados em informação, que tem significado, relevância e propósito".*

Perfeito! Basta lembrar Pirâmide DIKW: dados realmente são a matéria prima para informação – dados brutos são processados para gerar informação.

Gabarito: Correto

---

11. (CESPE / TCE-RJ – 2021) Um dado que está inserido em um contexto pode ser denominado de informação.

Comentários:

Perfeito! Um dado em um contexto é chamado de informação.

Gabarito: Correto

---

12. (CESPE / TCE-RJ – 2021) A informação pode ser facilmente obtida por meio de máquinas, além de ser transferível e frequentemente quantificável.

Comentários:

Opa... o enunciado descreveu características de dados, e não de informações.

Gabarito: Errado

---

13. (CESPE / TCE-RJ – 2021) O conceito de inteligência está relacionado à capacidade de julgamento, ou seja, o ser humano usa o seu conhecimento para tomar a melhor decisão em uma situação real.

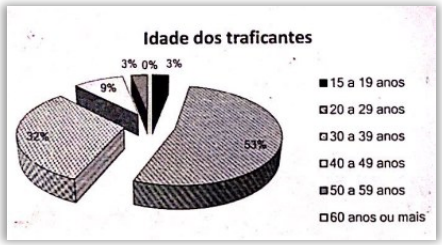
Comentários:

Perfeito! A inteligência realmente está associada à capacidade de julgamento, sendo uma habilidade essencialmente humana para tomar a decisão mais eficiente em uma situação concreta.

Gabarito: Correto

---

14. (CESPE / Polícia Federal – 2021) Considerando-se a classificação dados, informação, conhecimento e inteligência, é correto afirmar que o gráfico representa, por si só, a inteligência.



Comentários:

Opa... o gráfico – por si só – representa informação, e não inteligência!

Gabarito: Errado

---

15. (CESPE / Polícia Federal – 2021) O número 1.789 sozinho caracteriza uma informação, independentemente do contexto.

Comentários:

O número 1.789 por si só é um dado, não uma informação. Dados tornam-se informações quando são processados, organizados ou estruturados de maneira a serem úteis e terem significado. Sem contexto, não é possível determinar o que o número 1.789 representa; ele não transmite uma mensagem específica ou significativa. É o contexto que transforma esse dado (1.789) em informação, como por exemplo, indicando que ele é o número de traficantes presos, a quantidade de itens vendidos, ou qualquer outra métrica específica.

Gabarito: Errado

---

16. (CESPE / Polícia Federal – 2021) Considere que a Polícia Federal tenha registrado, em determinado período, a prisão de 1.789 traficantes de drogas pertencentes a facções criminosas, conforme faixas etárias mostradas no gráfico.

Com referência às informações e ao gráfico precedentes, julgue o item subsecutivo.

Considerando-se a classificação dados, informação, conhecimento e inteligência, é correto afirmar que o gráfico representa, por si só, a inteligência.

Comentários:

O gráfico apresenta uma distribuição de frequências das idades dos traficantes presos, o que constitui uma forma de informação. Inteligência, neste contexto, implicaria uma análise mais aprofundada desses dados, proporcionando insights, explicando padrões ou tendências, ou auxiliando na tomada de decisões estratégicas.

Gabarito: Errado

---

17. (CESPE / Polícia Federal – 2018) O atributo de inteligência depende mais da qualidade da informação disponível do que da sua quantidade, tendo, portanto, natureza qualitativa.

Comentários:

A inteligência é um atributo que realmente depende mais da qualidade da informação disponível do que da sua quantidade – ao contrário dos dados, que dependem mais da quantidade do que da qualidade em si.

Gabarito: Correto

---

18. (CESPE / Polícia Federal – 2018) Dados são fatos que descrevem os objetos de informação, por exemplo, eventos e entidades.

Comentários:

Dados realmente podem ser vistos como fatos que descrevem objetos de informação, isto é, eventos, acontecimentos, entidades, podendo se referir a mais de um fato – trata-se de qualquer registro ou indício relacionável a alguma entidade ou evento.

Gabarito: Correto

---

19. (CESPE / Polícia Federal – 2018) Na representação da informação, os atributos permitem que entidades e eventos possam ser reconhecidos, referidos e descritos. Um atributo relacional permite relacionar eventos e entidades.

Comentários:

Atributos são características que permitem que entidades e eventos sejam representados, reconhecidos, referidos e descritos, sendo que um atributo relacional, de fato, permite relacionar eventos e entidades.

Gabarito: Correto

---

20. (FCC / TCE/RS – 2018) Os conceitos de dados, informação e conhecimento são de grande importância no contexto de sistemas de informação.

Sobre eles, é correto afirmar que:

a) não são necessários os dados para que se obtenha o conhecimento.
b) a informação é obtida acrescentando-se significado aos dados.
c) a informação é obtida a partir do conceito de conhecimento.
d) o processo de tomada de decisão em um sistema de informação tem por base apenas os dados brutos.
e) os dados consistem do conhecimento analisado sob diferentes pontos de vista.

Comentários:

(a) Errado, dados e informações são necessários para que se obtenha conhecimento; (b) Correto, a informação realmente é obtida acrescentando-se significado aos dados; (c) Errado, a informação é obtida a partir do conceito de dados; (d) Errado, não são apenas dados brutos, são também a partir de informações; (e) Errado, o conhecimento consiste dos dados e informações analisados sob diferentes pontos de vista.

Gabarito: Letra B

---

21. (CESPE / Polícia Federal – 2018) A informação se caracteriza pela compreensão e internalização do conteúdo recebido, por meio do seu uso em nossas ações; o dado, por sua vez, é um elemento bruto dotado apenas de significado e relevância que visem fornecer uma solução para determinada situação de decisão.

Comentários:

Na verdade, é o conhecimento que se caracteriza pela compreensão e internalização do conteúdo recebido, por meio do uso em nossas ações. Além disso, dados não possuem nenhum significado que visem fornecer uma solução para determinada situação de decisão.

Gabarito: Errado

---

22. (CESPE / Polícia Federal – 2018) Informação é constituída por um conjunto de dados com características específicas. O ponto de análise é que os dados devem ser irrelevantes para o sistema a que se destinam.

Comentários:

Quando dados passam por algum tipo de relacionamento, avaliação, interpretação ou organização, tem-se a geração de informação. Logo, informação é realmente constituída por um conjunto de dados com características específicas, no entanto os dados são importantes e relevantes, visto que eles são o fundamento para a geração de informação que os sistemas utilizam.

Gabarito: Errado

---

23. (CESPE / TCE/SC – 2018) Define-se informação como significado, ou seja, como registros icônicos e simbólicos — fonéticos ou numéricos — e signos — linguísticos, lógicos ou matemáticos —, por meio dos quais se representam atos, conceitos ou instruções.

Comentários:

A questão trata de dados, e não de informação. Registros icônicos, simbólicos (fonéticos ou numéricos) e signos (linguísticos, lógicos ou matemáticos) são simplesmente dados que não possuem significado. Ao serem processados, eles adquirem significado e passam a ser informações.

Gabarito: Errado

---

24. (CESPE / Polícia Federal – 2018) O conceito de conhecimento é mais complexo que o de informação, pois conhecimento pressupõe um processo de compreensão e internalização das informações recebidas, possivelmente combinando-as.

Comentários:

Galera... que questão linda! Conhecimento é realmente mais complexo que informação porque pressupõe que haja um processo de compreensão e internalização das informações recebidas – vejam as palavras-chave: compreensão e internalização.

Gabarito: Correto

---

25. (CESPE / EMAP – 2018) No que se refere à espiral do conhecimento, socialização, externalização, modelagem e feedback são os diferentes modos de conversão do conhecimento.

Comentários:

Modelagem e feedback não são modos de conversão do conhecimento – os modos são socialização, externalização, combinação e internalização.

Gabarito: Errado

---

26. (CESPE / EMAP – 2018) Em uma instituição, o conhecimento explícito está relacionado ao capital intangível.

Comentários:

O conhecimento explícito está relacionado ao capital tangível.

Gabarito: Errado

---

27. (CESPE / EMAP – 2018) Na solução de problemas, o conhecimento tácito é associado ao conhecimento do expert.

Comentários:

O conhecimento tácito está realmente relacionado com o conhecimento do expert – aquele que vem da experiência.

Gabarito: Correto

---

28. (CESPE / TCE/PE – 2017) A informação caracteriza-se por ser frequentemente tácita, bem como por ser de estruturação e captura difíceis em máquinas.

Comentários:

Quem é frequentemente tácito é o conhecimento, e não a informação – que geralmente é explícita.

Gabarito: Errado

---

29. (CESPE / TCE-PA – 2016) O conhecimento tácito é pessoal, específico ao contexto e de difícil formulação e comunicação; o conhecimento explícito, por sua vez, é transmitido em linguagem formal e sistemática, propiciando ao indivíduo a apreensão de fatos.

Comentários:

O conhecimento tácito realmente é mais pessoal e específico de um contexto, sendo difícil de formular, formalizar e comunicar; já o conhecimento explícito é transmitido em linguagem mais formal, sistemática, estruturada e organizada, como livros, manuais, entre outros.

Gabarito: Correto

---

30. (IESES / BAHIAGÁS – 2016) De acordo com Marcos Magalhães e Rafael Sampaio, página 81, "a tecnologia disponível, os sistemas, a miríade de programas e aplicativos hoje existentes alimentam o Sistema de Informações e Inteligência de Marketing (SIM) das organizações de qualquer porte com uma avalanche de informações provenientes de fontes variadas. Se, antes, o desafio era 'obter informação', - o que podia ser sistematizado de modo relativamente fácil –, a questão que se coloca hoje é conseguir organizar e analisar uma imensa quantidade de dados que chegam de todos os lados. Ou seja, o problema agora é 'usar a informação'. Para entender melhor essa questão, é necessário adotar critérios para a compreensão de algumas definições e exemplos". Para isto, é preciso distinguir a diferença entre Dados, Informação e Conhecimento. Desta forma, é correto afirmar:

a) Conhecimento: são sequências de textos, fotos, figuras ou sons que podem ser manipulados e descritos.

b) Informação relaciona-se à prática, à ação. Por exemplo: 'A temperatura no interior do Rio Grande do Sul é de 18º Celsius e, nesta época do ano, deve provocar um aumento da colheita de soja naquela região'.

c) Dados são sequências de símbolos (letras ou números), textos, fotos, figuras ou sons que podem ser descritos, armazenados e manipulados. Por exemplo: '18º Celsius'.

d) Conhecimento são dados contextualizados que também podem ser armazenados e manipulados.

e) Representa um exemplo de 'dados': 'A temperatura média deste mês, no interior do Rio Grande do Sul, é de 18º Celsius'.

Comentários:

(a) Errado, isso é dado; (b) Errado, isso é conhecimento; (c) Correto, mas eu discordo do gabarito! Dados brutos não processados seria apenas 18 – quando a questão dá como exemplo 18º Celsius, isso já é informação. No entanto, essa alternativa é a menos errada; (d) Errado, isso é informação; (e) Errado, isso é informação.

Gabarito: Letra C

---

31. (FCC / CNMP– 2015) Os Sistemas de Informação (SI) são construídos com Dados, Informação, Conhecimento e Inteligência. Sobre o tema, considere:

I. Informação é coletada nos ambientes interno e externo e representa, por exemplo: fatos, textos, gráficos.

II. A inteligência é realizada por meio de síntese, baseada em experiência e intuição, sendo uma habilidade humana.

III. Conhecimento demanda análise e avaliação sobre a confiabilidade, relevância e importância de dados e informações para a construção de um quadro de situação.

Está correto o que consta APENAS em:

a) I e III.
b) II.
c) II e III.
d) I.
e) III.

Comentários:

(I) Errado, fatos são dados, e não informações; (II) Correto, dado é processado para se tornar informação, que é avaliada para se tornar conhecimento, que é sintetizado para se tornar inteligência baseado em experiência, intuição e complexidade; (III) Correto, para que informação se torne conhecimento, é necessário considerar aspectos como confiabilidade, relevância e importância.

Gabarito: Letra C

---

32. (CESPE / TC-DF – 2014) O conhecimento tácito é fruto de aprendizado e experiência de vida e é disseminado de maneira formalizada e declarada por meio de artigos e livros.

Comentários:

Questão polêmica! O conhecimento tácito é realmente fruto de aprendizado e experiência de vida, mas ele não é facilmente disseminado ou reproduzido de maneira formalizada e documentada como artigos e livros – sua reprodução é complexa. A questão não afirma que é fácil de ser disseminado, então caberia recurso! De toda forma, é o conhecimento explícito que geralmente é disseminado de maneira formalizada.

Gabarito: Errado

---

33. **(COPEVE-UFAL / UFAL – 2014)** Qual a denominação de um conjunto de dados devidamente ordenados e organizados de forma a terem significado?

a) Dado
b) Processamento
c) Informação
d) Memória
e) Cluster

Comentários:

Dados ordenados e organizados de forma a terem significado é a definição de informação.

Gabarito: Letra C

---

34. **(CESPE / MPU – 2013)** Considerando-se a espiral do conhecimento, é correto afirmar que situações em que os colaboradores de uma organização convertem o conhecimento explícito em tácito correspondem à etapa denominada combinação.

Comentários:

Quando se passa de conhecimento explícito para tácito é chamado de internalização, e não combinação.

Gabarito: Errado

---

35. **(CESPE / MPU – 2013)** Desejando externalizar o conhecimento tácito que determinado colaborador detenha sobre a elaboração de estudos técnicos, por exemplo, a organização deve documentar esse conhecimento de modo que seja possível a outros colaboradores reproduzi-lo facilmente. Nesse tipo de situação, observa-se a conversão do conhecimento tácito em explícito.

Comentários:

Observem palavras-chave como externalizar, conhecimento tácito, documentar, reproduzir, entre outros. A externalização converte conhecimento tácito em explícito por meio de estudos técnicos documentados e fáceis de reproduzir (Ex: livro).

Gabarito: Correto

---

36. **(CESPE / FUB – 2013)** Conhecimentos estruturados encontram-se materializados em bens materiais tangíveis.

Comentários:

Conhecimento estruturado é também conhecido como conhecimento explícito e, de fato, se materializa em bens materiais tangíveis.

Gabarito: Correto

---

37. **(FUMARC / PC-MG – 2013)** Sobre conceitos e definições de dados e informações no contexto de Sistemas de Informação, analise as seguintes afirmativas:

I. Dados se referem a uma descrição elementar de coisas, eventos, atividades e transações que são registrados, classificados e armazenados, mas não são organizados para transmitir significados específicos.

II. Informação se refere a dados que foram organizados para terem significado e valor para o receptor, que interpreta o significado e elabora conclusões sobre a informação.

III. Conhecimento consiste em informações organizadas para transmitir experiência, aprendizagem e prática aplicadas a um problema ou atividade empresarial.

Estão CORRETAS as afirmativas:

a) I, II e III.
b) I e II, apenas.
c) I e III, apenas.
d) II e III, apenas.

Comentários:

(I) Correto. Dados, ou dados brutos, são as descrições básicas de produtos, clientes, eventos, atividades e transações que são registradas, classificadas, armazenadas. Os dados são a matéria prima a partir da qual as informações são produzidas. Exemplos disso são o número de horas que um funcionário trabalhou em determinada semana ou o número de novos veículos que a Toyota vendeu no primeiro trimestre de 2010; (II) Correto. Informação é um dado que já foi processado, organizado ou colocado em contexto de forma a ter significado para a pessoa que recebe. Por exemplo: o número de vendas trimestrais de novos veículos da Toyota de 2008 até 2010 é uma informação, porque dará a ideia de como os recalls dos veículos de 2009 e 2010 causaram impacto nas vendas; (III) Correto. Conhecimento é um dado e/ou uma informação que foi processado, organizado e colocado em contexto para ter significado, e transmite um conhecimento, experiência, aprendizado acumulado e especialização, já que se aplica a um problema ou atividade atual. Saber como gerir o recall de veículos para minimizar os impactos negativos nas vendas é um exemplo de conhecimento.

Gabarito: Letra A

---

38. (CESPE / ANAC – 2012) Os modos de conversão do conhecimento tácito em explícito referem-se à socialização e à combinação.

Comentários:

A conversão de conhecimento tácito em explícito é chamada de externalização, e não socialização e combinação.

Gabarito: Errado

---

39. (CESPE / MEC – 2011) Considere que, em uma organização, os profissionais capazes de criar maquetes de prédios mostrem a seus aprendizes que, por meio da observação e imitação, eles obterão o conhecimento do assunto. Nessa situação, o processo de construção de conhecimento é realizado do conhecimento tácito para o explícito.

Comentários:

Trata-se de um caso em que profissionais mais experientes estão passando conhecimento para aprendizes sobre um determinado assunto. Nesse caso, temos conhecimento tácito para tácito, logo trata-se de uma socialização.

Gabarito: Errado

---

40. (CESPE / IFB – 2011) A transformação de informação em conhecimento envolve a realização de comparações e contrastes, a identificação de relacionamentos e a inferência de consequências e, portanto, é influenciada por especialização, experiência, valores e insights estruturados dos integrantes da equipe.

Comentários:

Perfeito, perfeito, perfeito! Nós vimos que ter informação é uma coisa e ter conhecimento é outra. Para converter informação em conhecimento, é necessário comparar, contrastar, identificar relações, inferir consequências, refletir, valorar e contextualizar. Para tudo isso, é necessário um conjunto de habilidades de especialização, experiência, valores, insights, know-how, entre outros.

Gabarito: Correto

---

41. (CESPE / EBC – 2011) Ocorre internalização quando parte do conhecimento tácito de uma pessoa converte-se no conhecimento tácito de outrem, tal como ocorre na realização de atividades práticas sob a supervisão de tutores.

Comentários:

A conversão do conhecimento tácito para tácito é chamada de socialização, e não internalização.

Gabarito: Errado

---

42. (CESPE / EBC – 2011) A socialização é a conversão de partes do conhecimento explícito da organização em conhecimento tácito do indivíduo.

Comentários:

A conversão de conhecimento explícito para tácito é chamada de internalização, e não socialização.

Gabarito: Errado

---

43. (CESPE / STM – 2011) O modelo SECI (sensibilização, externalização, combinação e informação) representa modos de criação de conhecimento organizacional.

Comentários:

SECI é o acrônimo para ~~Sensibilização~~ Socialização, Externalização, Combinação e ~~Informação~~ Internalização.

Gabarito: Errado

---

44. (FCC / TCE/RS – 2011) Com relação aos conceitos de dado e informação assinale a afirmação correta.

a) Uma lista com a quantidade de produtos vendidos é uma informação.
b) Um relatório sobre a produtividade dos funcionários é um dado.
c) Produtividade da mão de obra de uma empresa é um dado.
d) Custo da matéria prima é uma informação.
e) A capacidade de produção é uma informação.

Comentários:

(a) Errado, isso é um dado, visto que não há processamento; (b) Errado, isso é uma informação, visto que um relatório precisa ser processado; (c) Errado, isso é uma informação, visto que a produtividade precisa ser derivada a partir de dados; (d) Errado, isso é um dado, visto que – em princípio – não há processamento, mas é bem subjetivo; (e) Correto, isso é uma informação, visto que a capacidade de produção precisa ser processada e derivada de outros dados.

Gabarito: Letra E

---

45. **(CESPE / DPU – 2010)** Dados acerca de determinado assunto de interesse de uma ou mais pessoas, que possam ser interpretados ou tenham significado para o receptor, são considerados:

a) aprendizagem organizacional.
b) conhecimento.
c) capital intelectual.
d) informação.
e) sistema de gerenciamento.

Comentários:

Dados que possam ser interpretados ou que tenham um significado é o conceito básico de informação.

Gabarito: Letra D

---

46. **(CESPE / CGE-PB – 2008)** Informação é um conjunto de dados contextualizados que visam fornecer uma solução para determinada situação de decisão.

Comentários:

Informação é realmente um conjunto de dados contextualizados e processados para auxiliar na tomada de decisão.

Gabarito: Correto

QUESTÕES COMENTADAS – DIVERSAS BANCAS

1. **(CESPE / MPE-RO – 2023)** Em desenvolvimento de sistema de computação, os dados que oferecem a opção de serem utilizados livremente, reutilizados e redistribuídos são chamados de

a) dados móveis.
b) inteligência.
c) conhecimento.
d) dados estruturados.
e) dados abertos.

Comentários:

(a) Errado. "Dados móveis" geralmente se referem a dados acessados ou utilizados em dispositivos móveis, como smartphones e tablets, não necessariamente implicando livre uso, reutilização ou redistribuição.

(b) Errado. "Inteligência" em sistemas de computação pode referir-se a informações derivadas de análise de dados ou a sistemas de inteligência artificial, mas não especificamente a dados que podem ser livremente usados, reutilizados e redistribuídos.

(c) Errado. "Conhecimento" é um termo amplo que pode se referir a informações compreendidas e aplicadas por indivíduos ou sistemas, mas não denota especificamente a natureza livre e aberta dos dados para uso e distribuição.

(d) Errado. "Dados estruturados" descrevem um formato de dados, como dados organizados em tabelas de banco de dados, mas não indicam a permissão de uso livre, reutilização ou redistribuição desses dados.

(e) Correto. "Dados abertos" são dados que estão disponíveis para uso livre, reutilização e redistribuição por qualquer pessoa. Eles são frequentemente associados com iniciativas governamentais ou acadêmicas visando a transparência, colaboração e inovação.

Gabarito: Letra E

---

2. **(FGV / TCE-TO – 2022)** Dados abertos são dados publicados na internet, em formatos como csv e xml, para que qualquer pessoa possa acessar, modificar, analisar e compartilhar.

Os dados devem ser publicados na forma coletada da fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada, para atender ao princípio de dados abertos:

a) primários;

b) acessíveis;
c) completos;
d) processáveis por máquina;
e) acesso não discriminatório.

Comentários:

PRINCÍPIOS	DESCRIÇÃO
PRIMÁRIOS	Os dados são publicados na forma coletada na fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada.

Gabarito: Letra A

---

3. **(FGV / SEFAZ-AM – 2022) Com relação às características de dados abertos, analise as afirmativas a seguir.**

I. Reuso e redistribuição: os dados precisam estar disponíveis integralmente, devendo estar em um formato conveniente e modificável e sob custo não maior que um custo razoável de reprodução.

II. Disponibilidade e acesso: os dados precisam ser fornecidos em condições de reuso e redistribuição, podendo ser recombinados com outros dados.

III. Participação universal: todos os usuários podem usar, reusar e redistribuir o dado sem restrições de áreas de atuação, pessoas ou grupos.

Está correto o que se afirma em:

a) I, apenas.
b) II, apenas.
c) III, apenas.
d) I e II, apenas.
e) II e III, apenas.

Comentários:

DISPONIBILIDADE E ACESSO	Os dados devem estar disponíveis como um todo e sob custo não maior que um custo razoável de reprodução, preferencialmente possíveis de serem baixados pela internet. Os dados devem também estar disponíveis de uma forma conveniente e modificável.

REUTILIZAÇÃO E REDISTRIBUIÇÃO	Os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição – é permitida a combinação com outros conjuntos de dados não abertos (um conjunto de dados sobre um assunto específico também pode ser chamado de *data set*).
PARTICIPAÇÃO UNIVERSAL	Todos devem ser capazes de usar, reutilizar e redistribuir - não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos. Por exemplo, restrições de uso 'não-comercial' que impediriam o uso 'comercial', ou restrições de uso para certos fins (ex: somente educativos) excluem determinados dados do conceito de 'abertos'.

(I) Errado, essa é a descrição da característica de Disponibilidade e Acesso; (II) Errado, essa é a descrição da característica de Reutilização e Redistribuição; (III) Correto.

Gabarito: Letra C

---

4. **(CESPE / ISS-Aracaju – 2021)** Em decorrência do princípio de primariedade dos dados abertos governamentais, tais dados devem ser:

a) publicados na forma coletada na fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada.
b) razoavelmente estruturados para possibilitar o seu processamento automatizado.
c) disponibilizados em um formato cujo controle não seja exclusivo de nenhum ente.
d) disponibilizados o quão rapidamente seja necessário para preservar o seu valor.
e) disponibilizados para o público mais amplo possível e para os propósitos mais variados possíveis.

Comentários:

(a) Correto; (b) Errado, trata-se do princípio da processabilidade por máquina; (c) Errado, trata-se do princípio de formatos não proprietários; (d) Errado, trata-se do princípio da atualidade; (e) Errado, trata-se do princípio da acessibilidade.

PRINCÍPIOS	DESCRIÇÃO
COMPLETOS	Todos os dados públicos são disponibilizados. Dados são informações eletronicamente gravadas, incluindo, mas não se limitando a documentos, bancos de dados, transcrições e gravações audiovisuais. Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou controle de acesso, reguladas por estatutos.
PRIMÁRIOS	Os dados são publicados na forma coletada na fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada.
ATUAIS	Os dados são disponibilizados o quão rapidamente seja necessário para preservar o seu valor.
ACESSÍVEIS	Os dados são disponibilizados para o público mais amplo possível e para os propósitos mais variados possíveis.

PROCESSÁVEIS POR MÁQUINA	Os dados são razoavelmente estruturados para possibilitar o seu processamento automatizado.
ACESSO NÃO DISCRIMINATÓRIO	Os dados estão disponíveis a todos, sem que seja necessária identificação ou registro.
FORMATOS NÃO PROPRIETÁRIOS	Os dados estão disponíveis em um formato sobre o qual nenhum ente tenha controle exclusivo (ex: .csv, .html, .txt, .json, .xml, .ods, .rdf, .svg).
LIVRES DE LICENÇAS	Os dados não estão sujeitos a regulações de direitos autorais, marcas, patentes ou segredo industrial. Restrições razoáveis de privacidade, segurança e controle de acesso podem ser permitidas na forma regulada por estatutos (não é completamente livre de licenças).

Gabarito: Letra A

---

5. **(CESPE / SEFAZ-AL – 2021)** De acordo com os princípios de dados abertos, um dado disponível deve ser reutilizado, independentemente de estar disponível em formato aberto ou legível por máquina.

Comentários:

I - Se o dado não pode ser encontrado e indexado na Web, ele não existe;
II - Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e
III - Se algum dispositivo legal não permitir sua replicação, ele não é útil.

Bastava lembrar das leis de dados abertos: se os dados não estiverem abertos e disponíveis em formato compreensível por máquina, ele não pode ser reaproveitado.

Gabarito: Errado

---

6. **(CESPE / SEFAZ-AL – 2021)** A publicação e a catalogação de dados abertos demandam o uso de uma infraestrutura de software conhecida como ODP (Open Data Platform), ou plataforma de dados abertos, que contém componentes e interfaces.

Comentários:

A Plataforma de Dados Abertos (Open Data Platform) ou ODP é uma iniciativa compartilhada, formada por líderes da indústria na área de Big data, que realmente contém ferramentas, interfaces e outras tecnologias para a adoção do Apache Hadoop em organizações. No entanto, a publicação

e a catalogação de dados abertos não demandam necessariamente o uso dessa infraestrutura. Essa questão não faz o menor sentido e a banca deu o gabarito como... correto! Discordo veementemente!

Gabarito: Correto

---

7. **(FEPESE / CIASC – 2017)** Com relação ao assunto Dados Abertos, assinale a alternativa correta.

a) Um requisito importante para que o dado seja considerado aberto é o da reutilização e redistribuição no qual os dados devem estar disponíveis como um todo e sob custo razoável de reprodução. Tais dados preferencialmente devem ser possíveis de serem baixados pela internet e devem estar disponíveis de uma forma conveniente e modificável.

b) Um requisito importante para que o dado seja considerado aberto é o da disponibilidade e acesso no qual os dados devem ser fornecidos sob termos que permitam o seu fechamento, abertura ou a combinação com outros conjuntos de dados.

c) Dados abertos são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa. Estão sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras.

d) Participação Universal significa a capacidade de diversos sistemas e organizações trabalharem juntos. Nesse caso, trata-se da capacidade de interoperar ou combinar diferentes conjuntos de dados.

e) Entende-se Interoperabilidade como a capacidade de usar, reutilizar e redistribuir, não devendo haver discriminação contra áreas de atuação ou contra pessoas ou grupos. Por exemplo, restrições de uso não comercial que impediriam o uso comercial, ou restrições de uso para certos fins que excluem determinados dados do conceito de 'abertos'.

Comentários:

(a) Errado, essa descrição está relacionada com o requisito Disponibilidade e Acesso, e não Reutilização e Redistribuição; (b) Errado, a questão mistura o requisito da Disponibilidade e Acesso com o requisito da Reutilização e Redistribuição; (c) Correto, esse é o requisito da Participação Universal, em que todos devem ser capazes de usar, reutilizar e redistribuir – não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos; (d) Errado, a interoperabilidade significa a capacidade de diversos sistemas e organizações trabalharem juntos (interoperar). Neste caso, trata-se da capacidade de interoperar – ou combinar – diferentes conjuntos de dados; (e) Errado, esse é o requisito de Participação Universal.

Gabarito: Letra C

---

8. **(FEPESE / CIASC – 2017)** De acordo com a Política de Dados Abertos do Poder Executivo federal, os dados devem ser representados em meio digital e:

a) estruturados em formato aberto, disponibilizados em formato requerido pelo cidadão, de forma que permita utilização, consumo ou cruzamento de acordo com as normas da instituição detentora da informação.

b) estruturados em formato aberto, processáveis por máquina, disponibilizados sob licença aberta, de modo que permita livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte.

c) disponibilizados na Internet, processáveis por máquina, estruturados de forma que permita utilização, consumo ou cruzamento, de acordo com as normas da instituição detentora da informação.

d) disponibilizados na Internet, estruturados sob licença da instituição pública detentora da informação, de modo que permita utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte.

Comentários:

O Decreto 8.777/2016 afirma que os dados abertos são "*dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte*".

Gabarito: Letra B

---

9. **(CESPE / TCE-SC – 2016)** Dados abertos são os dados de livre utilização, reutilização e redistribuição, exigindo-se, no máximo, créditos à autoria e compartilhamento pela mesma licença.

Comentários:

Perfeito! Dados abertos são realmente de livre utilização, reutilização e redistribuição, exigindo-se, no máximo, créditos à autoria e compartilhamento pela mesma licença.

Gabarito: Correto

---

10. **(CESPE / TCE/RJ – 2021)** A filosofia de uso de dados abertos orienta que os dados sejam reutilizados amplamente, no entanto não é permitida a combinação de um data set aberto com conjuntos de dados não abertos.

Comentários:

Opa, é permitida – sim – a combinação de um *data set* aberto com um *data set* não aberto.

Gabarito: Errado

---

11. (CESPE / TCE/RJ – 2021) As três normas fundamentais que compõem o conceito de dados abertos são: disponibilidade e acesso; reuso e distribuição; e participação universal.

Comentários:

A questão menciona as características mínimas para um dado ser considerado aberto, e não as três normas fundamentais, que seriam:

LEIS DOS DADOS ABERTOS
I - Se o dado não pode ser encontrado e indexado na Web, ele não existe;
II - Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e
III - Se algum dispositivo legal não permitir sua replicação, ele não é útil.

De todo modo, essas três características podem ser interpretadas como as leis ou normas fundamentais dos dados abertos apresentadas acima.

Gabarito: Correto

---

12. (CESPE / ME – 2020) Um dos princípios de dados abertos é o acesso não discriminatório, ou seja, os dados estão disponíveis independentemente de identificação ou registro do usuário.

Comentários:

ACESSO NÃO DISCRIMINATÓRIO	Os dados estão disponíveis a todos, sem que seja necessária identificação ou registro.

Acesso não discriminatório realmente diz respeito aos dados que estão disponíveis a todos, sem que seja necessária identificação ou registro.

Gabarito: Correto

---

13. (CESPE / ME – 2020) A abertura de dados por organizações governamentais é uma prerrogativa das próprias organizações, que devem ditar normas internas sobre o assunto.

Comentários:

Opa... a abertura por organizações governamentais é uma obrigação, e não uma prerrogativa!

Gabarito: Errado

---

14. (CCV / UFC – 2013) A Instrução normativa Nº 4, de 12 de abril de 2012, instituiu a INDA, a qual disponibiliza um sítio eletrônico de referência para a busca e o acesso aos dados públicos, seus metadados, informações, aplicativos e serviços relacionados. Selecione a alternativa que descreve o significado da sigla INDA e o respectivo sítio eletrônico mencionado anteriormente:

a) Infraestrutura Nacional de Dados Acessíveis; Portal da Transparência.
b) Infraestrutura Nacional de Dados Abertos; Portal da Transparência.
c) Infraestrutura Nacional de Dados Abertos; Portal Brasileiro de Dados Abertos.
d) Infraestrutura Nacional de Dados Acessíveis; Portal Brasileiro de Dados Abertos.
e) Infraestrutura Nacional de Dados Acessíveis; Portal do Acesso à Informação Pública.

Comentários:

INDA é o acrônimo para Infraestrutura Nacional de Dados Abertos e seu sítio eletrônico é o Portal Brasileiro de Dados Abertos.

Gabarito: Letra C

---

15. (COVEST / UFPE – 2019) Quanto aos princípios de dados abertos governamentais, quando se fala em "não discriminatório", está se afirmando que os dados:

a) devem ser publicados na forma coletada na fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada.

b) devem estar disponíveis o quão rapidamente seja necessário para preservar o seu valor.

c) devem estar disponíveis com o objetivo de aumentar a participação de grupos de ativistas de dados cívicos.

d) devem estar disponíveis para todos, sem que seja necessária identificação ou registro.

e) devem estar razoavelmente estruturados para possibilitar o seu processamento automatizado.

Comentários:

ACESSO NÃO DISCRIMINATÓRIO	Os dados estão disponíveis a todos, sem que seja necessária identificação ou registro.

O acesso não discriminatório está associado à disponibilidade dos dados a todos, sem que seja necessária identificação ou registro. Analisando as outras alternativas: (a) Errado, trata-se do princípio: Primários; (b) Errado, trata-se do princípio: Atuais; (c) Errado, trata-se do princípio: Acessíveis; (e) Errado, trata-se do princípio: Processáveis Por Máquina.

Gabarito: Letra D

---

16. (CESPE / IPHAN – 2018) O indicador barômetro de dados abertos foi escolhido para medir o impacto das iniciativas dos dados abertos no Brasil.

Comentários:

Perfeito! Ele foi escolhido pela Estratégia de Governança Digital para medir o impacto das iniciativas dos dados abertos no Brasil.

Gabarito: Correto

---

17. (CESPE / TCE-RJ – 2021) O portal brasileiro de dados abertos (*www.dados.gov.br*) disponibiliza diversos conjuntos de dados em formatos PDF, CSV, HTML e JSON.

Comentários:

Perfeito! Ele realmente disponibiliza diversos *data sets* em formatos PDF, CSV, HTML e JSON. Lembrando que, desde 2008, PDF é um formato aberto.

Gabarito: Correto

QUESTÕES COMENTADAS – DIVERSAS BANCAS

1. **(CESPE / Polícia Federal – 2013)** Título, assunto, palavras-chave e comentários de um documento são metadados típicos presentes em um documento produzido por processadores de texto como o BrOffice e o Microsoft Office.

Comentários:

Os arquivos de processadores de texto realmente possuem metadados como título, assunto, palavras-chave, comentários, entre outros – BrOffice e Microsoft Office são suítes de escritório que possuem ferramentas específicas de processamento de texto (Writer e Word), mas a questão infelizmente ignorou esse detalhe.

Gabarito: Correto

---

2. **(CESPE / ABIN – 2018)** Os metadados são dados utilizados para a criação e o uso de documentos em sistemas informatizados indisponíveis aos usuários.

Comentários:

Metadados são dados utilizados para a criação e o uso de documentos em sistemas informatizados ~~indisponíveis~~ transparentes aos usuários, isto é, os metadados são gerados sem que necessariamente sejam notados pelos usuários.

Gabarito: Errado

---

3. **(CESPE / TRT-ES – 2013)** Os dados estruturados que descrevem e permitem encontrar, gerenciar, compreender e (ou) preservar documentos arquivísticos ao longo do tempo são conhecidos como metadados.

Comentários:

Perfeito, os dados estruturados que descrevem e permitem encontrar, gerenciar, compreender e (ou) preservar documentos arquivísticos ao longo do tempo são conhecidos como metadados.

Gabarito: Correto

---

4. **(CESPE / ANTAQ – 2009)** Metadados são dados estruturados e codificados de modo a descreverem características de entidades para auxiliarem na identificação, na descoberta e no gerenciamento das entidades descritas.

Comentários:

Mais uma vez, uma excelente descrição de metadados.

Gabarito: Correto

---

5. **(CESPE / MPE-CE – 2020)** Os metadados descrevem, explicam, localizam e facilitam a recuperação de um recurso informacional, permitindo que esse recurso esteja acessível futuramente.

Comentários:

Metadados realmente descrevem, explicam, localizam e facilitam a recuperação de um recurso informacional, permitindo que esse recurso esteja acessível futuramente (Definição da NISO – *National Information Standards Organization*).

Gabarito: Correto

---

6. **(FUNPAR / UFPR – 2010)** Em se tratando de gerenciamento da informação, dados estruturados que descrevem, identificam, explicam, localizam e, portanto, facilitam a recuperação, uso e gestão de recursos de informação, são chamados de:

a) informação.
b) sistema de informação.
c) conhecimento.
d) metadados.
e) dicionário de dados.

Comentários:

Dados estruturados que descrevem, identificam, explicam, localizam e, portanto, facilitam a recuperação, uso e gestão de recursos de informação, são chamados de metadados.

Gabarito: Letra D

---

7. **(CESPE / TCE-RJ – 2021)** Dublin Core é um esquema de metadados que auxilia na descrição de objetos digitais por meio da definição de diversos elementos de metadados, entre os quais se incluem título, autor, assunto, formato e fonte.

Comentários:

Ele realmente é um esquema de metadados que auxilia na descrição de objetos digitais por meio da definição de diversos elementos de metadados, entre os quais se incluem título, autor, assunto, formato e fonte.

Gabarito: Correto

---

8. **(CESPE / Polícia Federal – 2018)** Em arquivos no formato XML, as tags não são consideradas metadados.

Comentários:

As tags são, de fato, uma forma de metadados. Metadados são dados sobre dados, fornecendo informações adicionais sobre o conteúdo, contexto ou estrutura de um arquivo. Em XML, as tags desempenham exatamente esse papel. As tags em um documento XML definem a estrutura do documento e descrevem a natureza dos dados contidos entre elas. Elas ajudam a interpretar o significado e a estrutura dos dados contidos no documento, tornando-se metadados essenciais.

Gabarito: Errado

---

9. **(CESPE / TCE-RJ – 2021)** Os metadados sintáticos, estruturais e semânticos descrevem, respectivamente, a sintaxe, a estrutura e o conteúdo dos dados.

Comentários:

Perfeito! Metadados sintáticos descrevem a sintaxe dos dados; metadados estruturais descrevem a estrutura dos dados; e metadados semânticos descrevem o significado/conteúdo dos dados.

Gabarito: Correto

---

10. **(CESPE/ CNPQ – 2011)** As propriedades de um documento, também conhecidas como metadados, são detalhes de um arquivo que o descrevem ou o identificam. As propriedades incluem detalhes como título, nome do autor, assunto e palavras-chave que identificam o tópico ou o conteúdo do documento.

Comentários:

As propriedades realmente são conhecidas como metadados, que são capazes de descrever e identificar um arquivo, como título, nome do autor, palavras-chave, entre outros.

Gabarito: Correto

---

11. **(COPEVE / MPE-AL – 2012)** São elementos Metadados Dublin Core:

a) Título; Criador; Assunto; Descrição; Data; e Identificador.
b) Título; Responsabilidade; Assunto; Descrição; endereço web; e Tipo.
c) Criador; Assunto; Descrição; Data; Formato; e Notas.

d) Publicador; Contribuidor; Data; Formato; e série.
e) Autor; Assunto; Descrição; Publicador; Formato; e Etiqueta.

Comentários:

Entre os metadados do Dublin Core, temos: título, criador, assunto, descrição, produtor, colaborador, data, tipo, formato, identificador, fonte, idioma, relação, cobertura e direitos.

Gabarito: Letra A

---

12. (EXATUS / DETRAN-RJ – 2012) Os metadados podem ser definidos como:

a) Elementos de descrição/definição/avaliação de recursos informacionais armazenados em sistemas computadorizados e organizados por padrões específicos, de forma estruturada.

b) Conjunto de técnicas, práticas e procedimentos que estabelecem os fundamentos técnicos e metodológicos de um sistema digital.

c) Forma de estruturação da informação que permite a leitura não linear de um texto.

e) Sistema para identificar univocamente objetos digitais na WEB, tem como propósito principal a gestão da propriedade intelectual e o comércio digital dos objetos.

Comentários:

(a) Correto. No contexto de sistemas de informação, metadados podem ser definidos como elementos de descrição de recursos informacionais armazenados em sistemas computadorizados e organizados por padrões específicos de forma estruturada; (b) Errado, não há nenhuma relação com técnicas, práticas e procedimentos; (c) Errado, essa seria a definição de um hipertexto; (d) Errado, essa seria a definição de um DOI (Digital Object Identifier).

Gabarito: Letra A

---

13. (UEAP / CS-UFG – 2014) Metadados são dados sobre outros dados. No contexto de bancos de dados, metadados são usados para:

a) descrever a estrutura do banco de dados.
b) descrever o conteúdo do log do sistema.
c) resumir o conteúdo dos dados de cada relação.
d) resumir a estrutura dos programas de aplicação.

Comentários:

No contexto de bancos de dados, metadados descrevem a estrutura do banco de dados.

Gabarito: Letra A

---

14. (FGV / Fiocruz – 2010) Sobre os Metadados, assinale a afirmativa correta.

a) São um conjunto de informações estruturadas que descrevem os dados (bases) que representam, facilitando ao usuário o gerenciamento desses dados.
b) São o conjunto total de dados que integram um projeto de geoprocessamento.
c) São dados preliminares que necessitam de averiguação e eventuais correções para, posteriormente, serem utilizados.
d) São o conjunto de dados que já passaram por todas as averiguações e já estão prontos para utilização.
e) São dados que não foram utilizados e são arquivados para futuros trabalhos.

Comentários:

Metadados podem ser estruturados ou não estruturados e realmente descrevem os dados que representam – nenhum dos outros dados fazem qualquer sentido!

Gabarito: Letra A

---

15. (BIO-RIO / IF-RJ – 2015) Os metadados que facilitam a navegação e a apresentação de recursos eletrônicos são os:

a) descritivos.
b) estruturais.
c) administrativos.
d) operacionais.
e) interoperáveis.

Comentários:

METADADOS	DESCRIÇÃO
DESCRITIVOS	Estão relacionados à descrição de um objeto para torná-lo mais fácil de ser descoberto e identificado; podem incluir elementos, tais como título, autor, resumo, palavras-chave e identificador – dependem dos metadados estruturais para serem construídos;
ESTRUTURAIS	Estão relacionados à documentação de como um recurso é composto ou organizado – facilitando a navegação e apresentação de recursos. Um livro digitalizado, por exemplo, pode ser publicado como arquivos de imagens individuais das páginas, arquivos PDF ou HTML.
ADMINISTRATIVOS	Estão relacionados ao fornecimento de informações que apoiam os processos de gerenciamento de recursos informacionais. Incluem, por exemplo, informações sobre como e quando o recurso foi criado e a razão da sua criação.

O tipo de metadado que facilita a navegação e a apresentação de recursos eletrônicos são os metadados estruturais, isto é, descrevem como recursos são organizados estruturalmente.

Gabarito: Letra B

---

16. (CESPE / FUB – 2015) A informação que descreve a estrutura dos dados e sua relação com outros dados é o metadado.

Comentários:

Perfeito! Dados que descrevem a estrutura dos dados e sua relação com outros dados é o metadado.

Gabarito: Correto

---

17. (VUNESP / UNIFAI – 2019) No contexto da Ciência da Informação, os metadados são classificados, por função, em:

a) metadados modulares, metadados extensos e metadados refinados.
b) metadados descritivos, metadados estruturais e metadados administrativos.
c) metadados multilíngues, metadados aplicados e metadados principais.
d) metadados estruturais, metadados completos e metadados plurilíngues.
e) metadados teóricos, metadados aplicados e metadados administrativos.

Comentários:

METADADOS	DESCRIÇÃO
DESCRITIVOS	Estão relacionados à descrição de um objeto para torná-lo mais fácil de ser descoberto e identificado; podem incluir elementos, tais como título, autor, resumo, palavras-chave e identificador – dependem dos metadados estruturais para serem construídos;
ESTRUTURAIS	Estão relacionados à documentação de como um recurso é composto ou organizado – facilitando a navegação e apresentação de recursos. Um livro digitalizado, por exemplo, pode ser publicado como arquivos de imagens individuais das páginas, arquivos PDF ou HTML.
ADMINISTRATIVOS	Estão relacionados ao fornecimento de informações que apoiam os processos de gerenciamento de recursos informacionais. Incluem, por exemplo, informações sobre como e quando o recurso foi criado e a razão da sua criação.

Gabarito: Letra B

---

18. (COPEVE / MPE-AL – 2012) Metadados são:

a) dados sobre dados.
b) dados disponibilizados em redes sociais.
c) dados disponibilizados em redes cooperativas.

d) informações disponíveis na Internet.
e) dados e informações disponíveis na intranet.

Comentários:

Metadados são dados sobre dados – todos os outros itens não fazem o menor sentido!

Gabarito: Letra A

---

19. (CESPE / Polícia Federal – 2018) Na gestão de documentos digitalizados, os arquivos são criados em formato de imagens, e seus atributos são armazenados em metadados para facilitar a pesquisa e a localização desses arquivos.

Comentários:

O que é gestão de documentos digitalizados? É uma tecnologia que provê um meio de facilmente gerar, controlar, armazenar, compartilhar e recuperar informações existentes em documentos digitais. Os formatos mais comumente utilizados nessa tecnologia são .pdf, .jpeg, .bitmap e .tiff. *PDF é um formato de imagem?* Poucas pessoas sabem disso, mas ele realmente é um formato de imagem vetorial (aquele que pode ser redimensionado sem perder qualidade).

Dito isso, os atributos de arquivos em formato de imagem realmente são armazenados em metadados para facilitar a pesquisa e a localização desses arquivos (Ex: data, tamanho, criador, formato, comentário, título, direitos autorais, localização, etc).

Gabarito: Correto

---

20. (CESPE / Polícia Federal – 2021) A função do metadado de arquivo é descrever o destino final do arquivo definido pelo emissor da mensagem e proprietário do arquivo.

Comentários:

Metadados de arquivo são dados sobre dados, isto é, informações que descrevem o conteúdo, a qualidade, a condição ou outras características de um arquivo. Geralmente, os metadados incluem detalhes como o autor do arquivo, a data da criação, a última modificação, o tamanho e o tipo do arquivo. A função dos metadados não é descrever o destino final do arquivo, mas sim fornecer informações contextuais que podem ser usadas para gerenciar, organizar e localizar arquivos dentro de sistemas de armazenamento.

Gabarito: Errado

QUESTÕES COMENTADAS – DIVERSAS BANCAS

1. **(CESPE / SEPLAN-RR - 2023)** Com base nos conceitos de dados estruturados e não estruturados, de análise exploratória de dados e de dados abertos, julgue o item seguinte.

   Dados de redes sociais são exemplos de dados não estruturados.

Comentários:

Perfeito! Os dados das redes sociais, como posts, comentários, fotos, vídeos e mensagens, são exemplos típicos de dados não estruturados. Esses dados são gerados pelos usuários de forma livre e podem variar muito em termos de formato, conteúdo e contexto.

Gabarito: Correto

---

2. **(FGV / EPPGG - 2023)** Um analista de políticas públicas trabalha com diversos formatos e tipos de dados. Identificar corretamente essas características no começo das atividades é essencial para realizar a escolha correta das ferramentas e técnicas que serão utilizadas.

   Assinale a alternativa que traz apenas dados semiestruturados.

   a) Arquivo CSV (Comma Separated Values) e arquivos de áudio.
   b) Arquivo JSON (JavaScript Object Notation) e arquivos de imagens.
   c) Arquivo JSON (JavaScript Object Notation) e arquivo XML (eXtensible Markup Language).
   d) Arquivos PDF (Portable Document Format) e arquivos de vídeo.
   e) Tabela de um banco de dados relacional e arquivo XML (eXtensible Markup Language).

Comentários:

(a) Errado. Dado Semi-Estruturado e Dado Não-Estruturado;
(b) Errado. Dado Semi-Estruturado e Dado Não-Estruturado;
(c) Correto. Dado Semi-Estruturado e Dado Semi-Estruturado;
(d) Errado. Dado Semi-Estruturado e Dado Não-Estruturado;
(e) Errado. Dado Estruturado e Dado Semi-Estruturado;

Gabarito: Letra C

---

3. **(CESPE / DATAPREV – 2023)** Arquivos .xml, .csv e .json são exemplos de dados estruturados.

Comentários:

Arquivos .xml (eXtensible Markup Language), .csv (Comma-Separated Values) e .json (JavaScript Object Notation) são exemplos de formatos semiestruturados. Cada um desses formatos tem características específicas que os tornam adequados para armazenar e organizar dados de maneira estruturada:

- XML é um formato flexível que usa tags para definir elementos. É comumente usado para compartilhar dados na internet e em outros ambientes de rede devido à sua capacidade de representar dados complexos e hierárquicos de forma clara.

- CSV é um formato simples que usa vírgulas (ou outro delimitador) para separar valores. Cada linha do arquivo representa uma linha de dados, e cada valor separado por vírgula representa uma coluna. Este formato é amplamente utilizado devido à sua simplicidade e compatibilidade com muitos sistemas.

- JSON é um formato leve de troca de dados, fácil de ler e escrever para humanos, e fácil de analisar e gerar por máquinas. É baseado em um subconjunto da linguagem de programação JavaScript e é muito usado em APIs da web e configurações de aplicativos devido à sua facilidade de uso em ambientes de desenvolvimento web.

Gabarito: Errado

---

4. **(CONSULPLAN / MPE-PA – 2022)** Dados são um conjunto de fatos em estado bruto que podem ser utilizados para tomadas de decisão. Os dados podem ser divididos entre estruturados e não-estruturados. Sobre os dados estruturados, assinale a afirmativa correta.

a) Mensagens de e-mail são dados estruturados.
b) Dados estruturados são organizados em linhas e colunas.
c) Dados estruturados são armazenados em arquivos no formato ".TXT".
d) Mecanismos de busca possuem dificuldade para pesquisar dados estruturados.

Comentários:

(a) Errado. Mensagens de e-mail são geralmente consideradas dados não estruturados, pois consistem principalmente de texto livre e podem conter uma variedade de formatos e anexos;

(b) Correto. Dados estruturados são normalmente organizados em linhas e colunas, como em um banco de dados relacional. Esta estrutura permite fácil acesso, manipulação e análise dos dados;

(c) Errado. Embora seja possível armazenar dados estruturados em arquivos no formato ".txt", este não é o método predominante ou mais representativo;

(d) Errado. Mecanismos de busca, especialmente os projetados para bancos de dados, geralmente têm facilidade para pesquisar dados estruturados devido à sua organização clara e previsível.

Gabarito: Letra B

---

5. **(CONSULPLAN / MPE-PA – 2022)** Dados são um conjunto de fatos em estado bruto que podem ser utilizados para tomadas de decisão. Os dados podem ser divididos entre estruturados e não-estruturados. Sobre os dados estruturados, assinale a afirmativa correta.

a) Mensagens de e-mail são dados estruturados.
b) Dados estruturados são organizados em linhas e colunas.
c) Dados estruturados são armazenados em arquivos no formato ".TXT".
d) Mecanismos de busca possuem dificuldade para pesquisar dados estruturados.

Comentários:

(a) Errado. Mensagens de e-mail são geralmente consideradas dados não estruturados, pois consistem principalmente de texto livre e podem conter uma variedade de formatos e anexos;

(b) Correto. Dados estruturados são normalmente organizados em linhas e colunas, como em um banco de dados relacional. Esta estrutura permite fácil acesso, manipulação e análise dos dados;

(c) Errado. Embora seja possível armazenar dados estruturados em arquivos no formato ".txt", este não é o método predominante ou mais representativo;

(d) Errado. Mecanismos de busca, especialmente os projetados para bancos de dados, geralmente têm facilidade para pesquisar dados estruturados devido à sua organização clara e previsível.

Gabarito: Letra B

---

6. **(CESPE / Polícia Federal – 2021)** Os dados estruturados diferenciam-se dos dados não estruturados pela rigidez em seu formato e pelo fato de poderem ser armazenados em campos de tabelas de um banco de dados relacional.

Comentários:

Perfeito! Dados estruturados são aqueles que são organizados e representados por meio de uma estrutura rígida e formal; já os dados não-estruturados são aqueles que não possuem uma organização rígida e formal, isto é, não seguem uma estrutura padronizada, sendo totalmente flexível e dinâmica.

Gabarito: Correto

---

7. **(CESPE / SERPRO – 2021)** Os dados não estruturados constam de bancos de dados relacionais que são eminentemente pesquisáveis tanto por meio de consultas realizadas por humanos

quanto por meio daquelas realizadas por algoritmos que usam tipos de dados e nomes de campos, como alfabéticos ou numéricos, moeda ou data.

Comentários:

Na verdade, eles constam de bancos de dados não relacionais, não sendo eminentemente pesquisáveis por consultas realizadas por humanos ou por meio daquelas realizadas por algoritmos que usam tipos de dados e nomes de campos, como alfabéticos ou numéricos, moeda ou data.

Gabarito: Errado

---

8. **(CESPE/ ME – 2020)** Embora com características particulares, dados não estruturados podem ser classificados em sua totalidade, assim como os dados estruturados.

Comentários:

Galera, a questão não deixa muito claro o que quis dizer com essa classificação. De todo modo, podemos inferir que essa classificação se trata de padronização. Pois bem, dados não-estruturados não possuem uma estrutura definida e padronizada, logo não podem ser classificados em sua totalidade (como ocorre com dados estruturados).

Gabarito: Errado

---

9. **(FAURGS/ TJ-RS – 2018)** Os dados _____ são armazenados em _____. Os dados _____ misturam nomes de tipos de dados e valores de dados, mas nem todos eles precisam seguir uma estrutura predefinida fixa. Os dados _____ referem-se à informação exibida na Web, caso em que a informação sobre os tipos dos itens de dados não existe.

Assinale a alternativa que completa, correta e respectivamente, as lacunas do texto acima.

a) semiestruturados / bancos de dados tradicionais / estruturados / não estruturados
b) não estruturados / XML / semiestruturados /estruturados
c) estruturados / HTML / não estruturados / semiestruturados
d) estruturados / bancos de dados tradicionais / semiestruturados / não estruturados
e) semiestruturados / Web / estruturados / não estruturados

Comentários:

Os dados estruturados são armazenados em bancos de dados tradicionais. Os dados semiestruturados misturam nomes de tipos de dados e valores de dados, mas nem todos eles precisam seguir uma estrutura predefinida fixa. Os dados não-estruturados referem-se à informação exibida na Web, caso em que a informação sobre os tipos dos itens de dados não existe.

Gabarito: Letra D

---

10. (CESPE / Polícia Federal – 2018)

CPF
NOME
DATA DE NASCIMENTO
NOME DO PAI
NOME DA MAE
TELEFONE
CEP
NUMERO

As informações anteriormente apresentadas correspondem aos campos de uma tabela de um banco de dados, a qual é acessada por mais de um sistema de informação e também por outras tabelas. Esses dados são utilizados para simples cadastros, desde a consulta até sua alteração, e também para prevenção à fraude, por meio de verificação dos dados da tabela e de outros dados em diferentes bases de dados ou outros meios de informação.

Considerando essas informações, julgue o item que se segue.

Os dados armazenados na referida tabela são considerados não estruturados.

Comentários:

O enunciado afirma que as informações correspondem aos campos de uma tabela de um banco de dados, logo são dados estruturados. *Por quê?* Porque possuem uma estrutura rígida, formal, homogênea e previamente formatada.

Gabarito: Errado

---

11. (CESPE / TCE-SC – 2016) Em se tratando de dados estruturados, a informação de esquema está mesclada aos valores dos dados, e cada objeto de dados pode ter atributos diferentes, que não são conhecidos com antecedência. Essa característica os diferencia de dados não estruturados.

Comentários:

Dados mesclados com um esquema de representação parcialmente organizados são características de dados semiestruturados e, não, não estruturados.

Gabarito: Errado

---

12. **(CESPE / TCE-PA – 2016)** Em comparação aos dados não estruturados, os dados estruturados demandam mais espaço de armazenamento e um gerenciamento mais cauteloso, uma vez que constituem a maior parte dos dados corporativos.

Comentários:

A maior parte dos dados corporativos são não-estruturados, logo eles necessitam de maior espaço de armazenamento e um gerenciamento mais cauteloso.

Gabarito: Errado

---

13. **(FEPESE / SJC-SC – 2013)** Qual software, dentre os listados abaixo, permite a importação (e atualização) de dados estruturados de bancos de dados relacionais para posterior manipulação pelo aplicativo?

a) Word
b) Excel
c) Powerpoint
d) Internet Explorer
e) Outlook

Comentários:

Dados estruturados são aqueles que podem ser armazenados em... tabelas. Logo, é o MS-Excel.

Gabarito: Letra B

---

14. **(CESPE/ TRT-ES – 2013)** Os dados estruturados que descrevem e permitem encontrar, gerenciar, compreender e (ou) preservar documentos arquivísticos ao longo do tempo são conhecidos como metadados.

Comentários:

Perfeito, os dados estruturados que descrevem e permitem encontrar, gerenciar, compreender e (ou) preservar documentos arquivísticos ao longo do tempo são conhecidos como metadados.

Gabarito: Correto

QUESTÕES COMENTADAS – DIVERSAS BANCAS

1. **(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

---

2. **(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.

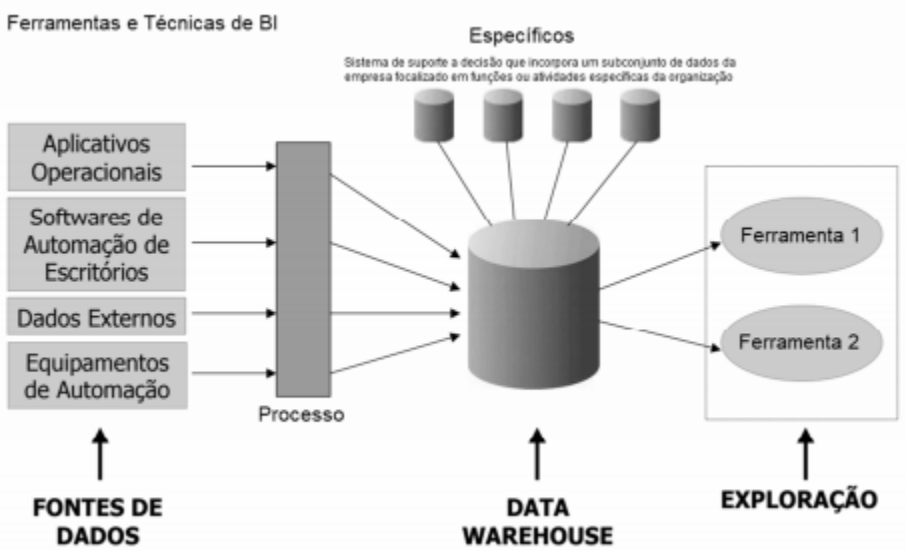
Comentários:

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

---

3. **(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

---

4. **(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

---

5. **(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

---

6. **(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 na 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

---

7. **(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

---

8. **(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

---

9. **(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

---

10. (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.

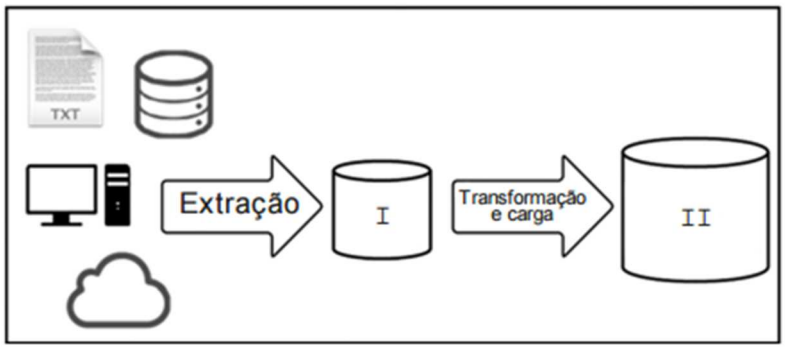
Comentários:

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

---

11. (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

---

12. (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

---

13. (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, como Sistemas ERP, Arquivos Texto, Planilhas Excel, entre outros.

Gabarito: Correto

---

14. (CESPE / ANATEL – 2014) As ferramentas de business inteligence, 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

---

15. (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

---

16. (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

---

17. (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

---

18. (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

---

19. (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

---

20. (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 carregá-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

---

21. (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

---

22. (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

---

23. (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

QUESTÕES COMENTADAS – DIVERSAS BANCAS

1. **(CESGRANRIO / BB – 2023)** No processo de aprendizagem de máquina, busca-se identificar padrões compreensíveis em bases de dados. Desse modo, supõe-se a existência de uma hierarquia de três níveis: a base, com os itens elementares, captados e armazenados por recursos de Tecnologia da Informação; o nível intermediário, no qual esses itens são processados, com significados e contextos bem definidos; e o nível mais alto, contendo os padrões ou os conjuntos cuja formulação pode envolver ou relacionar os níveis inferiores. Os três níveis dessa hierarquia, listados, respectivamente, da base para o nível mais alto, são:

   a) conhecimento, informação e dado
   b) dado, conhecimento e informação
   c) dado, informação e conhecimento
   d) informação, conhecimento e dado
   e) informação, dado e conhecimento.

2. **(CESPE / SEPLAN-RR - 2023)** Dados de redes sociais são exemplos de dados não estruturados.

3. **(CESPE / TCE-SC - 2022)** A gestão do conhecimento está relacionada à ação, à atitude e a uma intenção específica dos seres humanos.

4. **(CESPE / TCE-SC - 2022)** Informação pode ser definida como um conjunto de fatos diferentes relativos a uma transação e que é, geralmente, desprovida de significado; ela é considerada a matéria-prima dos sistemas de informação.

5. **(CESPE / TCE-SC - 2022)** O conhecimento serve para dar suporte ao processamento decisório; ele representa a informação tratada, confiável e íntegra.

6. **(FEPESE / ISS-Criciúma – 2022)** Analise as afirmativas abaixo com relação a dado, informação e conhecimento no domínio subjetivo.

   1. Dados são estímulos sensoriais, que nós percebemos através dos nossos sentidos.
   2. Informação é, mais do que ser somente um estágio intermediário entre dado e conhecimento, um conhecimento empírico em si.
   3. Conhecimento é o significado desses estímulos sensoriais que nós percebemos através de nossos sentidos.

   a) É correta apenas a afirmativa 2.
   b) São corretas apenas as afirmativas 1 e 2.
   c) São corretas apenas as afirmativas 1 e 3.
   d) São corretas apenas as afirmativas 2 e 3.
   e) São corretas as afirmativas 1, 2 e 3.

7. **(CESPE / Petrobrás - 2022)** A inteligência é a informação que é expressa na forma de números, letras ou símbolos, os quais são os insumos dos gráficos.

8. **(CESPE / PCDF – 2021)** A eficiência na representação de dados relaciona-se à codificação de fonte, cujo objetivo é remover redundâncias no conteúdo da informação e, com isso, reduzir o tamanho da mensagem.

9. **(CESPE / PF - 2021)** Os dados estruturados diferenciam-se dos dados não estruturados pela rigidez em seu formato e pelo fato de poderem ser armazenados em campos de tabelas de um banco de dados relacional.

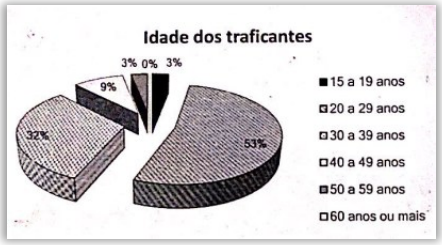
10. **(CESPE / PCDF – 2021)**. Embora não haja conceitos únicos para dados e informação, é possível afirmar que os dados são matéria-prima para a informação.

11. **(CESPE / TCE-RJ – 2021)** Um dado que está inserido em um contexto pode ser denominado de informação.

12. **(CESPE / TCE-RJ – 2021)** A informação pode ser facilmente obtida por meio de máquinas, além de ser transferível e frequentemente quantificável.

13. **(CESPE / TCE-RJ – 2021)** O conceito de inteligência está relacionado à capacidade de julgamento, ou seja, o ser humano usa o seu conhecimento para tomar a melhor decisão em uma situação real.

14. **(CESPE / Polícia Federal – 2021)** Considerando-se a classificação dados, informação, conhecimento e inteligência, é correto afirmar que o gráfico representa, por si só, a inteligência.



15. **(CESPE / Polícia Federal – 2021)** O número 1.789 sozinho caracteriza uma informação, independentemente do contexto.

16. **(CESPE / Polícia Federal – 2021)** Considere que a Polícia Federal tenha registrado, em determinado período, a prisão de 1.789 traficantes de drogas pertencentes a facções criminosas, conforme faixas etárias mostradas no gráfico.

Com referência às informações e ao gráfico precedentes, julgue o item subsecutivo.

Considerando-se a classificação dados, informação, conhecimento e inteligência, é correto afirmar que o gráfico representa, por si só, a inteligência.

17. **(CESPE / Polícia Federal – 2018)** O atributo de inteligência depende mais da qualidade da informação disponível do que da sua quantidade, tendo, portanto, natureza qualitativa.

18. **(CESPE / Polícia Federal – 2018)** Dados são fatos que descrevem os objetos de informação, por exemplo, eventos e entidades.

19. **(CESPE / Polícia Federal – 2018)** Na representação da informação, os atributos permitem que entidades e eventos possam ser reconhecidos, referidos e descritos. Um atributo relacional permite relacionar eventos e entidades.

20. **(FCC / TCE/RS – 2018)** Os conceitos de dados, informação e conhecimento são de grande importância no contexto de sistemas de informação.

Sobre eles, é correto afirmar que:

a) não são necessários os dados para que se obtenha o conhecimento.
b) a informação é obtida acrescentando-se significado aos dados.
c) a informação é obtida a partir do conceito de conhecimento.
d) o processo de tomada de decisão em um sistema de informação tem por base apenas os dados brutos.
e) os dados consistem do conhecimento analisado sob diferentes pontos de vista.

21. **(CESPE / Polícia Federal – 2018)** A informação se caracteriza pela compreensão e internalização do conteúdo recebido, por meio do seu uso em nossas ações; o dado, por sua vez, é um elemento bruto dotado apenas de significado e relevância que visem fornecer uma solução para determinada situação de decisão.

22. **(CESPE / Polícia Federal – 2018)** Informação é constituída por um conjunto de dados com características específicas. O ponto de análise é que os dados devem ser irrelevantes para o sistema a que se destinam.

23. **(CESPE / TCE/SC – 2018)** Define-se informação como significado, ou seja, como registros icônicos e simbólicos — fonéticos ou numéricos — e signos — linguísticos, lógicos ou matemáticos —, por meio dos quais se representam atos, conceitos ou instruções.

24. **(CESPE / Polícia Federal – 2018)** O conceito de conhecimento é mais complexo que o de informação, pois conhecimento pressupõe um processo de compreensão e internalização das informações recebidas, possivelmente combinando-as.

25. **(CESPE / EMAP – 2018)** No que se refere à espiral do conhecimento, socialização, externalização, modelagem e feedback são os diferentes modos de conversão do conhecimento.

26. **(CESPE / EMAP – 2018)** Em uma instituição, o conhecimento explícito está relacionado ao capital intangível.

27. **(CESPE / EMAP – 2018)** Na solução de problemas, o conhecimento tácito é associado ao conhecimento do expert.

28. **(CESPE / TCE/PE – 2017)** A informação caracteriza-se por ser frequentemente tácita, bem como por ser de estruturação e captura difíceis em máquinas.

29. **(CESPE / TCE-PA – 2016)** O conhecimento tácito é pessoal, específico ao contexto e de difícil formulação e comunicação; o conhecimento explícito, por sua vez, é transmitido em linguagem formal e sistemática, propiciando ao indivíduo a apreensão de fatos.

30. **(IESES / BAHIAGÁS – 2016)** De acordo com Marcos Magalhães e Rafael Sampaio, página 81, "a tecnologia disponível, os sistemas, a miríade de programas e aplicativos hoje existentes alimentam o Sistema de Informações e Inteligência de Marketing (SIM) das organizações de qualquer porte com uma avalanche de informações provenientes de fontes variadas. Se, antes, o desafio era 'obter informação', - o que podia ser sistematizado de modo relativamente fácil –, a questão que se coloca hoje é conseguir organizar e analisar uma imensa quantidade de dados que chegam de todos os lados. Ou seja, o problema agora é 'usar a informação'. Para entender melhor essa questão, é necessário adotar critérios para a compreensão de algumas definições e exemplos". Para isto, é preciso distinguir a diferença entre Dados, Informação e Conhecimento. Desta forma, é correto afirmar:

a) Conhecimento: são sequências de textos, fotos, figuras ou sons que podem ser manipulados e descritos.

b) Informação relaciona-se à prática, à ação. Por exemplo: 'A temperatura no interior do Rio Grande do Sul é de 18º Celsius e, nesta época do ano, deve provocar um aumento da colheita de soja naquela região'.

c) Dados são sequências de símbolos (letras ou números), textos, fotos, figuras ou sons que podem ser descritos, armazenados e manipulados. Por exemplo: '18º Celsius'.

d) Conhecimento são dados contextualizados que também podem ser armazenados e manipulados.

e) Representa um exemplo de 'dados': 'A temperatura média deste mês, no interior do Rio Grande do Sul, é de 18º Celsius'.

31. **(FCC / CNMP– 2015)** Os Sistemas de Informação (SI) são construídos com Dados, Informação, Conhecimento e Inteligência. Sobre o tema, considere:

I. Informação é coletada nos ambientes interno e externo e representa, por exemplo: fatos, textos, gráficos.

II. A inteligência é realizada por meio de síntese, baseada em experiência e intuição, sendo uma habilidade humana.

III. Conhecimento demanda análise e avaliação sobre a confiabilidade, relevância e importância de dados e informações para a construção de um quadro de situação.

Está correto o que consta APENAS em:

a) I e III.
b) II.
c) II e III.
d) I.
e) III.

32. **(CESPE / TC-DF – 2014)** O conhecimento tácito é fruto de aprendizado e experiência de vida e é disseminado de maneira formalizada e declarada por meio de artigos e livros.

33. **(COPEVE-UFAL / UFAL – 2014)** Qual a denominação de um conjunto de dados devidamente ordenados e organizados de forma a terem significado?

a) Dado
b) Processamento
c) Informação
d) Memória
e) Cluster

34. **(CESPE / MPU – 2013)** Considerando-se a espiral do conhecimento, é correto afirmar que situações em que os colaboradores de uma organização convertem o conhecimento explícito em tácito correspondem à etapa denominada combinação.

35. **(CESPE / MPU – 2013)** Desejando externalizar o conhecimento tácito que determinado colaborador detenha sobre a elaboração de estudos técnicos, por exemplo, a organização deve documentar esse conhecimento de modo que seja possível a outros colaboradores reproduzi-lo facilmente. Nesse tipo de situação, observa-se a conversão do conhecimento tácito em explícito.

36. **(CESPE / FUB – 2013)** Conhecimentos estruturados encontram-se materializados em bens materiais tangíveis.

37. **(FUMARC / PC-MG – 2013)** Sobre conceitos e definições de dados e informações no contexto de Sistemas de Informação, analise as seguintes afirmativas:

I. Dados se referem a uma descrição elementar de coisas, eventos, atividades e transações que são registrados, classificados e armazenados, mas não são organizados para transmitir significados específicos.

II. Informação se refere a dados que foram organizados para terem significado e valor para o receptor, que interpreta o significado e elabora conclusões sobre a informação.

III. Conhecimento consiste em informações organizadas para transmitir experiência, aprendizagem e prática aplicadas a um problema ou atividade empresarial.

Estão CORRETAS as afirmativas:

a) I, II e III.
b) I e II, apenas.
c) I e III, apenas.
d) II e III, apenas.

38. **(CESPE / ANAC – 2012)** Os modos de conversão do conhecimento tácito em explícito referem-se à socialização e à combinação.

39. **(CESPE / MEC – 2011)** Considere que, em uma organização, os profissionais capazes de criar maquetes de prédios mostrem a seus aprendizes que, por meio da observação e imitação, eles obterão o conhecimento do assunto. Nessa situação, o processo de construção de conhecimento é realizado do conhecimento tácito para o explícito.

40. **(CESPE / IFB – 2011)** A transformação de informação em conhecimento envolve a realização de comparações e contrastes, a identificação de relacionamentos e a inferência de consequências e, portanto, é influenciada por especialização, experiência, valores e insights estruturados dos integrantes da equipe.

41. **(CESPE / EBC – 2011)** Ocorre internalização quando parte do conhecimento tácito de uma pessoa converte-se no conhecimento tácito de outrem, tal como ocorre na realização de atividades práticas sob a supervisão de tutores.

42. **(CESPE / EBC – 2011)** A socialização é a conversão de partes do conhecimento explícito da organização em conhecimento tácito do indivíduo.

43. **(CESPE / STM – 2011)** O modelo SECI (sensibilização, externalização, combinação e informação) representa modos de criação de conhecimento organizacional.

44. **(FCC / TCE/RS – 2011)** Com relação aos conceitos de dado e informação assinale a afirmação correta.

a) Uma lista com a quantidade de produtos vendidos é uma informação.
b) Um relatório sobre a produtividade dos funcionários é um dado.
c) Produtividade da mão de obra de uma empresa é um dado.
d) Custo da matéria prima é uma informação.
e) A capacidade de produção é uma informação.

45. **(CESPE / DPU – 2010)** Dados acerca de determinado assunto de interesse de uma ou mais pessoas, que possam ser interpretados ou tenham significado para o receptor, são considerados:

a) aprendizagem organizacional.
b) conhecimento.
c) capital intelectual.
d) informação.
e) sistema de gerenciamento.

46. **(CESPE / CGE-PB – 2008)** Informação é um conjunto de dados contextualizados que visam fornecer uma solução para determinada situação de decisão.

GABARITO

1. LETRA C
2. CORRETO
3. CORRETO
4. ERRADO
5. CORRETO
6. LETRA B
7. ERRADO
8. CORRETO
9. CORRETO
10. CORRETO
11. CORRETO
12. ERRADO
13. CORRETO
14. ERRAD
15. ERRADO
16. ERRADO
17. CORRETO
18. CORRETO
19. CORRETO
20. LETRA B
21. ERRADO
22. ERRADO
23. ERRADO
24. CORRETO
25. ERRADO
26. ERRADO
27. CORRETO
28. ERRADO
29. CORRETO
30. LETRA C
31. LETRA C
32. ERRADO
33. LETRA C
34. ERRADO
35. CORRETO
36. CORRETO
37. LETRA A
38. ERRADO
39. ERRADO
40. CORRETO
41. ERRADO
42. ERRADO
43. ERRADO
44. LETRA E
45. LETRA D
46. CORRETO

LISTA DE QUESTÕES – DIVERSAS BANCAS

1. **(CESPE / MPE-RO – 2023)** Em desenvolvimento de sistema de computação, os dados que oferecem a opção de serem utilizados livremente, reutilizados e redistribuídos são chamados de

a) dados móveis.
b) inteligência.
c) conhecimento.
d) dados estruturados.
e) dados abertos.

2. **(FGV / TCE-TO – 2022)** Dados abertos são dados publicados na internet, em formatos como csv e xml, para que qualquer pessoa possa acessar, modificar, analisar e compartilhar.

Os dados devem ser publicados na forma coletada da fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada, para atender ao princípio de dados abertos:

a) primários;
b) acessíveis;
c) completos;
d) processáveis por máquina;
e) acesso não discriminatório.

3. **(FGV / SEFAZ-AM – 2022)** Com relação às características de dados abertos, analise as afirmativas a seguir.

I. Reuso e redistribuição: os dados precisam estar disponíveis integralmente, devendo estar em um formato conveniente e modificável e sob custo não maior que um custo razoável de reprodução.

II. Disponibilidade e acesso: os dados precisam ser fornecidos em condições de reuso e redistribuição, podendo ser recombinados com outros dados.

III. Participação universal: todos os usuários podem usar, reusar e redistribuir o dado sem restrições de áreas de atuação, pessoas ou grupos.

Está correto o que se afirma em:

a) I, apenas.
b) II, apenas.
c) III, apenas.
d) I e II, apenas.

e) II e III, apenas.

4. **(CESPE / ISS-Aracaju – 2021)** Em decorrência do princípio de primariedade dos dados abertos governamentais, tais dados devem ser:

a) publicados na forma coletada na fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada.
b) razoavelmente estruturados para possibilitar o seu processamento automatizado.
c) disponibilizados em um formato cujo controle não seja exclusivo de nenhum ente.
d) disponibilizados o quão rapidamente seja necessário para preservar o seu valor.
e) disponibilizados para o público mais amplo possível e para os propósitos mais variados possíveis.

5. **(CESPE / SEFAZ-AL – 2021)** De acordo com os princípios de dados abertos, um dado disponível deve ser reutilizado, independentemente de estar disponível em formato aberto ou legível por máquina.

6. **(CESPE / SEFAZ-AL – 2021)** A publicação e a catalogação de dados abertos demandam o uso de uma infraestrutura de software conhecida como ODP (Open Data Platform), ou plataforma de dados abertos, que contém componentes e interfaces.

7. **(FEPESE / CIASC – 2017)** Com relação ao assunto Dados Abertos, assinale a alternativa correta.

a) Um requisito importante para que o dado seja considerado aberto é o da reutilização e redistribuição no qual os dados devem estar disponíveis como um todo e sob custo razoável de reprodução. Tais dados preferencialmente devem ser possíveis de serem baixados pela internet e devem estar disponíveis de uma forma conveniente e modificável.

b) Um requisito importante para que o dado seja considerado aberto é o da disponibilidade e acesso no qual os dados devem ser fornecidos sob termos que permitam o seu fechamento, abertura ou a combinação com outros conjuntos de dados.

c) Dados abertos são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa. Estão sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras.

d) Participação Universal significa a capacidade de diversos sistemas e organizações trabalharem juntos. Nesse caso, trata-se da capacidade de interoperar ou combinar diferentes conjuntos de dados.

e) Entende-se Interoperabilidade como a capacidade de usar, reutilizar e redistribuir, não devendo haver discriminação contra áreas de atuação ou contra pessoas ou grupos. Por

exemplo, restrições de uso não comercial que impediriam o uso comercial, ou restrições de uso para certos fins que excluem determinados dados do conceito de 'abertos'.

8. **(FEPESE / CIASC – 2017)** De acordo com a Política de Dados Abertos do Poder Executivo federal, os dados devem ser representados em meio digital e:

a) estruturados em formato aberto, disponibilizados em formato requerido pelo cidadão, de forma que permita utilização, consumo ou cruzamento de acordo com as normas da instituição detentora da informação.

b) estruturados em formato aberto, processáveis por máquina, disponibilizados sob licença aberta, de modo que permita livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte.

c) disponibilizados na Internet, processáveis por máquina, estruturados de forma que permita utilização, consumo ou cruzamento, de acordo com as normas da instituição detentora da informação.

d) disponibilizados na Internet, estruturados sob licença da instituição pública detentora da informação, de modo que permita utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte.

9. **(CESPE / TCE-SC – 2016)** Dados abertos são os dados de livre utilização, reutilização e redistribuição, exigindo-se, no máximo, créditos à autoria e compartilhamento pela mesma licença.

10. **(CESPE / TCE/RJ – 2021)** A filosofia de uso de dados abertos orienta que os dados sejam reutilizados amplamente, no entanto não é permitida a combinação de um data set aberto com conjuntos de dados não abertos.

11. **(CESPE / TCE/RJ – 2021)** As três normas fundamentais que compõem o conceito de dados abertos são: disponibilidade e acesso; reuso e distribuição; e participação universal.

12. **(CESPE / ME – 2020)** Um dos princípios de dados abertos é o acesso não discriminatório, ou seja, os dados estão disponíveis independentemente de identificação ou registro do usuário.

13. **(CESPE / ME – 2020)** A abertura de dados por organizações governamentais é uma prerrogativa das próprias organizações, que devem ditar normas internas sobre o assunto.

14. **(CCV / UFC – 2013)** A Instrução normativa Nº 4, de 12 de abril de 2012, instituiu a INDA, a qual disponibiliza um sítio eletrônico de referência para a busca e o acesso aos dados públicos, seus metadados, informações, aplicativos e serviços relacionados. Selecione a alternativa que descreve o significado da sigla INDA e o respectivo sítio eletrônico mencionado anteriormente:

a) Infraestrutura Nacional de Dados Acessíveis; Portal da Transparência.
b) Infraestrutura Nacional de Dados Abertos; Portal da Transparência.
c) Infraestrutura Nacional de Dados Abertos; Portal Brasileiro de Dados Abertos.
d) Infraestrutura Nacional de Dados Acessíveis; Portal Brasileiro de Dados Abertos.
e) Infraestrutura Nacional de Dados Acessíveis; Portal do Acesso à Informação Pública.

15. **(COVEST / UFPE – 2019)** Quanto aos princípios de dados abertos governamentais, quando se fala em "não discriminatório", está se afirmando que os dados:

a) devem ser publicados na forma coletada na fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada.

b) devem estar disponíveis o quão rapidamente seja necessário para preservar o seu valor.

c) devem estar disponíveis com o objetivo de aumentar a participação de grupos de ativistas de dados cívicos.

d) devem estar disponíveis para todos, sem que seja necessária identificação ou registro.

e) devem estar razoavelmente estruturados para possibilitar o seu processamento automatizado.

16. **(CESPE / IPHAN – 2018)** O indicador barômetro de dados abertos foi escolhido para medir o impacto das iniciativas dos dados abertos no Brasil.

17. **(CESPE / TCE-RJ – 2021)** O portal brasileiro de dados abertos (*www.dados.gov.br*) disponibiliza diversos conjuntos de dados em formatos PDF, CSV, HTML e JSON.

GABARITO – DIVERSAS BANCAS

1. LETRA E
2. LETRA A
3. LETRA C
4. LETRA A
5. ERRADO
6. CORRETO
7. LETRA C
8. LETRA B
9. CORRETO
10. ERRADO
11. CORRETO
12. CORRETO
13. ERRADO
14. LETRA C
15. LETRA D
16. CORRETO
17. CORRETO

LISTA DE QUESTÕES – DIVERSAS BANCAS

1. **(CESPE / Polícia Federal – 2013)** Título, assunto, palavras-chave e comentários de um documento são metadados típicos presentes em um documento produzido por processadores de texto como o BrOffice e o Microsoft Office.

2. **(CESPE / ABIN – 2018)** Os metadados são dados utilizados para a criação e o uso de documentos em sistemas informatizados indisponíveis aos usuários.

3. **(CESPE / TRT-ES – 2013)** Os dados estruturados que descrevem e permitem encontrar, gerenciar, compreender e (ou) preservar documentos arquivísticos ao longo do tempo são conhecidos como metadados.

4. **(CESPE / ANTAQ – 2009)** Metadados são dados estruturados e codificados de modo a descreverem características de entidades para auxiliarem na identificação, na descoberta e no gerenciamento das entidades descritas.

5. **(CESPE / MPE-CE – 2020)** Os metadados descrevem, explicam, localizam e facilitam a recuperação de um recurso informacional, permitindo que esse recurso esteja acessível futuramente.

6. **(FUNPAR / UFPR – 2010)** Em se tratando de gerenciamento da informação, dados estruturados que descrevem, identificam, explicam, localizam e, portanto, facilitam a recuperação, uso e gestão de recursos de informação, são chamados de:

   a) informação.
   b) sistema de informação.
   c) conhecimento.
   d) metadados.
   e) dicionário de dados.

7. **(CESPE / TCE-RJ – 2021)** Dublin Core é um esquema de metadados que auxilia na descrição de objetos digitais por meio da definição de diversos elementos de metadados, entre os quais se incluem título, autor, assunto, formato e fonte.

8. **(CESPE / Polícia Federal – 2018)** Em arquivos no formato XML, as tags não são consideradas metadados.

9. **(CESPE / TCE-RJ – 2021)** Os metadados sintáticos, estruturais e semânticos descrevem, respectivamente, a sintaxe, a estrutura e o conteúdo dos dados.

10. **(CESPE/ CNPQ – 2011)** As propriedades de um documento, também conhecidas como metadados, são detalhes de um arquivo que o descrevem ou o identificam. As propriedades incluem detalhes como título, nome do autor, assunto e palavras-chave que identificam o tópico ou o conteúdo do documento.

11. **(COPEVE / MPE-AL – 2012)** São elementos Metadados Dublin Core:

a) Título; Criador; Assunto; Descrição; Data; e Identificador.
b) Título; Responsabilidade; Assunto; Descrição; endereço web; e Tipo.
c) Criador; Assunto; Descrição; Data; Formato; e Notas.
d) Publicador; Contribuidor; Data; Formato; e série.
e) Autor; Assunto; Descrição; Publicador; Formato; e Etiqueta.

12. **(EXATUS / DETRAN-RJ – 2012)** Os metadados podem ser definidos como:

a) Elementos de descrição/definição/avaliação de recursos informacionais armazenados em sistemas computadorizados e organizados por padrões específicos, de forma estruturada.

b) Conjunto de técnicas, práticas e procedimentos que estabelecem os fundamentos técnicos e metodológicos de um sistema digital.

c) Forma de estruturação da informação que permite a leitura não linear de um texto.

e) Sistema para identificar univocamente objetos digitais na WEB, tem como propósito principal a gestão da propriedade intelectual e o comércio digital dos objetos.

13. **(UEAP / CS-UFG – 2014)** Metadados são dados sobre outros dados. No contexto de bancos de dados, metadados são usados para:

a) descrever a estrutura do banco de dados.
b) descrever o conteúdo do log do sistema.
c) resumir o conteúdo dos dados de cada relação.
d) resumir a estrutura dos programas de aplicação.

14. **(FGV / Fiocruz – 2010)** Sobre os Metadados, assinale a afirmativa correta.

a) São um conjunto de informações estruturadas que descrevem os dados (bases) que representam, facilitando ao usuário o gerenciamento desses dados.
b) São o conjunto total de dados que integram um projeto de geoprocessamento.
c) São dados preliminares que necessitam de averiguação e eventuais correções para, posteriormente, serem utilizados.
d) São o conjunto de dados que já passaram por todas as averiguações e já estão prontos para utilização.
e) São dados que não foram utilizados e são arquivados para futuros trabalhos.

15. **(BIO-RIO / IF-RJ – 2015)** Os metadados que facilitam a navegação e a apresentação de recursos eletrônicos são os:

a) descritivos.
b) estruturais.
c) administrativos.
d) operacionais.
e) interoperáveis.

16. **(CESPE / FUB – 2015)** A informação que descreve a estrutura dos dados e sua relação com outros dados é o metadado.

17. **(VUNESP / UNIFAI – 2019)** No contexto da Ciência da Informação, os metadados são classificados, por função, em:

a) metadados modulares, metadados extensos e metadados refinados.
b) metadados descritivos, metadados estruturais e metadados administrativos.
c) metadados multilíngues, metadados aplicados e metadados principais.
d) metadados estruturais, metadados completos e metadados plurilíngues.
e) metadados teóricos, metadados aplicados e metadados administrativos.

18. **(COPEVE / MPE-AL – 2012)** Metadados são:

a) dados sobre dados.
b) dados disponibilizados em redes sociais.
c) dados disponibilizados em redes cooperativas.
d) informações disponíveis na Internet.
e) dados e informações disponíveis na intranet.

19. **(CESPE / Polícia Federal – 2018)** Na gestão de documentos digitalizados, os arquivos são criados em formato de imagens, e seus atributos são armazenados em metadados para facilitar a pesquisa e a localização desses arquivos.

20. **(CESPE / Polícia Federal – 2021)** A função do metadado de arquivo é descrever o destino final do arquivo definido pelo emissor da mensagem e proprietário do arquivo.

GABARITO – DIVERSAS BANCAS

1. CORRETO
2. ERRADO
3. CORRETO
4. CORRETO
5. CORRETO
6. LETRA D
7. CORRETO
8. ERRADO
9. CORRETO
10. CORRETO
11. LETRA A
12. LETRA A
13. LETRA A
14. LETRA A
15. LETRA B
16. CORRETO
17. LETRA B
18. LETRA A
19. CORRETO
20. ERRADO

LISTA DE QUESTÕES– DIVERSAS BANCAS

1. **(CESPE / SEPLAN-RR - 2023)** Com base nos conceitos de dados estruturados e não estruturados, de análise exploratória de dados e de dados abertos, julgue o item seguinte.

   Dados de redes sociais são exemplos de dados não estruturados.

2. **(FGV / EPPGG - 2023)** Um analista de políticas públicas trabalha com diversos formatos e tipos de dados. Identificar corretamente essas características no começo das atividades é essencial para realizar a escolha correta das ferramentas e técnicas que serão utilizadas.

   Assinale a alternativa que traz apenas dados semiestruturados.

   a) Arquivo CSV (Comma Separated Values) e arquivos de áudio.
   b) Arquivo JSON (JavaScript Object Notation) e arquivos de imagens.
   c) Arquivo JSON (JavaScript Object Notation) e arquivo XML (eXtensible Markup Language).
   d) Arquivos PDF (Portable Document Format) e arquivos de vídeo.
   e) Tabela de um banco de dados relacional e arquivo XML (eXtensible Markup Language).

3. **(CESPE / DATAPREV – 2023)** Arquivos .xml, .csv e .json são exemplos de dados estruturados.

4. **(CONSULPLAN / MPE-PA – 2022)** Dados são um conjunto de fatos em estado bruto que podem ser utilizados para tomadas de decisão. Os dados podem ser divididos entre estruturados e não-estruturados. Sobre os dados estruturados, assinale a afirmativa correta.

   a) Mensagens de e-mail são dados estruturados.
   b) Dados estruturados são organizados em linhas e colunas.
   c) Dados estruturados são armazenados em arquivos no formato ".TXT".
   d) Mecanismos de busca possuem dificuldade para pesquisar dados estruturados.

5. **(CONSULPLAN / MPE-PA – 2022)** Dados são um conjunto de fatos em estado bruto que podem ser utilizados para tomadas de decisão. Os dados podem ser divididos entre estruturados e não-estruturados. Sobre os dados estruturados, assinale a afirmativa correta.

   a) Mensagens de e-mail são dados estruturados.
   b) Dados estruturados são organizados em linhas e colunas.
   c) Dados estruturados são armazenados em arquivos no formato ".TXT".
   d) Mecanismos de busca possuem dificuldade para pesquisar dados estruturados.

6. **(CESPE / Polícia Federal – 2021)** Os dados estruturados diferenciam-se dos dados não estruturados pela rigidez em seu formato e pelo fato de poderem ser armazenados em campos de tabelas de um banco de dados relacional.

7. **(CESPE / SERPRO – 2021)** Os dados não estruturados constam de bancos de dados relacionais que são eminentemente pesquisáveis tanto por meio de consultas realizadas por humanos quanto por meio daquelas realizadas por algoritmos que usam tipos de dados e nomes de campos, como alfabéticos ou numéricos, moeda ou data.

8. **(CESPE/ ME – 2020)** Embora com características particulares, dados não estruturados podem ser classificados em sua totalidade, assim como os dados estruturados.

9. **(FAURGS/ TJ-RS – 2018)** Os dados _____ são armazenados em _____. Os dados _____ misturam nomes de tipos de dados e valores de dados, mas nem todos eles precisam seguir uma estrutura predefinida fixa. Os dados _____ referem-se à informação exibida na Web, caso em que a informação sobre os tipos dos itens de dados não existe.

   Assinale a alternativa que completa, correta e respectivamente, as lacunas do texto acima.

   a) semiestruturados / bancos de dados tradicionais / estruturados / não estruturados
   b) não estruturados / XML / semiestruturados /estruturados
   c) estruturados / HTML / não estruturados / semiestruturados
   d) estruturados / bancos de dados tradicionais / semiestruturados / não estruturados
   e) semiestruturados / Web / estruturados / não estruturados

10. **(CESPE / Polícia Federal – 2018)**

    CPF
    NOME
    DATA DE NASCIMENTO
    NOME DO PAI
    NOME DA MAE
    TELEFONE
    CEP
    NUMERO

    As informações anteriormente apresentadas correspondem aos campos de uma tabela de um banco de dados, a qual é acessada por mais de um sistema de informação e também por outras tabelas. Esses dados são utilizados para simples cadastros, desde a consulta até sua alteração, e também para prevenção à fraude, por meio de verificação dos dados da tabela e de outros dados em diferentes bases de dados ou outros meios de informação.

    Considerando essas informações, julgue o item que se segue.

    Os dados armazenados na referida tabela são considerados não estruturados.

11. **(CESPE / TCE-SC – 2016)** Em se tratando de dados estruturados, a informação de esquema está mesclada aos valores dos dados, e cada objeto de dados pode ter atributos diferentes, que não são conhecidos com antecedência. Essa característica os diferencia de dados não estruturados.

12. **(CESPE / TCE-PA – 2016)** Em comparação aos dados não estruturados, os dados estruturados demandam mais espaço de armazenamento e um gerenciamento mais cauteloso, uma vez que constituem a maior parte dos dados corporativos.

13. **(FEPESE / SJC-SC – 2013)** Qual software, dentre os listados abaixo, permite a importação (e atualização) de dados estruturados de bancos de dados relacionais para posterior manipulação pelo aplicativo?

a) Word
b) Excel
c) Powerpoint
d) Internet Explorer
e) Outlook

14. **(CESPE/ TRT-ES – 2013)** Os dados estruturados que descrevem e permitem encontrar, gerenciar, compreender e (ou) preservar documentos arquivísticos ao longo do tempo são conhecidos como metadados.

GABARITO

1. CORRETO
2. LETRA C
3. ERRADO
4. LETRA B
5. LETRA B
6. CORRETO
7. ERRADO
8. ERRADO
9. LETRA D
10. ERRADO
11. ERRADO
12. ERRADO
13. LETRA B
14. CORRETO

LISTA DE QUESTÕES – DIVERSAS BANCAS

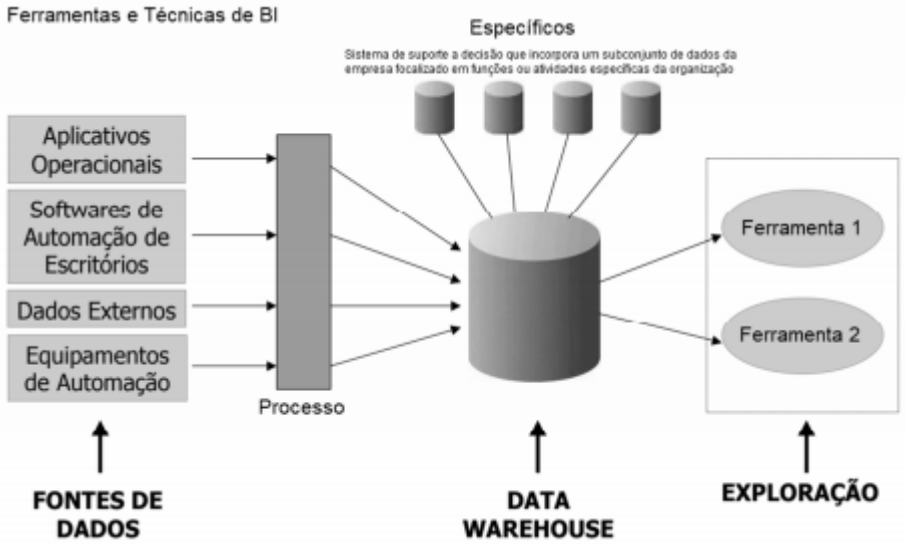
1. **(CESPE /TJ-PA – 2020)** Assinale a opção que indica um processo de extração e transformação de dados em um data Warehouse:

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

2. **(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.

3. **(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.

4. **(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.

5. **(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.

6. **(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.

7. **(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.

8. **(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

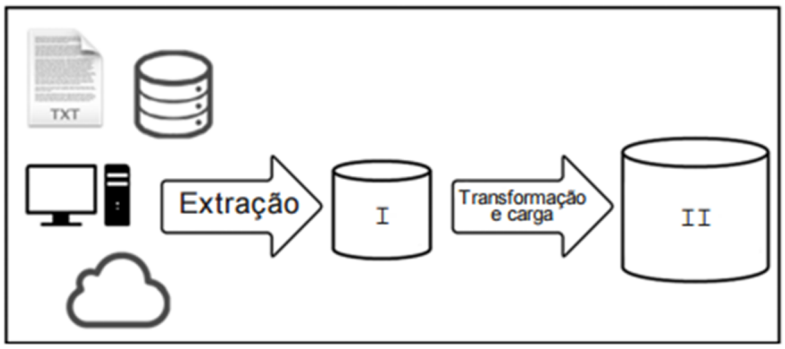
9. **(IBFC / TJ - PE – 2017)** Numa aplicação que necessita de *Data Warehouse* uma das fases mais críticas é a forma pela qual os dados são efetivamente carregados ou introduzidos em um Data Warehouse. As ferramentas de software cuja função é apoiar essa atividade são conhecidas pela sigla em inglês:

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

10. **(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.

11. **(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.

12. **(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.

13. **(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.

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

15. **(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.

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

17. **(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.

18. **(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.

19. **(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.

20. **(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.

21. **(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.

22. **(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.

23. **(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.

GABARITO – DIVERSAS BANCAS

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

ESSA LEI TODO MUNDO CONHECE:
PIRATARIA É CRIME.

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

1.

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

2.

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

3.

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

4.

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

5.

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

6.

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

7.

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

8.

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

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