



Aula 00

Análise de Dados e Informações para
TCE/RJ (Conhecimentos Básicos)

Prof. Arthur Mendonça

Sumário

SUMÁRIO	2
APRESENTAÇÃO	3
O PROFESSOR.....	3
NOSSO CURSO.....	3
TEORIA DA AULA	6
DADO, INFORMAÇÃO, CONHECIMENTO E INTELIGÊNCIA	6
<i>Dado</i>	6
<i>Informação</i>	7
<i>Conhecimento</i>	9
<i>Inteligência</i>	12
DADOS ESTRUTURADOS E NÃO ESTRUTURADOS	16
<i>Dados Estruturados</i>	16
<i>Dados Não Estruturados</i>	17
<i>Dados Semiestruturados</i>	18
DADOS ABERTOS	20
<i>Três Leis</i>	21
<i>Oito Princípios</i>	22
COLETA, TRATAMENTO, ARMAZENAMENTO, INTEGRAÇÃO E RECUPERAÇÃO DE DADOS	24
<i>ETL</i>	25
<i>Integração</i>	27
<i>Recuperação</i>	27
QUESTÕES COMENTADAS PELO PROFESSOR.....	29
LISTA DE QUESTÕES COMENTADAS.....	42
GABARITO	49
RESUMO DIRECIONADO	50
BIBLIOGRAFIA.....	53

Atenção! Se você irá prestar o concurso do TCE-RJ para as áreas de **Controle Externo ou Tecnologia da Informação**, recomendo que ignore este curso de conhecimentos básicos e vá direto às aulas presentes no nosso curso de **conhecimentos específicos**. Além dos assuntos presentes neste curso básico abordados de uma forma um pouco mais aprofundada, ele também conta com o conteúdo específico para o seu cargo.

Apresentação

O Professor



Olá concurseiro, tudo bem? Meu nome é **Arthur Mendonça** e é com grande satisfação que serei o seu professor de Análise de Informações neste curso preparatório para o **TCE-RJ**. Antes de introduzir a matéria, gostaria de falar um pouco sobre a minha trajetória no mundo dos concursos.

Sou recifense e me formei em Ciência da Computação pela UFPE em 2015. Já durante a graduação, cursei algumas matérias a respeito do assunto, comecei a estagiar com Bancos de Dados e *Business Intelligence* (BI) e terminei me apaixonando pela área! Reconheço que, à primeira vista, o assunto relacionado a essa minha área possa parecer meio árido, mas a ideia aqui é apresentá-lo da forma mais didática possível para que você seja capaz de resolver a maioria das questões de prova que virão por aí.

Seguindo a minha carreira, trabalhei durante algum tempo na iniciativa privada prestando serviços para diversos clientes e ganhando experiência e conhecimento, mas o serviço público acabou me conquistando. Soube que o edital do **TCE-PE** estava próximo, fiz meu pé de meia e pedi demissão para me preparar.

Foram meses de muito estudo e concentração, muita abdicção das saídas no final de semana, dos programas com os amigos e com a namorada e de tudo aquilo que você já deve estar cansado de saber (tamo junto!). Aproveitei as matérias que estava estudando e também me inscrevi para Analista Administrativo no **TJ-PE**, cuja prova seria um mês depois do certame do TCE.

Passaram as provas e saí confiante de que tinha feito o meu melhor. Talvez não desse pra passar, mas saí feliz pelo que tinha realizado até ali.

Aí veio o resultado e eu nem acreditei. **1º lugar para Auditor de Controle Externo do TCE-PE!** É impossível descrever em palavras a satisfação de ter seu esforço recompensado, foi uma felicidade muito grande que confesso que ainda sinto um pouquinho todo dia quando chego no trabalho. Um tempo depois chega outra notícia: **1º lugar para Analista Administrativo no TJ-PE!** Sinceramente, até hoje nem acredito que deu tudo tão certo.

Passou o tempo e chegaram minha nomeação e posse no TCE-PE. Olha, posso dizer o trabalho é muito gratificante! Fiscalizar os recursos públicos num país com tantas dificuldades como é o nosso é uma tarefa importantíssima. Espero que você esteja bastante motivado para trabalhar na área de controle externo e ajudar a fazer um País melhor.

Estude com confiança e dedicação que tenho certeza que a sua hora também vai chegar!

Agora que você já me conhece um pouquinho, vamos falar um pouco a respeito da matéria que estudaremos:

Nosso Curso

Análise de Informações é uma matéria que começou a ser cobrada em provas organizadas pelo **CESPE** a partir do concurso do **TCU** em 2015. Desde então, a matéria tem aparecido em diversos certames de Tribunais de Contas Estaduais, como o **TCE-PE** (2017), **TCE-PB** (2018), **TCM-BA** (2018) e mais recentemente o **TCE-RO** (2019).

Lembro que houve meio que um desespero nos grupos do concurso do TCE-PE que eu participava durante a minha preparação quando essa disciplina apareceu no edital. Espero que não seja o seu caso, mas, se for, não se preocupe! A matéria não é tão extensa e a maior parte dos assuntos não é tão técnica, então, mesmo que você tenha dificuldade com TI, vai ter plenas condições de adquirir um bom domínio do assunto e ter um bom desempenho na prova.

Claro que isso vai requerer um esforço da sua parte. Se não fosse assim, iria faltar vaga pra tanto aprovado. Enfim, acredito que Análise de Informações veio pra ficar, pois tem tudo a ver com a atividade de auditoria. Conhecer os dados, suas estruturas de armazenamento e técnicas de extração e análise é algo fundamental para a atividade de auditoria nos tempos modernos.

A matéria não é extensa, mas requer bastante atenção aos detalhes para que você não derrape na hora da prova. Nosso curso da parte de **conhecimentos básicos** estará estruturado da seguinte maneira:

Número da aula	Data de disponibilização	Assunto da aula
00	24/01/2019	1 Dado, informação, conhecimento e inteligência. 1.1 Dados estruturados e não estruturados. 1.2 Dados abertos. 1.3 Coleta, tratamento, armazenamento, integração e recuperação de dados.
01	26/01/2019	2 Bancos de dados relacionais. 2.1 Conceitos básicos e características. 2.2 Metadados. 2.3 Tabelas, visões (views) e índices. 2.4 Chaves e relacionamentos.
	26/01/2019	Teste a sua direção
02	28/01/2019	2 Bancos de dados relacionais. 2.1 Conceitos básicos e características. 2.2 Metadados. 2.3 Tabelas, visões (views) e índices. 2.4 Chaves e relacionamentos. (continuação)
03	30/01/2019	3 Noções de modelagem dimensional. 3.1 Conceito e aplicações.
	30/01/2019	Teste a sua direção

Além das **aulas em PDF contendo teoria e exercícios**, você também vai ter acesso ao nosso **curso completo em vídeo** e à ferramenta de **dúvidas**, onde estarei sempre disponível para responder as suas perguntas. A ideia é ser um material completo e totalmente focado no seu concurso, para seu foco ficar somente em estudar, sem perder tempo procurando material ou fazendo perguntas na internet.

Se ainda restou alguma dúvida a respeito da aquisição do curso, pode entrar em contato com nossa equipe de atendimento, que é bastante qualificada e tenho certeza que te ajudará.

* * *

E então? Pronto para começar? Na aula de hoje vamos abordar os seguintes tópicos do seu edital:

1 Dado, informação, conhecimento e inteligência. 1.1 Dados estruturados e não estruturados. 1.2 Dados abertos. 1.3 Coleta, tratamento, armazenamento, integração e recuperação de dados.

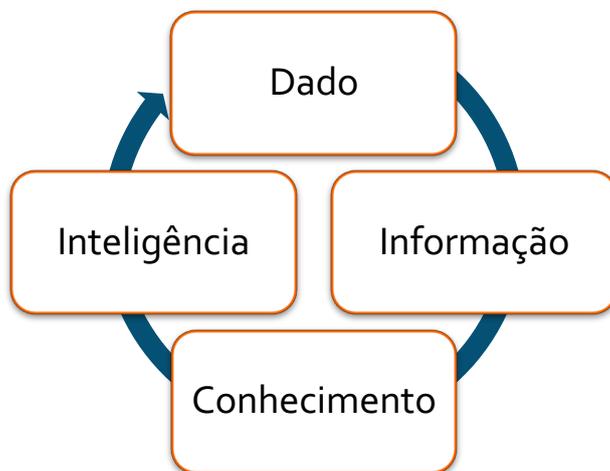
Bons estudos!



@analisedeinformacoes

Teoria da Aula

Dado, Informação, Conhecimento e Inteligência



Dado

Bom, caro aluno. Para começar, nada mais importante do que definir o que é o objeto de toda a nossa matéria, não é mesmo? Os **dados** são a base de tudo o que estudaremos.

Dado é nada mais nada menos que um **fato** a respeito do mundo. E sobre o mundo aqui quer dizer sobre praticamente qualquer coisa! Sua altura em metros? É um dado. A temperatura em °C do Polo Norte neste momento? Dado. A letra que corresponde ao gabarito de uma questão de concurso? Também é um dado.

Só que esses valores somente são dados se tomados isoladamente, de forma "crua". Ou seja, não há nenhum tipo de significado ou contexto associado a eles. Seguindo nossos exemplos acima, imagine que eu não dissesse do que se trata e só te apresentasse essa lista de números: "1,80", "-30", "C". Você iria entender alguma coisa de importante? Acho que não, né? Faltam **informações** para que possamos extrair algo de útil dessa lista.

Veja que o dado comumente assume a forma de um número ou de uma cadeia de caracteres, mas tenha em mente que da mesma maneira podemos considerar imagens, sons, etc. como dados. **O dado pode existir em qualquer formato.**

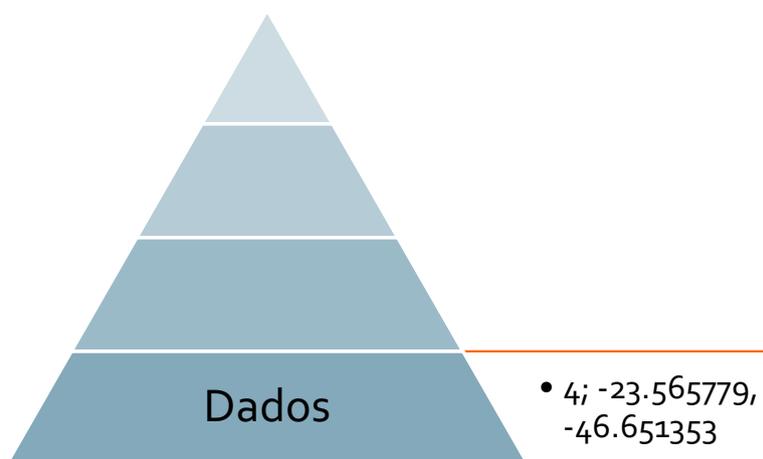
Atenção! *Dado é qualquer fato sobre o mundo apresentado em sua forma bruta*

Não precisa decorar essa definição ao pé da letra! Você pode definir de várias maneiras diferentes, o importante é entender o conceito. De fato, as bancas costumam apresentar as mais diversas definições, então não se prenda a formalismos.

Como o dado é dessa maneira, **simples** e **autocontido** (ele por si só não traz informações associadas), temos que ele apresenta algumas características interessantes para a computação: *dados são fáceis de estruturar, armazenar e quantificar*. Isso vai ficar mais claro ao longo da aula, não se preocupe!

Aqui vai um exemplo para que você possa entender melhor. O gráfico que vamos montar é conhecido como **Pirâmide do Conhecimento**. Para o nosso exemplo, vamos pensar em um aplicativo de GPS daqueles de celular. Acredito que você conheça o *Waze*, não é? Ele é um aplicativo de navegação que ajuda o motorista a fugir de engarrafamentos.

Então vamos lá, qual seria um exemplo de dado capturado pelo *Waze*? Imagine que você teve acesso aos dados que o aplicativo armazena internamente e encontrou o seguinte:



Isso não nos diz muita coisa. São apenas números soltos. Eles existem e podemos até especular no que consistem, mas sem contexto e significado não é possível afirmar nada.

Para que seja possível começar a extrair algo de interessante disso, é necessário apresentar o próximo conceito.

Informação

Bom, já temos o dado. Mas ficam várias perguntas: quem? Quando? Onde? O quê? A **informação** é o que vai nos ajudar a fazer sentido a partir dessas perguntas!

De acordo com McGee & Prusak (1995):

Informação são dados coletados, organizados, ordenados, aos quais são atribuídos significados e contexto.

Ou seja, os dados passam por um processo de enriquecimento, ganhando assim **qualidade**. Eles agora nos **informam** alguma coisa. Nesse sentido, há essa outra definição de informação que podemos apresentar. Esta foi criada por um austríaco chamado Peter Drucker (1988) e complementa a definição que já apresentamos:

Informação são os dados dotados de relevância e propósito.

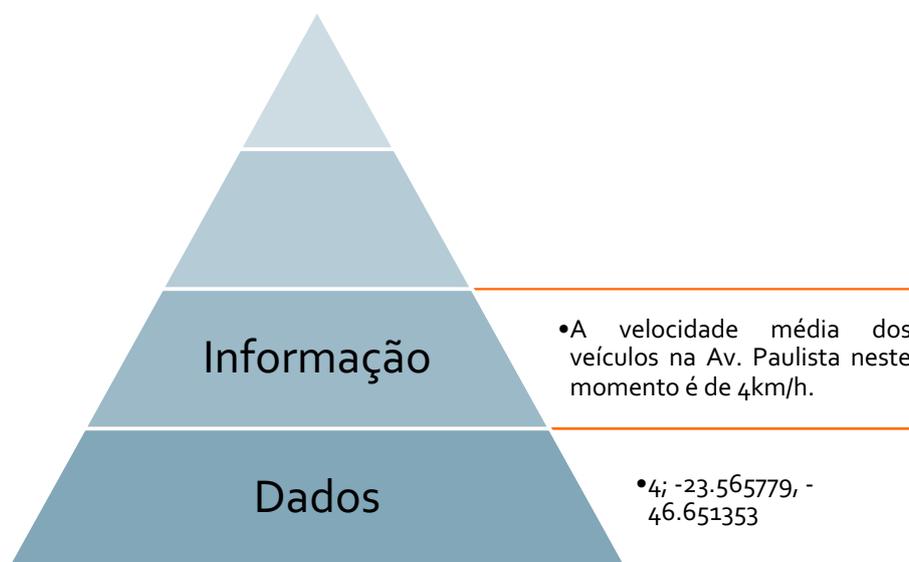
Bem simples, não é? Mas peraí, quem dota os dados de relevância e propósito? Uma máquina não é capaz de dizer isso, já que ela somente segue um programa predefinido, ela não tem vontades. Quem tem um objetivo para o uso da informação somos nós, os **seres humanos**. Assim, somente nós podemos dizer se a informação vai ser relevante para esses objetivos ou não. Grave aí: *a informação necessariamente precisa da intervenção humana!*

A informação exige um **consenso** a respeito do seu significado, já que eventuais ambiguidades e interpretações divergentes podem acabar por diminuir a sua relevância e a utilidade.

Tendo essas definições em mente, vamos processar os dados do nosso exemplo para obtermos uma informação:

Primeiro, apuramos que o número 4 se refere à velocidade média em determinada via, portanto, 4km/h. Também verificamos que os outros números se referem às coordenadas geográficas de um local no planeta, no nosso exemplo, um ponto bem no meio da Av. Paulista em São Paulo.

Assim, com o objetivo de encontrar o melhor caminho para o nosso destino, construímos a seguinte informação:



Ficou claro? A informação encapsula os dados e dá a eles elementos como contexto e significado. Agora não temos apenas um monte de números ou caracteres, mas sabemos também no que consistem aqueles valores.

Depois desse **processamento**, temos uma informação potencialmente útil que pode ajudar o aplicativo a indicar um caminho melhor para o motorista. Viu como os dados ganharam **relevância e propósito**? Assim, pode-se também dizer que *a informação dá subsídio à tomada de decisão*.

A partir disso, podemos dizer que a informação tem mais a ver com os atributos **qualitativos** dos dados do que com os **quantitativos**. Diferentemente dos dados, para a informação o foco é que se tenha significado e utilidade.

Isso traz algumas implicações para os sistemas computacionais. É mais difícil fazer um computador armazenar, classificar e realizar operações sobre a informação do que sobre os dados. Assim, a informação é mais apropriada para a compreensão e comunicação dos seres humanos.

E a partir das informações, o que é construído? É o que veremos a seguir.

Conhecimento

O conhecimento é um conceito um pouco mais abstrato, mas não é tão difícil de entender para resolver questões. O conhecimento tem relação com aquilo que nós humanos sabemos, então convidei você a fazer uma breve reflexão. *O que é saber algo?*

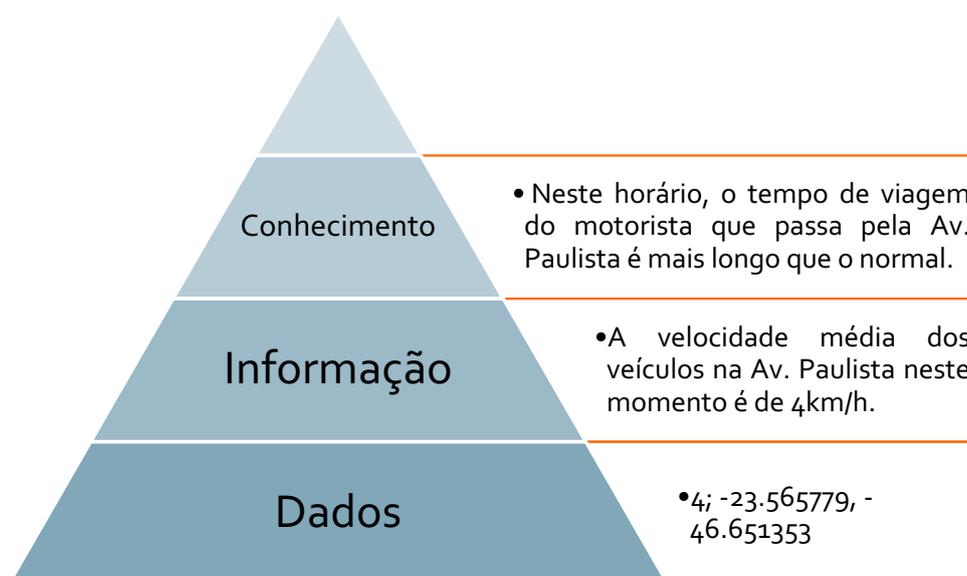
Complicado de definir, não é? O conhecimento envolve mais uma camada de enriquecimento do nosso modelo, dessa vez partindo da informação. A construção do conhecimento envolve uma mistura de vários fatores: **contexto, experiências passadas, significado**, etc.

Repare que novamente o processo de enriquecimento envolve contexto e significado. Sendo que desta vez se vê um **envolvimento humano** mais forte: a **interpretação** das pessoas é parte crucial desta etapa, levando em consideração a **confiabilidade** e a **importância** das informações e dados.

Note também que muitos autores afirmam que o conhecimento tem uma **aplicação**. Ele está atrelado a um objetivo prático e à intenção dos seres humanos de colocar as informações em ação.

Pode-se dizer que, ao absorver e ponderar muitas informações, você constrói conhecimento através do **aprendizado**. É só considerar esta aula: estou tentando te passar o máximo de informações que consigo, mas inevitavelmente vai cair alguma questão em alguma prova que a resposta não estará explícita aqui nos nossos PDFs. Assim, você precisará usar o seu conhecimento obtido durante o curso com a finalidade de analisar e responder. É esse o nosso objetivo! Queremos que você **aprenda** o assunto – assim construindo conhecimento – e não apenas decore uma lista de informações. Entendeu a diferença?

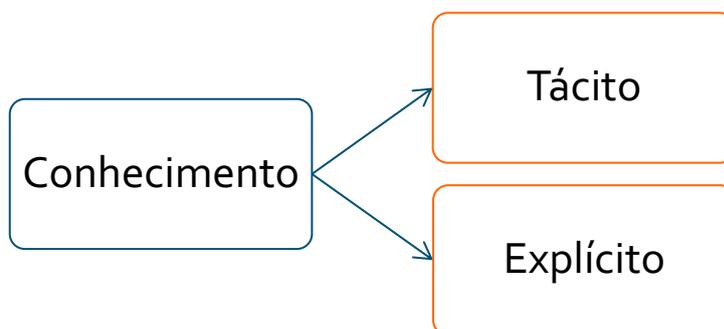
Agora que já temos uma ideia do que é conhecimento, podemos adicionar mais um nível na nossa Pirâmide:



Percebeu o raciocínio que está envolvido na transformação da informação em conhecimento? Há um processo de *interpretação e avaliação* para se chegar ao resultado final.

Conhecimento Tácito e Explícito

Uma coisa importante e que costuma cair em prova é a classificação do conhecimento, que costuma ser dividido em dois tipos:



O **conhecimento tácito** é aquele que está **implícito**, está internalizado nas mentes dos seres humanos. Por exemplo, imagine um mecânico com muitos anos de experiência. Chega um cliente em sua oficina, o mecânico apenas abre o capô do carro, dá uma olhadinha e já sabe qual é o defeito! Isso é um exemplo do uso do conhecimento tácito. É a experiência, o conhecimento acumulado que *é difícil de pôr em palavras*.

Assim, é fácil perceber que *o conhecimento tácito é de difícil estruturação e transmissão*. Uma coisa é ter anos de experiência no ramo, constituindo um vasto conhecimento, mas imagine o mecânico tentando organizar tudo o que ele sabe numa tabelinha, ou tentando ensinar um novo aprendiz em um curto espaço de tempo... Complicado, não é?

Já que o conhecimento tácito é difícil de se estruturar, é lógico concluir que também seria muito difícil processar em um computador. Observe que não estou falando apenas de escrever um documento com tudo o que alguém conhece e salvar o texto no computador. É armazenar de uma maneira que o computador "entenda" o seu conhecimento e seja capaz de processá-lo e aplicá-lo em uma situação prática.

Imagine o nosso exemplo do mecânico acima. Você poderia salvar vários livros de mecânica em um computador sem problemas. Porém, mesmo o melhor programador do mundo teria uma dificuldade enorme em desenvolver um software que conseguisse capturar a imagem do capô aberto e descobrir qual era o problema de primeira, assim como o mecânico fez. Ficou claro? Concluímos assim que *o conhecimento é de difícil captura em máquinas*.

Esse que discutimos agora é o tipo mais comum de conhecimento. Então fique ligado que já caiu em prova:

Atenção!

O conhecimento é frequentemente tácito

O outro tipo de conhecimento é o **conhecimento explícito**. Esse é o que está externalizado, registrado em algum lugar. Portanto, ele é muito mais fácil de armazenar, processar e transmitir. Por exemplo, é mais simples eu

te passar esta aula em PDF do que tentar te transmitir toda a minha experiência adquirida na área “de cabeça”, não é mesmo?

Esse é o tipo menos comum de conhecimento e é fácil perceber o porquê. Se juntarmos o conhecimento tácito combinado de todos os especialistas em determinado assunto, o resultado será um conhecimento muito maior do que cabe nos livros que eles porventura já tenham escrito.

Transferência do Conhecimento

Falamos bastante de transferência de conhecimento ao longo do último item, então é importante que apresentemos o modelo de espiral do conhecimento (Nonaka & Takeuchi, 1995), que define como funciona esse processo. Este assunto é bem simples e pode aparecer na sua prova.

- Quando a transferência ocorre de conhecimento tácito para conhecimento tácito, estamos falando da **socialização**. É a troca de experiências entre um servidor antigo e um recém empossado, por exemplo.
- Quando ocorre a conversão de conhecimento explícito para conhecimento explícito, diz-se que é realizada a **combinação**. Junta-se duas ou mais fontes para produzir novo conhecimento explícito. Isso acontece através da **sistematização** e **aplicação** do conhecimento.
- Já quando o conhecimento tácito é traduzido em conhecimento explícito, ocorre a **externalização**. O indivíduo externaliza suas ideias, ou seja, “tira” da mente e as transfere através de **diálogo** e **reflexão**.
- De maneira análoga, a conversão do conhecimento explícito em tácito é o processo inverso, portanto chamada de **internalização**. Geralmente está relacionado ao **aprender na prática**.

Assim, temos uma representação gráfica deste processo a seguir. Note que o gráfico é um ciclo, então o conhecimento está em constante transmissão e mutação.



* * *

Vamos sintetizar o que aprendemos até agora com uma tabela resumo. Ela vai te ajudar muito a classificar o que é dado, informação ou conhecimento na hora da prova, já que muitas vezes a distinção não é tão clara. Essa tabela pode ser cobrada na sua prova, então memorize bem.

Dado	Informação	Conhecimento
<i>Simples observações sobre o estado do mundo.</i>	<i>Dados dotados de relevância e propósito.</i>	<i>Informação valiosa da mente humana. Inclui reflexão, síntese, contexto</i>
- Facilmente obtido por máquinas	- Requer unidade de análise	- De difícil estruturação
- Frequentemente quantificado	- Exige consenso em relação ao significado	- De difícil captura em máquinas
- Facilmente transferível	- Exige necessariamente a mediação humana	- Frequentemente tácito
		- De difícil transferência

Fonte: (Davenport, 1998)

Um outro detalhe que você deve ter sempre em mente é que quando vamos adicionando contexto e significado, a necessidade do envolvimento humano vai crescendo. Quando isso ocorre, estamos ao mesmo tempo aumentando a complexidade e dificultando a captura e representação em máquinas.



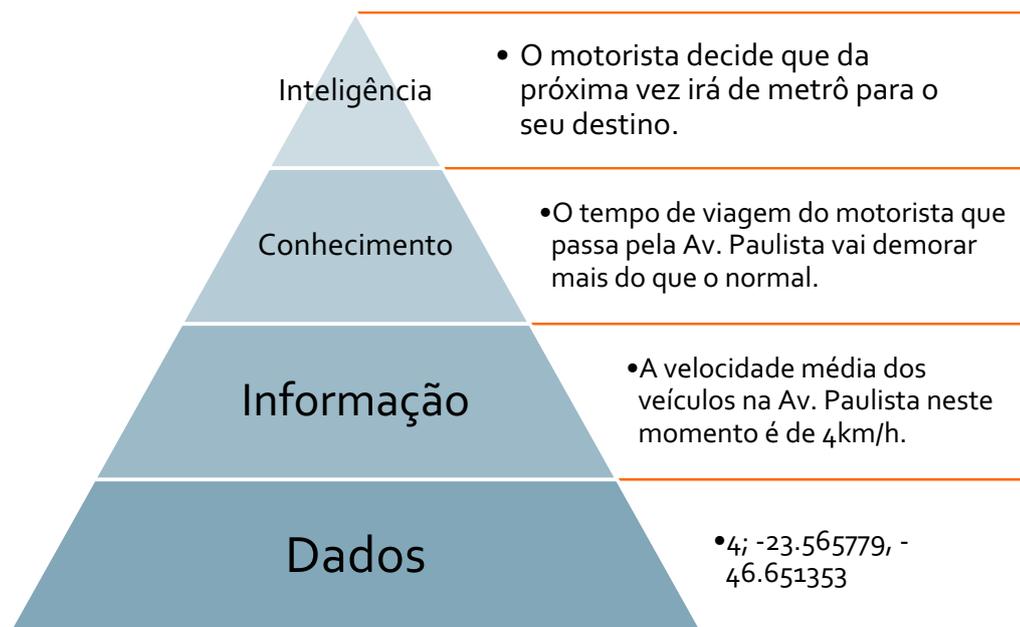
Inteligência

Por fim, sabemos que ter conhecimento se mostra útil em diversas situações. Quando usamos conhecimento para resolver algum problema complexo ou encontrar a resposta para uma questão de concurso, vemos a **inteligência** em ação.

A inteligência envolve a capacidade de **juízo**, é usar o conhecimento obtido para tomar a **melhor decisão** ou aplicá-lo a alguma situação concreta. É possível ver na inteligência uma influência muito forte da **intuição**, da **ética** e da **experiência**, sendo assim uma *habilidade essencialmente humana*.

A inteligência envolve noções de certo e errado, de bom ou ruim, envolve essa **ponderação** de qual é a melhor opção ou maneira de agir em uma determinada situação. Diz-se que é impossível capturar essa inteligência em máquinas (Ackoff, 1999).

Assim, no nosso exemplo, poderíamos ver a inteligência em ação:



Concluimos assim a nossa pirâmide de exemplo e a primeira parte da nossa aula! Para fixar o que aprendemos, seguem algumas questões comentadas:

(CESPE – 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.

RESOLUÇÃO:

Questão interessante! As vendas dos veículos devem ser organizadas em um relatório por faixa de preço, região, modelo e cor. Então, imagine possíveis dados contidos neste relatório:

30.000 a 40.000; Sudeste; Ford Fiesta; Vermelho...

Tomando como exemplo a palavra “vermelho”, está bem claro que se trata de um **dado**, correto? É um termo que, a não ser que você conheça o contexto, não tem muita relevância e significado. Sem maiores informações, a palavra “vermelho” pode estar caracterizando virtualmente qualquer coisa. O mesmo raciocínio serve para os demais itens.

Assim, o que se fez foi organizar, catalogar e ordenar esses dados em um relatório para que o resultado ganhasse relevância e se tornasse *subsídio para tomada de decisão*. Isso é o que definimos como sendo **informação**.

Gabarito: D

(CESPE – TCE/PE – 2017) Acerca dos conceitos de dado, informação e conhecimento, julgue o item a seguir:

A informação caracteriza-se por ser frequentemente tácita, bem como por ser de estruturação e captura difíceis em máquinas.

RESOLUÇÃO:

A segunda parte da afirmativa não é assim tão clara, já que informação realmente é mais difícil de estruturar e capturar em máquinas do que os dados, apesar de essa ser uma característica mais alinhada com o conhecimento (lembre-se da tabela). O que nos faz resolver a questão com bastante certeza é mesmo a primeira parte: o que se divide em tácito e explícito é o **conhecimento**.

Gabarito: E

(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.

RESOLUÇÃO:

Não considero essa uma questão fácil, partes das alternativas podem se encaixar em mais de um dos conceitos que já apresentamos. Vamos analisá-las:

I – Fatos, textos e gráficos, considerados individualmente, não têm significado e contexto, percebe? Então temos exemplos de dados. Alternativa **ERRADA**.

II – A chave desse item é lembrar que inteligência é, como vimos, uma habilidade essencialmente humana. Não há erro no restante da afirmativa, pode-se sim afirmar que é a síntese do conhecimento e é baseada em experiência e intuição. Alternativa **CERTA**.

III – Procure as palavras chave da questão. Conhecimento demanda **análise e avaliação** sobre a **confiabilidade, relevância e importância** de dados e informações, para a construção de um quadro de situação. Observe que o examinador está se referindo a um nível acima de dados e informações que, como sabemos, é o conhecimento. Alternativa **CERTA**.

Gabarito: C

(COPESE – UFT – 2018) Os dados são o elemento base para a produção/obtenção de um dos mais valiosos recursos que a organização dispõe, a informação. Esta, no contexto da Sociedade do Conhecimento, é imprescindível para a sobrevivência de uma empresa, sobretudo em função das constantes transformações a que se sujeita.

Assim, é **CORRETO** afirmar que a informação consiste em:

- a) fatos em sua forma primária.
- b) dados organizados, consolidados ou configurados de maneira significativa, para fornecer uma compreensão de um fato real.
- c) conjunto de regras e procedimentos usados para selecionar, organizar e manipular os dados.
- d) processo de decisões inter-relacionadas, executadas para atingir um resultado.

RESOLUÇÃO:

Outra questão interessante para fixarmos os conceitos.

- a) A definição apresentada nesta alternativa é a de **dados**. Alternativa **ERRADA**.
- b) Como já estudamos, a informação consiste em dados dotados de relevância e propósito. Em outras palavras, é o que alternativa afirma. É esta nossa resposta. **CERTA**.
- c) As regras e procedimentos utilizados não são objeto do nosso estudo. O resultado final é desse processo é que é a informação. **ERRADA**.
- d) A exemplo da letra C, o processo decisório não é objeto do nosso estudo. A informação é apenas um subsídio para a tomada de decisão. **ERRADA**.

Gabarito: B

(FGV – AL/RO – 2018) Após a entrada de novo contingente na ALE-RO, servidores mais antigos dão dicas para os novos em relação à execução mais adequada das tarefas de cada cargo, contribuindo para que os novos desenvolvam suas rotinas e práticas.

Conforme o modelo de espiral do conhecimento, a situação descrita acima, onde existe a conversão de um conhecimento tácito para outro conhecimento tácito, é chamada de:

- a) externalização.
- b) combinação.
- c) socialização.
- d) internalização.
- e) acomodação.

RESOLUÇÃO:

Questão bastante simples e objetiva! Conforme a espiral do conhecimento, a conversão de conhecimento tácito para outro conhecimento tácito é a socialização.

Gabarito: C

* * *

Dados Estruturados e Não Estruturados

Bom, agora que já vimos o que são dados, está na hora de apresentar as diferentes maneiras com que eles se apresentam nos sistemas computacionais. O assunto naturalmente vai ficar um pouco mais técnico, mas trarei exemplos simples e abstrações para que você possa entender melhor.

Vamos classificar os dados de acordo com a maneira como eles estão estruturados. Repare que até agora falamos bastante sobre a **natureza** dos dados, a partir de agora vamos tratar da maneira como eles podem ser gerenciados nos sistemas computacionais.

A palavra chave para entendermos o que é essa estruturação é **organização**: quanto mais estruturados, mais **rígida** é a maneira como os dados estão organizados e representados. Os dados estruturados apresentam estruturas fixas e seguem um conjunto mais restrito de regras, enquanto que os não estruturados podem se apresentar de diferentes maneiras.

Vamos ver os detalhes a respeito desses conceitos.

Dados Estruturados

A primeira categoria é a dos dados estruturados. Estes dados *respeitam uma estrutura rígida predefinida*, ou seja, eles estão sempre representados de acordo com um modelo que foi previamente estabelecido. É o tipo de dados comumente armazenado nos **bancos de dados convencionais**.

Os dados estruturados estão comumente armazenados em **blocos semânticos**, ou seja, elementos de natureza similar são agrupados nas mesmas estruturas. Por exemplo, num órgão público teríamos uma tabela para cadastrar servidores, outra tabela para cadastrar as licitações realizadas, outra para os contratos firmados e por aí vai.

Considere que esse órgão público precisa ter o cadastro dos servidores que lá trabalham para diversos propósitos, como disponibilizar no portal da transparência ou apurar o respeito aos limites com despesas de pessoal. Esse cadastro poderia estar organizado assim:

Nome	CPF	Salário
Clarice Lispector	987.654.321-00	R\$ 4.500
Jorge Amado	123.456.789-10	R\$ 3.000
Machado de Assis	879.645.132-05	R\$ 2.900
...

Repare que todos os valores para as colunas desta tabela possuem a mesma estrutura. Os nomes são compostos por cadeias de caracteres (várias letras em sequência) de tamanho variável, os CPFs seguem um padrão de 11 dígitos e os salários são valores numéricos em moeda corrente. Isso não é por acaso, os dados estruturados seguem uma estrutura **previamente definida** a que chamamos de **esquema**.

Esse **esquema** é definido ao projetar o banco de dados e compreende regras a respeito do tipo dos dados (número, cadeia de caracteres, bits...), tamanho máximo, se o valor em um campo pode estar em branco, dentre outras restrições.

Assim, se tentássemos incluir, por engano, um CPF de 7 dígitos ou colocássemos sem querer o nome do servidor na coluna destinada ao salário, o sistema provavelmente rejeitaria a entrada. Isso pode ser vantajoso em vários casos, principalmente quando queremos prevenir erros de cadastro.

A grande vantagem dos dados estruturados nos sistemas computacionais é que eles são *mais facilmente buscáveis e acessíveis*. Em um universo de um órgão público com centenas ou milhares de servidores, seria bastante complicado encontrar uma determinada pessoa caso os dados não seguissem algum tipo de rigor organizacional.

No nosso exemplo, como a estrutura já é conhecida, para obter o salário de um servidor seria só mandar um comando para o banco de dados que, traduzindo em linguagem humana, seria equivalente a "me mostre o valor **salário** da pessoa que tem o campo **nome** igual a 'Machado de Assis'".

Dados Não Estruturados

Ao contrário dos dados estruturados, os não estruturados *não têm estrutura definida*, simples assim. O examinador pode tentar te confundir dizendo que a estrutura está implícita, mas esse não é o caso, ela simplesmente não existe.

Assim, os dados não estruturados não seguem nenhum tipo de modelo ou organização predefinida. Como exemplos, podemos citar imagens, textos, vídeos, documentos... Os dados não estruturados são abundantes na internet e nas organizações, sendo *a maior parte dos dados no ambiente corporativo*.

Essa falta de estrutura torna esse tipo de dados de **difícil busca e processamento**. Já que eles não seguem um padrão, é como se os sistemas simplesmente ficassem sem saber pelo que procurar. Técnicas como mineração de dados e ferramentas que trabalham com *big data* geralmente possuem mecanismos para realizar uma preparação desses dados, dando uma certa organização para possibilitar que se trabalhe com eles.

Dados Semiestruturados

Os dados semiestruturados são algo intermediário entre os estruturados e não estruturados. Eles apresentam um certo grau de organização, contendo **marcadores**, espécies de rótulos que identificam os diferentes elementos de uma estrutura de dados, e podem incluir uma **hierarquia** entre as entidades representadas. Sendo assim, esse tipo de representação não é uma desordem total, mas ao mesmo tempo *apresenta mais flexibilidade do que os modelos estruturados*.

O exemplo mais difundido desse tipo de dados é o padrão **XML** (eXtensible Markup Language). A seguir trago um exemplo de código XML para que você possa compreender melhor:

```
<servidor nome="Clarice Lispector">
  <cpf>987.654.321-00</cpf>
  <funcaoGratificada>gerente</funcaoGratificada>
  <salario>4500</salario>
</servidor>
<servidor nome="Jorge Amado">
  <cpf>123.456.789-10</cpf>
  <salario>3000</salario>
</servidor>
<servidor nome="Machado de Assis">
  <cpf>879.645.132-05</cpf>
  <salario>2900</salario>
</servidor>
```

Não se preocupe em entender 100% do código, XML não cai no seu concurso! Só procure compreender o conceito de dados semiestruturados.

O início de cada elemento no XML possui uma tag (etiqueta) do tipo `<elemento>` e o final é delimitado com `</elemento>`. Esses são os **marcadores** que citamos há pouco.

Observe também que os elementos CPF e Salário estão mais à direita e entre as tags de início e fim de cada servidor. Isso significa que eles "pertencem" a um determinado servidor, sendo subelementos deste. Da mesma maneira, cada um desses subelementos poderia ter seus próprios subelementos e assim sucessivamente. Isso demonstra a capacidade de representação de uma **hierarquia** de dados de maneira simplificada.

Algumas características dos dados semiestruturados costumam ser cobradas em concurso, uma delas é que a estrutura deste tipo de dados é **irregular**. Isso se deve à ausência de regras rígidas de esquema, o que permite

que elementos semelhantes tenham conjuntos de atributos diferentes. É só observar o subelemento "funcaoGratificada" no exemplo acima. Um servidor possui esse subelemento, mas os demais não têm.

Não seria possível representar tal comportamento numa tabela de um banco de dados convencional, já que a tabela sempre teria as mesmas colunas ou atributos para todos os registros, mesmo que o valor estivesse em branco. Veremos mais sobre isso na aula a respeito de bancos de dados relacionais.

Ressalte-se que a estrutura dos dados semiestruturados também pode ser **parcial**. Ou seja, em um mesmo documento podemos ter parte dos dados seguindo uma determinada estrutura, mas ter também outros componentes não estruturados.

Finalmente, é importante lembrar que nos dados semiestruturados a estrutura é definida **a posteriori**. Isso significa que na maioria das vezes a estrutura está **implícita** nos próprios dados e pode ser extraída após uma análise, uma investigação. É esse o caso do nosso exemplo, não há nenhuma definição prévia que nos diga quais são os atributos que cada servidor pode ter ou qual o tipo de dados desses atributos, podemos apenas inferir ao analisar o código.

Por esse mesmo motivo, diz-se também que esse tipo de dados tem a característica de ser **autodescritivo**, já que as informações sobre a estrutura dos dados estão contidas nos próprios dados.

* * *

Terminamos assim mais uma parte da aula. Lembre-se que ao final teremos um resumo focado para que você se lembre das características que apresentamos acima. Vá beber um copo d'água e na volta resolva os seguintes exercícios para fixação:

(CESPE – TCE/SC - 2016) A respeito de dados estruturados, não estruturados e abertos, julgue o item subsequente.

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.

RESOLUÇÃO:

Como já vimos, os dados **semiestruturados** são **autodescritivos**, ou seja, a informação de esquema está contida nos valores dos dados, e têm **estrutura irregular**, podendo ter atributos diferentes para cada objeto. A estrutura (ou esquema) é definida **a posteriori**, então os atributos não são conhecidos com antecedência.

As características apresentadas se encaixam quase que perfeitamente na alternativa, só que o examinador troca o tipo de dado. Vamos corrigi-la:

Em se tratando de dados **semiestruturados**, 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 **estruturados**.

Gabarito: E

(CESPE – 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, conseqüentemente, 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.

RESOLUÇÃO:

- a) Dados não estruturados realmente podem ser caracterizados por arquivos diversos tipos, como aqueles apresentados pela questão. Até aí tudo bem. O final da questão é que tenta te complicar, porque contém uma informação incompleta, mas verdadeira. A estrutura desse tipo de dados não está **nem implicitamente nem explicitamente** descrita. **CORRETA**
- b) Conforme apresentamos durante a aula, os documentos XML são semiestruturados. **ERRADA**
- c) Não! Conforme conversamos, os dados não estruturados não seguem nenhum modelo ou padrão. É o oposto do que a questão afirma. Gostaria de lembrar a você que nem mesmo os dados semiestruturados se encaixariam nessas características, já que têm estrutura **irregular**. **ERRADA**
- d) Blocos semânticos são blocos de elementos com significado similar. Ou seja, elementos parecidos estão armazenados numa mesma estrutura. A afirmativa, então, está descrevendo os dados estruturados. **ERRADA**
- e) Pense comigo. Se os dados seguissem um padrão e fossem organizadinhos seria muito mais fácil de consultar, não é? Então a alta heterogeneidade dificulta as consultas aos dados não estruturados. **ERRADA**

Gabarito: A

Dados Abertos



Fonte: <http://www.governoaberto.cgu.gov.br>

Dando continuidade à nossa aula, chegamos aos dados abertos. Este conteúdo está intimamente relacionado com a noção de **transparência** da administração pública.

Muitos dados controlados pelos entes, principalmente aqueles relacionados ao patrimônio público, devem ser tornados amplamente acessíveis para a população. Assim, possibilita-se o **controle social** e a participação popular e aprimora-se a gestão pública. Nesse sentido, foi aprovada e entrou em vigor em 2012 a Lei de Acesso à Informação (nº 12.527/2011).

Uma definição completa do que são os dados abertos foi exarada pela organização Open Knowledge International e traduzida no portal dados.gov.br do Governo Federal:

Dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, a exigências que visem preservar sua proveniência e sua abertura.

Ou seja, vê-se que a abertura não diz respeito à natureza do dado em si, mas à sua acessibilidade e usabilidade. A regra geral aqui é colocar o mínimo de empecilhos possíveis para esse fim, sendo admitidas exigências para preservar a **proveniência** - por exemplo, exigência de que se credite o autor - e sua **abertura** - pode-se requerer que o indivíduo que deseje compartilhar o faça pelo mesmo tipo de licença que acessou, impedindo assim que os dados abertos sejam distribuídos através de licenças de uso mais restritivas.

Há ainda o Decreto nº 8.777/2016, que define a Política de Dados Abertos do Governo Federal. Essa política define dados abertos da seguinte maneira:

*Art. 2º, III - 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.***

Então, se vier na prova essa definição ao pé da letra você também pode considerar correta.

Três Leis

O autor David Eaves propôs três leis para os dados abertos governamentais. Cito as leis a seguir, colocando em negrito o texto original e em itálico os meus comentários, ok? Não confunda.

1. Se o dado não pode ser encontrado e indexado na Web, ele não existe;

Indexação diz respeito aos mecanismos de busca. Ou seja, se os dados não podem ser encontrados em pesquisa no Google ou similar, então são considerados inexistentes.

2. Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e

Por exemplo, não adianta muito publicar os dados a respeito das despesas de um determinado município somente através de fotos das notas de empenho. Se fosse assim, os sistemas utilizados para se analisar a despesa pública não seriam capazes de extrair dados importantes como valor, credor e dotação orçamentária.

3. Se algum dispositivo legal não permitir sua replicação, ele não é útil.

Acredito que este item seja autoexplicativo. Se o dado não pode ser replicado e, por consequência, compartilhado e difundido, ele perde sua utilidade.

Oito Princípios

Além das três leis, um grupo de especialistas se reuniu em 2007 nos EUA e definiu oito princípios fundamentais para os dados abertos. Estes princípios também são adotados e traduzidos no portal dados.gov.br (2018). Transcrevi o conteúdo do site pra você e teci alguns comentários:

1. 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.

É importante que os dados sejam disponibilizados em sua totalidade até mesmo para que se extraia as informações corretas. A omissão de partes do conjunto muitas vezes pode enviesar o significado, como seria o caso de uma frase fora de contexto em uma transcrição ou gravação, por exemplo.

2. Primários

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

Repare que o item se refere à publicação. Os usuários dos dados podem sim agregar, categorizar e realizar análises livremente.

3. Atuais

Os dados são disponibilizados o quão rapidamente seja necessário para preservar o seu valor.

4. Acessíveis

Os dados são disponibilizados para o público mais amplo possível e para os propósitos mais variados possíveis.

5. Processáveis por máquina

Os dados são razoavelmente estruturados para possibilitar o seu processamento automatizado.

6. Acesso não discriminatório

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

Alguns órgãos restringem o acesso a seus portais de transparência solicitando dados como nome e CPF para acessar. Isso não estaria de acordo com os princípios dos dados abertos.

7. Formatos não proprietários

Os dados estão disponíveis em um formato sobre o qual nenhum ente tenha controle exclusivo.

Formatos proprietários são os formatos de arquivos que não são livremente manipuláveis, mas pertencem a uma empresa ou entidade. O proprietário deste formato pode restringir o acesso ao cobrar uma taxa para utilização ou exigir que se use um software específico para manipulá-lo, por exemplo.

Hoje em dia a tendência geral é o estabelecimento de padrões abertos, porque facilitam a interoperabilidade com os diversos sistemas existentes.

8. 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.

Os dados abertos devem ser de domínio público ou estar sob licença aberta.

* * *

Muita coisa para memorizar, não é? Mas acho que você não deve se preocupar muito em saber listar de cabeça os oito princípios. O importante é que você entenda do que se tratam e absorva o espírito dos dados abertos. Você vai ver que isso e mais uma dose de bom senso são suficientes para resolver a maior parte das questões sobre o tema.

Vamos testar o que aprendemos?

(CESPE – TCE/SC – 2016) A respeito de dados estruturados, não estruturados e abertos, julgue o item subsequente.

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

RESOLUÇÃO:

Bom, isso é aquilo que já discutimos. Os dados abertos realmente são de livre utilização, reutilização e redistribuição. Esse final da afirmativa não corresponde à definição mais atual de dados abertos pela Open Definition (a atual foi transcrita na nossa parte teórica), então temos que considerar a realidade na época da prova.

Ainda assim, considere interessante trazer a assertiva, já que a parte dos créditos à autoria, apesar de se referir a um conceito menos amplo, corresponde à **proveniência**. Da mesma maneira, o compartilhamento pela mesma licença se destina a preservar a **abertura** dos dados, então o espírito da afirmação continua correto.

Gabarito: C

(CS-UFG – UFG – 2017) De acordo com a Política de Dados Abertos do Poder Executivo federal, os dados devem ser representados em meio digital e:

- 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.
- 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.

RESOLUÇÃO:

Podemos responder esta questão tendo em mente os oito princípios dos dados abertos. Assim, mesmo que não soubéssemos o texto do Decreto nº 8.777/2016, teríamos condições marcar a resposta correta. Veja:

a) Pense bem, não faz sentido que o dado seja disponibilizado no formato requerido pelo cidadão. A bem da verdade, os dados abertos não se destinam a atender requisições, mas a estar sempre disponíveis para o acesso amplo da sociedade. **ERRADA**

b) Todas essas características respeitam os princípios dos dados abertos. É essa a nossa resposta. **CERTA**

c) Como vimos, os dados abertos são de livre acesso, utilização, modificação e compartilhamento. Assim, é contraditório que a instituição defina normas para estas atividades. Os dados abertos, então, não devem estar sujeitos a nenhuma restrição, a não ser que o examinador especifique que tais normas se destinem a preservar a proveniência e a abertura. **ERRADA**

d) Os dados abertos devem ser de domínio público ou estar estruturados sob licença **aberta**, não do órgão detentor. **ERRADA**

Gabarito: B

Coleta, tratamento, armazenamento, integração e recuperação de dados

Para contextualizar esses vários termos presentes em um tópico do seu edital, precisamos entender o propósito dos sistemas que têm a ver com o processo de Inteligência de Negócio ou Business Intelligence (BI) que ocorre dentro das organizações.

BI é uma área da tecnologia da informação que tem como objetivo geral coletar e processar dados para que se extraia informações relevantes para **tomada de decisão**. Embora tenha a palavra *business* – ou *negócio* – no nome, as técnicas de BI são plenamente aplicáveis ao setor público e estão ganhando crescente importância nos mais diversos órgãos e entidades. Tome “negócio” aqui como sendo o ambiente de **qualquer organização**.

Esse processo é necessário porque, como vamos ver mais à frente, nem sempre os dados que servem às aplicações e operações diárias das organizações estão em um formato favorável à análise e à extração de conhecimento. É necessário realizar diversas tarefas e transformações que vão desde a coleta inicial até a análise dos dados.

Não há um processo rígido definido para Business Intelligence e para a formação de **sistemas analíticos** de modo geral, mas podemos resumir as principais tarefas e etapas no seguinte fluxograma. Creio que você não precisa memorizá-lo, mas acho importante que entenda, em linhas gerais, o caminho seguido pelos dados.



Fique sabendo que o conceito exato de BI pode variar bastante de acordo com o autor. Na prática, o que estamos tratando aqui é a formação de **sistemas analíticos** de modo geral, ou seja, aqueles que possibilitam a análise de informações com o objetivo de subsidiar a tomada de decisão.

ETL

No início desse processo que estamos descrevendo, temos a parte que envolve **coleta, tratamento e armazenamento** dos dados. Essa etapa é conhecida como ETL – Extract, Transform, Load. As tarefas efetivamente executadas em um processo de ETL variam bastante e vão depender da natureza dos dados e dos objetivos dos usuários.

Em linhas gerais, as fases do ETL são as seguintes:

Coleta

A **coleta** (ou extração) se destina a capturar os dados de **diversas fontes e sistemas** que serão transformados e modificados para se tornarem úteis para a tomada de decisão. Por exemplo, em uma empresa é comum que haja um ERP, o famigerado “sistema”, que é aquele software que gerencia a maior parte das atividades da empresa, possibilitando o controle de vendas, clientes, fornecedores, contas a pagar e a receber, RH, etc. Esse ERP é uma fonte muito valiosa para a coleta, já que por ele circula boa parte dos dados relevantes para a organização.

O ERP é um exemplo de sistema **transacional**, que processa as transações e permite a realização das operações do dia a dia da organização.

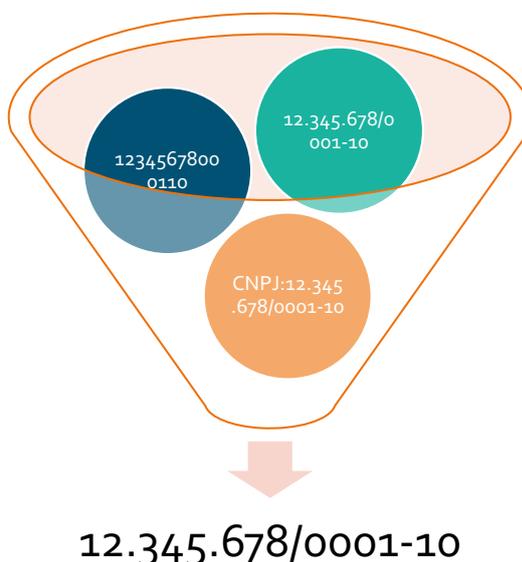
Também podem ser coletadas informações de planilhas, arquivos .txt, sites da internet, sistemas auxiliares, dentre outros. A escolha das fontes é um importante passo no âmbito de BI, pois esses dados inicialmente coletados é que serão a base para o resto do processo.

Transformação

Em seguida, na etapa de **transformação**, pode ocorrer a **limpeza** dos dados, corrigindo possíveis erros de entrada nos dados (erros de digitação na inserção, por exemplo), ajustes de tipo de dados e demais desconformidades com os padrões de qualidades definidos pelos usuários e projetistas do sistema. Nessa etapa, os dados são **tratados** para que se tornem mais corretos ou relevantes para os propósitos do BI.

Outra tarefa que é realizada no âmbito da transformação diz respeito à **padronização** dos dados para a futura integração. Essa tarefa é de fundamental importância quando estamos coletando dados de muitas fontes, já que os dados de diferentes sistemas ou arquivos podem assumir diferentes padrões para representar a mesma coisa.

Para exemplificar, imagine que você está trabalhando na fiscalização de uma empresa pública e precisa cruzar informações de diversos sistemas. Contudo, o CNPJ da entidade está escrito de forma diferente em cada um desses sistemas. A fase de transformação é responsável por tratar os dados dessas fontes e, dentre outras tarefas, convertê-los em um único padrão:



O mesmo processo poderia se aplicar, por exemplo, à conversão de diversas escalas de dados em uma só. Imagine que uma fonte registra uma determinada medida em **milímetros**, outra em **metros** e uma terceira em **centímetros**. Para facilitar a análise e, por consequência, a tomada de decisão, seria interessante construir uma representação unificada, com a escolha de uma só unidade (por exemplo, centímetros) e a conversão de todas as outras para essa escolhida.

Essas duas primeiras etapas do ETL, de **extração** e **transformação** geralmente se utilizam de uma base de dados temporária, chamada **staging area**. Esse banco de dados intermediário é considerado uma **área de transição** onde se pode inserir os dados oriundos da extração e realizar as transformações necessárias, tudo isso antes de carregá-los no data warehouse (que será o destino dos dados).

A utilização dessa staging area se dá de modo a **não sobrecarregar** nem o ambiente de produção (os sistemas operacionais da empresa) nem o próprio data warehouse com as operações de transformação.

Carregamento

Por fim, temos o **armazenamento** ou **carregamento**. Nesta etapa, os dados já coletados e tratados são transferidos para o banco de dados (ou qualquer estrutura que armazene dados) de destino, que geralmente é um BD especial voltado para análise chamado *data warehouse*.

Periodicidade do ETL

Um processo de ETL é criado e executado de forma periódica, geralmente de forma automatizada. O quão frequente ele será vai depender de muitos fatores, como a complexidade do processo, as necessidades dos usuários de enxergarem dados atualizados ou mesmo a frequência de atualização das diversas fontes de dados. Assim, podemos ter processos de ETL que serão executados a cada 5 minutos, 1 hora, 1 semana ou mesmo 1 mês, por exemplo.

Note que o desenvolvimento de um processo adequado de ETL costuma ser a tarefa **mais crítica e mais demorada** na formação de um sistema analítico.

Integração

O Data Warehouse (DW) ou Armazém de Dados é, na maioria das vezes, o responsável por conter essa **integração** de dados de diferentes fontes. O **DW** é um repositório que contém os dados dos mais diversos setores de uma organização, servindo de base para que se criem relatórios, *dashboards*, cubos OLAP ou se realize outras formas de análise. Não vamos aprofundar muito o conceito aqui, já que teremos um momento específico para isso.

O importante é que você saiba que a integração de dados nada mais é do que esse *processo de combinação de dados de diversas fontes em uma base de dados unificada* e tudo que o envolve. **Preste atenção:** a integração é um processo que se inicia já desde a coleta, com o planejamento das diferentes fontes de dados que vão compor o conjunto final, e passa pelas etapas subsequentes até que se chegue na base de dados final.

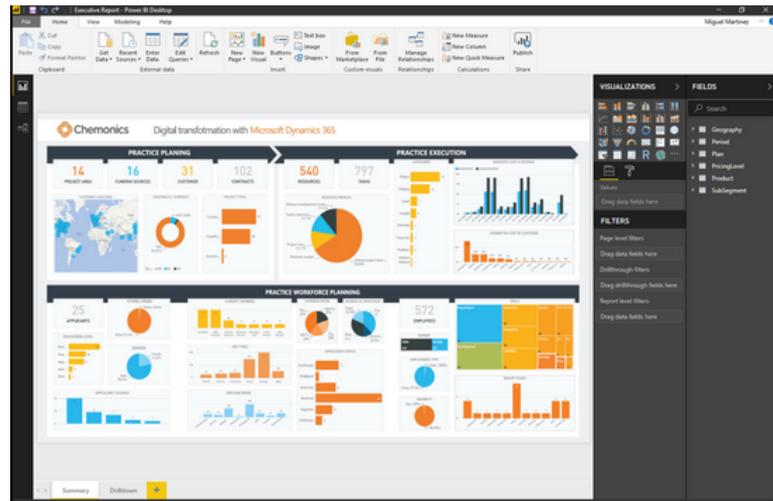
Recuperação

Os dados já coletados, tratados, armazenados e integrados devem ser tornados acessíveis para que sejam úteis. Neste processo de **recuperação** ou extração é possível alimentar diferentes sistemas de **visualização e análise de dados** para subsidiar a tomada de decisão.

Esta etapa requer um forte conhecimento a respeito dos **propósitos** que se têm para a utilização da informação e a respeito do negócio subjacente. Ou seja, para que possamos trabalhar com esses dados refinados de modo a torná-los úteis à tomada de decisão, precisamos saber primeiro **para que** queremos fazer isso e quais são os aspectos importantes do negócio que precisam ser representados.

Qual a decisão precisa ser tomada? Qual a aplicabilidade da informação recuperada? Como explicitar os dados de uma maneira útil e clara? Essas perguntas devem permear todo este processo.

Visualização se refere à produção de relatórios e dashboards que ajudam a compreender a dinâmica dos dados em um contexto de negócio. São documentos que contém muitas vezes gráficos, tabelas e dados em uma forma visualmente agradável.



Fonte: <https://powerbi.microsoft.com/pt-br/what-is-power-bi/>

A partir da integração dos dados transformados em um data warehouse, também é possível se utilizar de ferramentas mais avançadas de análise, como OLAP e de Mineração de Dados. Aqui, o usuário irá poder fazer o uso das mais diversas tecnologias e técnicas para extrair informações úteis, que irão, em última instância, subsidiar a tomada de decisão.

* * *

Prezado aluno, terminamos aqui nossa aula inaugural. Espero que o assunto tenha ficado claro! A seguir, segue uma bateria de questões comentadas para você fixar o conhecimento. Priorizei a banca CESPE, mas também selecionei questões de diversas bancas para chegarmos em uma quantidade adequada.

Bom proveito!

Questões comentadas pelo professor

1. (FCC – TCE/RS – 2018)

Julgue o item a seguir, acerca de dado, informação, conhecimento e inteligência.

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.

RESOLUÇÃO:

a) É o seguinte: os dados, dotados de relevância e propósito, passam a ser informações. Daí, adicionando-se a capacidade humana de síntese, contextualização e de um grau de reflexão, é possível obter o conhecimento. Dessa maneira, deduzimos que não é possível ter conhecimento sem que o dado, a unidade mais básica deste ciclo, exista. **ERRADA**

b) Existem algumas definições diferentes para informação. Contudo, o maior fator que permite distinguir esse conceito do conceito de dado é a ausência ou presença de significado. Os dados são valores brutos, sem significado, enquanto que a informação já conta com um contexto e com algum significado, podendo ser utilizada na tomada de decisão. Esta é a nossa resposta. **CERTA**

c) Na verdade geralmente acontece o contrário. O conhecimento está um nível acima da informação na pirâmide do conhecimento, sendo sua produção um processo que envolve síntese, reflexão e contexto a partir da informação.

Fique atento pois há autores que defendem que é necessário conhecimento para processar os dados e transformá-los em informações. Contudo, como a alternativa anterior está mais obviamente correta, podemos observar que a banca adotou o entendimento de que o processo da transformação de dados em informação e depois em conhecimento é algo linear. **ERRADA**

d) Falso! A tomada de decisão começa a se dar a partir da informação. Os dados brutos, por não possuírem significado ou contexto, não são tão relevantes para o gestor tomador de decisão. Por outro lado, são facilmente processáveis por máquina. **ERRADA**

e) Acredito que a alternativa mais correta para se encaixar nessa estrutura seria: o conhecimento consiste na informação analisada sobre diferentes pontos de vista. Os dados envolvem análise ou reflexão. **ERRADA**

Gabarito: B

2. (CESPE – SEFAZ/RS – 2019)

A respeito do BI (business intelligence), assinale a opção correta.

- a) O BI consiste na transformação metódica e consciente das informações exclusivamente prestadas pelos tomadores de decisão em novas formas de conhecimento, para evolução dos negócios e dos resultados organizacionais.
- b) ETL é o processo de análise de dados previsto pela arquitetura de BI.
- c) As técnicas do BI objetivam definir regras para a formatação adequada dos dados, com vista a sua transformação em depósitos estruturados de informações, sem considerar a sua origem.
- d) O repositório de dados analíticos de BI é representado pelas diversas bases de dados relacionais e por repositórios de dados que utilizem modelagens relacionais.
- e) A camada de apresentação de uma arquitetura de BI é aquela em que as informações são organizadas e centralizadas.

RESOLUÇÃO:

a) As informações presentes em um sistema de BI não são prestadas apenas pelos tomadores de decisão. Elas são oriundas de **diferentes fontes heterogêneas** e só depois são transformadas em estruturas que facilitam a tomada de decisão. **ERRADA**

b) O ETL é um processo de **extração, transformação e carregamento** dos dados. Esse processo visa preparar os dados para análise, não realizá-la. **ERRADA**

c) As técnicas e procedimentos de BI visam definir regras para a formatação adequada dos dados, com o propósito de armazená-los em um repositório estruturado e integrado chamado **data warehouse**.

É um pouco estranho o trecho que diz “sem considerar a sua origem”, já que, obviamente, o processo de ETL terá que lidar com as particularidades de cada fonte de dados. Como esse item é o mais correto da questão, podemos assumir que o examinador quis dizer que os dados, após o armazenamento no **data warehouse** (e, por consequência do ETL) ficam armazenados em uma só estrutura integrada, independentemente de suas fontes.

d) O repositório de dados analíticos do BI (que é o data warehouse) é representado por repositórios que utilizam a **modelagem dimensional**, não relacional! Esses repositórios utilizam uma estrutura **desnormalizada**, de modo a privilegiar a performance de leitura de dados para análise. **ERRADA**

e) A camada de apresentação do BI é aquela que permite a análise. Nela, as informações são dispostas em elementos como gráficos, dashboards e relatórios, ou ficam disponíveis para navegação e análise através do uso de OLAP. No entanto, a organização e centralização das informações é feita ainda durante o processo de ETL, na parte de **integração** dos dados. **ERRADA**

Gabarito: C

3. (CESPE – EMAP – 2018)

Julgue o próximo item, a respeito de gestão do conhecimento.

Na solução de problemas, o conhecimento tácito é associado ao conhecimento do expert.

RESOLUÇÃO:

É isso mesmo! O conhecimento tácito está internalizado nos seres humanos e é de difícil transferência. É o que acontece com um expert em determinado assunto.

Gabarito: C

4. (CESPE – EMAP – 2018)

Julgue o próximo item, a respeito de gestão do conhecimento.

Em uma instituição, o conhecimento explícito está relacionado ao capital intangível.

RESOLUÇÃO:

Se você já estudou um pouco de Contabilidade, você sabe que intangível quer dizer incorpóreo, sem substância física. Assim, o conhecimento explícito seria um exemplo de capital **tangível**, já que está registrado em algum lugar. O conhecimento tácito, por sua vez, tem a característica de ser intangível, já que está internalizado nas mentes das pessoas.

Gabarito: E

5. (CESPE – EMAP – 2018)

Julgue o próximo item, a respeito de gestão do conhecimento.

No que se refere à espiral do conhecimento, socialização, externalização, modelagem e feedback são os diferentes modos de conversão do conhecimento.

RESOLUÇÃO:

Na espiral do conhecimento temos os processos de socialização, combinação, externalização e internalização. Logo, modelagem e feedback não fazem parte deste modelo.

Gabarito: E

6. (CESPE – TCE/PA – 2016)

Julgue o item subsecutivo, referente à gestão do ciclo de vida da informação — ILM (Information Lifecycle Management).

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.

RESOLUÇÃO:

Como vimos durante a aula, os dados não estruturados costumam representar a maior parte dos dados em uma organização. Dessa maneira, eles irão requerer mais espaço para armazenamento e um gerenciamento mais cauteloso. A alternativa somente inverteu os conceitos.

Gabarito: E

7. (CESPE – TCE/SC – 2016)

Julgue o item a seguir, acerca de dado, informação, conhecimento e inteligência.

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.

RESOLUÇÃO:

Poderíamos considerar a primeira parte como correta, caso o examinador houvesse explicado corretamente o que queria dizer ao definir informação como significado. Contudo, após o “ou seja”, ele define de forma um tanto rebuscada o conceito de **dados**, já que estes são números, letras, palavras, etc. que representam fatos do mundo - na questão chamados de atos, conceitos ou instruções.

Questão um pouco difícil devido à redação, mas acredito que você teria condições de enxergar que não se tratava de uma definição aceitável para informação.

Gabarito: E

8. (CESPE – TCE/SC – 2016)

Julgue o item a seguir, acerca de dado, informação, conhecimento e inteligência.

O atributo de inteligência depende mais da qualidade da informação disponível do que da sua quantidade, tendo, portanto, natureza qualitativa.

RESOLUÇÃO:

É isso mesmo. Quando falarmos de natureza quantitativa estamos nos referindo aos dados. Em seguida, quando vamos subindo na pirâmide do conhecimento, os conceitos vão ficando mais qualitativos e mais dependentes da interação humana, já que se começa a adicionar contexto, significado, síntese, reflexão... Portanto, a inteligência, por ser uma habilidade **essencialmente humana**, se mostra muito mais ligada ao aspecto **qualitativo**.

Gabarito: C

9. (FUMARC – PC/MG – 2018)

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.

RESOLUÇÃO:

Vamos avaliar as alternativas:

I – É uma boa definição de dados. Uma descrição elementar das coisas que não está organizada de maneira a transmitir um significado específico. **CORRETA**

II – Informação está realmente organizada de modo a ter significado e valor para o receptor, servindo assim de subsídio para a tomada de decisão. Em outras palavras, é basicamente o que afirma a alternativa. **CORRETA**

III – Bom, esta não é uma definição ampla e exaustiva do conhecimento. Ela está aqui utilizada mais no âmbito das organizações. Mas, conforme estudamos, o conhecimento realmente envolve experiência, aprendizado e aplicação prática, estando relacionado a uma finalidade específica. Na afirmativa, a tal aplicação é um problema ou atividade empresarial. **CORRETA**

Gabarito: A

10. (FGV – TRT/SC - 2017)

Cada vez mais as empresas trabalham com dados, informações e conhecimentos. Uma organização cria e utiliza conhecimento transformando-o de tácito em explícito e vice-versa

Um dos 4 modos de conversão do conhecimento é:

- a) socialização: de explícito para tácito, procura compartilhar e criar conhecimento através de experiência direta;
- b) externalização: de tácito para explícito, procura articular conhecimento explícito através do diálogo e da reflexão;
- c) combinação: de explícito para explícito, procura sistematizar e aplicar o conhecimento explícito e a informação;
- d) internalização: de implícito para tácito, procura aprender e adquirir novo conhecimento tácito na prática;
- e) transferência: de explícito para tácito, procura sistematizar e aplicar o conhecimento explícito e a informação.

RESOLUÇÃO:

- a) Socialização, como já vimos, é a conversão de conhecimento tácito em tácito. **ERRADA**

b) Externalização é a conversão de conhecimento tácito para explícito e pode ocorrer sim a partir do diálogo e reflexão. Contudo, há um pequeno errinho que confesso que demorei a enxergar. O examinador diz que essa conversão procura articular conhecimento **explícito** através do diálogo e da reflexão. Na verdade, ele procura articular o conhecimento **tácito**, o explícito é o fruto dessa articulação. **ERRADA**

c) Combinação é a conversão de conhecimento explícito para explícito e ocorre através da sistematização e aplicação. Tudo certo. **CORRETA**

d) Internalização é a conversão de conhecimento explícito para tácito. O examinador tentou confundir ao colocar "implícito" no texto da questão, mas você sabe que conhecimento implícito é a mesma coisa que tácito. **ERRADA**

e) Transferência não é uma das maneiras de conversão de conhecimento. **ERRADA**

Gabarito: C

11. (CESPE – MEC – 2011)

Julgue o próximo item, relativo à gestão do conhecimento:

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.

RESOLUÇÃO:

A essa altura você já deve estar careca de saber! O exemplo de transferência mestre -> aprendiz é um exemplo clássico de socialização, ou seja, conversão de conhecimento tácito em tácito.

Gabarito: E

12. (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.

RESOLUÇÃO:

Colocando em outras palavras, o examinador quer saber o que é que se tem quando os dados passam a poder ser interpretados e ganham significado para o receptor. Vimos que esse é o conceito de **informação!**

Gabarito: D

13. (CESPE – DPU – 2010)

A teoria de Nonaka e Takeuchi trata do processo de criação do conhecimento e preconiza a existência do conhecimento tácito e explícito. Acerca da criação do conhecimento e do processo de conversão do conhecimento, assinale a opção correta.

- a) No processo de conversão do conhecimento, a internalização corresponde à transformação do conhecimento tácito em conhecimento explícito.
- b) A socialização pode ser entendida como a conversão do conhecimento tácito em novo conhecimento tácito.
- c) A externalização corresponde à articulação do conhecimento tácito em conceitos implícitos.
- d) Programas de treinamento formal não podem ser considerados ferramentas de combinação do processo de conversão do conhecimento.
- e) O método de ensino denominado aprender fazendo provoca o processo de criação do conhecimento conhecido como socialização.

RESOLUÇÃO:

- a) A internalização, na verdade, é o processo reverso. É a transformação do conhecimento explícito em tácito, muitas vezes através do aprendizado na prática. **ERRADA**
- b) A socialização é mesmo a conversão de conhecimento tácito em tácito, como uma troca de experiências. **CERTA**
- c) A externalização, como o nome já diz, é uma forma de externalizar ou seja, explicitar o conhecimento. **ERRADA**
- d) A afirmativa não faz muito sentido. Programas de treinamento formal envolvem sim conversão de conhecimento dos mais diversos tipos. **ERRADA**
- e) Aprender fazendo provoca a transformação do conhecimento explícito em tácito, ou seja, internalização. **ERRADA**

Gabarito: B

14. (FUNDATEC – IGP/RS - 2017)

Relacione a Coluna 1 à Coluna 2, associando o tipo de conhecimento (tácito ou explícito) às suas características.

Coluna 1

- 1. Conhecimento explícito.
- 2. Conhecimento tácito.

Coluna 2

- () Não é facilmente visível e explicável.

- () Pode ser expresso em palavras, números ou sons, compartilhado na forma de dados, fórmulas ou manuais.
- () Pode ser rapidamente transmitido aos indivíduos, formal ou sistematicamente.
- () Possui duas dimensões: a primeira é uma dimensão técnica, que engloba as habilidades informais e de difícil detecção, e a segunda consiste em crenças, percepções, ideais e valores.

A ordem correta de preenchimento dos parênteses, de cima para baixo, é:

- a) 1 – 1 – 2 – 2.
- b) 1 – 2 – 2 – 1.
- c) 2 – 1 – 1 – 2.
- d) 2 – 2 – 1 – 1.
- e) 2 – 1 – 2 – 1.

RESOLUÇÃO:

Questão para resumir esse assunto e não esquecer mais! Vamos lá:

- O conhecimento **tácito**, como está internalizado, não é facilmente visível e explicável. **(2)**
- O conhecimento **explícito** está expresso, ou seja, registrado em forma de palavras, números ou sons e pode ser compartilhado nas mais diversas mídias. **(1)**
- O conhecimento tácito é de difícil transmissão. Já o **explícito**, como está registrado, é muito mais facilmente transmitido, seja de maneira formal, seja de maneira sistemática. **(1)**
- Essa última alternativa é um pouco mais complexa, pois carrega um pouco mais na teoria da Ciência da Informação. Contudo, podemos ver que habilidades informais e de difícil detecção têm tudo a ver com o conhecimento **implícito**. **(2)**

Gabarito: C

15. (UFMT – UFSBA – 2017 – ADAPTADA)

Julgue a afirmativa a seguir:

Informação refere-se a dados apresentados em uma forma significativa e útil para seres humanos e dados são sequências de fatos brutos que representam eventos que ocorrem nas organizações ou no ambiente físico, antes de terem sido organizados e arranjados de uma forma que as pessoas possam entendê-los e usá-los.

RESOLUÇÃO:

Outra forma diferente de definir dado e informação. Você já deve estar percebendo que as bancas adotam diferentes formas para explicitar os conceitos. Contudo, o espírito permanece o mesmo: dados são fatos brutos a respeito do mundo, e informações são dotadas de contexto e significado, sendo mais úteis para a tomada de decisão.

Gabarito: C

16. (IESES – BAHIA GÁS – 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 18° 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: '18° 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 18° Celsius'.

RESOLUÇÃO:

- a) O examinador definiu o conceito de **dado**, não de conhecimento. Dados são fatos brutos a respeito do mundo, como textos, fotos, figuras ou sons, sem contexto ou significado próprio, que podem ser manipulados e descritos em máquina. **ERRADA**
- b) Essa alternativa pode causar alguma confusão. De fato, os conceitos de conhecimento e inteligência muitas vezes possuem alguns aspectos em comum. Contudo, veja que o examinador somente cita dados, informação e conhecimento como sendo o escopo da questão. Dito isso, temos que o **conhecimento** diz respeito à aplicabilidade das informações, sendo, portanto, o conceito mais adequadamente descrito pela alternativa. **ERRADA**
- c) Os dados são essas sequências de símbolos, figuras, sons, etc. que são mais facilmente processáveis por máquina. O exemplo de 18° Celsius deve estar se referindo a uma temperatura em algum local do mundo, mas veja que, sem contexto, ele não tem muito significado ou utilidade para o julgamento humano. Essa é a nossa resposta! **CERTA**
- d) Dados contextualizados dizem respeito à **informação**. Aparenta existir uma certa divergência entre pesquisadores a respeito da possibilidade de se armazenar e manipular informações em máquina. Contudo, de todo modo, já temos elementos suficientes para dizer que a alternativa não está correta. **ERRADA**
- e) Esse não pode ser um exemplo de dado, já que contém contexto e significado. O exemplo apresentado está mais alinhado à definição de **informação**. **ERRADA**

Gabarito: C

17. (FCC – TCE/GO – 2014)

No processo conhecido como “espiral do conhecimento”, a conversão do conhecimento de tácito para explícito é chamada de

- a) incorporação ou internalização.
- b) explicitação ou externalização.
- c) socialização ou compartilhamento.
- d) combinação ou mentalização.
- e) transformação ou reconfiguração

RESOLUÇÃO:

Na transferência de um meio implícito para um explícito, estamos tirando o conhecimento da cabeça de um indivíduo ou de uma cultura e formalizando em um documento ou meio físico, **explicitando-o**. Dessa maneira, mesmo que você não consiga se lembrar dos modos de transferência na espiral do conhecimento, vai conseguir deduzir que se trata da **externalização** ou **explicitação**.

Gabarito: B

18. (ESAF – MF – 2013)

Na Coluna A, à esquerda, estão os processos de conversão do conhecimento, e na Coluna B, à direita, como esses processos podem ser conceituados.

Coluna A**Coluna B**

AI – Socialização

BI – é o processo de sistematização de conjuntos diferentes de conhecimento explícito.

AII – Externalização

BII – é o processo de articulação do conhecimento tácito em conceitos explícitos, por meio de metáfora, analogias, símbolos.

AIII – Combinação

BIII – é o processo de compartilhamento de conhecimento tácito, por meio da experiência.

AIV – Internalização

BIV – está relacionado ao processo de “aprender fazendo”, por meio da aplicação do conhecimento formal nas atividades do trabalho.

Assinale a opção que indica as relações corretas.

- a) AI – BIV; AII – BIII; AIII – BII; AIV – BI.
- b) AI – BIII; AII – BII; AIII – BI; AIV – BIV.
- c) AI – BIV; AII – BIII; AIII – BI; AIV – BII.
- d) AI – BI; AII – BII; AIII – BIII; AIV – BIV.
- e) AI – BII; AII – BIV; AIII – BIII; AIV – BI.

RESOLUÇÃO:

AI – A socialização é o compartilhamento de conhecimento **tácito** para outra forma de conhecimento **tácito**. Assim, está associado ao treinamento de um servidor novo por parte de um mais experiente, por exemplo. Podemos deduzir que se trata da alternativa **BIII**.

AII – A externalização é a transferência do conhecimento **tácito** para **explícito**. Nesse processo podem ser utilizadas metáforas, analogias e símbolos. Logo, estamos falando da alternativa **BII**.

AIII – A combinação é a junção de conjuntos diferentes de conhecimento **explícito**. Isso ocorre através da **sistematização** e aplicação do conhecimento. Dessa maneira, podemos dizer que a opção correspondente é a **BI**.

AIV – A internalização é a conversão de conhecimento **explícito** em **tácito**. Está muito ligada ao aprendizado na prática, que ocorre muito nos ambientes de trabalho. É a nossa alternativa **BIV**.

Gabarito: B

19. (FAURGS – TJ/RS – 2018 – ADAPTADA)

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

RESOLUÇÃO:

Transformei a questão original em certo ou errado para dificultar um pouco. A primeira parte da questão está claramente correta. Os dados presentes nos bancos de dados tradicionais são um exemplo bem simples de dados estruturados.

Em seguida, a assertiva afirma que os dados não estruturados misturam tipos com valores de dados e que estes não precisam seguir uma estrutura fixa. Ora, essa definição diz respeito, na verdade, aos dados **semiestruturados**, cujo exemplo mais comum é o padrão XML. Por fim, o item diz que a informação acerca do tipo dos dados semiestruturados não existe, sendo esse tipo predominante na Web (internet).

Você pode ver que os conceitos de dados não estruturados e semiestruturados foram invertidos, tornando assim a afirmativa falsa.

Gabarito: E

20. (CESPE – CGM João Pessoa/PB – 2018)

Com relação à modelagem dimensional e à otimização de bases de dados para business intelligence, julgue o item subsequente.

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.

RESOLUÇÃO:

É isso mesmo. O ETL, colocando em termos gerais, extrai os dados de diversas fontes (ERP, arquivos txt, planilhas, etc.) e os insere em uma área de transição, ou staging area, onde é realizada a **transformação** desses dados. Essa transformação inclui limpeza, correções e a padronização dos dados para futura integração. Posteriormente, esses dados são carregados na base de dados ou qualquer estrutura de armazenamento de destino.

Gabarito: C

21. (CESPE – ANTAQ – 2014)

Acerca de ferramentas de BI (business intelligence), julgue o próximo item.

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.

RESOLUÇÃO:

A etapa de extração do módulo de ETL de um sistema de Business Intelligence se dedica a coletar dados de diversas fontes para posterior transformação e carregamento na base de destino. Você pode ter achado que a assertiva estava errada por dizer que são coletadas **informações** de diferentes fontes, mas a questão foi considerada correta.

Acontece que as bancas não costumam focar muito nesse aspecto formal da diferença entre dado, informação e conhecimento em questões que não dizem respeito a esse assunto. Então, minha recomendação é: a não ser que a questão trate explicitamente da diferença entre esses conceitos ou o erro seja muito óbvio, você pode considerar a diferença como sendo irrelevante para a resposta.

Gabarito: C

22. (CESPE – Banco da Amazônia – 2012)

Julgue o item que se segue:

O processo de extração, transformação e carga (ETL) ilustrado permite extrair dados de diversas fontes de dados e migrá-los para o data warehouse, mantendo sua estrutura e normalização originais.

RESOLUÇÃO:

Pense comigo. As fontes de dados em um processo de ETL costumam ser **heterogêneas**, variadas. Dessa maneira, não há como manter sua estrutura original. Aliás, o propósito do processo é justamente extrair os dados e transformá-los, preparando-os para a integração no seu destino, geralmente o data warehouse.

Gabarito: E

Lista de questões comentadas

1. (FCC – TCE/RS – 2018)

Julgue o item a seguir, acerca de dado, informação, conhecimento e inteligência.

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.

2. (CESPE – SEFAZ/RS – 2019)

A respeito do BI (business intelligence), assinale a opção correta.

- a) O BI consiste na transformação metódica e consciente das informações exclusivamente prestadas pelos tomadores de decisão em novas formas de conhecimento, para evolução dos negócios e dos resultados organizacionais.
- b) ETL é o processo de análise de dados previsto pela arquitetura de BI.
- c) As técnicas do BI objetivam definir regras para a formatação adequada dos dados, com vista a sua transformação em depósitos estruturados de informações, sem considerar a sua origem.
- d) O repositório de dados analíticos de BI é representado pelas diversas bases de dados relacionais e por repositórios de dados que utilizem modelagens relacionais.
- e) A camada de apresentação de uma arquitetura de BI é aquela em que as informações são organizadas e centralizadas.

3. (CESPE – EMAP – 2018)

Julgue o próximo item, a respeito de gestão do conhecimento.

Na solução de problemas, o conhecimento tácito é associado ao conhecimento do expert.

4. (CESPE – EMAP – 2018)

Julgue o próximo item, a respeito de gestão do conhecimento.

Em uma instituição, o conhecimento explícito está relacionado ao capital intangível.

5. (CESPE – EMAP – 2018)

Julgue o próximo item, a respeito de gestão do conhecimento.

No que se refere à espiral do conhecimento, socialização, externalização, modelagem e feedback são os diferentes modos de conversão do conhecimento.

6. (CESPE – TCE/PA – 2016)

Julgue o item subsecutivo, referente à gestão do ciclo de vida da informação — ILM (Information Lifecycle Management).

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.

7. (CESPE – TCE/SC – 2016)

Julgue o item a seguir, acerca de dado, informação, conhecimento e inteligência.

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.

8. (CESPE – TCE/SC – 2016)

Julgue o item a seguir, acerca de dado, informação, conhecimento e inteligência.

O atributo de inteligência depende mais da qualidade da informação disponível do que da sua quantidade, tendo, portanto, natureza qualitativa.

9. (FUMARC – PC/MG – 2018)

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.

10. (FGV – TRT/SC - 2017)

Cada vez mais as empresas trabalham com dados, informações e conhecimentos. Uma organização cria e utiliza conhecimento transformando-o de tácito em explícito e vice-versa

Um dos 4 modos de conversão do conhecimento é:

- a) socialização: de explícito para tácito, procura compartilhar e criar conhecimento através de experiência direta;
- b) externalização: de tácito para explícito, procura articular conhecimento explícito através do diálogo e da reflexão;
- c) combinação: de explícito para explícito, procura sistematizar e aplicar o conhecimento explícito e a informação;
- d) internalização: de implícito para tácito, procura aprender e adquirir novo conhecimento tácito na prática;
- e) transferência: de explícito para tácito, procura sistematizar e aplicar o conhecimento explícito e a informação.

11. (CESPE – MEC – 2011)

Julgue o próximo item, relativo à gestão do conhecimento:

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.

12. (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.

13. (CESPE – DPU – 2010)

A teoria de Nonaka e Takeuchi trata do processo de criação do conhecimento e preconiza a existência do conhecimento tácito e explícito. Acerca da criação do conhecimento e do processo de conversão do conhecimento, assinale a opção correta.

- a) No processo de conversão do conhecimento, a internalização corresponde à transformação do conhecimento tácito em conhecimento explícito.
- b) A socialização pode ser entendida como a conversão do conhecimento tácito em novo conhecimento tácito.
- c) A externalização corresponde à articulação do conhecimento tácito em conceitos implícitos.
- d) Programas de treinamento formal não podem ser considerados ferramentas de combinação do processo de conversão do conhecimento.
- e) O método de ensino denominado aprender fazendo provoca o processo de criação do conhecimento conhecido como socialização.

14. (FUNDATEC – IGP/RS - 2017)

Relacione a Coluna 1 à Coluna 2, associando o tipo de conhecimento (tácito ou explícito) às suas características.

Coluna 1

1. Conhecimento explícito.
2. Conhecimento tácito.

Coluna 2

- () Não é facilmente visível e explicável.
- () Pode ser expresso em palavras, números ou sons, compartilhado na forma de dados, fórmulas ou manuais.
- () Pode ser rapidamente transmitido aos indivíduos, formal ou sistematicamente.
- () Possui duas dimensões: a primeira é uma dimensão técnica, que engloba as habilidades informais e de difícil detecção, e a segunda consiste em crenças, percepções, ideais e valores.

A ordem correta de preenchimento dos parênteses, de cima para baixo, é:

- a) 1 – 1 – 2 – 2.
- b) 1 – 2 – 2 – 1.
- c) 2 – 1 – 1 – 2.

d) 2 – 2 – 1 – 1.

e) 2 – 1 – 2 – 1.

15. (UFMT – UFSBA – 2017 – ADAPTADA)

Julgue a afirmativa a seguir:

Informação refere-se a dados apresentados em uma forma significativa e útil para seres humanos e dados são sequências de fatos brutos que representam eventos que ocorrem nas organizações ou no ambiente físico, antes de terem sido organizados e arranjados de uma forma que as pessoas possam entendê-los e usá-los.

16. (IESES – BAHIA GÁS – 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’.

17. (FCC – TCE/GO – 2014)

No processo conhecido como “espiral do conhecimento”, a conversão do conhecimento de tácito para explícito é chamada de

- a) incorporação ou internalização.
- b) explicitação ou externalização.
- c) socialização ou compartilhamento.

- d) combinação ou mentalização.
- e) transformação ou reconfiguração

18. (ESAF – MF – 2013)

Na Coluna A, à esquerda, estão os processos de conversão do conhecimento, e na Coluna B, à direita, como esses processos podem ser conceituados.

Coluna A	Coluna B
AI – Socialização	BI – é o processo de sistematização de conjuntos diferentes de conhecimento explícito.
AII – Externalização	BII – é o processo de articulação do conhecimento tácito em conceitos explícitos, por meio de metáfora, analogias, símbolos.
AIII – Combinação	BIII – é o processo de compartilhamento de conhecimento tácito, por meio da experiência.
AIV – Internalização	BIV – está relacionado ao processo de “aprender fazendo”, por meio da aplicação do conhecimento formal nas atividades do trabalho.

Assinale a opção que indica as relações corretas.

- a) AI – BIV; AII – BIII; AIII – BII; AIV – BI.
- b) AI – BIII; AII – BII; AIII – BI; AIV – BIV.
- c) AI – BIV; AII – BIII; AIII – BI; AIV – BII.
- d) AI – BI; AII – BII; AIII – BIII; AIV – BIV.
- e) AI – BII; AII – BIV; AIII – BIII; AIV – BI.

19. (FAURGS – TJ/RS – 2018 – ADAPTADA)

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

20. (CESPE – CGM João Pessoa/PB – 2018)

Com relação à modelagem dimensional e à otimização de bases de dados para business intelligence, julgue o item subsequente.

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.

21. (CESPE – ANTAQ – 2014)

Acerca de ferramentas de BI (business intelligence), julgue o próximo item.

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.

22. (CESPE – Banco da Amazônia – 2012)

Julgue o item que se segue:

O processo de extração, transformação e carga (ETL) ilustrado permite extrair dados de diversas fontes de dados e migrá-los para o data warehouse, mantendo sua estrutura e normalização originais.

Gabarito

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

Resumo direcionado

Dado, informação, conhecimento, inteligência

Dado	Informação	Conhecimento
<i>Simples observações sobre o estado do mundo</i>	<i>Dados dotados de relevância e propósito</i>	<i>Informação valiosa da mente humana. Inclui reflexão, síntese, contexto</i>
- Facilmente obtido por máquinas	- Requer unidade de análise	- De difícil estruturação
- Frequentemente quantificado	- Exige consenso em relação ao significado	- De difícil captura em máquinas
- Facilmente transferível	- Exige necessariamente a mediação humana	- Frequentemente tácito
		- De difícil transferência

- **Inteligência:** Habilidade essencialmente humana, envolve exercício de ponderação para a tomada da melhor decisão. Envolve também noções de ética, bom e ruim, certo e errado.

Conversão do conhecimento



Dados estruturados, não estruturados e semiestruturados

Estruturados	Não Estruturados	Semiestruturados
Estrutura predefinida, esquema explícito	Sem estrutura definida	Estrutura definida <i>a posteriori</i> , irregular, parcial
Busca e processamento mais fáceis	Busca e processamento dificultados	Estrutura geralmente implícita nos próprios dados
Ex.: Bancos de dados tradicionais	Ex.: Imagens, áudio, vídeo, arquivos de texto	Ex.: Documentos XML

Dados abertos

Motivação

- Transparência;
- Obrigação legal;
- Melhoria do serviço público - serviços inovadores ao cidadão;
- Aprimoramento da qualidade dos dados governamentais;
- Viabilização de novos negócios.

Três Leis

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

Oito Princípios

1. Completos
2. Primários
3. Atuais
4. Acessíveis
5. Processáveis por máquina
6. Acesso não discriminatório
7. Formatos não proprietários
8. Livres de licença

Processo de Business Intelligence



Bibliografia

Ackoff, R. L. (1999). *Ackoff's Best*. New York: John Wiley & Sons.

Davenport, T. H. (1998). *Ecologia da Informação*. São Paulo: Editora Futura.

Drucker, P. (Janeiro de 1988). The Coming of the New Organization. *Harvard Business Review*.

Governo Federal - Ministério do Planejamento, Desenvolvimento e Gestão. (2018). *O que são dados abertos?* Fonte: Portal Brasileiro de Dados Abertos: <http://dados.gov.br/pagina/dados-abertos>

McGee, J. V., & Prusak, L. (1995). *Gerenciamento estratégico da informação: aumento a competitividade e a eficiência de sua empresa utilizando a informação como uma ferramenta estratégica*. Rio de Janeiro: Campus.

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge Creating Company*. Oxford: Oxford University Press.

Tauberer, J. (8 de Dezembro de 2007). *The Annotated 8 Principles of Open Government Data*. Fonte: The Annotated 8 Principles of Open Government Data: <https://opengovdata.org>