

## **Aula 09**

*Banco do Brasil (Escriturário - Agente de  
Tecnologia) Passo Estratégico de  
Tecnologia de Informação - 2023  
(Pós-Edital)*

Autor:

**Thiago Rodrigues Cavalcanti**

21 de Fevereiro de 2023

## 3. BIG DATA: FUNDAMENTOS; TÉCNICAS DE PREPARAÇÃO E APRESENTAÇÃO DE DADOS

### Sumário

Análise Estatística.....	1
Roteiro de revisão e pontos do assunto que merecem destaque .....	2
Big Data: Definições e conceitos .....	2
Falácias sobre Big Data .....	9
Aplicações de Big Data .....	11
Classificação de Big Data .....	12
Padrões atômicos e compostos de uma solução de Big Data .....	15
Hadoop.....	21
Aposta estratégica.....	23
Questões estratégicas .....	24

## ANÁLISE ESTATÍSTICA

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

Assunto	Quantidade	Grau de incidência em concursos similares
		CESGRANRIO
Modelagem conceitual de dados (a abordagem entidade-relacionamento); Modelo relacional de dados (conceitos básicos, normalização);	55	23,40%
Linguagem SQL2008;	49	20,85%



5. Estrutura de dados e algoritmos: Busca sequencial e busca binária sobre arrays; Ordenação (métodos da bolha, ordenação por seleção, ordenação por inserção), lista encadeada, pilha, fila e noções sobre árvore binária.	42	17,87%
6. Ferramentas e Linguagens de Programação para manipulação de dados: Ansible; Java (SE 11 e EE 8); TypeScript 4.0;	34	14,47%
Data Warehouse (modelagem conceitual para data warehouses, dados multidimensionais);	31	13,19%
Python 3.9.X aplicada para IA/ML e Analytics (bibliotecas Pandas, NumPy, SciPy, Matplotlib e Scikit-learn).	11	4,68%
2. Banco de Dados: Banco de dados NoSQL (conceitos básicos, bancos orientados a grafos, colunas, chave/valor e documentos);	4	1,70%
4. Desenvolvimento Mobile: linguagens/frameworks: Java/Kotlin e Swift. React Native 0.59; Sistemas Android api 30 e iOS xCode 10.	4	1,70%
3. Big data: Fundamentos; Técnicas de preparação e apresentação de dados.	3	1,28%
1. Aprendizagem de máquina: Fundamentos básicos; Noções de algoritmos de aprendizado supervisionados e não supervisionados	1	0,43%
Postgre-SQL;	1	0,43%
Noções de processamento de linguagem natural.	0	0,00%
Conceitos de banco de dados e sistemas gerenciadores de bancos de dados (SGBD); MongoDB;	0	0,00%

## ROTEIRO DE REVISÃO E PONTOS DO ASSUNTO QUE MERECEM DESTAQUE

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

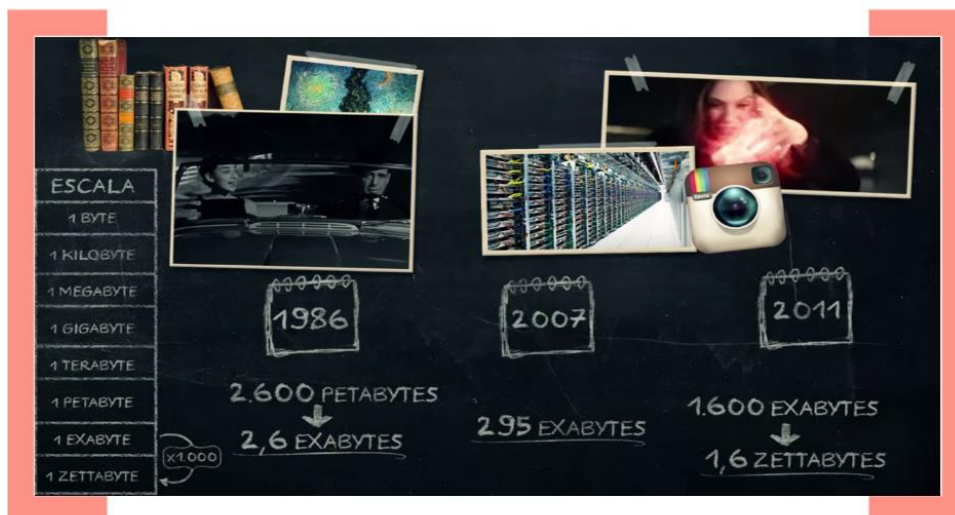
Para revisar e ficar bem preparado no assunto, você precisa, basicamente, seguir os passos a seguir:

### Big Data: Definições e conceitos

A humanidade, nos dias de hoje, produz uma quantidade diária de dados que é simplesmente improcessável pelos próprios seres humanos.

Para se ter uma ideia, a IBM, em 2013, estimou que **2,5 exabytes** (2.500.000.000.000.000.000) de bytes de dados são criados **por dia**. Se cada um dos 7 bilhões de habitantes tivesse que se debruçar sobre essas informações, seriam aproximadamente 300MB de dados diários para cada cidadão, incluindo bebês e idosos, rs.





São vídeos no youtube, postagens em redes sociais, blogs, portais de notícias, emails, dentre outros. E o que esses dados possuem em comum? São **dados não-estruturados**. Estima-se que 85% das informações com as quais as empresas lidam hoje não estão estruturadas. Desta forma, o SGBD tradicional e a modelagem relacional (datados da década de 60) não são mais suficientes para lidar com a realidade atual. É necessária uma nova abordagem. Surge então o conceito de Big Data!

**Big Data** pode ser entendido como a **captura, gerenciamento e análise de dados que vão além de dados estruturados** típicos, que podem ser consultados por sistemas de gerenciamento de banco de dados relacional — frequentemente em arquivos não estruturados, vídeo digital, imagens, dados de sensores, arquivos de log e, na verdade, qualquer dado não contido nos registros com campos pesquisáveis distintos.

Em um certo sentido, os dados não estruturados são dados interessantes, mas difíceis de sintetizar ou tirar conclusões deles, a menos que possam ser correlacionados a dados estruturados. Big data apresenta soluções para integrar os dados estruturados e desestruturados.

Em um primeiro momento, o Big Data pode até ser confundido com a *Business Intelligence*, mas difere na **ordem de grandeza** do volume de dados (que é muito maior), e **na natureza dos dados**. Enquanto as ferramentas de BI tradicionais extraem dados de fontes estruturadas, “abrindo exceções para a captura de dados não estruturados”, o Big Data entende que os dados não-estruturados são a “maioria”, por assim dizer.

Nossa ideia é começar o assunto apresentando, de forma sucinta, os conceitos básicos que permeiam o termo **Big Data**. Vamos começar pela definição que surgiu em 2001 no Meta Group (que viria a se juntar com a

Gartner<sup>1</sup> mais adiante) por meio do analista **Doug Laney**. Para compor o conceito ele se utilizou de três termos, conhecidos como os 3Vs: **Volume**, **Velocidade** e **Variedade**. Vejamos a definição de cada um deles.



**Volume.** Existem muitos fatores que contribuem para o aumento do volume de dados armazenados e trafegados. Podemos citar como exemplo: dados de transações armazenados ao longo de vários anos, dados de texto, áudio ou vídeo disponíveis em streaming nas mídias sociais e a crescente quantidade de dados coletados por sensores. No passado o volume de dados excessivo criou um problema de armazenamento. Mas, com os atuais custos de armazenamento decrescentes, outras questões surgem, incluindo, **como determinar a relevância** entre grandes volumes de dados e **como criar valor** a partir dessa relevância.

**Velocidade.** De acordo com o Laney, velocidade significa o quão **rápido** os dados estão sendo **produzidos** e o quão rápido os dados devem ser **tratados** para atender as demandas. Etiquetas de RFID e sensores inteligentes estão impulsionando uma necessidade crescente de lidar com dados quase em tempo real. Reagir rápido o suficiente para lidar com a velocidade é um desafio para a maioria das organizações.

**Variedade.** Os dados de hoje vêm em todos os **tipos de formatos**. Sejam bancos de dados tradicionais, hierarquias de dados criadas por usuários finais e sistemas OLAP, arquivos de texto, e-mail, medidores ou sensores de coleta de dados, vídeo, áudio, dados de ações do mercado e transações financeiras. Em algumas estimativas, 80% dos dados de uma organização não são numéricos! Mas, estes dados também precisam ser incluídos nas análises e nas tomadas de decisões das empresas.

Big Data é importante porque permite que as organizações recolham, armazenem, administrem e manipulem grandes quantidades de dados na velocidade certa, no tempo certo para conseguir os conhecimentos corretos. A novidade é que, pela primeira vez, o custo de ciclos de **computação e armazenamento** alcançou uma relação interessantes de custo x benefício. Por que isso é importante? Há alguns anos atrás, as empresas normalmente armazenariam “fotos” ou subconjuntos de informações importantes, **porque o custo de armazenamento e a limitação do processamento os impedia de armazenar tudo o que queriam analisar.**

Além da melhoria na capacidade de processamento e armazenamento, melhorias na velocidade de rede e confiança removeram outras limitações físicas da capacidade de administrar quantidades massivas de dados a

---

<sup>1</sup> Gartner é uma empresa de consultoria fundada em 1979 por Gideon Gartner. A Gartner desenvolve tecnologias relacionadas a introspecção necessária para seus clientes tomarem suas decisões todos os dias.

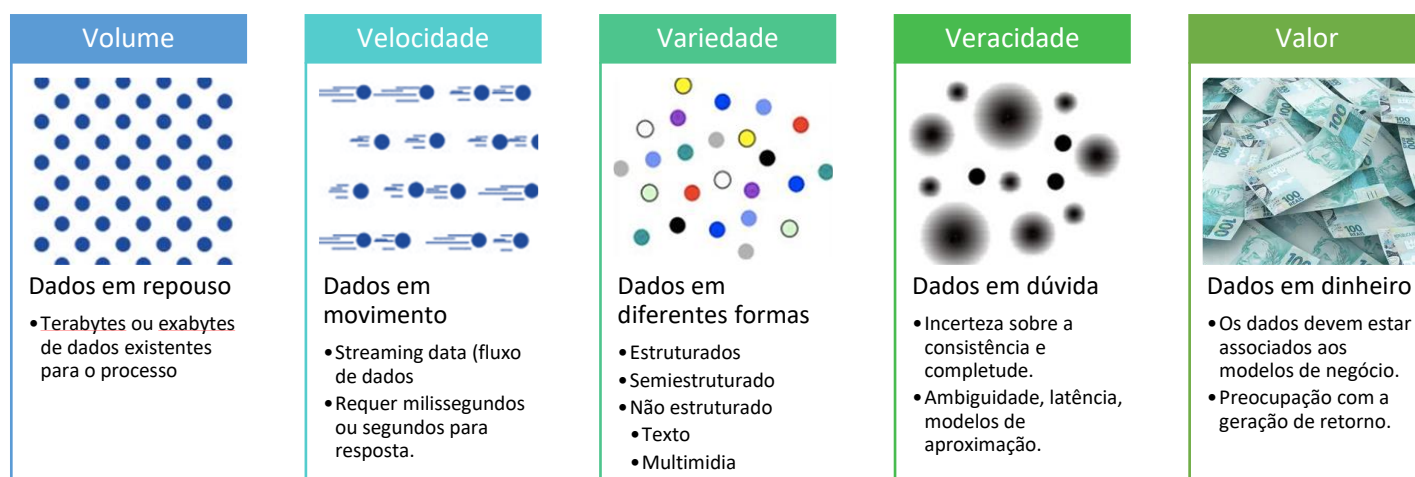


um passo aceitável. Assim, as organizações querem ser capazes de compreender e acionar resultados de diferentes tipos de dados na velocidade certa — não importa quantos dados estejam envolvidos.

*Se as empresas podem analisar petabytes de dados (o equivalente a 20 milhões de gabinetes, com quatro gavetas cheias de arquivos de texto ou 13.3 anos de conteúdo de HDTV) com desempenho aceitável para discernir padrões e anomalias, elas podem começar a compreender dados de novas maneiras.*

*Judith Hurwiz, et al - Big Data Para Leigos*

A mudança para Big Data **não é exclusivamente em negócios**. Ciência, pesquisa e atividades do governo também ajudaram a impulsioná-la. Pense sobre analisar o genoma humano ou lidar com todos os dados astronômicos coletados em observatórios para avançar nosso conhecimento do mundo à nossa volta. Considere também a quantidade de dados que o governo coleta em suas atividades antiterroristas e você entenderá a ideia de que Big Data não é só sobre negócios.



Vejam a figura apresentada acima. Ela representa uma indicação do conceito de Big Data relacionando os três V's sobre os quais tratamos até agora, adicionando duas características: **veracidade e valor**. Contudo, existem ainda outros V's, são eles **Visibilidade, Variabilidade e Visualização** que complementam o conceito, detalhando um pouco mais algumas características de Big Data. Vamos apresentar a definição destes termos no quadro abaixo.





**Veracidade.** A veracidade foi um termo cunhado pela IBM, considerado o quarto V, que representa **a falta de confiabilidade inerente** em algumas fontes de dados. Por exemplo, medir os sentimentos dos clientes em mídias sociais é incerto por natureza, já que implicam uso do juízo humano. No entanto, eles contêm valiosas informações. Assim, a necessidade de lidar com dados imprecisos e incertos é outra faceta de Big Data, geralmente resolvida usando ferramentas e análises desenvolvidas para gerenciamento e mineração de dados imprecisos.

É necessário avaliar as inconsistências, incompletudes, ambiguidades, latência e possíveis modelos de aproximação utilizados. Os dados podem ainda perder a vigência. Verificar se os dados são consistentes é extremamente necessário para qualquer análise de dados.

**Visibilidade.** É a relevância dos dados. A organização está ciente de todos os dados que ele gera? Estes poderiam ser (aparentemente) registros de dados inconsequentes. Em outras palavras tentamos entender se todos os dados gerados estão disponíveis, e se são de fato armazenados e ficam **visíveis para os analistas de dados**.

**Valor.** A Oracle introduziu valor como um atributo na definição de Big Data. Com base na definição da Oracle, *Big Data* é, muitas vezes, caracterizado por uma "densidade de valor relativamente baixa". Isto é, os dados recebidos na forma original, geralmente tem um valor baixo em relação ao seu volume. Entretanto, um valor elevado pode ser obtido pela análise de grandes volumes destes mesmos dados. Assim, as informações geradas **devem produzir algum valor** para as organizações.

**Variabilidade (e complexidade).** A SAS apresentou variabilidade (e complexidade) como duas dimensões adicionais para Big Data. Variabilidade refere-se **à variação nas taxas de fluxo de dados**. Muitas vezes, a velocidade de Big Data não é consistente e tem picos e depressões periódicas. **Complexidade** refere-se ao fato de Big Data gerar ou receber informações através de uma multiplicidade de fontes. Isso impõe um desafio crucial: a necessidade de se conectar, integrar, limpar e transformar os dados recebidos de diferentes fontes.

**Visualização.** É o modo como aqueles dados complexos serão representados ou apresentados.

Big data é um termo utilizado para descrever **grandes volumes de dados** e que ganha cada vez mais relevância à medida que a sociedade se depara com um aumento sem precedentes no número de informações geradas. As dificuldades em armazenar, analisar e utilizar grandes conjuntos de dados têm sido um considerável gargalo para as organizações. Vamos fazer mais uma questão recente sobre esses conceitos:



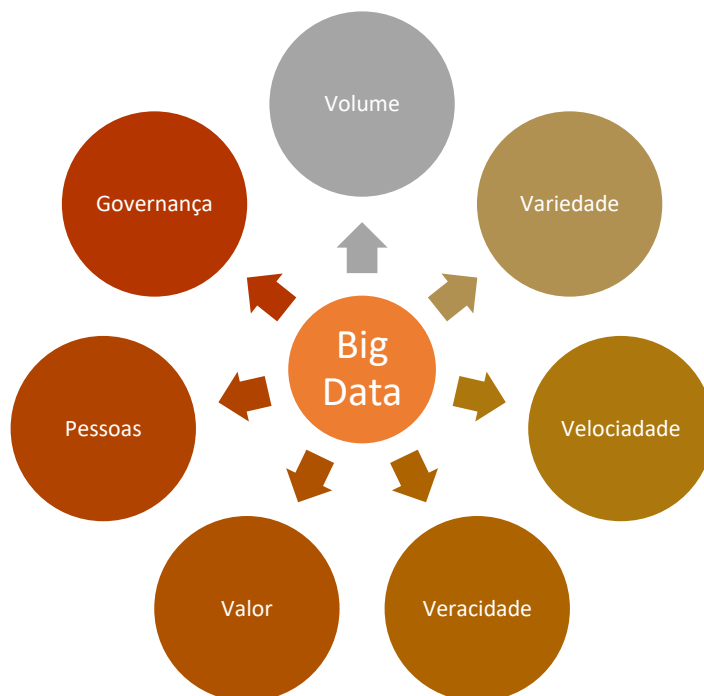
## Dimensões sobre os dados

A IBM cita atualmente 7 dimensões sobre os dados. Essas dimensões são uma outra forma de apresentar as características que vimos até o momento. As definições de velocidade, variedade, volume, valor e variedade são as mesmas neste contexto. Incluímos mais duas percepções que não foram listadas acima: **governança** e **pessoas**. Vejamos suas definições:

**Governança** - Ao decidir implementar ou não uma plataforma de big data, uma organização pode estar olhando novas fontes e novos tipos de elementos de dados nos quais a propriedade não está definida de forma clara. Por exemplo, no caso de assistência médica, é legal acessar dados de paciente para obter insight? É correto mapear as despesas do cartão de crédito do cliente para sugerir novas compras? Regras semelhantes regem todos os segmentos de mercado. Além da questão da governança de TI, também pode ser necessário redefinir ou modificar os processos de negócios de uma organização para que ela possa adquirir, armazenar e acessar dados externos.

**Pessoas** – É necessário ter pessoas com aptidões específicas para entender, analisar os requisitos e manter uma solução de Big Data. Envolve conhecimento do segmento de mercado, domínio técnico sobre as ferramentas de Big Data e conhecimentos específicos de modelagem, estatística e outros.

Essas duas dimensões, na percepção da IBM, juntamente com volume, variedade, velocidade, veracidade e valor dão viabilidade a um projeto de Big Data, podemos observar de forma organizada esses termos na figura abaixo.



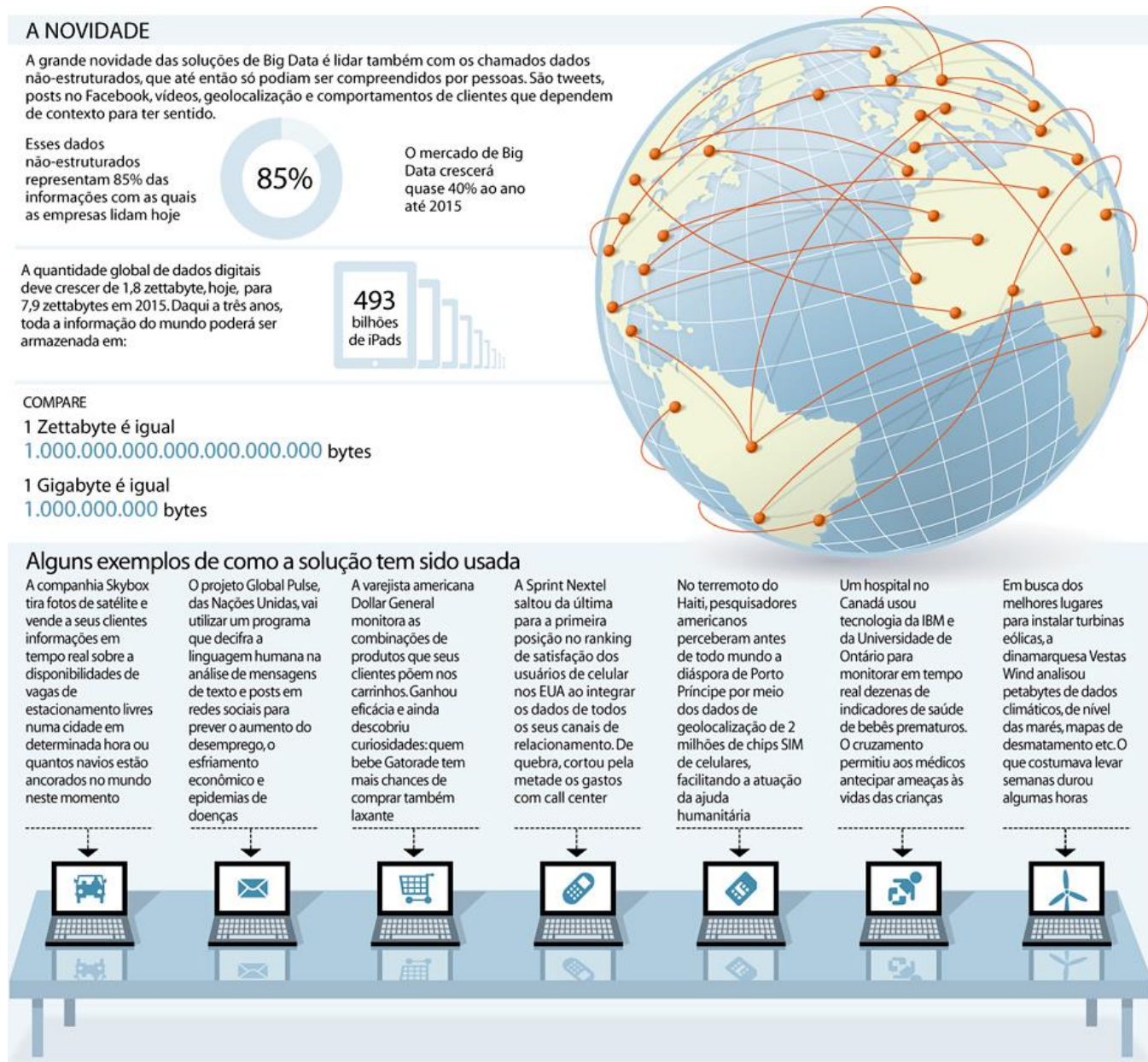


Big data trata, portanto, de grandes volumes, alta velocidade e variedade dos ativos de informação, procurando formas inovadoras e rentáveis de processamento da informação, visando uma melhor percepção dos fatos e uma tomada de decisão mais consistente.

Outra definição da **TechAmerica** define big data da seguinte forma:

*"Big Data é um termo que descreve grandes volumes de dados de alta velocidade, complexos e variáveis que exigem técnicas e tecnologias avançadas para permitir a captura, armazenamento, distribuição, gestão e análise da informação".*

Antes de seguirmos em frente veja no infográfico do Jornal O Globo algumas soluções corporativas que envolvem Big Data:



Big Data pelo mundo. Fonte: Jornal O Globo.



Entendido os conceitos básicos vamos avançar no assunto. Antes de falar sobre as premissas e aplicações, gostaria de tecer alguns comentários sobre falácias ou mitos associados a Big Data.

## Falácias<sup>2</sup> sobre Big Data

Quando pensamos em premissas sobre Big Data imaginamos uma caixa preta que vai receber dados de um lado e entregar algo pronto do outro. Nesta linha de raciocínio, provavelmente falaciosa, antes de apresentarmos as condições ideais para funcionamento dos sistemas ou projetos de Big Data, nós mostraremos os erros que muitos assumem como verdade. Chamaremos de falácias ou mitos sobre Big Data.

### Falácia 01 – Big Data engloba somente dados estruturados.

Com o crescente volume de dados, o banco de dados relacional precisou ser complementado com outras estruturas de armazenamento, devido principalmente à escalabilidade e flexibilidade das novas soluções tecnológicas. Entretanto, os dados relacionais continuam sendo valiosos e são muito utilizados em soluções de Big Data. O que mudou de fato foi a inclusão de mais tipos de dados, além dos estruturados. Lembre-se do conceito de **variedade**.

### Falácia 02 – Big Data refere-se somente a soluções com petabytes de dados

Embora o volume de dados seja o fator que impulsionou o fenômeno Big Data, aplicações que utilizam conjuntos de dados em uma escala menor do que petabytes também podem se beneficiar das tecnologias de Big Data. Afinal, **o mais importante nessas aplicações é a capacidade de extrair valor dos dados**.

### Falácia 03 – Big data é aplicado somente às empresas do Vale do Silício (Califórnia)

Quando se fala em Big Data, é normal associarmos os termos às grandes empresas de tecnologia que prestam serviços na Web, tais como Facebook, Twitter, Netflix, Google. Embora elas tenham sido as primeiras a serem desafiadas com o grande volume, variedade e velocidade de dados, atualmente empresas de outros domínios, como agricultura, varejo e logística, também se beneficiam das tecnologias de Big Data.

### Falácia 04 – Big Data é aplicado somente em grandes organizações

Ainda existe essa percepção de que Big Data oferece valor exclusivamente para grandes organizações. Entretanto, pequenas e médias empresas também podem obter vantagem competitiva por meio de soluções de Big Data, oferecendo uma melhor experiência aos seus clientes, otimizando processos, reduzindo custos ou criando novos produtos e serviços.

### Falácia 05 – Big Data requer uso de dados externos

<sup>2</sup> O termo falácia deriva do verbo latino *fallere*, que significa enganar. Designa-se por falácia um raciocínio errado com aparência de verdadeiro. Na lógica e na retórica, uma falácia é um argumento logicamente inconsistente, sem fundamento, inválido ou falho na tentativa de provar eficazmente o que alega.



Embora a adoção de dados de diferentes fontes seja uma prática muito adotada em soluções de Big Data, a aquisição de dados externos é um requisito obrigatório. Na verdade, a sugestão para quem inicia um projeto de Big data é buscar extrair valor primeiramente dos dados internos para somente depois ampliar suas jornadas a dados de terceiros.

#### Falácia 06 - Big Data pode prever o futuro

Big data e todas as suas ferramentas de análise, comentários, experiências científicas e visualizações não podem dizer o que vai acontecer no futuro. Por quê? Os dados que você coletar vem inteiramente do passado. Temos ainda de atingir um grau de evolução em que será possível coletar dados e os valores do futuro. Sendo assim, nós podemos analisar o que aconteceu no passado e tentar desenhar **as tendências** entre as ações e os pontos de decisão, e as suas consequências, baseadas nos dados. Podemos usar isso para **adivinhar** que, em **circunstâncias semelhantes**, se uma decisão semelhante for tomada, resultados semelhantes ocorreriam como resultado. Mas **não podemos prever o futuro**.

#### Falácia 07 - Big Data pode substituir seus valores ou os da sua organização

Big Data é pobre para substituir valores ou aqueles costumes e padrões pelos quais você vive sua vida e sua empresa se esforça para operar. Suas escolhas sobre essas questões podem ser bem cristalinas, e pode ser mais fácil e claro resolver as vantagens e desvantagens de diferentes cursos da ação, mas os dados em si não podem ajudá-lo a interpretar como as decisões certas se comparam com os padrões que você definiu para si e para a sua empresa.

Os dados podem descrever todos os tipos de cenários, tanto os próprios números quanto com a ajuda de software de visualização. Sua equipe pode criar muitas projeções de cenários sobre um determinado assunto, mas esses resultados são simplesmente isso - uma projeção. O trabalho de um executivo, como um CIO, é utilizar as ferramentas e pessoal disponível dentro de seu negócio, e realmente **reconciliar os dados contra os valores** da sua empresa.

#### Falácia 08 - Big Data pode resolver problemas não quantificáveis

Eis o velho ditado: Quando você só tem um martelo, tudo parece um prego. Uma vez que você começa a ter algum sucesso usando big data para prever e resolver problemas de negócios, haverá inevitavelmente uma tentação para "perguntar aos dados" toda vez que você tiver um problema ou um item sobre o qual a resolução não está clara.

Como mencionado anteriormente, os dados podem apresentar mais e melhores opções e, talvez, deixar claro o que pode acontecer com cada uma dessas escolhas. Às vezes, porém, os dados não são bons e isso ocorre quando ele é usado de forma individual.

Por quê? É quase impossível de quantificar o comportamento de um indivíduo. As pessoas têm seus próprios conjuntos de circunstâncias, os seus próprios universos, suas próprias razões e contextos. É impossível aplicar a matemática para um único indivíduo. Em vez disso, você tem que olhar para um grupo de indivíduos, de preferência um subgrupo com características semelhantes. Só então você pode observar as tendências de comportamento que se aplicam a todo o grupo.



Agora que sabemos o que pode e o que **não** pode ser feito com Big Data, vamos entender quais as premissas que devem ser verificadas quando decidimos pela implementação de um projeto de Big Data. Primeiro precisamos considerar os elementos fundamentais para o crescimento de Big Data tais como o aumento da **capacidade de armazenamento**, aumento do **poder de processamento** e **disponibilidade de dados**.

## Aplicações de Big Data

As aplicações de *Big Data* e análise de dados são as mais variadas. Uma lista retirada do livro do Aguinaldo Aragon nos traz as seguintes opções de uso: desenvolvimento de mercado, inovação, desenvolvimento de produtos e serviço, eficiência operacional, previsões de demanda de mercado, detecção de fraudes, gerenciamento de riscos, previsão de concorrência, vendas, campanhas de marketing, avaliação de desempenho de funcionários, alocação de orçamento anual, estabelecimento de previsões financeiras, gestão de planos de saúde, identificação de potenciais compradores e entendimento da base de cliente. Essa lista extensa nos mostra apenas parte das oportunidades geradas pela utilização de *Big Data*.

O uso de *Big Data* por empresas como **Amazon** e **Netflix** tem demonstrado como a mineração de dados pode gerar resultados surpreendentes. A partir destes dados, é possível conhecer melhor as escolhas dos usuários. Implementações dos conceitos de Big Data permitem, hoje, possibilidades quase infinitas, utilizando, por exemplo, mineração de dados conseguimos verdadeiros insights sobre os dados.

Um exemplo disso foi a pesquisa feita pela Consultoria MGI. Ela estudou dados em cinco grandes domínios - saúde nos Estados Unidos, o setor público na Europa, varejo nos Estados Unidos, dados de produção e dados de localização de pessoas em nível mundial. Big Data poderia gerar valor em cada um deles. Por exemplo, um varejista pode utilizar do conceito para aumentar sua margem operacional em mais de 60%.

O aproveitamento de **Big Data no setor público tem um potencial enorme** também. Se o programa de saúde dos Estados Unidos fosse usar big data de forma criativa e eficaz poderia impulsionar a eficiência e qualidade, o setor poderia criar mais de US\$ 300 bilhões em valor a cada ano. Dois terços dos quais seriam sob a forma de **redução nas despesas** de saúde dos EUA em cerca de 8 por cento. Nas economias desenvolvidas da Europa, os administradores do governo poderiam economizar mais de € 100 bilhões em melhorias de eficiência operacional usando big data, não incluindo o uso para reduzir a fraude e erros ou aumentar a cobrança das receitas fiscais.

A quantidade de dados em nosso mundo está explodindo, e a análise de grandes conjuntos de dados - os chamados big data - se tornará uma base essencial na concorrência pelo mercado ou na prestação de um serviço público de qualidade, apoiando novas ondas de crescimento na produtividade, inovação e expectativa dos consumidores. Líderes em todos os setores terão de lidar com as implicações de Big Data, não apenas os gestores orientados a dados, ou cientistas de dados.

O crescente volume e detalhamento das informações capturadas por empresas, o aumento de dados multimídia, mídia social, e a Internet das Coisas vão alimentar o crescimento exponencial dos dados no futuro.

A minha experiência pessoal no estudo de Big Data é que quando começamos a ler sobre o assunto esses conceitos teóricos vistos até aqui são apresentados de forma semelhante em vários artigos disponíveis na internet e em livros especializados, mas aí você deve estar fazendo a seguinte pergunta: E como eu implemento esse trem? (Homenagem aos meus amigos mineiros). E alguém vai te responder: usa uma base **NoSQL** com





a infraestrutura do **Hadoop**! Se você observar esses tópicos estão presentes nesta aula ... tenha um pouco de paciência que chegaremos neles.

## Classificação de Big Data

É possível categorizar problemas de negócios em **tipos** de problemas de big data. Quando problemas de big data são categorizados por tipo, é mais fácil ver as características de cada tipo de dados. Essas características ajudam a entender como os dados são obtidos, como são processados para o formato apropriado e com que frequência novos dados estão disponíveis.

Dados de diferentes fontes possuem características diferentes; por exemplo, dados de mídia social podem ter vídeos, imagens e texto não estruturado, como postagens de blog, entrando continuamente.

Quer conferir alguns exemplos? A tabela a seguir contém problemas comuns de negócios e atribui um tipo de big data a cada um.



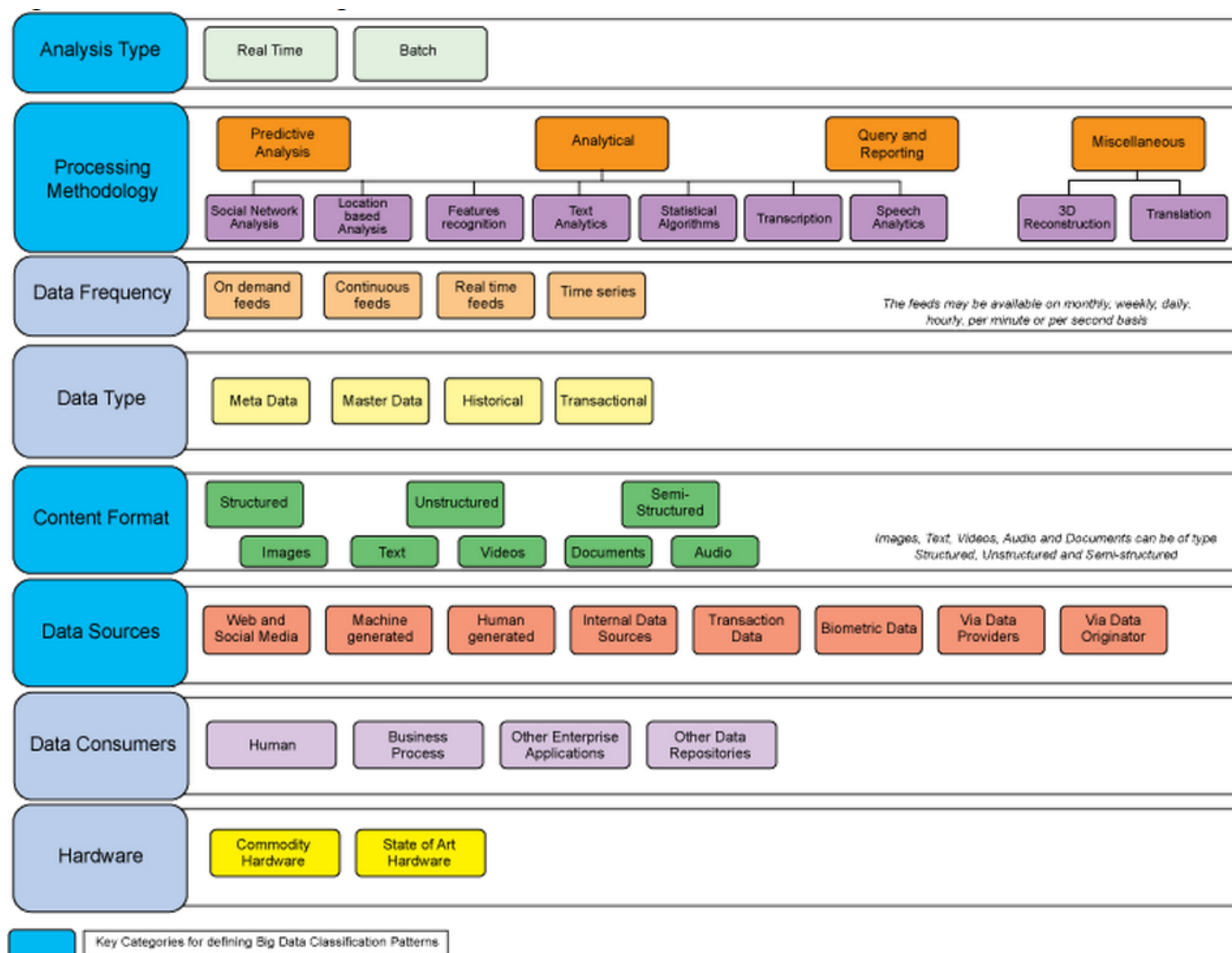
Problemas de negócios	Tipo de big data	Descrição
<b>Serviços públicos:</b> <b>Prever o consumo de energia</b>	Dados gerados por máquina	Concessionárias de serviços públicos implementaram medidores inteligentes para medir o consumo de água, gás e eletricidade a intervalos regulares de uma hora ou menos. Esses medidores inteligentes geram enormes volumes de dados de intervalo que precisam ser analisados.  Para ter eficiência operacional, a empresa precisa monitorar os dados entregues pelo sensor. Uma solução de big data pode analisar dados de geração de energia (fornecimento) e de consumo de energia (demanda) usando medidores inteligentes.
<b>Telecomunicações:</b> <b>Análítica de perda de clientes</b>	Dados da web e sociais  Dados de transação (operacionais)	Operadores de telecomunicações precisam criar modelos detalhados de perda de clientes que incluam dados de mídias sociais e de transação, para estar à frente da concorrência.  Provedores que implementam uma estratégia de analítica preditiva podem gerenciar e prever a perda analisando os padrões de chamada dos assinantes.
<b>Varejo: Sistema de mensagens personalizado com base</b>	Dados da web e sociais	Varejistas podem usar tecnologia de reconhecimento facial combinada a uma foto da mídia social para



em reconhecimento facial e mídia social	Biométrica	<p>fazer ofertas personalizadas a clientes com base no comportamento de compra e na localização.</p> <p>Esse recurso pode ter um impacto tremendo nos programas de fidelização dos varejistas, mas há sérias considerações sobre a privacidade. Os varejistas precisariam ser transparentes com relação à privacidade para implementar esses aplicativos.</p>
---	------------	---

Problemas de negócios de big data por tipo. Fonte: <http://www.ibm.com/developerworks/br/library/bd-archpatterns1/>

A figura a seguir mostrará as várias categorias ou taxonimias que podemos usar para classificar Big Data, e as possíveis divisões ou grupos em cada categoria. As categorias mais relevantes estão em azul turquesa.



Falando um pouco mais sobre as classificações:





**Tipo de análise** — Se os dados são analisados em **tempo real** ou agrupados para análise posterior. Essa escolha afeta várias outras decisões sobre produtos, ferramentas, hardware, fontes de dados e a frequência estimada dos dados. Para alguns casos de uso é necessária uma mistura dos dois tipos.

**Metodologia de processamento** — O tipo de técnica a ser aplicada para processar dados (por exemplo, preditiva, analítica, consulta ad hoc e relatórios). As necessidades de negócios determinam a metodologia de processamento apropriada. É possível usar uma combinação de técnicas. A escolha de metodologia de processamento ajuda a identificar as ferramentas e técnicas apropriadas para uso na solução de big data.

**Frequência e tamanho dos dados** — O volume estimado de dados e a frequência com que chegam. Saber a frequência e o tamanho ajuda a determinar o mecanismo de armazenamento, formato de armazenamento e as ferramentas necessárias de pré-processamento. Frequência e tamanho de dados dependem das fontes.

- Sob demanda, como dados de mídia social
- Feed contínuo, em tempo real (dados de clima ou transacionais)
- Série temporal (dados com base em tempo)

**Tipo de dados** — Tipo dos dados a serem processados — transacionais, históricos, principais e outros. Saber o tipo de dados ajuda a segregar os dados no armazenamento.

**Formato de conteúdo** — Formato dos dados recebidos — estruturados (SGBDR, por exemplo), não estruturados (áudio, vídeo e imagens, por exemplo) ou semiestruturados. O formato determina como os dados recebidos precisam ser processados e é essencial para escolher ferramentas e técnicas e definir uma solução de uma perspectiva de negócios.

**Fonte de dados** — Fontes de dados (onde os dados são gerados) — web e mídia social, gerados por máquina, gerados por humanos, etc. Identificar todas as fontes de dados ajuda a determinar o escopo de uma perspectiva de negócios.

**Consumidores de dados** — Uma lista de todos os possíveis consumidores dos dados processados:

- Processos de negócios
- Usuários corporativos
- Aplicativos corporativos
- Pessoas individuais em várias funções de negócios
- Parte dos fluxos do processo
- Outros repositórios de dados ou aplicativos corporativos

**Hardware** — O tipo de hardware no qual a solução de big data será implementada — hardware barato ou de ponta. Entender as limitações do hardware ajuda na escolha da solução big data.



## Padrões atômicos e compostos de uma solução de Big Data

Os padrões auxiliam a definir os parâmetros, quando da adoção de uma solução de big data. Veremos dois tipos principais: os padrões **atômicos** descrevem as abordagens típicas para o consumo, processamento, acesso e armazenamento de big data; os padrões **compostos**, que são formados por padrões atômicos, são classificados de acordo com o escopo da solução de big data.

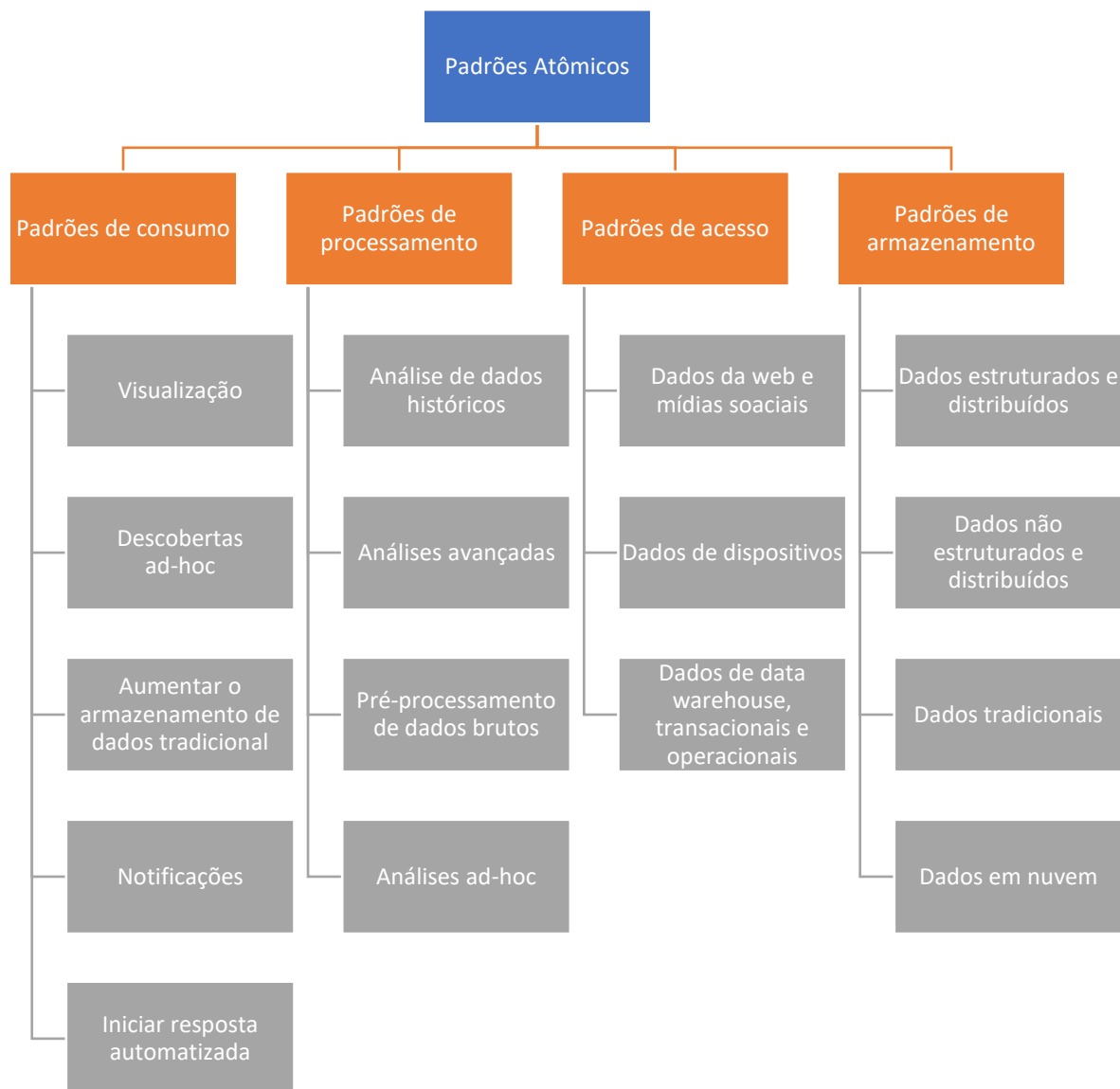
Por apresentarem as ideias mais relevantes acerca do Big Data, exploraremos os padrões atômicos.

### PADRÕES ATÔMICOS

Os **padrões atômicos** ajudam a identificar a forma que os dados são consumidos, processados, armazenados e acessados por problemas de big data. Eles também podem ajudar a identificar os componentes necessários.

Cada padrão lida com requisitos específicos — visualização, análise de dados históricos, dados de mídia social e armazenamento de dados não estruturados, por exemplo. Os padrões atômicos podem trabalhar em conjunto para criar um padrão composto. Não há camadas ou sequência para esses padrões atômicos. Por exemplo, os padrões de visualização podem interagir com os padrões de acesso a dados para mídia social diretamente e os padrões de visualização podem interagir com o padrão de processamento de análise avançada.





**Padrões atômicos de Big Data. Fonte:** <http://www.ibm.com/developerworks/br/library/bd-archpatterns4/>

Vejamos um pouco de cada padrão atômico:

## PADRÕES DE CONSUMO

Lidam com as várias formas em que o resultado da análise de dados é consumido. Inclui padrões de consumo de dados para atender a diversos requisitos. Vejamos os principais padrões de consumo a seguir:

### Visualização

A forma tradicional de visualizar dados se baseia em gráficos, painéis e relatórios de resumo. Essas abordagens tradicionais não são sempre a melhor maneira de visualizar os dados.

Os requisitos típicos para visualização de big data, incluindo os requisitos emergentes, são listados abaixo:



- Realizar análise em tempo real e exibição de dados de fluxo
- Extrair dados de forma interativa, com base no contexto
- Executar procuras avançadas e obter recomendações
- Visualizar informações paralelamente
- Ter acesso a hardware avançado para necessidades de visualização futuristas

A pesquisa para determinar como os insights de big data podem ser consumidos por humanos e máquinas está em andamento. Os desafios incluem o volume de dados envolvido e a necessidade de associar contexto a eles. O insight deve ser apresentado no contexto adequado.

#### Descoberta ad hoc

Criar de relatórios padrão que sejam adequados para todas as necessidades de negócios, via de regra, não é viável, pois as empresas têm requisitos de consultas de dados de negócios diversas. Os usuários precisam da capacidade de enviar consultas *ad hoc*, ou seja, consultas criadas “na hora”, ao procurar por informações específicas, dependendo do problema.

#### Aumentar os armazenamentos de dados tradicionais

Aumentar os armazenamentos de dados existentes ajuda a ampliar o escopo de dados disponível para a analítica atual para incluir dados que residem dentro e fora dos limites organizacionais, como dados de mídia social, que podem melhorar os dados principais. Ao ampliar o escopo para incluir novas tabelas de fatos, dimensões e dados principais nos armazenamentos existentes e adquirir dados de clientes a partir de mídia social, uma organização pode obter um insight mais profundo do cliente.

#### Notificação

Os insights de big data permitem que as pessoas, negócios e máquinas ajam instantaneamente usando notificações para indicar eventos. A plataforma de notificação deve ser capaz de lidar com o volume antecipado de notificações a serem enviadas de maneira oportuna. Essas notificações são diferentes das malas diretas ou do envio em massa de mensagens SMS, pois o conteúdo geralmente é específico para o consumidor. Por exemplo, os mecanismos de recomendação podem fornecer insights sobre a enorme base de clientes em todo o mundo, e as notificações podem ser enviadas para tais clientes.

#### Iniciar uma resposta automatizada

Os insights de negócios derivados do big data podem ser usados para acionar ou iniciar outros processos de negócios ou transações.

### **PADRÕES DE PROCESSAMENTO**

O big data pode ser processado quando os dados estão em repouso ou em movimento. Dependendo da complexidade da análise, os dados podem não ser processados em tempo real. Esse padrão lida com como o



big data é processado em tempo real, quase em tempo real ou em lote (rotinas *batch*, processadas em horários pré-determinados).

Vejamos um pouco mais sobre esses padrões a seguir:

### Análise de dados históricos

A análise de dados históricos tradicional é limitada a um período predefinido de dados, que normalmente depende das políticas de retenção de dados. Após desse período, geralmente os dados são arquivados ou limpos em virtude de limitações de armazenamento e processamento.

A análise histórica envolve analisar as tendências históricas para um determinado período, conjunto de períodos e produtos e compará-las aos dados atuais disponíveis.

### Analítica Avançada

O big data fornece enormes oportunidades de obter insights criativos. É possível correlacionar diferentes conjuntos de dados em muitos contextos. A descoberta desses relacionamentos requer técnicas e algoritmos complexos inovadores.

A análise avançada inclui previsões, decisões, processos inferenciais, simulações, identificações de informações contextuais e resoluções da entidade. Os aplicativos de analítica avançada incluem análise de dados biométricos, por exemplo, análise de DNA, análise espacial, analítica baseada em localização, análise científica, pesquisa e muitas outras. A analítica avançada requer a computação exigente para gerenciar a enorme quantidade de dados.

### Pré-processar dados brutos

A extração de dados a partir de dados não estruturados, como imagens, áudio, vídeo, feeds binários ou até mesmo texto, é uma tarefa complexa e precisa de técnicas como aprendizado de máquina e processamento de idioma natural, etc. O outro grande desafio é como verificar a precisão e a exatidão do resultado de tais técnicas e algoritmos.

**Para executar a análise em quaisquer dados, eles devem estar em algum tipo de formato estruturado.** Os dados não estruturados acessados de várias fontes podem ser armazenados como estão e, em seguida, transformados em dados estruturados e novamente armazenados nos sistemas de armazenamento de big data. O texto não estruturado pode ser convertido em dados estruturados ou semiestruturados. Da mesma forma, os dados de imagem, áudio e vídeo precisam ser convertidos nos formatos que podem ser usados para análise. Além disso, a precisão e exatidão da analítica avançada que usa algoritmos preditivos e estatísticos dependem da quantidade de dados e algoritmos usados para treinar os modelos.

### Análise ad hoc

O processamento de consultas ad hoc no big data traz desafios diferentes daqueles incorridos ao realizar consultas ad hoc em dados estruturados pelo fato de as fontes e formatos dos dados não serem fixos e exigirem mecanismos diferentes para recuperá-los e processá-los.



Embora as consultas ad hoc simples possam ser resolvidas pelos provedores de big data, na maioria dos casos, elas são complexas porque os dados, algoritmos, formatos e resoluções da entidade devem ser descobertos dinamicamente. O conhecimento dos cientistas de dados e dos usuários corporativos é necessário para definir a análise exigida para as seguintes tarefas:

- Identificar e descobrir os cálculos e algoritmos
- Identificar e descobrir as fontes de dados
- Definir os formatos necessários que podem ser consumidos pelos cálculos
- Executar os cálculos nos dados paralelamente

## PADRÕES DE ACESSO

Existem muitas fontes de dados e formas em que os dados podem ser acessados em uma solução de big data, vejamos as mais comuns:

### Web e mídias sociais

A Internet é a fonte de dados que fornece muitos dos insights produzidos atualmente. A web e a mídia social são úteis em praticamente todas as análises, mas são necessários mecanismos de acesso diferentes para obter esses dados.

A web e a mídia social são a fonte de dados mais complexa de todas em virtude de sua enorme variedade, velocidade e volume. Há aproximadamente de 40 a 50 categorias de websites e cada uma exigirá um tratamento diferente para acessar esses dados.

### (gerados por) Dispositivos

O conteúdo gerado por dispositivos inclui dados de sensores. Os dados são detectados a partir das origens de dados, como informações sobre o clima, medições elétricas e dados sobre poluição, e capturados pelos sensores. Os dados podem ser fotos, vídeos, texto e outros formatos binários.

### Dados transacionais, operacionais e de Warehouse

É possível armazenar os dados operacionais e transacionais em warehouse existentes para evitar a limpeza ou o arquivamento deles (em virtude de limitações de armazenamento e processamento) ou para reduzir a carga no armazenamento tradicional quando os dados são acessados por outros consumidores.

Os dados transacionais podem ser inseridos no armazenamento de warehouse usando conectores padrão disponibilizados por diversos fornecedores de banco de dados. O pré-processamento de dados transacionais é muito mais fácil, pois a maior parte deles é estruturada. Os processos de extração, transformação e carregamento simples podem ser usados para mover os dados transacionais para o armazenamento em um *data warehouse*.

## PADRÕES DE ARMAZENAMENTO





Os padrões de armazenamento auxiliam a determinar o armazenamento adequado para diversos formatos e tipos de dados. Os dados podem ser armazenados como estão, com relação a pares de valores de chave ou em formatos predefinidos. Vejamos os principais padrões:

### Dados não estruturados e distribuídos

A maior parte do big data não é estruturada, já sabemos, e pode conter informações que podem ser extraídas de diferentes formas para diferentes contextos. Na maioria das vezes, os dados não estruturados devem ser armazenados como estão, em seu formato original.

Tais dados podem ser armazenados em sistemas de arquivos distribuídos, como **HDFS (Hadoop Distributed File System)**, e em armazenamento de documentos **NoSQL (Not Only SQL)**, como o **MongoDB**. Esses sistemas fornecem uma maneira eficiente de recuperar dados não estruturados.

### Dados estruturados e distribuídos

Os dados estruturados incluem aqueles que chegam da fonte de dados e já estão em um formato estruturado e os dados não estruturados que foram pré-processados. Esses dados convertidos devem ser armazenados para evitar a frequente conversão de dados brutos para dados estruturados.

Tecnologias como **BigTable** do Google são usadas para armazenar dados estruturados. O BigTable é um sistema de autogerenciamento tolerante a falhas de grande escala que inclui terabytes de memória e petabytes de armazenamento.

### Armazenamento de dados tradicionais

O armazenamento de dados tradicional não é a melhor opção para armazenar big data, mas nos casos em que as empresas estão realizando a exploração de dados inicial, elas podem optar por usar o data warehouse, o sistema **RDBMS** (sistemas relacionais) e outros armazenamentos de conteúdo existentes. Esses sistemas de armazenamento existentes podem ser usados para armazenar os dados que são compilados e filtrados usando a plataforma de big data. **Os sistemas de armazenamento de dados tradicionais não são adequados para o big data.**

### Armazenamento na nuvem

Muitos provedores de infraestrutura da nuvem possuem recursos de armazenamento estruturado e não estruturado distribuídos. As tecnologias de big data são um pouco diferentes das perspectivas de configurações, manutenção, gerenciamento de sistemas e programação e modelagem tradicionais. Além disso, as qualificações necessárias para implementar as soluções de big data são raras e caras. As empresas explorando as tecnologias de big data podem usar soluções de nuvem que fornecem o gerenciamento de sistemas, manutenção e armazenamento de big data.

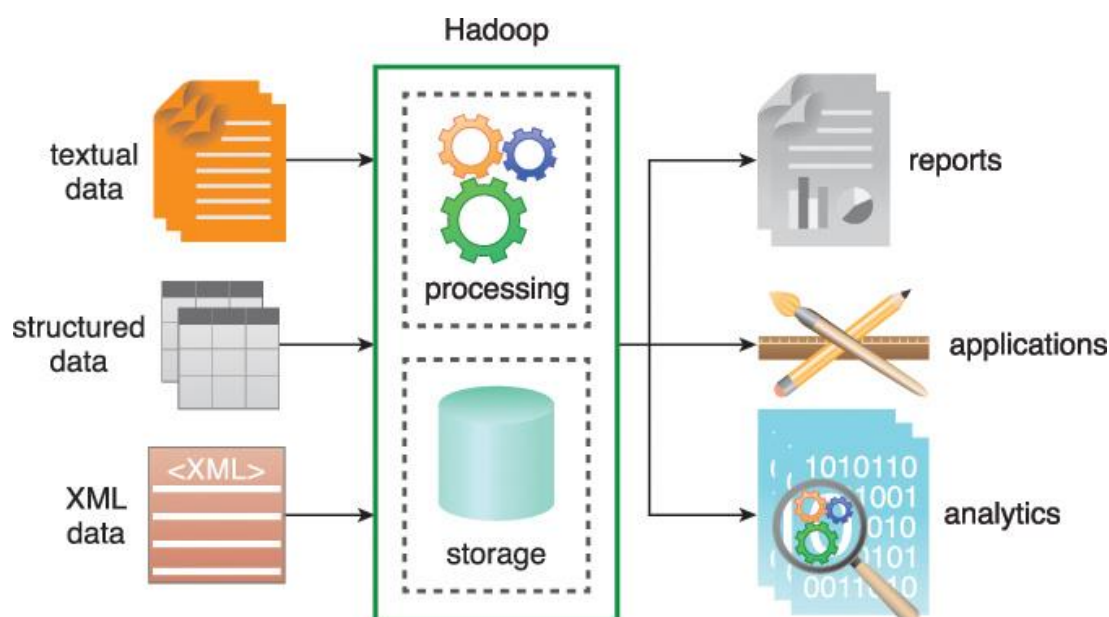
Contudo, não-raro, os dados a serem armazenados são confidenciais, incluindo dados biométricos e registros médicos. **A segurança de dados, o compartilhamento de dados, a governança de dados e outras políticas relacionadas aos dados, são aspectos a serem considerados ao ponderar a nuvem como um repositório de**



armazenamento para big data. A capacidade de transferir enormes quantidades de dados também é outra consideração fundamental para o armazenamento em nuvem.

## Hadoop

Hadoop é um framework de código aberto para armazenamento de dados em grande escala e processamento de dados que seja compatível com o hardware específico. O framework Hadoop se estabeleceu como uma plataforma da indústria de fato para soluções de Big Data contemporâneas. Ele pode ser usado como um motor de ETL ou como um mecanismo de análise para o processamento de grandes quantidades de dados estruturados, semiestruturados e não estruturados. A partir de uma perspectiva de análise, Hadoop implementa a estrutura de processamento de MapReduce. A figura abaixo ilustra algumas das características do Hadoop.



Hadoop é usado amplamente na indústria para processamento de grande escala, massivamente paralelo e distribuído. Hadoop é altamente tolerante a falhas e configurável para tantos níveis quanto precisarmos. O que tem um impacto direto no número de vezes que os dados são armazenados.

Como já abordado, em sistemas de Big Data, a arquitetura gira em torno de dois componentes principais: a **computação distribuída** e **processamento paralelo**. No Hadoop, a computação distribuída é tratada por meio do **HDFS** e processamento paralelo é tratado pelo **MapReduce**. Em suma, podemos dizer que o Hadoop é uma combinação de HDFS e MapReduce, como mostrado na imagem a seguir:





Hadoop tem uma série de vantagens, e algumas delas são:

**Baixo custo** - Funciona em hardware simples (commodity): Hadoop pode ser executado em hardware simples e não requer um sistema de alto desempenho, o que pode ajudar no controle de custos e, mesmo assim, alcançar escalabilidade e desempenho. Adicionar ou remover nós do cluster é simples. O custo por terabyte é mais baixo para armazenamento e processamento no Hadoop.

**Flexibilidade de armazenamento:** Hadoop pode armazenar dados em formato plano (raw) em um ambiente distribuído. Hadoop pode processar os dados não estruturados e dados semiestruturados melhor que a maioria das tecnologias disponíveis.

**Comunidade de código aberto:** Hadoop é de código aberto e apoiado por muitos contribuintes com uma crescente rede de desenvolvedores em todo o mundo. Muitas organizações, tais como Yahoo, Facebook, Hortonworks, e outros têm contribuído imensamente para o progresso do Hadoop e outros subprojetos relacionados.

**Tolerante a falhas:** Hadoop é altamente escalável e tolerante a falhas. Hadoop é confiável em termos de disponibilidade de dados, e até mesmo se alguns nós falharem, Hadoop pode recuperar os dados. A arquitetura assume que os nós podem falhar e o sistema deve ser capaz de continuar a processar os dados.

**Análise de dados complexos:** Com o surgimento de Big Data, a ciência de dados também tem crescido aos trancos e barrancos, e nós temos de algoritmos intensivos complexos e pesados para análise de dados. Hadoop pode processar tais algoritmos para um conjunto de dados muito grande.

Alguns exemplos onde o Hadoop pode ser utilizado são os seguintes: pesquisa ou mineração de texto, processamento de registros, sistemas de recomendação, inteligência de negócios, análise de vídeo e imagem, reconhecimento de padrões, avaliação de risco e análise de sentimentos.

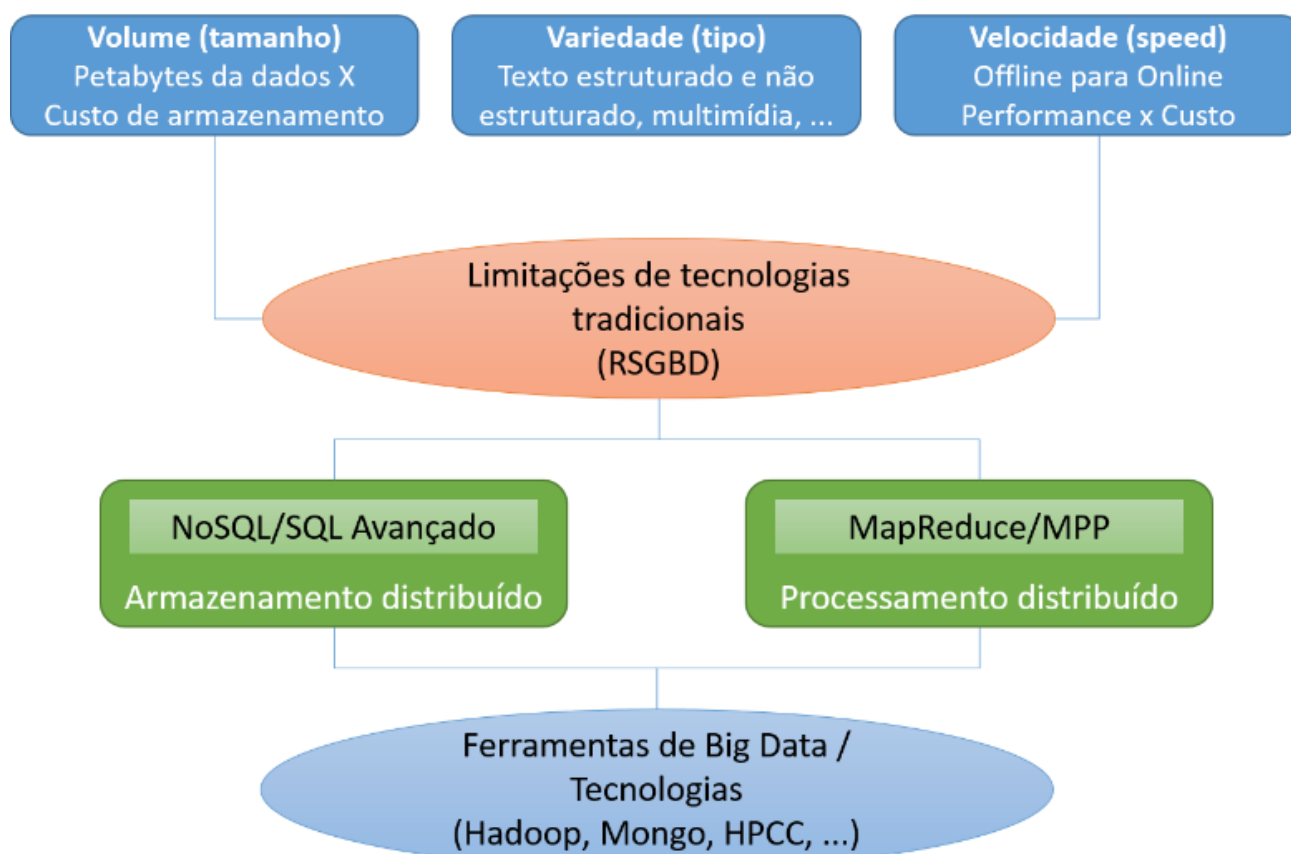


## APOSTA ESTRATÉGICA

A ideia desta seção é apresentar os pontos do conteúdo que mais possuem chances de serem cobrados em prova, considerando o histórico de questões da banca em provas de nível semelhante à nossa, bem como as inovações no conteúdo, na legislação e nos entendimentos doutrinários e jurisprudenciais<sup>3</sup>.



### Conceitos de Big Data



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



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

## QUESTÕES ESTRATÉGICAS

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

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



### 1. CESGRANRIO - Administrador (PETROBRAS)/Júnior/2018

O termo Big Data é bastante conhecido pelos profissionais de tecnologia da informação, especialmente aqueles envolvidos com bancos de dados, inteligência de negócios, sistemas de informações e sistemas de apoio à decisão.

Uma característica inerente a esse conceito é a da

- A) complexidade das suas fontes de informação, o que demanda a necessidade de sua prévia limpeza, integração e transformação.
- B) estabilidade da taxa de geração desses dados, o que garante sua utilização confiável na geração analítica de informação com independência temporal.
- C) heterogeneidade do conjunto de dados, empregada em dados originalmente estruturados ou semiestruturados.
- D) qualidade das fontes de dados, por conta dos padrões de expansão e de retenção reveladores da ordem existente nos dados.
- E) escalabilidade, que, na sua forma original, possui alto valor granular quando comparado ao de seu volume.

### Comentários

Big Data pode ser entendido como a captura, gerenciamento e análise de dados que vão além de dados estruturados típicos, que podem ser consultados por sistemas de gerenciamento de banco de dados relacional



— frequentemente em arquivos não estruturados, vídeo digital, imagens, dados de sensores, arquivos de log e, na verdade, qualquer dado não contido nos registros com campos pesquisáveis distintos.

Em um certo sentido, os dados não estruturados são dados interessantes, mas difíceis de sintetizar ou tirar conclusões deles, a menos que possam ser correlacionados a dados estruturados. Big data apresenta soluções para integrar os dados estruturados e desestruturados.

Big data é um termo utilizado para descrever grandes volumes de dados e que ganha cada vez mais relevância à medida que a sociedade se depara com um aumento sem precedentes no número de informações geradas. As dificuldades em armazenar, analisar e utilizar grandes conjuntos de dados têm sido um considerável gargalo para as organizações.

Diante disso, vamos analisar as alternativas:

A) É comum existirem diversas fontes de informação, o que traz uma grande complexidade. Muitas vezes, são necessárias atividades de limpeza (para remover ruídos e outliers, por exemplo), integração (para integrar os dados, afinal existem fontes diferentes) e transformação (pois nem todos os dados chegam padronizados). **GABARITO.**

B) Não existe nenhuma necessidade de estabilidade da taxa geração de dados. Os dados são gerados de acordo com a realidade do negócio em questão, sendo muito comum, inclusive, que os dados não sejam gerados a uma taxa estável e constante. **ERRADO.**

C) Não existe a necessidade de os dados serem heterogêneos (diversos, variados) nem que sejam estruturados ou semiestruturados. Cada aplicação tem uma realidade diferente, podendo existir casos em que o conjunto de dados é homogêneo e que os dados não estejam estruturados. **ERRADO.**

D) De fato, o ideal é que as fontes de dados sejam confiáveis e de boa qualidade. No entanto, a alternativa faz uma confusão ao dizer “por conta dos padrões de expansão e de retenção reveladores da ordem existente nos dados”, não tendo relação alguma com o Big Data. **ERRADO.**

E) O que pode ou não ser escalável é uma aplicação que utiliza uma quantidade massiva de dados, por exemplo. Mas isso é uma preocupação posterior, não fazendo sentido falar de escalabilidade no conceito de Big Data. **ERRADO.**

**Gabarito: alternativa A.**

---

## 2. CESGRANRIO - Analista (PETROBRAS)/Sistema Júnior/2018

A principal definição de Big Data parte de três características, conhecidas como 3 V do Big Data, a saber: velocidade, variedade e volume.

O termo velocidade refere-se, principalmente, à

A) necessidade das aplicações de gerar respostas rapidamente, a partir de grandes massas de dados.

B) existência de um alto fluxo de dados na entrada.





- C) necessidade de gerar aplicações rapidamente, em função da demanda do negócio.
- D) importância da facilidade de manipular cubos de visualização de dados, rapidamente.
- E) rapidez com que os dados se tornam inválidos com o tempo.

### Comentários

De acordo com o Laney, velocidade significa o quão **rápido** os dados estão sendo **produzidos** e o quão rápido os dados devem ser **tratados** para atender as demandas. Etiquetas de RFID e sensores inteligentes estão impulsionando uma necessidade crescente de lidar com dados quase em tempo real. Reagir rápido o suficiente para lidar com a velocidade é um desafio para a maioria das organizações.

Portanto, a alternativa correta é a letra B.

**Gabarito: alternativa B.**

---

...

Forte abraço e bons estudos.

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

(Bill Gates)

## Thiago Cavalcanti



**Face:** [www.facebook.com/profthiagocavalcanti](https://www.facebook.com/profthiagocavalcanti)

**Insta:** [www.instagram.com/prof.thiago.cavalcanti](https://www.instagram.com/prof.thiago.cavalcanti)

**YouTube:** [youtube.com/profthiagocavalcanti](https://youtube.com/profthiagocavalcanti)



# ESSA LEI TODO MUNDO CONHECE: PIRATARIA É CRIME.

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



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



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



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



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



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



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



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



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



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