Segurança da Informação aplicada em Big Data

Lara Vitória Silva Scaramuzza, Curso Superior de Tecnologia em Segurança da Informação, Fatec Ministro Ralph Biasi - Americana, lara.scaramuzza@fatec.sp.gov.br
Henri Alves de Godoy, Fatec Ministro Ralph Biasi - Americana, henri.godoy@fatec.sp.gov.br

Resumo

Os dados são o novo petróleo (KERSHNER, 2021). *Big data* é um conjunto extremamente volumoso de dados, que podem ser classificados como dados estruturados e não estruturados. A segurança da informação é intimamente relacionada à big data, pois a segurança da informação é fundamental a todos os processos da *big data*, como coleta, armazenamento, gerenciamento de acesso, descarte de dados, políticas de gerenciamento do *big data*, na assertividade e a confidencialidade dos resultados e interpretação de dados. Tudo isso ligado integralmente com a legislação que trata o assunto, como a Lei Geral de Proteção de Dados, número 13.709/2018. Os dados agregam valor para as empresas dependendo da maneira que são tratados e utilizados, uma vez que a empresa saiba manter todos esses dados e informações de modo seguro. Contudo, o armazenamento, tratamento e utilização indevida e insegura dos dados irá acarretar riscos e prejuízos para as empresas, colaboradores e até mesmo clientes. O objetivo deste trabalho é elucidar as necessidades de segurança da informação do *big data*, e avaliar o conhecimento dos profissionais de tecnologia da informação (TI) sobre os riscos e ameaças a *big data*. Este trabalho fundamenta-se em uma pesquisa bibliográfica da literatura atual de *big data* e Segurança da Informação (SI). E também foi realizada uma pesquisa quantitativa por meio de questionário eletrônico. Os resultados dessa pesquisa demonstraram que os participantes tinham consciência da maioria dos assuntos tratados.

Palavras-chave: *Big Data*, Segurança da Informação, LGPD

Abstract

Data is the new petroleum (KERSHNER, 2021). Big Data is a set of highly voluminous data, which may be classified as structured data or non-structured data. Information security is intimately related to big data because information security is fundamental to all extensive data processes, such as collecting data, storage, access management, data discard, management policy, assertiveness and confidentiality of results, and data interpretation. All of this is related to the legislature that deals with the issue, such as General Data Protection Law, number 13.709/2018. Data brings value to an organization depending on the way that data is treated and utilized, given that the organization knows how to keep all this data safe and secured. However, the unsafe storage, treatment, and usage of data will lead to risks and losses for the organization, its collaborators, and clients. The objective of this work is to elucidate the information security needs of big data and to assess the knowledge of information technology (IT) professionals about the risks and threats to big data. This present text is based on bibliographical research on

the existing literature on big data and information security. Also, quantitative research was realized through an electronic questionnaire. The research results show that participants had consciousness of the majority of topics.

***Keywords**: Big Data, Information Security, LGPD*

1. Introdução

Em 1997 o termo *big data* aparece pela primeira vez na *ACM digital library* (PRESS, 2013) sendo um termo utilizado para definir grandes quantidades de dados, podendo ser analisados e processados que, como resultado geram e mostram tendências de padrões e associações.

Big data pode ser definido como um grande conjunto de dados, complexos e de novas fontes de dados. Esses grandes volumes de dados podem ser usados para resolver problemas de negócios (ORACLE,2022). O conceito do *big data* está dividido em 7 Vs, e cada um tem suas particularidades.

O primeiro V é sobre o volume dos dados que são criados a cada minuto, e o *big data* agrupa todo esse volume, podendo ser dos mais variados tipos, como: *feeds* de dados do Twitter, fluxos de cliques em uma página web ou em um aplicativo para dispositivos móveis, e-mails, fotos, vídeos, mensagens ou ainda um equipamento habilitado para sensores. Esse volume pode ser *terabytes* ou *pentabytes* (FIA, 2021).

O segundo V é sobre velocidade, a taxa mais rápida na qual os dados são recebidos e administrados. Normalmente, a velocidade mais alta dos dados é transmitida diretamente para a memória, em vez de ser gravada no disco, o *big data* vai analisar no instante em que os dados são criados, como publicações em sites e blogs, transações de cartão de crédito e viralização de mensagens em redes sociais (FIA, 2021).

O terceiro V trata do assunto de variedade, refere-se a abundância de dados disponíveis. Tipos de dados tradicionais foram estruturados e se adequam perfeitamente a um banco de dados relacional, mas com o aumento do *big data*, os dados e as informações vêm em novo formato como os não estruturados. Tipos de dados não estruturados, como, arquivos em texto, áudio, vídeo, e-mail, cotações e transações financeiras. Esses exigem um pré-processamento adicional para obter significado (FIA, 2021).

O quarto V é sobre valor que vem das análises da grande quantidade de dados e informações estruturadas e não estruturadas recebidas pela big data. A partir desses é possível

extrair valor dos dados recebidos, *insights* e gerar estatísticas, acrescentando valor e sentido aos dados, para serem utilizados pela empresa (FIA, 2021).

O quinto V é sobre veracidade, refere-se a importância de tudo que está sendo reunido para análise, para que seja verdadeiro, para que os dados possam ter confidencialidade, por esse motivo o local onde eles foram extraídos deve ser de fontes confiáveis, em função de que não foram alterados ou manipulados. Esses são fatores que determinam a veracidade desses dados e os riscos associados na utilização deles para as análises e decisões do negócio, pois quanto mais confiável for a fonte de dados e informações, maior assertiva ela será (FIA, 2021).

O sexto V é sobre volatilidade, por quanto tempo os dados são úteis? O que é relevante hoje, será imprescindível ou descartável amanhã? A volatilidade é o processo pelo qual a pertinência da informação é medida. Com o fluxo crescente na velocidade e variedade as vezes pode ser difícil gerenciar a volatilidade, ainda mais com os não estruturados (FIA, 2021).

O sétimo V é sobre visualização onde os dados precisam ser apresentados de forma acessível e elegível, com o objetivo de compreender tendências, comportamentos e resultados para melhor atender as necessidades dos clientes e empresas, dos mais variados mercados. Dessa forma busca extrair proveito dos resultados apresentados (FIA, 2021).

Por conseguinte, o *big data* possui dados estruturadas e não estruturadas, essa diferencia impacta no processo dentro do *big data*, por isso é importante fazer a distinção deles.

Os dados estruturados são aquelas possíveis de se categorizar com mais facilidade, já que possuem um padrão mais rígido, e, são dados que podem ser colocados em linhas e colunas, como os formulários por isso, quando o usuário preenche um cadastro, ele precisa apenas completar os campos, como nome, idade, e-mail, entre outros, e, porventura, responde uma pergunta de sim ou não.

Os dados não estruturados aproximadamente são 80% do conteúdo disponível na Internet, como vídeos, imagens, áudios, e-mails ou outro tipo que não se tem uma estrutura padrão.

Os dados são o novo petróleo (KERSHNER, 2021), estamos vivendo em um uma época em que a velocidade que os dados são criados é cada vez maior e quem detém os dados mais confiáveis, assertivos e íntegros tem maiores vantagens competitivas no mercado, pois um poste em uma rede social pode causar alterações no mercado financeiro e na valorização ou desvalorização de ações de empresas, tudo graças a dados, informações, tecnologia e conectividade (ORACLE, 2022).

Por intermédio dos dados e as informações que são criadas, divulgadas e em muitos casos vazadas no mundo digital tem o poder de afetar o mundo físico, como um *tweet* do empresário Elon Musk dono de algumas empresas como a Tesla.

No dia 07/06/2016 ele escreveu um esclarecimento no seu Twitter: “Gostaria de esclarecer que a Tesla está trabalhando exclusivamente com a Panasonic para células Model 3. Artigos de notícias que afirmam o contrário estão incorretos” (MUSK, 2016 *apud* NOVAK, 2016).

Com apenas um tweet Elon Musk fez desaparecer US$ 580 milhões da capitalização de mercado da Samsung SDI e as ações caíram 8%. Já a Panasonic adicionou cerca de US$ 800 milhões ao seu capital no mesmo período. (NOVAK ,2016)

O objetivo desse trabalho é conhecer a história do *big data*, sua origem, significado, proposito, identificar fatos relevantes, como a segurança da informação está ligada ao *big data,* eventos importantes nessa trajetória, citar e exemplificar acontecimentos marcantes, identificar pontos chaves de segurança da informação e a importância de política de segurança da informação no *big data*.

2. Referencial Teórico

Em aproximadamente 6 minutos o mundo gera 9,1 mil *terabytes* de dados (EXAME,2021) e, sendo esse um volume de dados expressivos que são gerados e uma vez que as pessoas estão mais conectadas, sendo em uma reunião, que pode ser marcada de modo virtual por inúmeras plataformas diferentes e diversas pessoas de diferentes localidades podem participar da reunião sem sair de casa, graças à tecnologia.

Toda essa interação e facilidade também vem acompanhada de riscos, porque tudo o que se faz na internet, seja por meio de uma rede social, navegador, *site* ou aplicativo é coletado, armazenado, analisado e usado.

Cada vez mais utiliza-se inúmeros aplicativos, sites, aplicativos de bancos para pagar contas, fazer e realizar transações, aplicativos voltados a saúde para agendar consultas médicas, receber resultados de exames, também as redes sociais de conversas instantâneas de trabalho, aplicativos voltados a vídeos rápidos, aplicativos de conversas, e entre outros. Tudo o que se faz na Internet deixa rastros (AZAMBUJA,2022).

Essa quantidade expressiva de dados são coletados, armazenados, analisados pelo *big data* e geram resultados confiáveis, decisões com maior precisão, é quase como se o *big data*

conseguisse permitir antecipar ao futuro com direcionamentos e recomendações tirados de uma base de dados volumosa e variada, podendo ser utilizado em diversos setores por diferentes áreas como, e-commerce, segurança da informação, entretenimento, *marketing* digital, mídias sociais, serviços financeiros, energia, saúde, astronomia e entre muitos outros. E trabalhando com dados, a Lei Geral de Proteção de Dados (13.709/2018) tem como principal objetivo proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural, a política de segurança da informação deve estar em conformidade com a lei vigente.

A Lei n.º 13.709/2018 (BRASIL, 2018) determina que é um compromisso de todos os cidadãos proteger os dados, tanto a administração pública como empresas que utilizam esses dados, e, o primeiro dado classificado na lei é o dado pessoal, que possibilita a identificação direta ou indireta da pessoa natural todos esses dados devem ter um grau elevado de proteção definido e aplicado dentro da política de segurança e implementado no *big data*. Alguns dados de identificação pessoal são: "RG, CPF, retrato em fotografia, endereço de e-mail, número de cartão bancário, hábitos de consumo, dados de localização, endereço de IP (protocolo de internet), testemunhos de conexão (*cookies*) e entre outros." (BRASIL, 2018).

Diante disso deve-se aplicar política de segurança da informação no *big data,* pois ele armazena e utiliza dados e informações, com um alto volume, uma alta taxa de velocidade e alta variedade que exigem formas econômicas e inovadoras de processamento de informações que permitem insights aprimorados, tomada de decisões e automação de processos (GARTNER,2022).

Uma vez que a administração e gerenciamento de todos os dados que entram e saiam, as estatísticas que são geradas ou qualquer outro dado ou informação extraído desses centros gigantes de dados deve ser tratado com responsabilidade de quem detém esses dados.

É necessário implementar políticas de segurança da informação nas etapas do *big data*, pois muitas organizações afirmam que enfrentam problemas com a segurança de dados. (MARCON,2021). Com isso faz-se necessário implementar políticas de segurança da informação em *big data.*

3. Materiais e Métodos (ou Metodologia)

O método de pesquisa utilizado nesse trabalho foi a pesquisa bibliográfica, utilizando *sites* como o google acadêmico, livros, publicações eletrônicas em fontes como, jornais, *site* do

governo federal, instituição de ensino e *sites* de empresas voltadas para tecnologia, com o objetivo de aprofundar os conhecimentos do assunto, fundamentar a história sobre o tema, basear técnicas e soluções de segurança e utilizando *strings* como, segurança da informação, *big data*, LGPD, dados, 7 Vs.

Foi utilizado questionário no Google para alunos da Faculdade de Tecnologia do Estado de São Paulo (FATEC) da cidade de Americana e de Campinas que cursam Segurança da Informação (SI), Análise e Desenvolvimento de Sistemas (ADS), Jogos Digitais (JD), Gestão da Tecnologia da Informação (GTI) e profissionais que trabalham com TI.

4. Resultados e Discussões

De acordo com a classificação de dados da Lei n. 13.709, de 14 de agosto de 2018 (BRASIL,2018) com dados sensíveis exigem mais atenção no tratamento, os quais são: "aqueles relacionados a crianças e adolescentes; e os sensíveis, que são os que revelam origem racial ou étnica, convicções religiosas ou filosóficas, opiniões políticas, filiação sindical, questões genéticas, biométricas e sobre a saúde ou a vida sexual de uma pessoa."

A lei aborda sobre os dados de menores de idade, o qual devem ter o consentimento específico de pelo menos um dos pais ou responsável legal.

Dados públicos podem ser disponibilizados considerando a finalidade, interesse público que se justificaram a disponibilização e boa-fé. Podendo a organização tornar público sem pedir um novo consentimento.

Os dados anonimizados removem ou modificam informações que podem identificar a pessoa, desse modo sem vinculação dos dados a um indivíduo. Nesse caso a LGPD não se aplica.

Caso for possível reconstruir o caminho de volta e revelar a titularidade dos dados a LGPD se aplica.

4.1 Segurança em fontes de dados

Os ativos de uma empresa vão muito além de ativos físicos, com o avanço crescente da tecnologia em todas as áreas, os ativos lógicos de uma empresa são da mesma importância ou até mesmo mais importante que os ativos físicos, acrescentando mais importância de política de segurança da informação aplica ao *big data* pois, para as empresas em relação a continuação dos negócios, estratégias e tomadas de decisões, os ativos lógicos são de grande relevância. É

preciso ter cuidado na coleta dos dados, que se tornarão ativos pois, a fonte de coleta dos dados não estruturados pode conter vírus ou links falsos e a partir do momento que começa a fazer do sistema de análises do *big data* a empresa pode estar em riscos de estar trabalhando com dados que foram adulterados, podem conter os mais variados vírus (HURWITZ, 2013).

4.2 Proteção dos dados no *Big Data*

A lei LGPD e Hurwitz (2013) sugerem soluções para a proteção dos dados:

1-) Anonimização de dados: Na anonimização os dados passam por etapas que garantiram a desvinculação dele a essa pessoa é retirado qualquer dado que possa exclusivamente identificar um indivíduo.

Os dados anonimizados são essenciais para o crescimento da inteligência artificial, da internet das coisas, do aprendizado das máquinas, das cidades Inteligentes, da análise de comportamentos, entre outros. Quando os dados de uma organização, pública ou privada, são anonimizados isso aperfeiçoa a segurança da informação na organização e gera, assim, mais confiança em seus serviços e para seus públicos.

2-) Tokenização: Esta técnica protege os dados sensíveis substituindo-os com *tokens* aleatórios ou valores de que não significam nada para alguém que obtém acesso não autorizado a esses dados. Com essa técnica sendo utilizada as diminui chance de que os ladrões pudessem fazer alguma coisa com os dados. A tokenização pode proteger informações de cartão de crédito, senhas, informações pessoais e em breve. Alguns especialistas argumentam que é mais seguro do que a criptografia.

3-) Controles de banco de dados em nuvem: Os controles de acesso são construídos no banco de dados para proteger todo o banco de dados, o gerenciamento de acesso, grau de privilegio para acesso as informações, apagar, alterar, sobrescrever, baixar e enviar (HURWITZ, 2013).

4.3 Armazenamento de dados

A facilidade de armazenar dados parece inofensiva, mas, traz um alto custo para a empresa, e estima-se que as informações e dados armazenadas digitalmente no mundo, até 2025 será de 175 *zetabytes*, principalmente por dados sem utilização (CAETANO, 2020).

Todo esse armazenamento tem um custo para as empresas e para o planeta terra, uma vez que a energia aplicada para armazenar dados inúteis lançara 6,4 milhões de toneladas de carbono (CAETANO, 2020).

Além disso, o armazenamento em massa dos dados que vão de informações pessoais de clientes e funcionários, são armazenados também documentos estratégicos para o negócio da empresa, e, com todos esses fatores em jogo cada vez mais empresas buscam maneiras de se proteger contra ameaças internas e externas.

Empresas de todos os ramos podem sofrer com incidentes de segurança da informação, como a Netshoes que entre o período de 2017 e 2018 sofreu com vazamentos de dados de quase dois milhões de clientes, os dados vazados foram nome, CPF e-mail e compras realizadas.

Em decorrência, o Ministério Público aplicou uma multa de R$ 500 mil para a empresa de comércio eletrônico, e em caso de reincidência seria aplicada uma infração. Com base nesses acontecimentos novas ações foram implementadas para mitigar os problemas de segurança da informação no futuro (GLICFAS, 2020).

Os dados redundantes podem ser removidos, facilitando a busca por informações, liberando espaço, para novos dados, menor sobrecarga de trabalho pois o dado será analisado uma vez.

Os dados obsoletos devem ser descartados. Porque esses dados não agregam mais valor, e são um risco em potencial de vazamento. Qualquer dado, incompleto, desatualizado ou incorreto é considerado obsoleto.

As informações que circulam pelos sistemas da empresa não têm objetivo comercial, portanto são triviais.

4.4 Descarte de dados

Antes de realizar o descarte deve-se classificar o nível de importância dos dados, a dificuldade de reverter informações e o veículo que estão passando, pois, os dados podem estar tanto em papel como em meios eletrônicos e o descarte deles deve ser realizado de maneira que não em mãos maliciosas, devem ser totalmente destruídos de forma que não seja possível fazer qualquer tipo identificação, devendo ressaltar que dados pessoais necessitam de mais atenção em todas as etapas que passarem. Fazer a subscrição de dados no hardware que estava armazenado ou eliminando o hardware de forma adequada (ANADD, 2021).

4.5 Análise e interpretação dos resultados

Os participantes da pesquisa mencionada no capítulo 3 puderam ter acesso ao questionário eletrônico através de link de acesso que foi disponibilizado aos alunos e profissionais de TI. Os participantes da pesquisa foram alunos do ensino superior e profissionais de TI que tem maior poder cognitivo e familiaridade em responder questionário.

As questões elaboradas e direcionadas ao público-alvo foram levadas em consideração aos estudos abordados no trabalho confidencialidade dos dados utilizados pelo *big data,* seu armazenamento, dados sensíveis, acesso às informações, descarte de dados e importância da aplicação da LGPD e segurança da informação em todas essas etapas.

A pesquisa quantitativa se deu entre o dia 03 de outubro de 2022 até o dia 21 de outubro e 2022, com um total de 85 respostas do grupo de participantes.

As respostas obtidas através do questionário estão de acordo com o esperado, levando em consideração o público-alvo do trabalho de pesquisa, apenas a questão "Qualquer informação deve ser armazenada?" os alunos e profissionais demostraram dificuldade na resolução da questão, evidenciada pela baixa porcentagem de acertos, portanto é interessante estudar os motivos desse resultado.

Tabela 1: Questionário e porcentagem de acertos por grupo

Questões	Acertos (%)
1-) A confidencialidade dos dados para serem utilizados pelo *big data* é relevante? ( ) As informações devem ser apenas coletadas sem verificar as fontes dos dados. ( ) Os dados devem ser coletados de fontes oficiais. ( ) Os dados podem ser coletados de qualquer fonte. ( ) Não, os dados podem vir de qualquer fonte disponibilizada na internet.	86%
2-) Sobre o armazenamento dos dados. Deve-se ( ) Armazenar os dados em qualquer lugar com memória disponível. ( ) Criar políticas de arquivamento de dados. ( ) Armazenar os dados sem políticas de armazenamento.	94%
3-) Os dados sensíveis precisam ser mais protegidos? ( ) Os dados sensíveis necessitam de um alto grau de confidencialidade. ( ) Todos os dados precisam ser protegidos da mesma forma. ( ) Apenas dados relacionados a renda e saúde precisam ser mais protegidos.	72%
4-) Os dados que foram coletados e não forem utilizados, como devem ser tratados? ( ) Todos os dados que não forem ser utilizados devem ser apenas descartados. ( ) Os dados devem ser descartados de maneira que não sejam utilizados por pessoas maliciosas.	94%

5-) Quem deve ter acesso aos dados do *big data*? ( ) Todas as pessoas da empresa; ( ) Apenas os gerentes e conter restrições de acesso; ( ) Pessoas que estão trabalhando com os dados e conter restrições de acesso.	79%
6-) Qualquer informação deve ser armazenada? ( ) Apenas as informações com valor devem ser armazenadas; ( ) Toda informação é importante e deve ser armazenada; ( ) Qualquer informação que possa ter valor deve ser armazenada; ( ) Apenas as informações relevantes e que vão ser utilizadas devem ser armazenadas.	41%
7-) As políticas do *Big Data* devem estar de acordo com a LGPD? ( ) As políticas do *Big Data* precisam estar de acordo com LGPD; ( ) As políticas do *Big Data* devem estar de acordo somente com as políticas da empresa e não com a LGPD; ( ) As políticas do *Big Data* não precisam estar de acordo com a LGPD.	91%

Fonte: Próprio autor

Com a resolução do questionário é possível perceber que os alunos e profissionais de tecnologia da informação tiveram boas porcentagens de acerto, mostrando conhecimentos sobre segurança da informação.

5. Considerações Finais

Os dados que uma empresa armazena devem ser somente os que serão utilizados e que realmente tem valor para o negócio, dados armazenados sem um objetivo ou proposito para o negócio é um risco que a empresa corre sem necessidade.

Uns dos principais problemas em *big data* é a falta de segurança da informação nos dados e informações, tanto recebidas quanto as que estão sendo armazenadas e utilizadas para, por exemplo estratégias de decisão.

A tendência geral dos resultados é positiva, tendo em vista que os alunos e profissionais que participaram do questionário tiveram um desempenho positivo na resolução das perguntas.

Dado a pesquisa bibliográfica realizada é possível perceber a importância da conscientização dos profissionais que estão trabalhando com *big data*, uma vez que não são apenas profissionais de TI que trabalham com segurança da informação ou com tecnologia da informação estão em contato com essa tecnologia, são profissionais das áreas de saúde, marketing, serviço financeiro e muitas outras áreas de atuação que não são ligadas a tecnologia

e por isso não tem relação a segurança da informação, a privacidade, a lei geral de proteção de dados, e todos os aspectos que estão envolvidos com o profissional que está em contato com dados e informações dos mais variados tipos e níveis, explicando conceitos, esclarecendo dúvidas da área de segurança da informação como os possíveis riscos, vulnerabilidades, engenharia social como no dia a dia é possível evitar muitas vulnerabilidades e uma frase popularmente conhecida no meio de segurança da informação é "o ser humano é o elo mais fraco", mas com conscientização e treinamento o ser humano passa ser o elo mais forte.

Dado a baixa porcentagem de acerto da questão "Qualquer informação deve ser armazenada?", é interessante estudar os motivos desse resultado. Campanhas e programas de conscientização devem focar em explanar os diferentes tipos de informações e quais devem ser coletados, armazenados e/ou descartados e exemplificando os riscos da falta de segurança da informação.

Referências

ANADD. **A proteção de dados pessoais e o descarte adequado em meio físico** Disponível em: https://anadd.org/blog/f/a-protecao-de-dados-pessoais-e-o-descarte-adequado-em-meio-fisico Acesso em: 16 out. 2022.

AZAMBUJA G. J. A.; GRANVILLE Z. L.; SARMENTO M.G.A **A privacidade, a segurança da informação e a proteção de dados no Big Data Disponível** em: http://seer.cgee.org.br/index.php/parcerias_estrategicas/article/viewFile/914/831 Acesso em: 08 out. 2022

BRASIL. **Classificação dos Dados** Disponível em: https://www.gov.br/cidadania/pt-br/acesso-a-informacao/lgpd/classificacao-dos-dados Acesso em: 09 out. 2022

BRASIL. **Lei n. 13.709, de 14 de ago**. De 2018 Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm Acesso em: 25 out 2022

BRASIL. **O que são dados anonimizados, segundo a LGPD** Disponível em: https://www.serpro.gov.br/lgpd/menu/protecao-de-dados/dados-anonimizados-lgpd Acesso em: 14 out 2022

BRASIL. **dados sensíveis** Disponível em: https://www.serpro.gov.br/lgpd/menu/protecao-de-dados/dados-sensiveis-lgpd Acesso em: 23 out. 2022

CAETANO, R.; **Big data**: armazenamento de dados inúteis tem custo e afeta o meio ambiente Disponível em: https://exame.com/tecnologia/armazenamento-de-dados-inuteis-gera-custos-e-prejudica-o-meio-ambiente/ Acesso em: 21 nov 2022

EXAME. **Temos mais dados do que nunca. Como usá-los a nosso favor?** Disponível em: https://exame.com/carreira/dados-uso-favor/ Acesso em: 08 jun. 2022

FERREIRA T.; VICENTIN T. **LGPD**: qual a diferença entre dados pessoais, sensíveis e anonimizados? Disponível em: https://olhardigital.com.br/2021/08/17/tira-duvidas/lgpd-qual-a-diferenca-entre-dados-pessoais-sensiveis-e-anonimizados/ Acesso em: 20 jun. 2022

FIA. **Big Data**: como funciona, exemplos, importância e desafios Disponível em: https://fia.com.br/blog/big-data/ Acesso em: 18 out. 2022

GLICFAS. **Big Data e armazenamento de dados:** risco ou oportunidade? Disponível em: https://glicfas.com.br/big-data-e-armazenamento-de-dados-risco-ou-oportunidade/ Acesso em: 30 out. 2022

GARTNER. **Gartner Definition of Big Data – IT Glossary** Disponível em: https://www.gartner.com/en/information-technology/glossary/big-data Acesso em: 16 jun. 2022

GUIMARÃES L**. Big Data VS: você sabe quais são os Vs do Big Data?** Disponível em: https://www.knowsolution.com.br/big-data-vs/ Acesso em: 16 jun. 2022

HURWITZ, Judith. et al. **Big data for dummies.** Nova Jersey: John Wiley & Sons, 2013. Disponível em: https://jan.newmarch.name/IoT/BigData/Big%20Data%20For%20Dummies.pdf Acesso em: 10 out. 2022

KERSHNER, **Michael Data Isn't The New Oil — Time Is Disponível** em: https://www.forbes.com/sites/theyec/2021/07/15/data-isnt-the-new-oil--time-is/?sh=614bf3d635bb Acesso em: 17 set. 2022

MARCON F. J. **6 maiores desafios em big data enfrentado pelas empresas** Disponível em: https://digital.br.synnex.com/6-maiores-desafios-em-big-data-enfrentado-pelas-empresas Acesso em: 31 out. 2022

NOVAK, **Matt One Tweet From Elon Musk Made $580 Million Disappear** Disponível em: https://gizmodo.com/one-tweet-from-elon-musk-made-580-million-disappear-1781261277 Acesso em: 18 set. 2022

ORACLE. **O que é Big Data** Disponível em: https://www.oracle.com/br/big-data/what-is-big-data/ Acesso em: 08 jun. 2022

ORACLE. **O que é Gerenciamento de Dados?** Disponível em: https://www.oracle.com/br/database/what-is-data-management/#data-management-defined Acesso em: 10 out. 2022

PRESS, **Gil. A Very Short History Of Big Data** Disponível em: https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/?sh=526379a165a1 Acesso em: 18 set. 2022