



XX ENANCIB

21 a 25 Outubro/2019 – Florianópolis

A Ciência da Informação e a era da Ciência de Dados

GT-4 – GESTÃO DA INFORMAÇÃO E DO CONHECIMENTO

ENSAIO EXPLORATÓRIO SOBRE AS CONDICIONANTES DO BIG DATA NO CONTEXTO ORGANIZACIONAL

EXPLORATORY TEST ON BIG DATA CONDITIONERS IN THE ORGANIZATIONAL CONTEXT

Ricardo Rodrigues Barbosa (Universidade Federal de Minas Gerais – UFMG)

Luander Cipriano de Jesus Falcão (Universidade FUMEC)

Brenner Lopes (Universidade Federal de Minas Gerais – UFMG)

Modalidade: Trabalho Completo

Resumo: O atual crescimento na geração de dados tem criado um oceano de dados, de alta variedade, e que transitam em alta velocidade. Esse fenômeno tem sido chamado de *Big Data*, que se baseia na coleta, processamento e análise de um conjunto tão grande e diverso de dados, que não podem ser suportados pelas ferramentas tradicionais de processamento e análise. Diante de todo o impacto e relevância do *Big Data*, essa pesquisa, de cunho exploratório, tem como objetivo, constituir uma visão inicial sobre algumas das principais vantagens identificadas na literatura, condicionantes e dificuldades na utilização do *Big Data*. Para tanto, realizou-se uma pesquisa estruturada com base em metodologia mista, através da aplicação de questionários semiestruturados na fase quantitativa, seguida de uma fase qualitativa, baseada em análise documental. Dentre os resultados dessa pesquisa foi identificado uma considerável confusão conceitual em torno do tema *Big Data*, corroborando a literatura sobre o tema. Também foi identificado o efeito da questão da cultura organizacional, sobre a adoção de *Big Data* nas empresas pesquisadas, e a fase inicial de implementação da abordagem de big data em que a maioria delas está.

Palavras-Chave: Big Data; Analytics; Vantagens; Condicionantes; Dificuldades.

Abstract: The current growth in data generation has created an ocean of high-variety, high-speed data. This phenomenon has been called Big Data, which is based on the collection, processing, and analysis of such a large and diverse set of data that cannot be supported by traditional processing and analysis tools. Given all the impact and relevance of Big Data, this exploratory research aims to provide an initial view of some of the main advantages identified in the literature, constraints and difficulties in using Big Data. For this, a structured research based on mixed methodology was carried out through the application of semi-structured questionnaires in the quantitative phase, followed by a qualitative phase, based on document analysis. Among the results of this research was identified a considerable conceptual confusion around the Big Data theme, corroborating the literature on the subject. We also identified the effect of the organizational culture issue on the adoption of big data in the surveyed companies, and the early implementation phase of the big data approach that most of them are in.

Keywords: Big Data; Analytics; Advantages; Constraints; Difficulties.

1 INTRODUÇÃO

O crescimento da disponibilidade de dados em magnitudes impensáveis até algumas décadas atrás, juntamente com os avanços na tecnologia e principalmente na capacidade de processamento e armazenamento, que tiveram suas capacidades aumentadas, ao mesmo tempo que seus custos reduzidos, tem trazido um cenário de contornos ainda não possíveis de se delinear.

Na atualidade, como afirmaram Jin et al (2015, p. 60, tradução nossa), onde o mundo emergiu complementemente na chamada “era da informação”, onde “o uso extensivo da internet, da internet das coisas, da computação em nuvem e de outras tecnologias de TI emergentes, fez com que várias fontes de dados” aumentassem em quantidade e complexidade a uma taxa sem precedentes.

Essa nova realidade foi chamada pelos pesquisadores do Microsoft Research, como quarto paradigma para a ciência, que segundo tais pesquisadores, está “baseado em computação intensiva em dados”. Na visão desses profissionais, “estamos num estágio de desenvolvimento que é análogo a quando a impressora foi inventada. A impressão levou mil anos para se desenvolver e evoluir para as muitas formas que toma hoje”. Esse novo paradigma, essa nova ciência intensiva em dados está ancorada em três processos que lhe sustentam: captura, curadoria e análise (HEY et al, 2009, p. xi, tradução nossa).

Uma das poucas coisas possíveis de se prever nesse cenário, é que temos dados, muitos dados disponíveis, como pontuado acima, e que teremos mais ainda, muito mais. Brands (2014, p. 65, tradução nossa), citando informações disponibilizadas pela IBM, aponta que “2,5 quintilhões de bytes (2,5 milhões de *terabytes*) de dados são criados diariamente – isto é 2.500.000.000.000.000 bytes a cada dia, o que representa 250.000 vezes o tamanho da Biblioteca do Congresso Americano”.

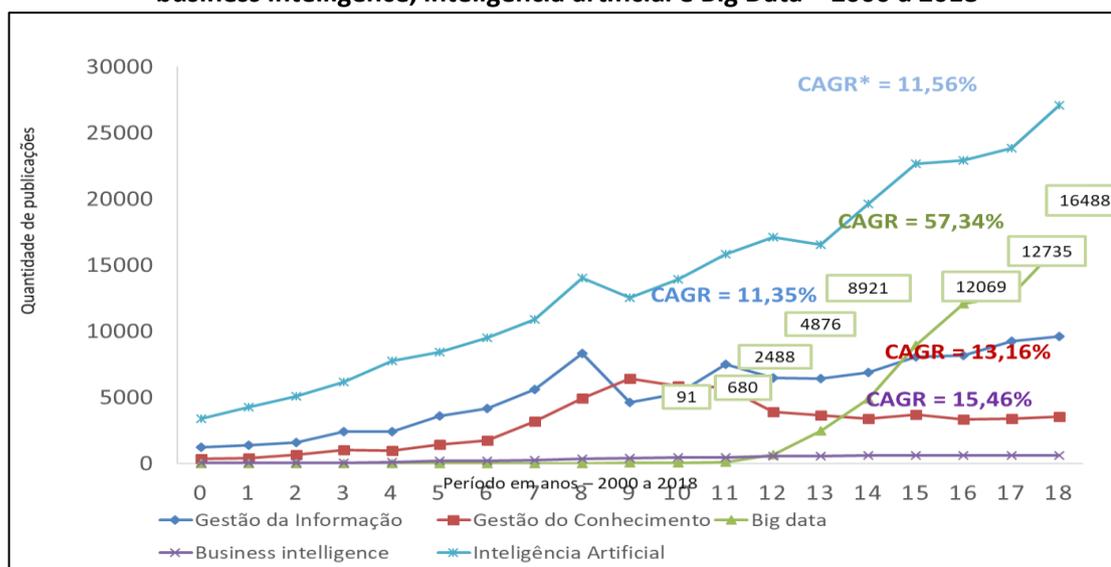
Uma característica fundamental desse oceano de dados, é que ele cresce não só quanto ao seu volume e quantidade, mas também em termos de variedade, sendo criado e transitando em alta velocidade. Se antes o foco e a disponibilidade de dados tinham como interface prioritária os chamados dados estruturados, hoje eles estão em quantidade e importância bem menor, e os ajustes e aprimoramentos nas tecnologias e análises foram em parte realizados para se adaptarem a essa nova realidade, e convencionou-se chamar de *Big Data*.

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

Além do crescimento exponencial da quantidade de dados disponíveis, é possível verificar a importância que a temática *Big Data*, vem abrangendo através do incremento no número de publicações sobre o tema. Levantamento realizado por Barbosa (2018) na base Scopus (www.scopus.com), aponta para tal fato. A base Scopus é considerada o maior banco de dados de resumos e citações da literatura revisada por pares: periódicos científicos, livros e anais de congressos.

O levantamento realizado considerou as publicações realizadas sobre as temáticas, gestão da informação, gestão do conhecimento, *business intelligence*, inteligência artificial e *big data*, nos últimos dezenove anos, e estão apresentadas no Gráfico 1, abaixo:

Gráfico 1: Análise temporal na base Scopus do número de publicações, nas temáticas, gestão da informação, gestão do conhecimento, business intelligence, inteligência artificial e Big Data – 2000 a 2018



Fonte: elaborado pelo autor a partir do levantamento de Barbosa, 2018

De forma a se obter uma melhor visualização da evolução e análise de cada tema ao longo dos anos, foram calculadas as taxas anuais compostas de crescimento (*compound annual growth rate* - CAGR) de cada um dos temas pesquisados.

No gráfico 1 pode-se verificar que no período analisado, a partir do ano 2000, a temática *Big Data* se origina do patamar mais baixo, e cresce ao longo dos anos. Mas principalmente a partir dos anos 2011-2012, há o início de uma acentuada curva de crescimento. Ela passa a apresentar a maior taxa de crescimento anual composto (57,34%), dentre as demais temáticas pesquisadas. Se considerarmos um período menor, delimitado pelos últimos nove anos, que coincide com o início de um grande salto no número de

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

publicações sobre *Big Data*, no ano de 2011, até o ano de 2018, encontramos as seguintes taxas de crescimento, segundo os termos pesquisados:

Tabela 1: Análise temporal na base Scopus do número de publicações, nas temáticas, gestão da informação, gestão do conhecimento, business intelligence, inteligência artificial e Big Data – 2011 a 2018

Termos Busca	CAGR (%)
Gestão do Conhecimento (GC)	-5,85
Gestão da Informação (GI)	3,18
Business Intelligence (BI)	3,27
Inteligência Artificial	6,92
Big Data	91,54

Fonte: elaborado pelo autor a partir do levantamento de Barbosa, 2019

Pode-se constatar baseando-se na análise dos resultados apresentados pela Tabela 1, que de todas as temáticas analisadas, “*Big Data*” apresenta um crescimento exponencial a partir de 2011, ao passo que as demais temáticas parecem estar perdendo fôlego, apresentando tendência decrescente.

Para além da análise da taxa de crescimento, é importante também ressaltarmos, que em termos de quantidade absoluta de publicações, “*Big Data*” está a caminho de se equiparar à temática “Inteligência Artificial”.

Diante de todo o impacto e relevância do *Big Data*, essa pesquisa tem como objetivo, dentro de uma visão exploratória, constituir uma visão inicial sobre algumas das principais vantagens identificadas na literatura, as condicionantes e as principais dificuldades na utilização do *Big Data*.

2 BACKGROUND

Ao que parece, “*Big Data*” é um termo cuja definição ainda não encontrou, e talvez não encontre em qualquer outro momento, uma proposição que seja consensual, entre os diversos autores, pesquisadores e profissionais que dele se valem.

Big Data, portanto, é um termo em construção, que surgiu pela primeira vez em 1997, no artigo *Application Controlled Demand Paging for Out-of-Core Visualization*, publicado pela Administração Nacional da Aeronáutica e Espaço dos Estados Unidos (NASA – *National Aeronautics and Space Administration*), de autoria de Michael Cox e David Ellsworth, pesquisadores do MRJ Technology Solutions da NASA. Esses autores discorriam sobre a questão de alternativas de visualização na chamada fluidodinâmica computacional (ou CFD –

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC**

Computational Fluid Dynamics), área do conhecimento que envolve a simulação numérica relativa a escoamentos de fluidos, transferência de calor e outros fenômenos relacionados.

Para Cox e Ellsworth (1997, p.2, tradução nossa), a questão da visualização “fornece um desafio interessante para sistemas de computadores, onde os conjuntos de dados são geralmente muito grandes, sobrecarregando a memória principal, disco local e até disco remoto”. Os autores, portanto, chamaram isso de “problema de Big Data” (COX e ELLSWORTH, 1997).

As primeiras tratativas de delimitarem um conceito para *Big Data*, se concentraram nos chamados três “V”, quais sejam, volume, velocidade e variedade. Mas à medida que as discussões e aplicações do *Big Data*, se aprofundavam, foi possível verificar um aumento no número de “V” (valor, veracidade, variabilidade, vocabulário e local – do inglês “venue”).

Jain et al (2016, p.1), Big Data “refere-se especificamente a um conjunto de dados tão grandes ou complexos que os aplicativos tradicionais de processamento de dados não são suficientes”, para fazer frente ao seu processamento e análise. Os autores ainda complementam seu raciocínio pontuando que “toda a enorme quantidade de dados produzidos a partir de várias fontes em vários formatos com alta velocidade é referida como Big Data” (JAIN et al, 2016, p. 1, tradução nossa).

Muito próximo do entendimento exposto por Jain et al (2016), Davenport (2014, p.1), aponta que *Big Data* pode ser entendido como um “termo genérico, para dados que não podem ser contidos nos repositórios usuais”, sendo “volumosos demais para caber em um único servidor; não estruturados demais” para se adequarem aos bancos de dados tradicionais (relacionais, com organização em linhas e colunas), e “fluidos demais para serem armazenados em um data warehouse estático”.

Davenport (2014) ainda destaca que o principal fator no âmago da problematização e diferenciação do *Big Data*, frente a outras denominações, mais do que a questão da quantidade de dados, está na falta de estrutura desses. Esses fatores ficam nítidos na comparação feita pelo autor entre *Big Data* e o *analytics* tradicional:

Quadro 2: Big Data x analytics tradicional

Característica	Big Data	Analytics Tradicional
Tipo de dados	Formatos não estruturados	Dados formatados em linhas e colunas
Volume de dados	100 terabytes e petabytes ou superior	Dezenas de terabytes ou menos
Fluxo de dados	Fluxo constante de dados	Pool estático de dados
Métodos de análise	Aprendizado de máquina	Baseado em hipóteses
Objetivo principal	Produtos baseados em dados	Suporte ao processo decisório interno

Fonte: elaborado pelo autor a partir de Davenport, 2014, p. 4

Em um levantamento realizado com o intuito de entender e conseguir diferenciar e conceituar de forma contundente, os conceitos de BI (*business intelligence*), BA (*business analytics*) and BD (*big data*), Mashingaidze e Backhouse (2015, p. 489, tradução nossa), colocam que apesar de serem termos muito citados na literatura e do crescente investimento e pesquisa atrelada aos mesmos, esses termos “não possuem definições únicas na literatura acadêmica e profissional”, o que tem na visão dos autores causado grande confusão de significado relativo aos mesmos.

Mashingaidze e Backhouse (2015), estabeleceram critérios de busca, e realizarem uma busca ampla nas bases EBSCO (de cunho mais prático) e Scopus (de cunho mais acadêmico), e analisarem os conteúdos selecionados. Eles concluíram e definiram como resultado desse trabalho, que *Big Data* deveria ser entendido como “dados em alto volume, obtidos a partir de uma variedade de fontes, que são gerados e analisados em alta velocidade”. Os autores complementam e reforçam sua definição, colocando ainda que, “o *Big Data* é muito grande e complexo para o gerenciamento de tecnologias convencionais”, requerendo portando, “tecnologias e técnicas avançadas para seu armazenamento e análise” (MASHINGAIDZE e BACKHOUSE, p. 499, tradução nossa).

A visão de Taurion (2015, p. 31), é esclarecedora, quando afirma que “*Big Data* não é tecnologia. Não é hadoop. É claramente uma mudança de *mindset*. Envolve novas e antigas tecnologias”; sendo os impactos em termos de transformação que pode provocar nas organizações e seus respectivos negócios, o seu verdadeiro valor. Ainda na visão do autor, “dados são os recursos naturais da sociedade da informação, como é o petróleo para a sociedade industrial. Têm valor apenas se tratados, analisados e usados para tomada de decisões”, relevantes, que podem trazer impactos positivos aos negócios ou evitar algum grande problema ou risco (TAURION, 2015, p. 39).

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC**

Gupta e George (2016, p. 2, tradução nossa), destacam que Big Data, é “frequentemente usado para descrever dados massivos, complexos e transmitidos em tempo real, que exigem técnicas sofisticadas de gerenciamento, análise e processamento para extrair insights”. Uma visão diferente, que poderíamos classificar no campo da heterodoxia, é trazida por Machado (2018, p. 36), quando o autor aponta que Big Data, poderia ser entendido, “considerado, em parte” como a “junção de BI e *Data Mining*”. Parte dessa afirmação, consta também como resultado da pesquisa de Mashingaidze e Backhouse (2015, p. 499, tradução nossa), onde os autores declaram que “*Big Data* é um tipo de dado que é usado em análise de negócios avançada”.

Mesmo num ambiente de certa incompreensão de conceitos e até mesmo desconfiança de alguns especialistas que veem o *Big Data* como mais um modismo, olhando-se para os resultados esperados pela sua aplicação é possível termos um vislumbre que parece dizer algo mais. E ele aponta para a direção de mudança de realidades e efetivação de resultados, buscados há um bom tempo, mas que não haviam conseguido obter até então resultados mensuráveis e/ou diretamente relacionados à aplicação das “soluções de coleta, tratamento, análise e uso da informação” até então utilizadas.

Jin et al (2015, p.60, tradução nossa), avaliam que “devido ao seu grande valor, o *Big Data* está essencialmente mudando e transformando a forma como vivemos, trabalhamos e pensamos”. Nesse sentido os autores apontam para algumas realidades, alguns campos de elevada significância, que vem sendo impactados pelo *Big Data*, como: desenvolvimento nacional; maior produtividade e competitividade industrial; as pesquisas científicas e as pesquisas interdisciplinares emergentes, e outros. Esses campos têm contribuído para que as pessoas consigam ter uma melhor percepção do presente, assim como vislumbres preditivos mais adequados do futuro (JIN ET AL, 2015, p.60).

Big Data, tende a ser uma função comum a todas as organizações. Gupta e George (2016), apontam para a concordância com essa visão, ao mesmo tempo em que a nosso ver se equivocam ao advogarem que por esse motivo ele não poderia ser considerado como uma fonte de vantagem competitiva para as organizações. Na visão deles “é improvável que o *Big Data* seja uma fonte de vantagem competitiva, uma vez que todas as empresas (com tamanhos comparáveis), provavelmente coletarão hordas de dados de uma variedade de fontes” (GUPTA e GEORGE, 2016, p.1, tradução nossa).

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

Uma comparação interessante é pensarmos nas facetas da estratégia ou do marketing, no campo da gestão, ou até mesmo nas ferramentas de BI (do inglês *business intelligence*), no campo da tecnologia; ou seja, não é o fato de possuir acesso aos conceitos, construtos e ferramentas de estratégia e marketing, ou da melhor solução de BI, que uma organização será tão competitiva quanto um líder de mercado.

É justamente a capacidade analítica, de entendimento dos problemas e do mercado, de cultura, gestão e outros, que darão o adequado diferencial competitivo às organizações. Como os próprios autores discutem em sua pesquisa, no contexto das capacidades de análise de *Big Data*.

Nesse contexto Wamba et al (2017, p.1, tradução nossa), apontam para o que denominam de capacidade analítica em *Big Data*, que definem como “uma abordagem holística para gerenciamento, processamento e análise” das dimensões do *Big Data* (nesse caso 5 “V”: volume, variedade, velocidade, veracidade e valor), para “criar ideias acionáveis para fornecer valor sustentado, medir o desempenho e estabelecer vantagens competitivas”, para as organizações.

Quanto aos resultados esperados pela aplicação do *Big Data* no contexto das organizações, citamos abaixo alguns exemplos citados na literatura do tema e apontado pela prática nas empresas, sem a preocupação de construirmos uma listagem exaustiva, que perpassa todas as suas possíveis aplicações e resultados nas mais diversas indústrias: novos recursos, novos produtos e serviços para os clientes, aumento de vendas, aumento vendas cruzadas, aumento de receita, aumento do ROI, redução de custos, decisões mais rápidas, melhores decisões, inovações produto/serviço, melhores processos de negócios, aumento da lucratividade, performance financeira superior, novos mercados, novos negócios, retenção clientes, maior eficiência operacional, melhoria relacionamento com os clientes, melhoria desempenho organizacional, maior efetividade no monitoramento do ambiente organizacional, acesso mais rápido aos dados, melhor gerenciamento dos preços, redução tempo e aumento da assertividade em processos seletivos, identificação e redução de fraudes, desenvolvimento de novos medicamentos, otimização de contratos de serviços e intervalos de manutenção, otimização tráfego em grandes cidades, dentre outros (DAVENPORT , 2014; DAVENPORT e HARRIS, 2007; TAURION, 2015; FOREMAN, 2016; BRAGANZA et al, 2017; WAMBA et al, 2017).

Em linhas gerais, tendo como pano de fundo a contínua ampliação da disponibilidade de grandes e variadas massas de dados diariamente, provenientes de múltiplas fontes, principalmente externas, podemos entender que num futuro próximo a maior parte dos problemas nas organizações e em outros campos, serão solucionados ou entendidos, de alguma forma, com utilização do *Big Data* e soluções no campo do *Big Data Analytics*. Por isso mesmo, torna-se crucial o entendimento sobre as principais condicionantes de desenvolvimento e aumento da efetividade e qualidade das análises e soluções no campo do *Big Data*.

3 METODOLOGIA

A metodologia que suportou essa pesquisa, se deu em dois estágios consecutivos e complementares; um quantitativo, através da aplicação de questionários semiestruturados, junto a profissionais de empresas públicas e privadas de diversos portes e segmentos econômicos que integram o mailing do Movimento ComInteligência; e outro qualitativo, que se baseou na transcrição e análise das discussões ocorridas durante evento específico sobre a temática *Big Data*, que centralizou os esforços de levantamento dos dados, e nos materiais das palestras realizadas no mesmo evento.

Para atendimento ao raciocínio exposto acima, este trabalho de pesquisa se utilizará da denominada “estratégia exploratória sequencial”. Segundo Creswell (2010, p. 248, tradução nossa), tal estratégia “envolve uma primeira fase de coleta e de análise de dados qualitativos, seguida de uma segunda fase de coleta e análise de dados quantitativos que é desenvolvida sobre os resultados da primeira fase qualitativa”.

Vieira e Zouain (2005, p. 124. Org.), em uma síntese dos pressupostos da abordagem quantitativa e qualitativa de pesquisa, sob a ótica dos pressupostos metodológicos, apontam que a abordagem quantitativa está centrada na causa e efeito (processo dedutivo, da teoria para os dados), em que as “generalizações levam à predição, explanação e ao entendimento” (descontextualização), “mediante validade e confiabilidade” (acurácia e consistência).

No caso da abordagem qualitativa, o foco é nas “inter-relações de fatores” (processo indutivo dos dados para a teoria), com base em “padrões e teorias desenvolvidas para o entendimento” (contextualização), “mediante verificação e força da argumentação teórica” (acurácia e consistência) (VIEIRA e ZOUAIN, 2005, p. 124. Org.).

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

O levantamento de dados que sustentou esse trabalho, se deu dentro do esforço de uma pesquisa científica, que teve como uma de suas etapas mais importantes e estruturadora (por possibilitar as condições suficientes para congregação, reflexão e discussão entre acadêmicos e profissionais envolvidos com a temática, coleta de informações primárias e secundárias), a realização de um seminário de cunho acadêmico e prático, com foco em pesquisa e na criação de pontes entre a academia e empresas, no âmbito do Programa de Pós-Graduação em Gestão e Organização do Conhecimento (PPG-GOC - //ppggoc.eci.ufmg.br), da Escola de Ciência da Informação (ECI), da Universidade Federal de Minas Gerais (UFMG).

Com a temática Inteligência Organizacional e *Big Data*, o seminário foi estruturado em torno de quatro palestras, com especialistas de empresas privadas, de diferentes portes e setores e com graus variados de maturidade de implementação da abordagem de *Big Data*, seguidas de um debate entre o público presente e os profissionais palestrantes e professores e alunos de mestrado e doutorado do PPGGOC, após a apresentação ao público dos resultados iniciais da pesquisa realizada.

No processo de divulgação e sensibilização do referido seminário, foram enviados aos convidados, por meio eletrônico, a partir do mailing do Movimento ComInteligência (www.facebook.com/ComInteligencia), organização privada e sem fins lucrativos, parceira e co-realizadora do evento, questionários semiestruturados, baseados na escala *likert* de cinco pontos, que responderam pela fase quantitativa da pesquisa.

Os questionários foram enviados para a base de contatos do Movimento ComInteligência, sem qualquer tipo de estratificação quanto a setores, funções ou porte das empresas, dentro de uma lógica de conveniência visando a obtenção do maior número de respostas possíveis, entendendo que a base de contatos do Movimento ComInteligência, com atividades contínuas quanto a realização de eventos, seminários, workshops e debates sobre as temáticas inteligência competitiva, estratégia e *advanced data analytics*, desde o ano de 2012, condensa profissionais atuantes e interessados nessas temáticas.

A fase qualitativa se baseou na abordagem de análise documental, tendo como insumos, os documentos das apresentações realizadas pelos palestrantes, documento com o conteúdo da transcrição da gravação das discussões entre o público presente, os profissionais palestrantes e os professores e alunos de mestrado e doutorado do PPGGOC, e as anotações

realizadas pelo pesquisador no transcorrer das apresentações e discussões, ocorridas no âmbito do referido seminário.

De acordo com Gil (2008, p. 51), “a pesquisa documental vale-se de materiais que não receberam ainda um tratamento analítico, ou que ainda podem ser reelaborados de acordo com os objetivos da pesquisa”. Ainda na visão desse mesmo autor, para a finalidade de pesquisa científica “são considerados documentos não apenas os escritos utilizados para esclarecer determinada coisa, mas qualquer objeto que possa contribuir para a investigação de determinado fato ou fenômeno” (GIL, 2008, p. 147).

A partir da execução do processo de revisão da literatura, foram constituídas três dimensões, que sustentaram a estruturação do questionário e que foram analisadas nessa pesquisa: 1) vantagens (de utilização do *Big Data*); 2) condicionantes (fatores que suportam o alcance das vantagens de utilização do *Big Data*), e 3) as principais dificuldades, para implementação e utilização do *Big Data*.

Essa estrutura utilizada para a construção do questionário quantitativo, também foi o insumo que permitiu a construção de um “quadro de referência analítica” que sustentou, conduziu e direcionou as análises realizadas nos documentos acessados.

4 ANÁLISE RESULTADOS

No âmbito da aplicação dos questionários semiestruturados, na fase quantitativa, foram obtidas 91 respostas válidas. Para análise das informações do questionário, foi utilizado o *software* SPSS 20, e foi realizada uma análise descritiva dos dados, através da distribuição de frequência das informações e cruzamento de variáveis selecionadas. Na sequência se explorou os principais resultados dessa fase da pesquisa, começando pela caracterização dos respondentes e suas respectivas organizações.

No que tange ao grau de instrução dos respondentes, a quase totalidade destes (97,8%) possuíam no mínimo graduação completa, o que mostra um grupo selecionado, com um excelente potencial de entendimento para temáticas mais complexas e recentes, como é o caso do *Big Data*.

A distribuição dos cargos/funções dos respondentes, trouxe também uma visão de elevado valor para os objetivos dessa pesquisa e análise de seus resultados, onde obteve-se uma excelente representatividade tanto do grupo que apoia o planejamento das diversas iniciativas nas organizações no nível estratégico (*c-level*: 15,4% dos entrevistados), quanto

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

daqueles que controlam e gerenciam essas iniciativas no nível tático (média gerência: 27,5% dos entrevistados) e aqueles que executam as estratégias e iniciativas definidas (analistas e especialistas: 45,1% dos entrevistados). Nesse contexto mais estratégico consideramos as iniciativas no campo do *Big Data*.

As organizações de todos os setores que integraram a pesquisa foram unânimes em concordar que todos os itens apresentados configuram vantagens trazidas pela utilização do *Big Data*, com destaque para a vantagem de auxiliar “na criação de novas oportunidades de negócio”, e para as empresas do setor de serviços com percentuais mais elevados na avaliação de quase todos os itens.

Tabela 2: Vantagens utilização Big Data x Setores

		Principal setor organização			
		Indústria	Comércio	Serviços	Média
Redução de custos	Discordo	0,0%	14,3%	6,5%	6,9%
	Concordo	85,7%	85,7%	67,6%	79,7%
Auxilia na criação de novas oportunidades de negócio	Discordo	0,0%	0,0%	1,3%	0,4%
	Concordo	85,7%	85,7%	94,8%	88,7%
Permite melhorias no desempenho organizacional	Discordo	0,0%	0,0%	1,3%	0,4%
	Concordo	71,4%	85,8%	92,2%	83,1%
Possibilita acesso mais rápido aos dados	Discordo	0,0%	14,3%	9,1%	7,8%
	Concordo	85,7%	71,5%	77,9%	78,4%
Oferece dados de forma mais tempestiva e acionáveis	Discordo	0,0%	0,0%	2,6%	0,9%
	Concordo	85,7%	85,7%	85,8%	85,7%
Permite melhorar o monitoramento do ambiente organizacional	Discordo	0,0%	0,0%	3,9%	1,3%
	Concordo	71,4%	71,5%	78,0%	73,6%
Auxilia na melhoria do relacionamento da organização com os clientes	Discordo	0,0%	0,0%	3,9%	1,3%
	Concordo	71,4%	85,7%	84,5%	80,5%

Fonte: elaborada pelos autores, pesquisa realizada 2019

Quando analisados os graus de concordância ou discordância das organizações frente aos posicionamentos e recursos da organização sobre *Big Data* (Tabela 3), verifica-se uma queda acentuada dos graus de concordância, concomitante a um aumento acentuado dos graus de discordância, assim como daqueles que “não souberam avaliar”, ou “não concordaram, nem discordaram”.

Os dados demonstram para esse extrato de empresas, que apesar de reconhecerem de maneira nítida as possíveis vantagens da utilização do *Big Data*, existe um hiato na prática

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

em como utilizar e implementar soluções de *Big Data*, assim como do entendimento e conhecimento de várias de suas ferramentas e quesitos básicos.

Como constatação das afirmações acima, destacamos o grau de discordância e a ausência de conhecimento necessário para realizar a avaliação solicitada, no âmbito de algumas das principais facetas que sustentam o *Big Data*. Por exemplo, um dos pontos cruciais na mudança de paradigma entre “o antes” e “o depois”, no cerne do aparecimento do chamado Big Data, estão as bases de dados “NoSQL”, ou seja, aquelas que não se enquadram no formato de banco de dados relacionais, assentado na lógica de tabelas em linhas e colunas normalizados. Cerca de 35,2% dos entrevistados, disseram “não saber avaliar” essa questão. A análise por porte de empresas encontrou o mesmo percentual de entrevistados que disseram “não saber avaliar”, essa questão.

Conforme destacado por Machado (2018, p. 57), um banco de dados “NoSQL”, “é projetado para aumentar a sua escala em sentido horizontal por meio de clusters distribuídos em *hardwares* de baixo custo para aumentar as taxas de transferência de dados, sem aumentar a latência”.

Vale destacar também, o grau de discordância sobre a utilização frequente de abordagens analíticas no contexto dos algoritmos de aprendizagem de máquina e inteligência artificial, que dentre outras, compõe o escopo mais direto do *Big Data Analytics*.

Sem ferramentas, abordagens e tecnologias, como as apontadas acima, a possibilidade de processamento e análise de um grande conjunto de dados, de múltiplos formatos (imagem, *tweets*, *likes*, linguagem natural, etc.), tende a ficar bastante comprometida.

Ou seja, esses resultados apontam, para algumas possibilidades:

- a) As empresas que compõe o extrato desse levantamento, em sua maioria efetivamente não implementaram de maneira plena, em todo o seu potencial, o *Big Data*;
- b) A maioria das empresas que compõe o extrato desse levantamento, tem um conhecimento teórico de *Big Data*, mas de alguma forma viesado, ou seja, continuam utilizando apenas ferramentas, abordagens e tecnologias “antigas”, ou seja, soluções que sem a utilização concomitante com as novas soluções, não tem capacidade para fazer frente aos “V” do *Big Data*.

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

Como reforço dessa visão, inserimos no grupo de questões desse bloco, uma questão teste. A questão afirmava que “no seu entendimento *Big Data* e inteligência competitiva / de mercado, são nomes distintos para a mesma abordagem”. O resultado a nosso ver foi surpreendente e reforçou a constatação da falta de entendimento dos conceitos e aplicações para a maior parte das empresas pesquisadas nesse estudo. Apenas 48% das empresas discordaram da afirmação, ou seja, compreendem que são abordagens distintas. Já 28% das empresas, consideraram as abordagens como sendo nomes distintos para a mesma abordagem; e aqueles que “não concordaram, nem discordaram”, somados aos que “não souberam avaliar”, somaram 24%, ou seja, o bloco das empresas que não tem claro os conceitos avaliados somaram, mais da metade do conjunto de todas as empresas pesquisadas (52%).

c) A maioria das empresas que compõe o extrato desse levantamento, praticam o chamado “*analytics*”, ou até mesmo “*advanced data analytics*”, e pensam ter adentrado no mundo do *Big Data* e de sua abordagem analítica de alto nível, chamada de “*Big Data Analytics*”.

d) Apesar de acreditarem que entendem os conceitos, novas demandas, aplicações, ferramentas, tecnologias e capacidades requeridas pelo *Big Data*, elas na verdade não possuem em sua maioria um pleno entendimento.

A nosso ver, os dados apontam para indícios com alguma consistência de que não é difícil a ocorrência de mais de uma dessas alternativas de forma concomitante.

Tabela 3: Posicionamentos e recursos x Setores

		Principal setor organização			
		Indústria	Comércio	Serviços	Média
Utilizamos novos formatos de bases de dados como NoSQL, MongoDB, Cassandra, Hbase, myXMLB, Redis, dentre outros	Discordo	28,6%	42,9%	19,5%	30,3%
	Concordo	0,0%	28,6%	36,4%	21,7%
Integramos dados internos e externos para analisar nosso ambiente externo	Discordo	28,6%	28,6%	18,2%	25,1%
	Concordo	42,9%	71,4%	63,6%	59,3%
Temos grande entendimento das áreas que podem se beneficiar das análises de big data	Discordo	42,9%	28,6%	27,3%	32,9%
	Concordo	28,6%	71,4%	51,9%	50,6%
Utilizamos softwares de código aberto (ex.: Hadoop, R, Spark, Hive, Pig, dentre outros) para as análises de big data	Discordo	57,1%	28,6%	23,4%	36,4%
	Concordo	0,0%	42,9%	39,0%	27,3%

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

Temos funcionários capacitados para análise de big data	Discordo	14,3%	42,9%	32,5%	29,9%
	Concordo	28,6%	42,9%	32,5%	34,7%
Temos fácil acesso a múltiplas fontes de dados internos	Discordo	57,1%	57,1%	24,7%	46,3%
	Concordo	28,6%	42,9%	50,6%	40,7%
Utilizamos com frequência em nossas análises diversos algoritmos no contexto da aprendizagem de máquina, inteligência artificial, dentre outros	Discordo	71,4%	42,9%	35,1%	49,8%
	Concordo	14,3%	42,9%	41,6%	32,9%

Fonte: elaborada pelos autores, pesquisa realizada 2019

Quando analisado o tempo em que a organização trabalha/utiliza *Big Data*, na comparação com os setores, é possível reforçar as constatações anteriores. Como a maior parte das organizações ainda estão num estágio inicial de utilização do *Big Data*, é natural que algumas idiosincrasias estejam presentes entre o discurso e a prática do *Big Data* nas empresas pesquisadas.

Na análise do cruzamento do porte das empresas pelo tempo de utilização do *Big Data*, a diferença mais nítida é a quantidade de empresas de médio e grande porte que utilizam o *Big Data* a mais de três anos.

Fica evidente também que a maioria das empresas dos dois extratos, são organizações iniciantes na utilização do *Big Data*, o que de alguma forma, aponta para a dificuldade desse grupo majoritário no acesso e utilização plena dos principais posicionamentos e recursos no contexto do *Big Data*.

Tabela 5: Tempo utilização Big Data x Setores

Porte	Tempo utilização Big Data		
	Menos de 1 ano	De 1 ano até 3 anos	Mais de 3 anos
Micro e Pequenas	60,9%	34,8%	4,3%
Médias e Grandes	52,9%	26,5%	20,6%

Fonte: elaborada pelos autores, pesquisa realizada 2019

Na Tabela 6 (abaixo), podemos visualizar as principais dificuldades no processo de utilização do *Big Data*, indicadas pelas empresas que compõe o extrato analítico dessa pesquisa. De maneira destacada e alinhada às diversas indicações da literatura existente sobre o tema, temos a questão da “cultura organizacional”, que pesquisas anteriores em diversos campos da ciência, tem mostrado ser uma grande barreira para implementação de diversos projetos e iniciativas, principalmente aquelas de cunho inovador, que tratam de questões novas em diversos matizes no âmbito das organizações.

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC**

Suppiah e Sandhu (2010, p. 462, tradução nossa), numa pesquisa realizada sobre a influência da cultura organizacional no comportamento de compartilhamento de conhecimento tácito, apontam como resultados do trabalho que, “os tipos de cultura organizacional influenciam o comportamento tácito de compartilhamento de conhecimento”, onde essa influência tende a ser positiva ou negativa em função da cultura organizacional existente.

Por sua vez, Shahbaz et al (2019), pontuam em pesquisa realizada sobre o papel da resistência à mudança no âmbito da adoção do *Big Data Analytics*, pontual que, apesar de não terem investigado diretamente a questão da cultura, ela tem um efeito sobre o nível de adoção do BDA, apontando ainda que a cultura organizacional pode variar de setor para setor.

A segunda dificuldade mais citada, “conhecimento inexistente ou insuficiente quanto aos modelos analíticos (algoritmos) e como e onde aplica-los”, guarda uma profunda relação com as respostas no âmbito da investigação sobre “os posicionamentos e recursos da organização sobre *Big Data*”, onde o quesito “utilizamos com frequência em nossas análises diversos algoritmos no contexto da aprendizagem de máquina, inteligência artificial, dentre outros”, obteve o índice mais alto de “discordância” com a afirmação apresentada.

O que mais chama a atenção para essa questão são os níveis de “nem concordo, nem discordo” e “não sei avaliar”, terem sido relativamente baixos, com 11% das indicações cada um, ou seja, isso reforça o nível de discordância encontrado para esse quesito.

De maneira geral as demais dificuldades encontradas, também reforçam o grau de discordância encontrado em todos os quesitos avaliados quanto “aos posicionamentos e recursos de sua organização sobre *Big Data*”.

Quando foram avaliadas as principais dificuldades encontradas no processo de utilização do Big Data, segundo o porte das empresas, foi possível identificar que as empresas de menor porte (micros e pequenas), tiveram scores mais baixos na indicação das dificuldades elencadas.

Parte dessa visão pode ser explicada pelo tempo menor de utilização do *Big Data*, pelas empresas de menor porte, ou seja, 96% dessas empresas utilizam o Big Data a menos de três anos, sendo que desse percentual, mais da metade, 61% a menos de um ano.

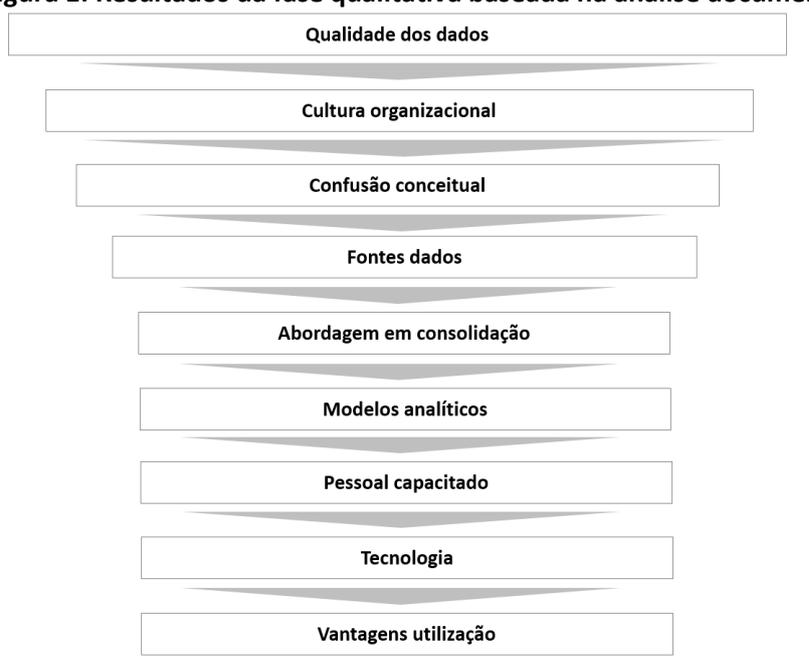
No que tange à parte qualitativa dessa pesquisa, essa esteve baseada na abordagem de análise documental, contemplada pela análise dos seguintes documentos: apresentações

dos palestrantes durante o evento; apontamentos registrados pelos pesquisadores durante o evento; transcrição da sessão de perguntas e respostas e apontamentos dos respondentes em campo aberto, disponibilizado para considerações adicionais, no corpo do questionário.

Com base nessa etapa qualitativa, foi possível aprofundar algumas visões, ampliar o escopo de entendimento e reforçar questões importantes que surgiram nessa fase.

A representação dos principais resultados, foi consolidada num quadro com a utilização de uma escala de recorrência de temas, baseada nos apontamentos, reunidos a partir da “contagem” e análises individuais dos pesquisadores, das colocações, desenhos, frases ou outros extratos de informação que se vinculavam e/ou reforçavam as questões que estruturaram a construção desse estudo.

Figura 1: Resultados da fase qualitativa baseada na análise documental



Fonte: Elaborada pelos autores, pesquisa realizada 2019

De maneira complementar, visando uma maior profundidade analítica e efetivo relacionamento com primeira fase desse estudo, faremos a seguir uma rápida análise de cada dimensão.

Qualidade dos dados: foi a dimensão mais recorrente na análise documental, sendo entendida como o ponto principal de qualquer abordagem de *Big Data*. Ou seja, como garantir a qualidade dos dados, sua validade, confiabilidade e disponibilidade são pilares sustentadores das abordagens de *Big Data*. A questão de segurança dos dados, também foi

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

considerada nessa dimensão. Destaque ainda para o fato de que no quadro de avaliação das dificuldades encontradas pelas organizações no processo de utilização do *Big Data*, essa dimensão foi ranqueada como a terceira maior dificuldade, com 24,2% do total de citações.

Cultura organizacional: nessa dimensão destaque para os múltiplos apontamentos à questão de organizações “*data-driven*” e do papel das lideranças (facilitando ou dificultando os processos). Destaque também para um quesito complementar, que surgiu nessa fase de análise documental, que foi a questão da importância “de se fazer as perguntas certas/pertinentes”. Nessa dimensão ainda encontramos vínculo com duas das dificuldades listadas no que tange ao escopo de utilização do Big Data, sendo: a) Cultura organizacional (perceber valor no uso da ferramenta, mudar processos antigos, etc.); b) Estruturar informações, ou seja, definir que dados devem ser coletados e analisados. Nesse último item, a ideia é de que, sem a pergunta adequada, não há possibilidade de estruturar o problema, e portanto, não há como se definir adequadamente quais dados deverão ser coletados e analisados.

Confusão conceitual: apesar do nítido entendimento de que, como apontado pela literatura, e pelos resultados da fase quantitativa, ainda há uma considerável confusão sobre o conceito de *Big Data*. Foi possível detectar que uma parte fundamental da grande maioria das conceituações sobre o tema está bem entranhada na cabeça dos profissionais que integram as empresas pesquisadas. E essa quase unanimidade está na questão dos “V” do *Big Data*, onde destacou-se nesse estudo, os seguintes: volume, variedade, valor e velocidade.

Fontes de dados: nessa dimensão destaque para a necessidade e utilização de múltiplas fontes de dados, internas e externas. Reforço do escopo pesquisado na primeira fase, no âmbito dos “posicionamentos e recursos da organização sobre *Big Data*”: a) integramos dados internos e externos para analisar nosso ambiente externo; b) temos fácil acesso a múltiplas fontes de dados internos.

Abordagem em consolidação: essa dimensão válida e reforça a constatação de que a maioria das empresas que integraram o escopo dessa pesquisa, estão em estágio inicial de adoção/utilização do *Big Data*.

Modelos analíticos: na análise documental destaque para aqueles modelos vinculados ao escopo da aprendizagem de máquina e da inteligência artificial. O ruído detectado está no contexto do distanciamento entre a citação / a visão teórica desses modelos e algoritmos e

seu efetivo entendimento e visão prática, em destaque no rol de dificuldades avaliadas no contexto das dificuldades encontradas para utilização pelas organizações do *Big Data*.

Pessoal capacitado: na etapa quantitativa, essa dimensão foi analisada dentro do rol de dificuldades de utilização do Big Data e também no escopo dos posicionamentos e recursos das organizações sobre *Big Data*. O apontamento se dá no sentido de que esse é um considerável gargalo que as organizações devem envidar esforços para suplantá-lo.

5 CONCLUSÕES

Como detectado na literatura e nos resultados dessa pesquisa, ainda temos uma considerável confusão conceitual em torno do tema *Big Data*. Uma parte considerável de influência sobre essa situação, parece estar ligado à questão de ser um tema relativamente novo de utilização recente pela maior parte das empresas, que se encontram em estágios iniciais de uso de utilização. As empresas de maneira geral, objetivando queimar etapas e avançar de forma mais ágil na implementação e utilização do *Big Data*, deveriam pensar na concretização de ações e investimentos visando o acesso a recursos com as capacidades adequadas. E isso envolve, não só o desenvolvimento de habilidades analíticas, como também de utilização de ferramentas e identificação de problemas relevantes (entendimento dos principais drivers do negócio).

Identificação e acesso a dados de qualidade, de fontes múltiplas e diferentes estruturas em quantidades que possam viabilizar os modelos de Big Data, são fundamentais. Acesso a tecnologias que possibilitem processamento de grandes massas de dados e que contemplem o armazenamento e processamento de dados de múltiplas fontes e formatos, é um outro ponto crítico.

Por fim, uma questão, que não é exclusiva da temática *Big Data*, mas de diversas outras temáticas e iniciativas, que é a questão da cultura organizacional. Ações no sentido de mudança do “*mindset*” da organização, no sentido de se conformar organizações “*data-driven*” e um cuidado e ações efetivas no que tange às lideranças são a base de sustentação de qualquer iniciativa no âmbito do Big Data.

Uma das limitações dessa pesquisa foi o número de empresas participantes. Para futuros trabalhos sugerimos a realização de um esforço adicional no sentido de se conseguir um volume maior de participantes, que possam reforçar, principalmente a estrutura analítica, possibilitando a utilização de modelos de análise mais robustos, no contexto da análise

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

multivariada, por exemplo. Assim, como um possível maior equilíbrio, talvez a partir da predefinição de extratos, entre os setores e portes das empresas.

6 REFERÊNCIAS

BARBOSA, Ricardo Rodrigues. Levantamento base Scopus, 2018

BRAGANZA, Ashley; BROOK, Laurence; NEPELSKI, Daniel; ALI, Maged; MORO, Russ. **Resource management in Big Data initiatives: processes and dynamic capabilities**. Journal of Business Research, 2017.

BRANDS, Kritine. **Big Data and Business Intelligence for Management Accountants**. Strategic Finance: Tech Practices, June 2014.

COX, Michael; ELLSWORTH, David. **Application Controlled Demand Paging for Out-of-Core Visualization**. NASA, Ames Research Center, 1997. Acesso em: <https://www.nas.nasa.gov/assets/pdf/.../1997/nas-97-010.pdf>, acesso em 10/07/2019.

CRESWELL, John W. **Research design: qualitative, quantitative, and mixed methods approaches**. 4. ed. Thousand Oaks: Sage, 2014.

DAVENPORT, Thomas H. **Big Data no trabalho: derrubando mitos e descobrindo oportunidades**. Rio de Janeiro: Elsevier, 2014.

DAVENPORT, Thomas H; HARRIS, Jeanne G. **Competição analítica: vencendo através da nova ciência, como decisões baseadas em fatos e business intelligence transformaram os negócios**. Rio de Janeiro: Elsevier, 2007.

FOREMAN, John W. **Data Smart: usando data Science para transformar informação em insight**. Rio de Janeiro: Alta Books, 2016.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social**. 6 ed. São Paulo: Editora Atlas, 2008, 200 p.

GUPTA, Manjul; GEORGE, Joey F. **Toward the development of a Big Data analytics capability**. Information & Management, v. 53, n. 8, p. 1049–1064, 2016.

HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin. **The fourth paradigm. Data-intensive scientific Discovery. 2 v**. Redmond, Wahsington: Microsof Research, 2009.

JAIN, Priyank; GYANCHANDONI, Manasi; KHARE, Nilay. **Big Data privacy: a technological perspective and review**. Journal of Big Data, 2016.

JIN, Xialong; WAH, Benjamin W.; CHENG, Xueqi; WANG, Yuanzhuo. **Significance and Challenges of Big Data Research**. Elsevier, 2015; p. 59-64.

MACHADO, Felipe Nery Rodrigues. **Big Data: o futuro dos dados e aplicações**. São Paulo: Érica, 2018.

MASHINGAIDZE, Kenneth; BACKHOUSE, Judy. **The relationships between definition of BD, BI and BA**. Int. J. Business Information Systems, 2017; vol. 26; n. 4.

SHAHBAZ, Muhammad; GAO, Changyuan; ZHAI, Lili; SHAHZAD, Fakhar; HU, Yanling. **Investigating the adoption of Big Data analytics in healthcare: the moderating role of resistance to change**. Journal of Big Data, 2019.

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019
21 a 25 de outubro de 2019 – Florianópolis – SC

SUPPIAH, Visvalingam; SANDHU, Manjit Singh. **Organizational culture's influence on tacit knowledge sharing behaviour**. Journal of Knowledge Management, 2010, p. 462-477.

WAMBA, Samuel Fosso; GUNASEKARAN, Anagappa; AKTER, Shahriar; REN, Steven Ji-fan; DUBEY, Rameshwar; CHILDE, Stephen J. **Big Data analytics and firm performance: effects of dynamic capabilities**. Journal of Business Research, 2017.

TAURION, Cezar. **Big Data**. Rio de Janeiro: Brasport, 2015.

VIEIRA, Marcelo Milano Falcão; ZOUAIN, Deborah Moraes.(Org.). **Pesquisa qualitativa em administração: teoria e prática**. Rio de Janeiro: Editora FGV, 2005. 237 p.