



Campus São Mateus  
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO



Pró-Reitoria de Pesquisa e Pós-Graduação  
Universidade Federal do Espírito Santo

ARTIGO ORIGINAL

OPEN ACCESS

## ANÁLISE DE COCITAÇÕES EM MÚLTIPLAS PERSPECTIVAS SOBRE O TEMA *BIG DATA ANALYTICS*

*MULTIPLE PERSPECTIVE CO-CITATION ANALYSIS ON THE BIG DATA ANALYTICS THEME*

*ANÁLISIS DE COCITACIONES EN MÚLTIPLES PERSPECTIVAS SOBRE EL TEMA *BIG DATA ANALYTICS**

Arthur Moura Vaz<sup>1</sup>, Stella Jacyszyn Bachega<sup>2\*</sup>, & Dalton Matsuo Tavares<sup>3</sup>

<sup>1,2</sup> Engenharia de Produção, <sup>3</sup> Ciência da Computação, Universidade Federal de Catalão

<sup>1</sup> [arthur.moura.eng@gmail.com](mailto:arthur.moura.eng@gmail.com) <sup>2</sup> [stella@ufcat.edu.br](mailto:stella@ufcat.edu.br) <sup>3</sup> [dalton\\_tavares@ufcat.edu.br](mailto:dalton_tavares@ufcat.edu.br)

### ARTIGO INFO.

**PALAVRAS-CHAVE:** *Big Data Analytics*; CiteSpace®; Análise de Cocitações em Múltiplas Perspectivas; Indústria 4.0.

**KEYWORDS:** *Big Data Analytics*; CiteSpace®; Multiple Perspective Co-Citation Analysis; Industry 4.0.

**PALABRAS CLAVE:** *Big Data Analytics*; CiteSpace®; Análisis de Co-citaciones en Múltiples Perspectivas; Industria 4.0.

\*Autor Correspondente: Bachega, S. J.

### RESUMO

Uma das tecnologias da indústria 4.0 é o *Big Data Analytics*, que permite transformar a grande quantidade de diversos tipos de dados em informações de valor e com potencial para tomada de decisões, inovação, entre outros. Neste contexto, o objetivo do artigo é apresentar uma análise de cocitações em múltiplas perspectivas sobre o tema *Big Data Analytics*. Para tanto, utilizou-se a explicação científica hipotético-dedutiva, a abordagem de pesquisa mista qualitativa e quantitativa e o método de análise de cocitações em múltiplas perspectivas, com o uso do software CiteSpace®. Considerou-se 11 anos completos de publicações e as publicações já disponibilizadas no ano corrente da pesquisa na Web of Science (WoS). Foram verificadas as principais linhas de pesquisa sobre o tema, os principais autores, as palavras-chave de destaque, a linha do tempo de realização dos estudos nas principais áreas de pesquisa e os países líderes em publicações. O estudo evidenciou a existência de uma forte relação entre o *Big Data Analytics* e o *Supply Chain Management*, sendo um grande indicativo da influência desta tecnologia no setor produtivo. Dentre as contribuições, esta pesquisa aponta possibilidades e vantagens de aplicação da análise de grande volume de dados e desafios da mesma.

### ABSTRACT

*One of the technologies of industry 4.0 is Big Data Analytics, which allows transforming a large number of different types of data into valuable information with potential for decision-making, innovation, among others. In this context, the objective of the article is to present an analysis of co-citations in multiple perspectives on the theme Big Data Analytics. For that, we used the hypothetical-deductive scientific explanation, a mixed qualitative and quantitative research approach and the method of analysis of co-citations in multiple perspectives, using the CiteSpace® software. We considered 11 full years*

*of publications, and the ones already available in the current year of the Web of Science (WoS) search. We verified the main lines of research on the subject, the main authors, the prominent keywords, the timeline of studies in the main areas of research and the leading countries in publications. The study showed the existence of a strong relationship between Big Data Analytics and Supply Chain Management, being a great indication of the influence of this technology in the productive sector. Among the contributions, this research points out possibilities and advantages of applying large data analysis and its challenges.*

### RESUMEN

*Una de las tecnologías de la industria 4.0 es Big Data Analytics, que permite transformar la gran cantidad de diferentes tipos de datos en información valiosa con potencial para la toma de decisiones, innovación, entre otros. En este contexto, el objetivo del artículo es presentar un análisis de cocitaciones en múltiples perspectivas sobre el tema Big Data Analytics. Para ello, se utilizó la explicación científica hipotética-deductiva, el enfoque de investigación mixto cualitativo y cuantitativo y el método de análisis de cocitaciones en múltiples perspectivas, utilizando el software CiteSpace®. Se consideraron 11 años completos de publicaciones y las publicaciones ya disponibles en el año actual de la búsqueda en Web of Science (WoS). Se verificaron las principales líneas de investigación sobre el tema, los principales autores, las palabras clave destacadas, la cronología de los estudios en las principales áreas de investigación y los países líderes en publicaciones. El estudio mostró la existencia de una fuerte relación entre Big Data Analytics y Supply Chain Management, siendo un gran indicador de la influencia de esta tecnología en el sector productivo. Entre las contribuciones, esta investigación señala las posibilidades y ventajas de aplicar el análisis de grandes datos y sus desafíos.*



## INTRODUÇÃO

A implantação do modelo de indústria 4.0 no Brasil tem sido uma preocupação existente em diversos setores da sociedade, como empresas, universidades e governo. De acordo com um estudo da Agência Brasileira de Desenvolvimento Industrial (ABDI, 2020), se a indústria nacional adotasse tecnologias 4.0, a economia poderia chegar a R\$ 73 bilhões por ano. No entanto, esta revela que cerca de 43% das empresas não veem as novas tecnologias como meio de alavancar a indústria.

Dentro desse tema, encontra-se o conceito de *Big Data Analytics*, que sucede o termo *Big Data*, um grupo de dados de difícil ou até mesmo impossível processamento por meio do uso de métodos tradicionais, devido ao seu grande tamanho, rapidez e complexidade (SAS, 2020a). Nesse sentido, o *Big Data Analytics* refere-se à técnica de análise de *Big Data* para gerar estatísticas passadas, atuais e futuras, e informações úteis a serem utilizadas na tomada de melhores decisões de negócios (Ankam, 2016).

Com base nesse contexto, surge a seguinte questão de pesquisa: Quais são os principais artigos que contribuem para o tema *Big Data Analytics* nos últimos 11 anos? Logo, o objetivo geral deste artigo é realizar uma análise de cocitações em múltiplas perspectivas sobre o referido tema.

Para tanto, foram estabelecidos os seguintes objetivos específicos: i) analisar as referências em comum, utilizadas pelas publicações que se encontram na base de dados empregada neste trabalho e assim, elucidar as principais linhas de pesquisa acerca do conceito de *Big Data Analytics*; ii) identificar as palavras-chave mais frequentes dentro das publicações mapeadas; iii) estabelecer os países que lideram os estudos pertencentes ao assunto.

Portanto, o presente trabalho justifica-se pela importância do tema discutido, dado que as organizações podem converter seus dados coletados e armazenados diariamente em conhecimento relevante por meio de *Big Data Analytics*, como exposto por Kühne (2020). Além disso, 93% dos gestores de 32 principais indústrias do país preveem que os dados tenham considerável influência nas suas decisões, conforme aponta pesquisa da PricewaterhouseCoopers Brasil (PwC, 2016). Laudon e Laudon (2020) também discorrem sobre a análise de dados e a sua apresentação de forma significativa e útil, como informações que apoiem a tomada de decisões.

## REFERENCIAL TEÓRICO

A International Business Machines Corporation (IBM, 2020) define o *Big Data* como um conjunto de dados que está além da capacidade que bancos de dados tradicionais possuem de administrar e processá-lo, devido ao seu tamanho ou tipo. Estes dados podem vir de diversas fontes, como: sensores, dispositivos, arquivos de áudio e vídeo, redes de computadores e mídia social.

O conceito de *Big Data* pode ser representado pelos cinco “v’s”, que tratam de características ligadas à forma como os dados são gerados e à sua importância. Estas características, de acordo com a definição de Sagioglu et al. (2016), são: i) volume:



existência de um número elevado de registros, sendo necessária uma ampla capacidade de armazenamento; ii) velocidade: transferência e geração de dados com grande frequência; iii) variedade: diversificada rede de fontes e formatos de dados; iv) veracidade: característica ligada à confiabilidade e qualidade dos dados; v) valor: os dados devem fornecer padrões e insights úteis.

De acordo com Ankam (2016), o *Big Data Analytics* é majoritariamente classificado em duas categorias. O *Data Analytics* trata da coleta e interpretação dos dados, tipicamente com foco em estatísticas passadas e presentes, tem a natureza do trabalho de relatar e otimizar, e tem como saídas relatórios e dashboards. Por outro lado, o *Data Science* foca no futuro, por meio da execução de análise exploratória para prover informações baseadas em modelos identificados pelos dados passados e presentes, e tem como saída o *data product*.

Em síntese, Schwab (2016) sustenta que o uso do *Big Data* viabilizará processos decisórios melhores e mais rápidos em diversas indústrias e outras aplicações. No entanto, a exploração deste grande volume de dados traz tanto oportunidades, quanto riscos, de forma significativa, podendo gerar impactos positivos (ex.: redução de custos, mais tomada de decisões em tempo real etc.); negativos (ex.: perdas de trabalho, preocupações com a privacidade etc.) ou até mesmo ainda desconhecidos (ex: mudanças nas estruturas comerciais e normativas).

## METODOLOGIA

A explicação científica usada na pesquisa foi hipotético-dedutiva (Carvalho, 2000). Na realização da análise de cocitações em múltiplas perspectivas foram consideradas as proposições ligadas ao objetivo geral desta pesquisa. Ainda, foi empregada a abordagem de pesquisa mista qualitativa e quantitativa (Creswell, 1994). Na tabulação dos resultados da análise, foi adotada a perspectiva quantitativa, enquanto a abordagem qualitativa foi utilizada na análise da literatura, na busca de se inteirar do fenômeno investigativo e obter informações.

Para a realização deste estudo foi aplicada a análise de cocitações em múltiplas perspectivas (Chen, 2006). Este procedimento faz uso do software CiteSpace®, para visualização e análise de tendências de crescimento, ou mudanças, em redes de cocitações dentro do campo da literatura científica (Chen, 2004; Chen, Ibekwe-Sanjuan, & Hou, 2010).

Para a aplicação do método, foram empregados os passos descritos por Li, Ma e Qu (2017):

1) Construção de um banco de dados: a base de dados empregada neste trabalho foi a Web of Science (WoS), que possui integração com o software CiteSpace®, além de ser internacionalmente reconhecida pela sua relevância.

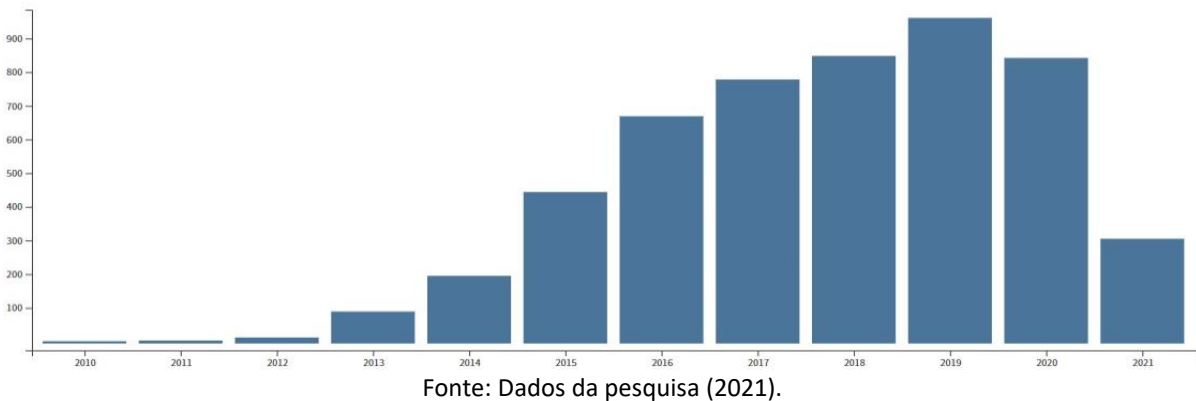
Para o refinamento do estudo, três filtros foram aplicados para selecionar os tipos de documentos utilizados no trabalho, sendo eles “artigos”, “artigos de congressos” e “*early access*” (artigos de 2021 já liberados para acesso), pois estes contribuíram para adquirir artigos mais afins ao objetivo da pesquisa.



2) Utilização das palavras-chave adequadas: a palavra-chave empregada foi “*Big Data Analytics*”. O termo foi selecionado por pertencer ao campo do “*Big Data*”, que é uma área mais abrangente, por ser inserido na indústria 4.0 e por sua relevância.

Os artigos selecionados foram publicados entre os anos de 2010 e 2021. O motivo desta escolha justifica-se pelo fato de o ano de 2010 refletir o início de publicações sobre o tema na WoS (Figura 1). Como resultado, obteve-se 5.147 publicações, as quais posteriormente constituíram a base de dados expandida.

**Figura 1.** Quantitativo de publicações relacionadas ao tema *Big Data Analytics*.



3) Extensão da base de dados: são inseridas no banco de dados, as redes de cocitações apontadas pelas referências dos artigos escolhidos (Chen *et al.*, 2010). Houve a obtenção da base de dados por meio da seleção das opções de “registro completo e referências citadas” para a gravação do conteúdo, e “texto sem formatação” para a escolha do formato dos arquivos.

4) Mapeamento do desenvolvimento do tema com o uso do CiteSpace®: o processo de análise baseou-se primeiro na categorização dos conjuntos de referências cocitadas (*clusters*) do tema *Big Data Analytics*. Depois, a identificação dos estudos predominantes que contribuíram significativamente com o desenvolvimento do tema, e por último, a realização de uma análise cronológica dos estudos em relação à origem, desenvolvimento e *status* atual.

5) Identificação do desenvolvimento da pesquisa mais recente: buscou-se identificar os tópicos de pesquisa relevantes e pesquisas líderes no tema estudado, com base na frequência de palavras-chave populares empregadas nos artigos de periódicos.

6) Mapeamento dos países citados: este passo contribuiu com a identificação e visualização dos principais países que colaboraram com o desenvolvimento do mapa do conhecimento sobre *Big Data Analytics*.

## RESULTADOS OBTIDOS

### Mapeamento, Modularidade e Silhueta

O Identificou-se a existência de 90 *clusters*, sendo apenas 17 válidos. Cada um destes *clusters* representa uma linha de pesquisa dentro da área do *Big Data Analytics*. Dentre os conjuntos válidos encontrados, é possível citar os cinco principais em ordem decrescente de tamanho, conforme a nomenclatura utilizada pelo CiteSpace®: #0 *Big Data Analytics*



*application*, #1 *mediating role*, #2 *Big Data-driven knowledge management system*, #3 *industrial internet* e #4 *Supply Chain Management*.

Como forma de sustentar os resultados gerados pela função de agrupamento, o CiteSpace® disponibiliza duas importantes métricas. A modularidade Q indica a extensão em que a rede obtida pode ser decomposta em outros módulos, e deve possuir valores entre 0 e 1, em que valores próximos a 1 implicam em maior conexão entre os *clusters* (Chen *et al.*, 2010). A silhueta S mede a qualidade de configuração do *cluster* e seu valor deve variar entre -1 e 1, em que o valor mais alto expressa uma solução perfeita (Chen *et al.*, 2010). Para o tema do *Big Data Analytics*, obteve-se o valor de Q = 0,6464, ou seja, um valor aceitável para a qualidade de conexão entre os *clusters*, enquanto a silhueta apresentou o valor de S = 0,8383. Logo, este resultado mostra ótima qualidade de configuração dos *clusters* apresentados.

### Detecção de Explosão (*Burst*)

A detecção de explosão está relacionada a duração específica em que houve uma mudança repentina na frequência de pesquisa em determinada área, o que pode evidenciar, por exemplo, que uma publicação atraiu maior grau de interesse pela comunidade científica, e em qual espaço de tempo isto aconteceu (Kleinberg, 2002). Por meio da Tabela 1, é possível identificar que a publicação de Chen, Chiang e Storey (2012) mostrou a explosão de citação mais significativa (Força = 52,47), tendo perdurado por um período de três anos. As referências estão apresentadas conforme explicitado no relatório do CiteSpace®.

Tabela 1. Top 10 bibliografias com explosão de citações (autores).

Referência	Força	Início	Fim	2010 - 2021
Manyka J, 2011, BIG DATA NEXT FRONTI, V0, P0	33,61	2012	2016	
Chen HC, 2012, MIS QUART, V36, P1165	52,47	2014	2017	
McAfee A, 2012, HARVARD BUS REV, V90, P60	36,08	2014	2017	
Zikopoulos P, 2011, UNDERST BIG DA, V0, P0	14,72	2014	2016	
Wu XD, 2014, IEEE T KNOWL DATA EN, V26, P97	19,46	2015	2017	
Boyd D, 2012, INFORM COM SOC, V15, P662	16,82	2015	2017	
Raghupathi W, 2014, HEALTH INF SCI SYST, V2, P0	13,29	2016	2018	
Chen M, 2014, MOBILE NETW APPL, V19, P171	13,06	2016	2018	
Chen CLP, 2014, INFORM SCIENCES, V275, P314	17,11	2017	2018	
Waller MA, 2013, J BUS LOGIST, V34, P77	16,5	2017	2018	

Fonte: Dados da pesquisa (2021).

Além disso, ao ser considerada a fase mais recente do período analisado (início 2017 e fim 2018), o estudo com a maior força de explosão de citação é o de Chen e Zhang (2014), tornando-o uma fonte importante para pesquisas atuais.



Na Tabela 2, é possível perceber que a palavra-chave com o maior número de citações, entre as publicações do banco de dados, é *Hadoop*.

*Hadoop* faz referência à estrutura de software de código aberto para armazenamento de dados em grande volume (SAS, 2020b). Ainda, ressalta-se que dentre as palavras-chave encontradas, o *Cloud Computing* vem sendo um tema crescente tanto de estudo, quanto de aplicação corporativa, onde se prevê um aumento de 18% no dispêndio de recursos financeiros pelas organizações para atender demandas de computação em nuvem, ainda em 2021 (GARTNER, 2021).

**Tabela 2.** Top 5 palavras-chave com explosão de citações (palavras-chave).

Palavras-chave	Força	Início	Fim	2010 – 2021
<i>Map Reduce</i>	22,73	2010	2016	
<i>Cloud Computing</i>	17,99	2010	2016	
<i>Hadoop</i>	27,93	2012	2017	
<i>Ontology</i>	7,21	2014	2017	
<i>Resource Based View</i>	6,11	2019	2021	

Fonte: Dados da pesquisa (2021).

### CENTRALIDADE, RANKING DE CITAÇÕES E SIGMA

Uma outra importante métrica é a centralidade, um cálculo da extensão em que determinado nó se encontra na metade de um caminho que se conecta a outros nós da rede (Brandes, 2001; Freeman, 1977). Isto faz com que valores mais altos de centralidade indiquem publicações chave ou de grande relevância (Chen, 2005). Na Tabela 3 há os maiores valores de centralidade encontrados. As quatro primeiras publicações em ordem decrescente de centralidade, de Gunasekaran *et al.* (2017), Chen, Preston e Swink (2015), Papadoulos *et al.* (2017) e Wang *et al.* (2016) e mais outros três estudos da tabela, estão relacionados ao *Supply Chain Management* (Gestão da Cadeia de Suprimentos), revelando assim que este tem sido um tema significativo quando se trata da aplicabilidade do *Big Data Analytics*.

**Tabela 3.** Maiores valores de centralidade encontrados.

Centralidade	Referências	Cluster#
67	Gunasekaran A, 2017, J BUS RES, 70, 308	2.5
56	Chen DQ, 2015, J MANAGEMENT INFORM SYST, 32, 4	4.5
56	Papadoulos T, 2017, J CLEAN PROD, 142, 1108	2.5
55	Wang G, 2016, INT J PROD ECON, 176, 98	2.5
54	Wamba SF, 2017, J BUS RES, 70, 356	2.5
54	Wamba S, 2015, INT J PROD ECON, 165, 234	3.5
52	Hazen BT, 2014, INT JPROD ECON, 154, 72	3.5
52	Akter S, 2016, INT J PROD ECON, 182, 113	3.5
50	Waller, MA, 2013, J BUS LOGIST, 34, 249	4.5
50	Chae B, 2015, INT J PROD ECON, 165, 247	3.5

Fonte: Dados da pesquisa (2021).

Já o *ranking* de citações fornece um resumo das publicações que foram mais citadas no período analisado. Na Tabela 4 é possível verificar que a publicação mais citada entre 2010 e 2021 é a de Gandomi e Haider (2015). Este estudo, assim como a maior parte das dez



publicações mais citadas, trata do *Big Data Analytics* de forma mais ampla, diferentemente do que foi visto na métrica de centralidade, e discorrem aspectos como definição do tema, impactos e análise de capacidade.

O sigma é um indicador que mede a força de duas propriedades analisadas anteriormente, a explosão de citação e a centralidade, de forma que juntas revelem as publicações de maior destaque no meio científico, e que tiveram maior relevância dentro da linha de pesquisa da qual fazem parte (Chen *et al.*, 2009). Dentre as publicações com maior valor de sigma (Tabela 5), observou-se a presença equilibrada de artigos de abordagem mais ampla com aqueles que versam acerca da junção entre os temas do *Big Data Analytics* e *Supply Chain*.

**Tabela 4.** Referências com maiores quantidades de citações.

Frequência	Referências	Cluster#
232	Gandomi A, 2015, INT J INFORM MANAGE, 35, 137	3.5
191	Wamba SF, 2017, J BUS RES, 70, 356	2.5
164	Wang G, 2016, INT J PROD ECON, 176, 98	2.5
155	Wamba SF, 2015, INT J PROD ECON 165, 234	3.5
145	Akter S, 2016, INT J PROD ECON, 182, 113	3.5
123	Chen HC, 2012, MIS QUART, 36, 1165	3.5
116	Raghupathi W, 2014, HEALTH INF SCI SYST, 2, 0	3.5
115	Hashem IAT, 2015, INFORM SYST, 47, 98	2.5
111	Chen CLP, 2014, INFORM SCIENCE, 275, 314	3.5
111	Gupta M, 2016, INFORM MANAGE-AMSTER, 53, 1049	3.5

Fonte: Dados da pesquisa (2021).

**Tabela 5.** Sigma.

Sigma	Referências	Cluster#
0,13	Waller Ma, 2013 J BUS LOGIST, 34, 249	4.5
0,12	Labrinidis A, 2012 PROC VLBD ENDOW, 5, 2032	4.5
0,11	Condie T, 2010 C NETW SYST DES IM, 0, 21	4.5
0,07	Dittrich J, 2010, PROC VLBD ENDOW, 3, 518	3.5
0,06	Gunasekaran A, 2017, J BUS RES, 70, 308	2.5
0,06	Papadopoulos T, 2017, J CLEAN PROD, 142, 1108	2.5
0,06	Wamba SF, 2015, INT J PROD ECON, 165, 234	3.5
0,06	Kwon O, 2014, INT J INFORM MANAGE, 34, 387	3.5
0,06	Manyka J, 2011, BIG DATA NEXT FRONT, 0, 0	3.5
0,06	Chen Y, 2012, PROC VLDB ENDOW, 5, 1802	1.5

Fonte: Dados da pesquisa (2021).

### LINHA DO TEMPO E RANKING DE CITAÇÕES

Quanto à disposição dos *clusters* na forma de linha do tempo (*timeline*), é possível visualizar os períodos que tiveram maior número de pesquisas realizadas, além de possibilitar a identificação das publicações mais relevantes (autores-chave).

Realizando a análise da *timeline*, observou-se que o período de maior volume de publicações relacionadas ao *Big Data Analytics* (entre 2013 e 2019), é o ano de 2015, o qual se mostra como um marco de transição entre os dois tipos de abordagem citados anteriormente. Entretanto, não tão somente as publicações que se relacionam ao *supply chain* se destacam, mas também, aquelas ligadas ao campo de pesquisa em *mediating role* (papel mediador). Assim, as pesquisas que se encontram dentro dos *clusters* #1 *mediating role* e #4 *Supply Chain Management*, se mostram importantes para estudos atuais e futuros, em especial este



último. No *cluster #0 big data analytics application*, nota-se que as publicações tiveram maior relevância nos primeiros seis anos, e depois as áreas de pesquisa foram evoluindo para outros tópicos de discussão. Pode-se relacionar o fato de que neste intervalo de tempo, o assunto se encontrava em desenvolvimento e sua investigação se iniciava no meio científico, com foco maior na aplicabilidade. Logo, os estudos apresentam um gênero mais amplo. Apesar disso, as publicações ligadas a este *cluster* se revelam um bom material para interessados em conhecer de forma holística o assunto do *Big Data Analytics* e suas principais ferramentas, características e para o aprofundamento em tópicos ainda pouco explorados.

Ainda, levantou-se os países com maior número significativo de pesquisas relacionadas ao tema do *Big Data Analytics*. Identificou-se que os Estados Unidos da América lideram as pesquisas no tema, seguidos da China e Índia. Inglaterra e Austrália seguem estes países, compondo os top 5 de países líderes no tema.

### CONSIDERAÇÕES FINAIS

Neste trabalho, realizou-se uma análise de cocitações em múltiplas perspectivas sobre o tema *Big Data Analytics*, portanto, o objetivo almejado foi alcançado. Dentre os resultados, observou-se uma relação entre o relatório de *timeline* para os cinco principais *clusters* e outros resultados da pesquisa. Em particular, o ano de 2015 se mostrou um marco de transição de pesquisas de cunho mais amplo para mais específico. Um exemplo disso, é que as pesquisas ligadas ao *cluster #0 big data analytics application* praticamente deixaram de existir nesta época, enquanto aquelas ligadas ao *cluster #1 mediating role* e *#4 supply chain management* chamaram consideravelmente a atenção da comunidade científica.

Na identificação das palavras-chave mais frequentes, o aparecimento de temas como o *Hadoop* e o *Cloud Computing* correspondeu ao que realmente aconteceu na prática, uma vez que ganharam muita notoriedade quando o assunto é a análise e armazenamento de grandes volumes de dados, especialmente no meio corporativo.

Desse modo, o presente trabalho possui suas contribuições acadêmicas ao estruturar de forma sistemática a literatura concernente ao *Big Data Analytics*, podendo direcionar outros pesquisadores às linhas de pesquisa com maior tendência de crescimento e relevância, assim como aos principais autores. Além disso, o mesmo traz sua colaboração ao campo da gestão da informação, que se encontra dentro da área de engenharia organizacional da engenharia de produção (Associação Brasileira de Engenharia de Produção [ABEPRO], 2023).

Para o ramo empresarial, esta pesquisa contribui ao revelar as tendências atuais sobre o tema analisado, apontando a importância em se considerar a transformação de inúmeros dados em informação relevante que possibilite a elaboração de estratégias de mercado, decisões mais assertivas e de forma rápida, além de ter evidenciado o uso da mesma na gestão da cadeia de suprimentos. Salienta-se que a expansão do uso das ferramentas de armazenamento e leitura de dados podem trazer implicações sociais em seu aspecto ético, diante do controle e privacidade dos mesmos, e ocupacional diante da retirada de mão de obra do mercado.

Sugere-se para pesquisas futuras, aplicar a técnica estudo de escopo, ou seja, a investigação intrínseca às principais publicações apontadas, em especial, as que tangem a gestão da cadeia de suprimentos, como também a análise das instituições líderes no tema.

---





## REFERÊNCIAS

- Agência Brasileira de Desenvolvimento Industrial. (2020). *Indústria 4.0 – Uma Jornada de Transformação da Indústria*. Recuperado de <https://www.abdi.com.br/postagem/industria-4-0-uma-jornada-de-transformacao-da-industria>
- Associação Brasileira de Engenharia de Produção. (2023). A profissão. Retirado de <https://portal.abepro.org.br/profissao/>
- Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, 182, 113-131.
- Ankam, V. (2016). *Big Data Analytics*. 1a ed. Birmingham: Packt Publishing Ltd.
- Boyd, D. & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662-679.
- Brandes, U. (2021). A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology*, 25(2), 163-177. <https://doi.org/10.1080/0022250X.2001.9990249>
- Carvalho, M. C. M., de. (2000). *A construção do saber científico: algumas proposições*. Cap. 4, 63-86. Construindo o saber. Campinas: Papirus.
- Chae, B. (2015). Insights from hashtag #supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, 165, 247-259.
- Chen, C. (2004, April). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences*, Washington, DC, USA.
- Chen, C. (2005, January). The centrality of pivotal points in the evolution of scientific networks. *Proceedings of the International Conference on Intelligent User Interfaces*, San Diego, CA, USA.
- Chen, C. (2006). CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359-377. <https://doi.org/10.1002/asi.20317>
- Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, 3(3), 191-209. <https://doi.org/10.1016/j.joi.2009.03.004>
- Chen, C., Ibekwe-Sanjuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: a multiple-perspective cocitation analysis. *J. Am. Soc. Inf.Sci. Technol.* 61(7), 1386-1409. <https://doi.org/10.1002/asi.21309>
- Chen, C. L. P. & Zhang, C. Y. (2014). Data-intensive: applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314-347.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From *Big Data* to Big Impact. *MIS Quarterly*, 34(4), 1165-1188. <https://doi.org/10.2307/41703503>
- Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the Use of *Big Data Analytics* Affects Value Chain Creation in *Supply Chain Management*. *Journal of Management Information Systems*, 32(4), 4-39. <https://doi.org/10.1080/07421222.2015.1138364>
- Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, 19(2), 171-209.
- Chen, Y., Alspaugh, S., & Katz, R. (2012). Interactive analytical processing in big data systems. In *VLDB ENDOWMENT*, 38, 1802-1813.
- Condie, T., Conway, N., Alvaro, T., & Hellerstein, J. M. (2010). Map Reduce Online. In *USENIX Symposium on Networked Systems Design and Implementation*, 7, 1-15.
- Creswell, J. W. (1994). *Research design: qualitative & quantitative approaches*. 1a ed. Londres: Sage,
- Freeman, L.C. (1977). A set of measuring centrality based on betweenness. *Sociometry*, 40(1), 35-41. <https://doi.org/10.2307/3033543>
- Dittrich, J., Quiané-Ruiz, J. A., Jindal, A., Kargin, Y., Setty, V. & Schad, J. (2010). Hadoop++: Making a yellow elephant run like a cheetah (without it even noticing). In *VLDB ENDOWMENT*, 36, 518-529.
- Gandomi, A. & Haider, M. (2015). Beyond the hype: *Big Data* concepts, methods and analytics. *International Journal of Information Management*, 35(2), 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Gartner. (2021, julho 4). *Gartner forecasts worldwide public cloud end-user spending to grow 18% in 2021*. Retirado de <https://www.gartner.com/en/newsroom/press-releases/2020-1117-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-grow-18-percent-in2021>
- Gunasekaran, A., Papadoulos, T., Dubey, R., Wamba, S.F., Childe, S. J., Hazen, B., & Akter, S. (2017). *Big Data* and predictive analytics for supply chain and organizational performance. *Journal of Business Research*, 70(1), 308-317. <https://doi.org/10.1016/j.jbusres.2016.08.004>
- Gupta, M. & George, J. F. (2016). Toward the development of a *Big Data Analytics* capability. *Information & Management*, 53(8), 1049-1064. <https://doi.org/10.1016/j.im.2016.07.004>
- Hashem, I. A. T., Yaqook, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of “big



- data" on cloud computing: Review and open research issues. *Information Systems*, 47, 98-115.
- Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and *Big Data* in *Supply Chain Management*: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, 154(1), 72-80. <https://doi.org/10.1016/j.ijpe.2014.04.018>
- International Business Machines Corporation. (2020, agosto 4). What is *Big Data Analytics*? Retirado de <http://www.ibm.com/analytics/Hadoop/big-data-analytics>.
- Kleinberg, J. *Bursty and hierarchical structure in streams*. (2002, July). *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, Edmonton, Canada, 8.
- Kuhne, I. E. (2020). *Aplicação de técnicas de Big Data Analytics às Smart Grids como forma de descoberta de padrões relevantes* (Dissertação de Mestrado). Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Ijuí, RS, Brasil. Recuperado de <https://bibliodigital.unijui.edu.br:8443/xmlui/handle/123456789/6769>
- Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, 34(3), 387-394.
- Labirinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. In *VLDB ENDOWMENT*, 38, 2032-2033.
- Laudon, K. C. & Laudon, J.P. (2020). *Management Information Systems: Managing the Digital Firm*. 16a ed. New York: Pearson.
- Li, X., Ma, E., & Qu, H. (2017). Knowledge mapping of hospitality research – A visual analysis using CiteSpace. *International Journal of Hospitality Management*, 60(1), 77-93. <https://doi.org/10.1016/j.ijhm.2016.10.006>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute.
- McAfee, A. & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, 90(10), 60-66, 68, 128.
- Papadoulos, T., Gunasekaran, A., Dubey, R., Altay, N., Childe, S. J., & Wamba, S. F. (2017). The role of *Big Data* in explaining disaster resilience in supply chains for sustainability. *Journal of Cleaner Production*, 142(1), 1108-1118. <https://doi.org/10.1016/j.jclepro.2016.03.059>
- PricewaterhouseCoopers Brasil. (2016). *Indústria 4.0: Digitalização como vantagem competitiva no Brasil*. Pesquisa Global Indústria 4.0: Relatório Brasil, 1-38, 2016. Retirado de <https://www.pwc.com.br/pt/publicacoes/servicos/assets/consultoria-negocios/2016/pwc-industry-4-survey-16.pdf>
- Raghupathi, W. & Raghupati, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 1-10.
- Sagiroglu, S., Terzi, R., Canbay, Y., Colak, I. (2016, August). *Big Data* Issues in Smart Grid Systems. *Proceedings of the International Conference On Renewable Energy Research and Applications*, Birmingham, England, 5.
- SAS. (2020a, setembro 27). *Big Data*: What it is and why it matter. Retirado de [https://www.sas.com/en\\_us/insights/big-data/what-is-big-data.html](https://www.sas.com/en_us/insights/big-data/what-is-big-data.html)
- SAS. (2020b, setembro 27). *Hadoop*: What it is and why it matters. Retirado de [https://www.sas.com/en\\_us/insights/big-data/Hadoop.html](https://www.sas.com/en_us/insights/big-data/Hadoop.html)
- Schwab, K. (2016). *The Fourth Industrial Revolution*. 1a ed. New York: Penguin Random House LLC.
- Waller, M. A., & Fawcett, S. E. (2013). Click Here for a Data Scientist: Big Data, Predictive Analytics, and Theory Development in the Era of Maker Movement Supply Chain. *Journal of Business Logistics*, 34(4), 249-252.
- Waller, M. A. & Fawcett, S. E. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2), 77-84.
- Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How "big data" can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165, 234-246.
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356-365.
- Wang, G., Gunasekaran, A., Ngai, E. W. T., & Papadoulos, T. (2016). *Big Data Analytics* in logistics and *Supply Chain Management*: Certain investigations for research and applications. *International Journal of Productions Economics*, 176(1), 98-110. <https://doi.org/10.1016/j.ijpe.2016.03.014>
- Wu, X., Zhu, X., Wu, G.Q., & Ding, W. (2014). Data mining with big data. *IEE Transactions on Knowledge and Data Engineering*, 26(1), 97-107.
- Zikopoulos, P. & Eaton, C. (2011). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. 1a ed. Toronto: McGraw-Hill.

