

DOI:

## BIG DATA EXPECTATIONS AND INTERACTIONS

### EXPECTATIVAS E INTERAÇÕES DE BIG DATA

**Larissa Pereira**

UNIVERSIDADE DO EXTREMO SUL CATARINENSE (UNESC) – CRICIÚMA – SC - ORCID:

<https://orcid.org/0000-0002-3483-7154>

**Paulo João Martins**

UNIVERSIDADE DO EXTREMO SUL CATARINENSE (UNESC) – CRICIÚMA – SC - ORCID:

<https://orcid.org/0000-0001-5018-3069>

#### Abstract

We live in the information age. Data is more valuable and we produce it in an impressive way. All interactions we have with technology generate astronomical amounts of data and all of this, until recently, was discarded. From the development of technologies in the fields of information processing, data storage and computational algorithms, we can now work on these data, which were previously discarded, and transform them into very useful information in various sectors, both commercial and social. This study aims to demonstrate the motivations for the emergence of Big Data, showing an overview of the current scenario regarding the volume, speed and variety of data that are generated today. Also showing what we currently have to work with this huge amount of data, since the big challenge is to handle this infinity of data and extract relevant information from them, we will show some tools and concepts that become essential in working with Big Data. Recently a North American espionage program was released, which is only possible thanks to the advent of Big Data, the American government intercepts and analyzes daily absurd amounts of data, which are continuously generated, coming from different sources, our study tries to leave a little clearer how this process works that is being practiced by many today. To carry out the study, books on the subject of Big Data, printed and online articles, surveys with companies from all over the world, magazines and the Internet were consulted as a reference.

**Key words:** Big Data, Structured Data, Unstructured Data, Hadoop, Information Systems

#### Resumo

Vive-se a era da informação. Os dados são mais valiosos e os produzimos de uma forma impressionante. Todas as interações que temos com a tecnologia geram quantidades astronômicas de dados e tudo isso, até pouco tempo atrás, era descartado. A partir do desenvolvimento de tecnologias nos ramos de processamento de informação, armazenamento de dados e algoritmos computacionais, podemos agora trabalhar em cima desses dados, que antes eram descartados, e transformá-los em informação muito útil em diversos setores tanto comercial quanto social. Este estudo tem como objetivo demonstrar as motivações para o surgimento do Big Data, demonstrando uma visão geral do cenário atual no que tange ao volume, velocidade e variedade dos dados que são gerados hoje. Também expor o que dispomos no momento para trabalhar com essa quantidade enorme de dados, já que o grande desafio é manusear essa infinidade de dados e extrair informação relevante a partir dos mesmos, mostraremos algumas ferramentas e conceitos que se tornam imprescindíveis no trabalho com Big Data. Recentemente foi divulgado um programa de espionagem Norte Americana, que só é possível graças ao advento do Big Data, o governo Americano intercepta e analisa diariamente quantidades absurdas de dados, que são gerados continuamente, vindos de diversas fontes, o nosso estudo tenta deixar um pouco mais claro como funciona esse processo que está sendo praticado por muitos atualmente. Para realização do estudo foram consultados como referência, livros voltados ao assunto Big Data, artigos impressos e online, pesquisas com empresas de todo o mundo, revista e Internet.

**Palavras-chave:** Big Data, Dados Estruturados, Dados Não Estruturados, Hadoop, Sistemas de Informação

# Iniciação Científica (PIC 170)

<b>Título do Projeto</b>	Expectativas e interações de Big Data
<b>Nome do Bolsista</b>	Larissa Pereira
<b>Nome do Orientador</b>	Paulo João Martins
<b>Grupo de Pesquisa</b>	Redes de Comunicação
<b>Palavras-chave</b>	Big Data, Dados Estruturados, Dados Não Estruturados, Hadoop
<b>Período de Vigência da Bolsa</b>	Abril de 2019 a fevereiro de 2020

## Resumo (máximo 250 palavras)

---

Vive-se a era da informação. Os dados são mais valiosos e os produzimos de uma forma impressionante. Todas as interações que temos com a tecnologia geram quantidades astronômicas de dados e tudo isso, até pouco tempo atrás, era descartado. A partir do desenvolvimento de tecnologias nos ramos de processamento de informação, armazenamento de dados e algoritmos computacionais, podemos agora trabalhar em cima desses dados, que antes eram descartados, e transformá-los em informação muito útil em diversos setores tanto comercial quanto social. Este estudo tem como objetivo demonstrar as motivações para o surgimento do Big Data, demonstrando uma visão geral do cenário atual no que tange ao volume, velocidade e variedade dos dados que são gerados hoje. Também expor o que dispomos no momento para trabalhar com essa quantidade enorme de dados, já que o grande desafio é manusear essa infinidade de dados e extrair informação relevante a partir dos mesmos, mostraremos algumas ferramentas e conceitos que se tornam imprescindíveis no trabalho com Big Data. Recentemente foi divulgado um programa de espionagem Norte Americana, que só é possível graças ao advento do Big Data, o governo Americano intercepta e analisa diariamente quantidades absurdas de dados, que são gerados continuamente, vindos de diversas fontes, o nosso estudo tenta deixar um pouco mais claro como funciona esse processo que está sendo praticado por muitos atualmente. Para realização do estudo foram consultados como referência, livros voltados ao assunto Big Data, artigos impressos e online, pesquisas com empresas de todo o mundo, revista e Internet.

---

## Introdução (máximo 500 palavras)

---

Mídias Sociais, sensores em celulares, sensores em automóveis, câmeras de filmagens, chamadas de *call centers* - nunca se produziu tantos dados como atualmente. A possibilidade de extrair *insights* a partir dessa imensidão de fontes faz com que o Big Data, aliado à computação em nuvem (*cloud computing*), seja um dos assuntos mais debatidos hoje na área da Tecnologia da Informação (DAVENPORT, 2006; SCHMARZO, 2013). Muitos dados não estruturados e complexos, anteriormente descartados, tornam-se úteis no processo decisório de negócios. Os dados são correlacionados a fim de buscar padrões que não seriam evidentes ao serem analisados com sistemas de banco de dados ou ferramentas de software tradicionais. Este novo processo, dá-se o nome Big Data Analytics (SIMON, 2013). *Big Data* são ativos de informação de grande volume, velocidade e variedade que exigem formas custo-efetivas e inovadoras de processamento de informações para uma melhor percepção e tomada de decisões (GARTNER, 2013). O grande volume de dados que caracteriza o Big Data advém da quantidade de transações, eventos e história que as novas tecnologias nos possibilitam capturar e armazenar. Porém, Gandomi e Haider (2015) salientam a dificuldade de se estabelecer um parâmetro de qual o tamanho que os dados devem ter para serem considerados “Big”. Primeiro, porque o que é considerado grande hoje, no futuro com o avanço tecnológico, tende a ficar pequeno. Segundo, porque determinado tamanho de um conjunto de dados considerado grande para certos setores é consideravelmente pequeno para outros. Este estudo tem como objetivo demonstrar as motivações para o surgimento do Big Data e expor o que dispomos no momento para trabalhar com essa quantidade enorme de dados.

## Metodologia Utilizada (máximo 300 palavras)

---

O objetivo desse estudo é identificar e representar os elementos gerais do processo de tomada de decisão no contexto da big data. Trata-se de um estudo de caráter exploratório e natureza teórica e dedutiva, pois pretende investigar de forma preliminar o fenômeno da big data associada a processos decisórios, tendo em vista a gestão da informação. Triviños (1987) descreve o estudo de tipo exploratório como aquele que permite ao investigador aumentar sua experiência em torno de determinado problema. Parte-se de uma hipótese e aprofunda seu estudo nos limites de uma realidade específica, buscando antecedentes, maior conhecimentos para, em seguida, planejar uma pesquisa descritiva ou de tipo experimental. Sampieri, Collado e Lucio (1994), justificam o uso do estudo exploratório quando o objetivo é examinar um tema que tenha sido pouco ou nada estudado anteriormente, permitindo que se obtenha um maior grau e familiaridade com os fenômenos envolvidos. No caso desse estudo, tendo em vista a escassez de produção relacionada aos termos big data e decisão no campo da ciência da informação, caracteriza-se como pesquisa básica, uma vez que algumas investigações são fundamentais para sustentar futuras aplicações. Segundo Schwartzman (1979), a pesquisa básica acumula informações e conhecimentos que podem eventualmente levar a resultados acadêmicos ou aplicados importantes, mas sem fazê-lo diretamente. No que se refere a fonte de dados, esse estudo é bibliográfico com seleção intencional de referencial teórico baseado no conhecimento e análise geral da literatura. Para Oliveira (2007) a pesquisa bibliográfica é uma modalidade de estudo e análise de documentos de domínio científico tais como livros, periódicos, enciclopédias, ensaios críticos, dicionários e artigos científicos. Como característica diferenciadora ela pontua que é um tipo de estudo direto em fontes científicas, sem precisar recorrer diretamente aos fatos/fenômenos da realidade empírica.

## Atividades Realizadas (máximo 300 palavras)

---

<b>Período</b>	<b>Atividades realizadas</b>
Abril 2019 a julho 2019	Realização de pesquisa bibliográfica
Julho 2019 a novembro 2019	Coleta de dados sobre big data e MapReduce
Novembro 2019 a janeiro 2020	Análise dos dados e possibilidades de aplicação
Janeiro 2020 a fevereiro 2020	Escrita de resultados e discussão
Agosto 2019 até os dias atuais	Divulgação dos resultados em congressos e periódicos indexados

## Resultados Obtidos

---

Resultou na identificação dos principais elementos envolvidos no processo decisório no ambiente da big data e na sua representação sistemográfica.

O estudo objetivou identificar e representar os elementos gerais de um contexto de tomada de decisão que envolva big data. Verificou-se que a literatura sobre o tema é recente e escassa no Brasil no campo da ciência da informação há uma identificação de sua relevância, entretanto, a área carece de estudos sobre o assunto.

Identificou-se que entre os principais problemas relacionados à big data estão a questão do tamanho, desestrutura e variedade de fontes de dados, bem como as formas comumente utilizadas para tomada de decisão baseadas na intuição ou em modelos quantitativos e computacionais. Foi apresentado aspectos básicos dos conceitos de Business Intelligence e Big Data Analytics que são ferramentas que tem emergido como uma área de sistemas de apoio à decisão. Analisou-se a sua importância, tendo em vista a racionalidade limitada do tomador de decisão. Resgatou alguns conceitos fundamentais da estruturação do contexto para o processo de decisão, particularmente a necessidade de consideração dos elementos objetivos e subjetivos; a importância do decisor e seus objetivos e interesses; e o papel da sistemografia na modelagem preliminar do contexto decisional. Foi apresentado um modelo geral de representação sistemográfica que evidencia, sintetiza e explicita os elementos anteriores da discussão, bem como sua utilidade e possibilidades de desenvolvimento. Futuros estudos deverão ampliar e detalhar a partir de problemas reais, 'dados e decisão' e seus respectivos inputs, bem como selecionar ferramentas computacionais para simulação, conforme o modelo Mcgrath (1982), para prever o comportamento particular do sistema.

## Sugestões para desdobramentos da pesquisa: (opcional)

---

A pesquisa deverá ser continuada em função de realizar diagnóstico em empresas de Criciúma e região. Futuros estudos deverão ampliar e detalhar a partir de problemas reais, 'dados e decisão' e seus respectivos inputs, bem como selecionar ferramentas computacionais para simulação, conforme o modelo Mcgrath (1982), para prever o comportamento particular do sistema.

## Referências

---

- Ansarinejad, A., Amalnick, M.-S., Ghadamyari, M., Ansarinejad, S., & Hatami-Shirkouhi, L. (2011). **Evaluating the critical success factors in ERP implementation using fuzzy AHP approach.** *International Journal of Academic Research*, 3(1), 65-80.
- Bingi, P., Sharma, M. K., & Godla, J. K. (1999). **Critical issues affecting an ERP implementation.** *Information Systems Management*, 16(3), 7-14.
- Bruce, W., Lenita, D., & Paul, D. B. (2013). **Perspectives on big data.** *Journal of Marketing Analytics*, 1(4), 187-201.
- Burton, R. M., Mastrangelo, D., & Salvador, F. (2014). **Introduction.** *Journal of Organizational Design*, 3(1), 1.
- Castro, S. (2014). **Optimizing your data management for big data.** *Journal of Direct, Data and Digital Marketing Practice*, 16(1), 15-18.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). **Business intelligence and analytics: From big data to big impact.** *Mis Quarterly*, 36(4), 1-22.
- Davenport, T. (2014). **Big data at work: Dispelling the myths, uncovering the opportunities.** Boston: Harvard Business Press.
- Davenport, T., Barth, P., & Bean, R. (2012). **How 'Big data' is different.** *MIT Sloan Management Review*, 54(1), 43-46.
- George, G., Hass, M., & Pentland, A. S. (2014). **Big data and management.** *Academy of Management Journal*, 57(2), 321-326.
- Gupta, R., Gupta, S., & Singhal, A. (2014). **Big data: An overview.** *International Journal of Computer Trends and Technology*, 9(5), 1-3.
- International Data Corporation. (2014). *Predictions Brazil 2014: Top trends presentation.* Link: <http://www.brasscom.org.br/brasscom/Portugues/download.php?cod=567>
- Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). **Big data applications in the government sector.** *Communications of the ACM*, 57(3), 78-85.
- Leeflang, P. S. H., Verhoef, P. C., Dahlström, P., & Freundt, T. (2014). **Challenges and solutions for marketing in a digital era.** *European Management Journal*, 32(1), 1-12.
- Maçada, A. C. G., & Canary, V. P. (2014). **A tomada de decisão no contexto do big data: Estudo de caso único.** *Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Rio de Janeiro, RJ, Brasil, 38.
- McAfee, A., & Brynjolfsson, E. (2012). **Big data: The management revolution.** *Harvard Business Review*, 90(10), 60-68.
- Mcgrath, Joseph E. **Dilemmatics: the study of research choices and dilemmas.** In: McGrath, J.E.; Martin, J.; Hulka, R. A. (Ed.). *Judgement calls in research.* Beverly Hills: Sage, 1982. p. 69-102
- Minelli, M., Chambers, M., & Dhiraj, A. (2013). **Big data big analytics: Emerging business intelligence and analytic trends for today's businesses.** Hoboken: John Wiley & Sons.
- Novo, R., & Neves, J. M. S. D. (2013). **Inovação na inteligência analítica por meio do Big data: característica de diferenciação da abordagem tradicional.** *Workshop de Pós-Graduação e Pesquisa do Centro Paula Souza.* São Paulo, SP, Brasil, 8. p. 32-44.

- Ohlhorst, F. (2013). **Big data analytics: Turning big data into big money**. Hoboken, N.J.: John Wiley & Sons.
- Oliveira, Maria Marly de. **Como fazer pesquisa qualitativa**. Petropolis: Vozes, 2007.
- Sampieri, Roberto Hernández; COLLADO, Carlos Fernández; LUCIO, Pilar Baptista. **Metodología de la investigación**. México: McGraw Hill, 1994.
- Schmarzo, B. (2013). **Big data: Understanding how data powers big business**. Indianapolis, IN: John Wiley & Sons.
- Silva, I. M. D., & Campos, F. C. D. (2014). **New perspectives using big data: A study of bibliometric 2000-2012**. *Proceedings of the International Conference on Information Systems and Technology Management - CONTECSI*, São Paulo, Brasil, 11.
- Simon, P. (2013). **Too big too ignore**. Hoboken: John Wiley & Sons.
- Somers, T. M., Nelson, K., & Ragowsky, A. (2001). **Enterprise resource planning (ERP) for the next milenium: development of an integrative framework and implications for research**. *Proceedings of the Americas Conference on Information Systems - AMCIS*, New York, EUA.
- Sumner, M. (1999). **Critical success factors in enterprise wide information management systems projects**. *Proceedings of the Americas Conference on Information Systems - AMCIS*, New York, EUA.
- TRIVIÑOS, Augusto Nibaldo Silva. **Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação**. São Paulo: Atlas, 1987.
- Wigan, M. R., & Clarke, R. (2013). **Big data's big unintended consequences**. *Computer*, 46(6), 46-53.
- Yeoh, W., & Koronius, A. (2010). **Critical success factors for business intelligence systems**. *Journal of Computer Information Systems*, 50(3), 23-32.