

UMA REVISÃO SISTEMÁTICA SOBRE BIG DATA

Emanoel R. Parente da Cruz^{1*} (IC), Elisabete T. Kowata²

emanuelparente2015@gmail.com

^{1,2}Câmpus Anápolis de Ciências Socioeconômicas e Humanas

Resumo: O Big Data é um fenômeno que acontece atualmente na era digital ao qual estamos inseridos. De modo geral significa uma grande quantidade de dados estruturados e não estruturados sendo gerados a todo momento e o desafio das empresas é capturar esses dados, processá-los e gerar valor agregado, ou seja, transformar dados brutos em oportunidades de inovação e melhorias em produtos e serviços. O objetivo desta pesquisa é realizar uma revisão sistemática sobre o tema Big Data voltada para aplicações às pequenas e médias empresas. Esta revisão sistemática usa como instrumentos de pesquisa a string de busca submetida no Google acadêmico e demais base de dados como Scielo e periódicos da Capes, seleção de material encontrado, e análise. Os resultados encontrados foram a identificação de boa parte dos estudos serem provenientes de artigos e monografias, o que mostra que o Big Data ainda se mantém de forma significativa no campo da teoria, sendo os estudos e casos práticos concentrados na área da saúde, engenharia e tecnologia, o que mostra uma necessidade maior no campo das ciências humanas e sociais, ao qual está inserido a administração de empresas, o foco deste trabalho são as empresas.

Palavras-chave: Pequenas e médias empresas. Análise de dados. String de busca. Revisão de literatura.

Introdução

Dados são gerados a todo o momento por todo e qualquer dispositivo conectado à internet, eles podem ser explícitos. Os dados explícitos e representado por postagens em redes sociais, de áudio, texto, vídeo e-mails. Já os implícitos é trilha digital que usuários deixam na internet, como atividades em sites de compras, tempo de permanência em determinada página, número de cliques, etc (ANSELMO, 2015).

É o exemplo dos automóveis que a cada segundo coleta dados de consumo, como frequência e intensidade de frenagem, giros no volante, abastecimento, reparos e manutenção, e entre outros dados. Esses dados são enviados ao fabricante para que possam por meio deles identificar falhas que podem vir a ocorrer, identificar o perfil consumidor para que possam aperfeiçoar os produtos e serviços a fim de atender melhor às necessidades do cliente (DAVENPORT, 2014).

Agora imagine esses dados coletados de um único carro, multiplique-o por milhares e milhares de carros transitando pelas ruas no mundo, sendo produzidos a cada segundo. É uma quantidade gigantesca e exponencial de dados (MAYER-SCHÖNBERGER e CUKIER, 2013).

Esse é o Big Data, tendo ele uma definição ampla e dividida por vários autores, mas com a mesma finalidade, entender e explorar o potencial que os dados têm para criar soluções no setor público e privado em uma sociedade cada vez mais conectada a internet.

Para Davenport (2014) Big Data se trata de uma grande e variada quantidade de dados que não cabem em bancos de dados usuais, e não estruturados demais para se adequar a um banco de dados relacional organizado em linhas e colunas. Porém para Diamandis (2016) se antes havia uma limitação no processamento de dados por conta do pouco espaço de armazenamento, hoje não existe mais, pelo contrário, a capacidade de armazenamento e processamento tem aumentado com o passar do tempo, logo o que antes era visto como um problema, hoje é visto como uma oportunidade de entender o mercado consumidor e entregar valor por meio dos produtos e serviços oferecidos. Isso facilita para que as empresas captem a quantidade mais alta que seja de dados e os processem a fim de extrair informações de grande valor para a tomada de decisões.

O Big Data tem suas características e peculiaridades, é composto por 5 Vs, chamados de volume, velocidade, variedade, veracidade e valor (DAVENPORT, 2014). A parte de volume descreve a quantidade de dados que são gerados todos os dias. Para entendermos melhor veja o Quadro 1:

Quadro 1 – Unidade de medida dos dados gerados

Unidade	Capacidade	Unidade	Capacidade
1 byte (B)	8 bits	1 Petabyte (PT)	1024 TB
1 Kilobyte (kB)	1024 bytes	1 Exabyte (EB)	1024 PT
1 Megabyte (MB)	1024 KB	1 Zettabyte (Zbytes)	1024 (EB)
1 Gigabyte (GB)	1024 MB	1 Yottabytes (Ybytes)	1024 (Zbytes)
1 Terabyte (TB)	1024 GB		

Fonte: Martins (2009).

Estima-se que diariamente sejam produzidos dados na faixa dos petabytes (MAYER-SCHONBERGER e CUKIER, 2013).

Quadro 2 – Estimativa de dados gerados por dia e por operação

Quantidade	Onde	Fonte
864.000	Horas de vídeo são enviados ao YouTube pelos usuários	Schönberger-Mayer (2013)
3 bilhões	De interações são realizadas no Facebook somando as curtidas e comentários dos usuários	
15 petabytes	De dados estruturados e não estruturados são gerados no mundo, segundo projeções da IBM	Oliveira (2013)
24 petabytes	São processados pelo Google	Tecmundo (2011)
1 exabyte	Circula pela Internet mundial	
14 exabytes	São gerados pelo maior acelerador de partículas do mundo, mas apenas 1% é analisado e processado, para posteriormente ser armazenado.	CERN (2015)

Fonte: Anselmo (2015).

Esses dados são gerados a uma velocidade muito grande, fala-se do “V” de velocidade, referente a velocidade em que os dados são gerados e a velocidade com que os mesmos são tratados, com o intuito de obter informação em tempo hábil para tomada de decisão.

Quando se fala em big data, os termos mais mencionados são relativos às tecnologias utilizadas para armazenamento e processamento, porém, já se fala em *stream computing*, ou seja, o processamento em tempo real de um fluxo contínuo de dados (BRETERNITZ e SILVA, 2013), é o caso da Netflix e Spotify.

A variedade é composta pelas diversas fontes de dados, por cada dispositivo conectado a internet. Apesar de existirem muitos dados disponíveis, é necessário que se questione a origem dos dados. Muitos sites de vendas geram cliques falsos, seja por falha no sistema ou por ataques cibernéticos (MIRSHAWKA, 2016). Mas do que adianta ter um volume considerável de dados, processá-los em tempo hábil, se certificar da origem dos dados e garantir sua veracidade, se não há relevância para o problema que se deseja resolver?

É preciso fazer perguntas coerentes aos dados que se possui, caso contrário, uma pergunta errada pode resultar em interpretações completamente errôneas e causar prejuízos para a empresa. Em um momento onde há fartura de dados, há

também fatura de correlações inúteis, apesar de a ideia principal ser a correlação com base de dados heterogêneas é preciso verificar se as bases de dados irão gerar resultados coerentes (HENRIQUES e COSTA, 2015). O valor dos dados se trata disso, qual o valor que esta base de dados tem para o problema que eu quero resolver? Que relação há?. Há autores que apontam um sexto “V”, a validade dos dados. Por quanto tempo é necessário guardar os dados? Para Mayer-Schönberger e Cukier (2013), os dados têm valor latente, isso significa que é possível usá-los várias vezes, e de maneiras diferentes para extrair valor. “Independentemente do tamanho da empresa, o volume, a velocidade a veracidade e a variedade dos dados são insumos para um novo mundo de oportunidades.” (HENRIQUES e COSTA, 2015, p. 2).

Em se tratando de oportunidades que a utilização do Big Data pode oferecer, os cientistas de dados podem peneirar densos volumes de dados utilizando técnicas avançadas de correlação, obtendo assim informações capazes de revelar tendências e probabilidades mais exatas do que pode vir a ocorrer, evitando acontecimentos indesejados ou adiantando metas almejadas (NESELLO e FACHINELLI, 2014).

Material e Métodos

Para a realização da revisão sistemática neste trabalho foi consultado o Google Acadêmico e Scielo utilizando a *string* de busca para a seleção de documentos a serem analisados, apresentados no quadro abaixo:

Quadro 3 - String de busca usada

String de busca	Retorno
("Big Data" or "Massa de dados" or "muitos dados" or "grande volume de dados") and (organização or instituição or firma or companhia or estabelecimento or corporação or negócio) and (tratamento or processamento or análise) and ("como são tratados")	0 resultados
("Big Data") and (organização or instituição or firma or companhia or estabelecimento or corporação or negócio) and (tratamento or processamento or análise) and ("como são tratados")	0 resultados
("Big Data") and (organização or empresa) and (processamento or análise) and ("como são tratados")	1 resultado
("Big Data") and (empresa or instituição) and (tratamento or análise) and (uso or aplicação or resultado)	1060 resultados

Fonte: Autor.

Posteriormente foi selecionado documentos, sendo eles: artigos, monografias, trabalhos apresentados em congressos, do ano de 2013 até o presente momento, excluindo aqueles que não mencionasse o uso de Big Data em qualquer que fosse a área, ou que não contivesse o termo Big Data. Foi consultado inclusive obras de autores que descrevem os conceitos de Big Data, bem como seus desafios e oportunidades. Após a leitura e análise de documentos foi realizada uma categorização dos documentos selecionados, que foram preenchidos no formulário criado no recurso disponibilizado pelo Google e que foram explicitados em gráficos para melhor compreensão.

Resultados e Discussão

Após a leitura e análise dos documentos identificou-se que, dentre os artigos que foram publicados de 2013 a 2017 o maior número de publicação se concentrava no ano de 2015. A partir da análise de documentos foram produzidos os seguintes gráficos: no primeiro gráfico mostra o tipo de texto e identificou-se que trabalhos de conclusão de cursos e reflexão teórica tiveram o maior número de publicação, ficando à frente de artigos e trabalhos de revisão de literatura.

Gráfico 1 – Tipos de publicações sobre Big Data

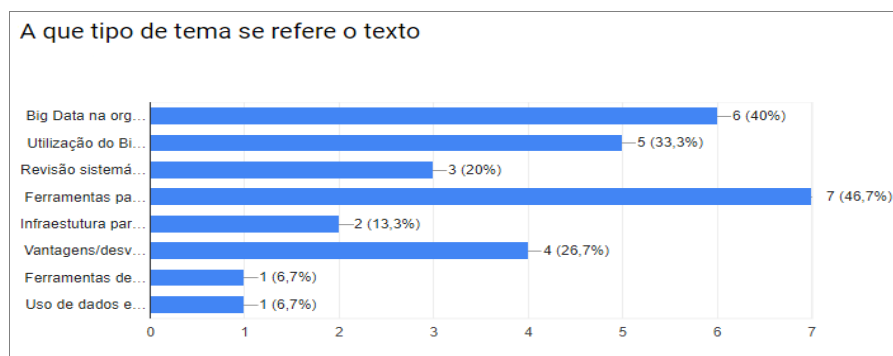


Fonte. Própria dos autores.

Quando feita a pergunta, “a que tipo de tema o texto se refere”, há um resultado quase unânime entre “Big Data na organização” e “ferramentas de Big Data”, isso porque Big Data representa o conjunto de dados não estruturados, por isso faz-se necessário ferramentas específicas para o tratamento dos dados, que

não poderia ser armazenados em bancos de dados convencionais como MySQL, Oracle, PostgreSQL etc.

Gráfico 2 – Temas abordados nos textos

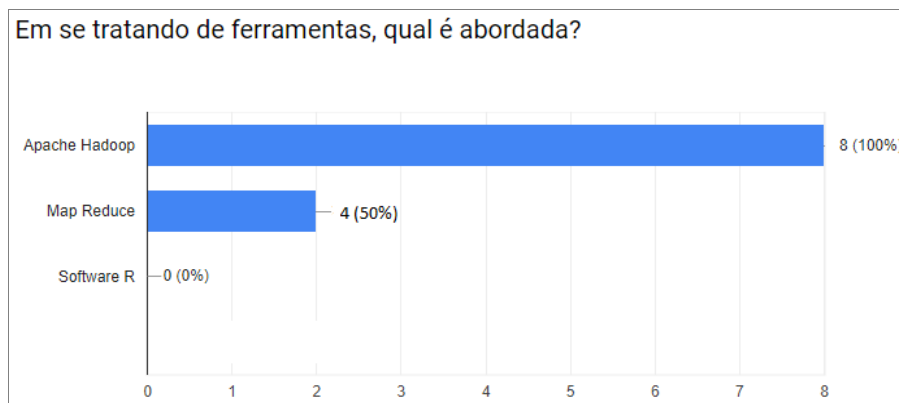


Fonte. Própria dos autores.

Das ferramentas de uso para Big Data mencionadas, obteve maior pontuação o uso da ferramenta Hadoop, por ser gratuita e amplamente usada, os demais *softwares* também são gratuitos porém menos populares. Segundo Davenport (2014, p. 113) o Hadoop é

Um conjunto de ferramentas de software de código aberto e um framework para distribuir os dados em vários computadores; trata-se de um ambiente unificado de armazenamento e processamento altamente escalável para grandes e complexos volumes de dados. O Hadoop as vezes é chamado de Apache Hadoop, porque sua versão mais comum é um projeto da The Apache Software Foundation. No entanto, como costuma acontecer com projetos de código aberto, muitos fornecedores comerciais também criaram suas próprias versões do Apache Hadoop. Em consequência temos o Cloudera Hadoop, Hortonworks Hadoop, EMC Hadoop, Intel Hadoop, Microsoft Hadoop, entre vários outros (DAVENPORT, 2014, p. 113).

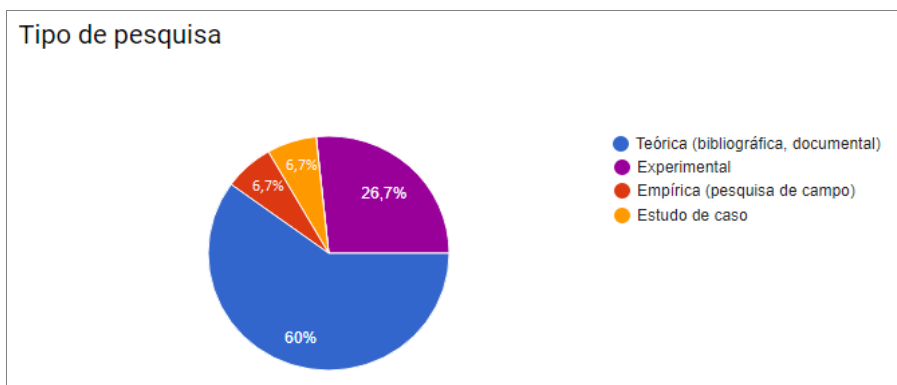
Gráfico 3 – Ferramentas abordadas nos textos



Fonte. Própria dos autores.

Em relação ao tipo de pesquisa feita nos trabalhos, em sua maioria predominou a pesquisa bibliográfica, seguida por experimental, empírica, e estudo de caso.

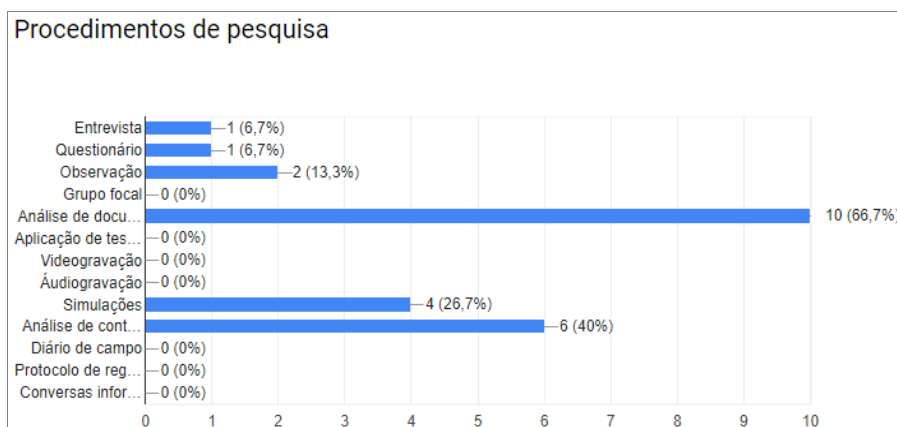
Gráfico 4 – Tipos de pesquisas realizadas pelos pesquisadores



Fonte. Própria dos autores.

Os instrumentos de pesquisa utilizados, resultou em 66,7% de análise de documentos e 40% em análise de conteúdo para pesquisas empíricas e estudos de caso, o restante se manteve concentrado em entrevistas, questionários e simulações.

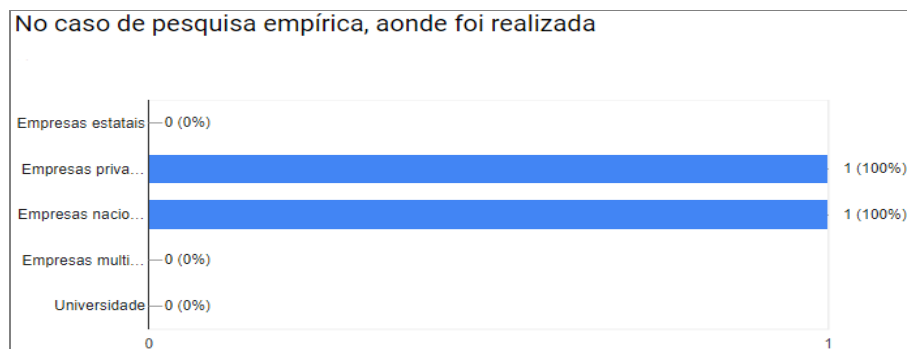
Gráfico 5 – Procedimentos de pesquisa adotados pelos pesquisadores



Fonte. Própria dos autores.

Houve apenas um caso de pesquisa empírica que foi realizado em empresa privada nacional e no setor administrativo.

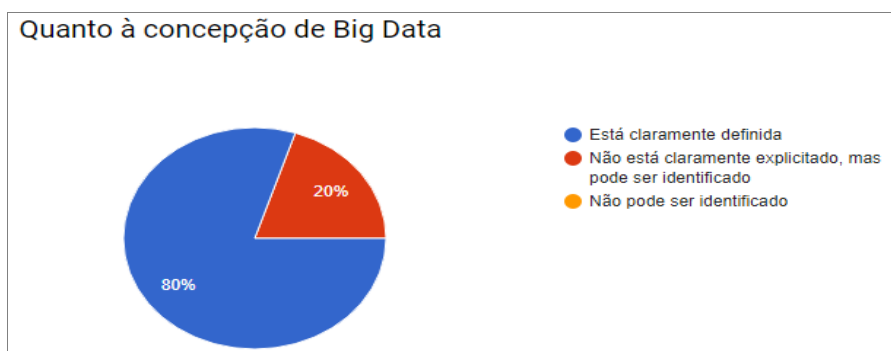
Gráfico 6 – Locais onde foram realizadas as pesquisas empíricas



Fonte. Própria dos autores.

Em relação aos textos lidos, a concepção de Big Data se manteve clara em 80% dos documentos analisados e em 20% não estavam claros mas puderam ser identificados.

Gráfico 7 – Concepção de Big Data apresentado nos textos



Fonte. Própria dos autores.

Considerações Finais

Mediante a análise de documentos, conclui-se que, o Big Data é um tema emergente nas organizações e universidades, sendo mais amplamente estudado em universidades, sua aplicação se restringe às empresas de grande porte e multinacionais, é pouco conhecido e utilizado em empresas de pequeno e médio porte.

A proposta da revisão sistemática foi alcançada, o conhecimento acerca do tema foi adquirido, para prosseguimento no projeto de pesquisa "Big Data,

transformando grande quantidade de dados em valor agregado para as empresas”, sendo o próximo passo desta, a criação de um portfólio voltado para as pequenas e médias empresas.

Os resultados apontaram que ainda é necessário mais pesquisas empíricas e estudo de caso, principalmente nas ciências humanas e sociais aplicadas, o uso de análise de Big Data na saúde e setores da engenharia e tecnologia já é evidente em alguns resultados encontrados nos artigos analisados.

Em uma sociedade cada vez mais conectada a internet, e produzindo dados a todo o momento, faz-se necessário a criação de mecanismos de aproveitamento desses dados a fim de gerar melhorias no setor público, e redução de custos e maximização de lucros no setor privado, bem como o engajamento de instituições de ensino na pesquisa e aplicação de conhecimentos acerca do tema nos diversos campos do conhecimento.

Agradecimentos

À Universidade Estadual de Goiás por fazer parte da Bolsa de Iniciação Científica Voluntário (VIC/UEG).

Referências

ANSELMO, Felipe. Big Data: Uma análise conceitual e funcional. 2015. 87 f. TCC (Graduação) - Curso de Sistemas de Informação, Universidade do Planalto Catarinense, Lages, 2015. Cap. 3. Disponível em: <http://revista.uniplac.net/ojs/index.php/tc_si/article/view/1933>. Acesso em: 10 jun. 2017.

BRETERNITZ, Vivaldo José; SILVA, Leandro Augusto. Big Data: Um novo conceito gerando oportunidades e desafios. *Revista Eletrônica de Tecnologia e Cultura*, Jundiaí, v. 13, n. 0, p.106-113, out. 2013. Semestral. Disponível em: <<http://201.55.32.167/retc/index.php/RETC/article/view/74/pdf>>. Acesso em: 29 maio 2017.

DAVENPORT, Thomas H. *Big Data no trabalho*: Derrubando mitos e descobrindo oportunidades. Rio de Janeiro: Elsevier, 2014. 221 p.

DIAMANDIS, Peter H. *Oportunidades exponenciais*: Um manual prático para transformar os maiores problemas do mundo nas maiores oportunidades de negócio e causar um impacto positivo na vida de bilhões. São Paulo: HSM, 2016. 320 p.

HENRIQUES, Daniela Aparecida; COSTA, Helder Rodrigues da. Big Data: Como utilizar a extraordinária quantidade de informação coletada por novas tecnologias para obter vantagens competitivas. *Revista Pensar Tecnologia*, Belo Horizonte, v. 1, n. 1, p.1-11, jan. 2014. Semestral. Disponível em: <<http://revistapensar.com.br/tecnologia/artigo/no=a72.pdf>>. Acesso em: 24 jul. 2017.

MIRSHAWKA, Victor. *Economia criativa: Fonte de novos empregos*. São Paulo: Dvs Editora, 2016. 408 p.

MAYER-SCHÖNBERGER, Viktor. CUKIER, Kenneth. *BIG DATA: Como extrair volume, variedade, velocidade, e valor da avalanche de informação cotidiana*. Tradução de Paulo Polzonoff Junior. 1. ed., Rio de Janeiro: Elsevier, 2013.

NESELLO, Priscila; FACHINELLI, Ana Cristina. Big Data: O novo desafio da gestão. *Revista Inteligência Competitiva*, São Paulo, v. 4, n. 1, p.1-21, fev. 2014. Bimestral. Disponível em: <<http://www.inteligenciacompetitivarev.com.br/ojs/index.php/rev/article/view/76>>. Acesso em: 27 maio 2017.