

# UM PANORAMA DE 20 ANOS DO BIG DATA:

Definição, conceitos e aplicações na engenharia.

Paulo Egreja Sartorelli  
Centro de Ciências e Tecnologia  
Universidade Presbiteriana Mackenzie  
Campinas - SP, Brasil  
[paulo.sartorelli1@live.com](mailto:paulo.sartorelli1@live.com)

Mariana Z. T. de Lima  
Centro de Ciências e Tecnologia  
Universidade Presbiteriana Mackenzie  
Campinas - SP, Brasil  
[mariana.lima@mackenzie.br](mailto:mariana.lima@mackenzie.br)

Massaki de Oliveira Igarashi  
Centro de Ciências e Tecnologia  
Universidade Presbiteriana Mackenzie  
Campinas - SP, Brasil  
[massaki.igarashi@mackenzie.br](mailto:massaki.igarashi@mackenzie.br)

**Resumo** - Nos últimos anos, houve um crescimento exponencial na geração de dados de diferentes fontes, principalmente devido aos avanços na tecnologia da informação e comunicação. Assim, as organizações têm potencial para obter vantagens competitivas com as análises desses dados, alterando-as em informações que, sem as ferramentas de Big Data, não poderiam ser obtidas. Nesse contexto, este trabalho traz uma pesquisa sobre Big Data e explica sobre esse conceito que mudou ao longo dos anos. Este artigo tem como objetivo desenvolver um panorama da pesquisa nos últimos vinte anos de Big Data e aplicações em diferentes áreas da engenharia, como: civil, elétrica, mecânica, manufatura, materiais, química e engenharia de software.

**Palavras-chave** - Big Data, pesquisa, vinte anos, áreas de engenharia,

**Abstract**-In recent years there was an exponential growth in data generation from different sources especially due to advances in information and communication technology. Thus, organizations have seen potential to gain competitive edges from the analyses of this data, changing it in information that, without Big Data tools, cannot be obtained. In this context, this work brings a survey about Big Data and explains a about this concept that have changed during the years. This paper aims to develop a survey of the last twenty years of Big Data and applications in different areas of engineering, such as: civil, electrical, mechanical, manufacturing, materials, chemical and software engineering.

**Keywords** – Big Data, survey, twenty years, engineering areas,

## I. INTRODUÇÃO

Nos últimos 20 anos, houve um aumento expressivo na utilização da internet pela população mundial, além do massivo uso das tecnologias da informação, o que fez com que a quantidade de dados e a velocidade com a qual estes dados são gerados passasse a exigir novas ferramentas e tecnologias.

Cada vez mais, as pessoas, as empresas e as organizações estão conectadas e compartilhando informações pela internet através de diversos dispositivos (celular, computador, tablet,

etc.) o que tem contribuído para este aumento da geração de dados; este pode ser ainda mais expressivo ao se considerar a internet das coisas e os dados analíticos que ainda serão gerados [1].

A maior parte das ações realizadas na internet atualmente gera dados e tem-se visto que a partir da análise destes dados é possível identificar as preferências dos clientes, padrões de comportamentos, avaliar tendências e, até mesmo, detectar possíveis crises e fraudes [2].

A importância da geração de dados e suas aplicações aumenta ao se considerar que as estimativas de crescimento do volume de dados gerados para 2020 era de 40 Zettabytes em 2012 [2]. Este número certamente foi alcançado porque cada vez mais pessoas passam cada vez mais tempo navegando e disponibilizando dados na internet [3].

Dentre as possibilidades para se extrair valor desta grande quantidade de dados (estruturados ou não); uma das formas destacadas é identificar padrões existentes nas bases de dados através das informações utilizadas com maior frequência. Outra forma é que as empresas podem criar e armazenar dados e através deles obter informações detalhadas sobre diversas áreas, como previsões de estoque, perspectivas de demanda nos próximos meses e com isto utilizar estas informações para tomar melhores decisões e melhorar o desempenho da organização [4].

O termo *Big Data* é utilizado para definir um conjunto de dados tão grande e complexo cujas técnicas tradicionais de banco de dados, ferramentas e *softwares* já não são mais eficientes. Portanto, o escalonamento destes dados, a diversidade e a complexidade exigem novas técnicas, arquiteturas e algoritmos para seu gerenciamento e análise; permitindo, portanto, a tão desejada extração de valor e conhecimento.

As ferramentas e técnicas *Big Data* auxiliam a extração de valor e informações úteis para melhores tomadas de decisões nas mais distintas áreas e possibilidades (Tabela 1).

Com base nestas diversas possibilidades, realizar uma pesquisa de revisão bibliográfica para expor um panorama das diversas possibilidades *Big Data* em diferentes áreas da engenharia pode contribuir para estimular pesquisas futuras na engenharia num mundo movido a dados.

Tabela 1 - Aplicações do Big Data em diferentes áreas

ÁREA	REFERÊNCIA	DESCRIÇÃO
AGRICULTURA DE PRECISÃO	[5]	O papel das tecnologias da informação e da comunicação na agricultura
	[6]	Desafios e oportunidades para a silvicultura de precisão
BANCOS/CRÉDITO E SEGURADORAS	[7]	O impacto do Big Data no setor bancário
SEGUROS	[8]	Algoritmo de análise de Big Data para seguros
EDUCAÇÃO	[9]	O conceito de Big Data na educação
JUDICIAL	[10]	Algumas regras de Big Data da justiça criminal
JOGOS E REDES SOCIAIS	[11]	Jogos e Big Data: um modelo escalável de previsão de rotatividade multidimensional
GOVERNO E GESTÃO PÚBLICA	[12]	Aplicações de Big Data no setor governamental
	[13]	Uma proposta de ecossistema de Big Data para a análise de dados abertos governamentais conectados
SETOR FINANCEIRO	[14]	O papel dos dados no setor financeiro
LOGÍSTICA E SUPPLY CHAIN	[15]	Análise de Big Data em logística e gerenciamento da cadeia de suprimentos: algumas investigações para pesquisas e aplicações
	[16]	Data Science, análise preditiva e Big Data: uma revolução que transformará o design e o gerenciamento da cadeia de suprimentos
MARKETING	[17]	Análise do consumidor de Big Data e a transformação do marketing
	[18]	Efeitos da análise do Big Data e análise de marketing tradicional no sucesso de novos produtos
MEDICINA DE PRECISÃO	[19]	Conciliando medicina baseada em evidências e medicina de precisão na era do Big Data: desafios e oportunidades
	[20]	Métodos integrativos para análise de Big Data em medicina de precisão
SAÚDE	[21]	A aplicação inevitável do Big Data aos cuidados de saúde
	[22]	Análise de Big Data na área da saúde: promessa e potencial
TOMADA DE DECISÃO	[23]	Data Science e sua relação com Big Data e tomada de decisão orientada por dados
	[24]	Fatores que influenciam a qualidade da tomada de decisão de big data
TRANSPORTE	[25]	Big Data para Transporte Social
	[26]	Previsão de fluxo de tráfego com big data

## II. REFERENCIAL TEÓRICO

De acordo com Hsinchun Chen [27] a utilização dos dados como informação não é uma novidade, já que desde a década de 50 áreas como administração e tecnologia da informação já utilizavam conceitos e sistemas de informação nas empresas.

As primeiras pesquisas relacionadas à grande quantidade de dados começaram a ser estudadas na década de 70; dando origem aos métodos de processamento de dados. A partir dos anos 90 surgiram os *softwares* que eram utilizados para tratar a grande quantidade de dados existentes na época [28], algo muito aquém do atual; quando se quer se cogitava o termo *Big Data*.

Entretanto foi a partir dos anos 2000 que o termo *Big Data* começou a ganhar destaque, com o surgimento de novos *softwares*, novas técnicas de processamento, de armazenagem e transmissão de dados [28]. Ainda nesta mesma época, Hsinchun Chen [27], Thomas Davenport [29] e seus colaboradores mencionam o surgimento de novas tecnologias e a disseminação da *internet* e do comércio eletrônico global; como o início de uma nova era da geração de dados que são transmitidos através desta rede mundial. Os autores citam alguns exemplos de ações realizadas na *internet* que começaram a conduzir um fluxo constante e crescente de dados, como padrões de navegação na rede, quantidade de cliques, comércio *online*, conteúdos que os próprios usuários geram nas mídias sociais, *sites*, *blogs* e plataformas.

Ou seja, o termo *Big Data* não está relacionado à um conceito totalmente novo, mas à atualização de várias tecnologias, ferramentas e técnicas que já existiam; esta união proporciona um entendimento mais claro do *Big Data* atualmente [28].

### A. Os 5 V's do Big Data

Não há um consenso bem estabelecido sobre a definição do termo *Big Data*. Diferentes autores de diferentes países em livros, artigos e posts na *internet* definem, cada um à sua revelia, o termo *Big Data*. Fazendo referência à grande quantidade de dados que são gerados, diariamente, em todo o mundo, advindos de diferentes fontes.

Nas primeiras tratativas de definir o termo, um dos principais autores da época o classificava a partir de três pilares: volume, velocidade e variedade [30]. Porém, outros autores, ao definirem o *Big Data*, citam cinco pilares (os cinco V's do Big Data), são eles: o volume, velocidade, veracidade, variedade e valor.

- 1- Volume:** está relacionado com a grande quantidade de dados que são gerados diariamente em todo o mundo, advindos de fontes como: áudios, vídeos, conteúdos de redes sociais, imagens, entre outros [31].
- 2- Velocidade:** com a popularização do *Big Data*, a velocidade com que os dados são gerados e armazenados é cada vez maior, principalmente a forma rápida com que os concorrentes coletam e analisam dados obtidos dos consumidores, a agilidade e o dinamismo que são exigidos pelo mercado para a tomada de decisão [32].
- 3- Variedade:** assim como o volume, a diversidade dos dados é muito grande, devido às diversas fontes que existem atualmente. Existem dados estruturados, que são os utilizados pelas organizações, em arquivos sequenciais, bancos de dados e os dados não estruturados (a principal matéria prima do *Big Data*), que são os conteúdos advindos das diversas mídias sociais, vídeos, imagens, áudios, entre outros [33].
- 4- Veracidade:** devido ao grande volume e variedade dos dados, saber a sua autenticidade é imprescindível para as organizações [34]. Ela está relacionada com a qualidade das informações disponíveis para a tomada de decisão [31].
- 5- Valor:** é uma das dimensões mais importantes do *Big Data*, pois refere-se ao valor agregado obtido pelas tomadas de decisão baseadas em dados previamente analisados pelas organizações [31].

A partir das definições da bibliografia e para sistematizar este aumento expressivo da quantidade de letras V's nos pilares que definem o termo *Big Data* foi criada uma linha do tempo que relaciona os anos das publicações com as diferentes definições e a respectiva quantidade de letras V's (Fig. 1).

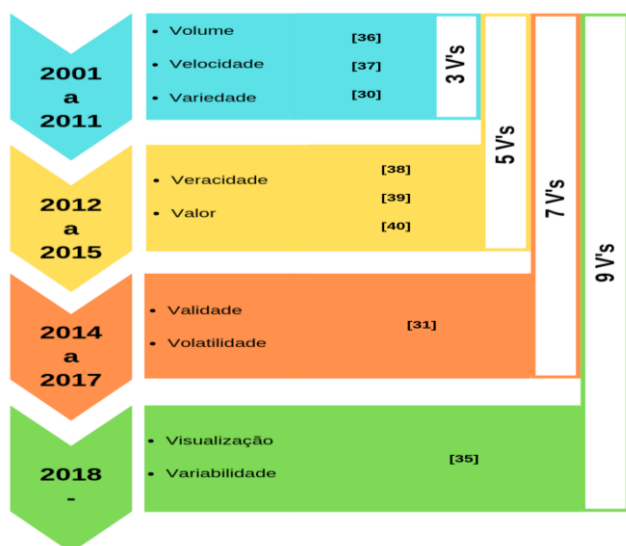


Fig 1 - Linha do tempo sobre a evolução do Big Data

Apesar da maioria dos autores citarem apenas os cinco pilares (volume, velocidade, variedade, veracidade e valor), existem outros autores que adicionam mais pilares. Por exemplo, M. Ali-ud-din Khan [31] adicionou outros dois pilares: a validade e a volatilidade. Não contentes, Monali Sahoo [35] e colaboradores destacam outros dois: a visualização e a variabilidade. Já existem, até mesmo, algumas tentativas atuais em blogs de definir o termo através de 10 V's, porém, sem reconhecimento acadêmico.

### III. METODOLOGIA

A presente pesquisa tem uma abordagem qualitativa (cujo objetivo é produzir novas informações aprofundadas e ilustrativas sobre o assunto *Big Data*), de natureza aplicada (já que objetiva gerar conhecimentos para aplicação prática, dirigidos à solução do problema de não existir uma coesão nas definições do termo *Big Data*) e cujo objetivo é ser uma pesquisa explicativa sobre o potencial de aplicação do *Big Data* nas diversas áreas da engenharia.

Foram utilizadas publicações acadêmicas da literatura nacional e internacional a partir das bases de acadêmicas como *Google Scholar*, *ScienceDirect*, *Elsevier Journals*, *ISI Web of Science*, entre outras. O método de revisão bibliográfica, em função da ampla quantidade de informações sobre o tema em questão indica diferentes caminhos que podem ser seguidos a fim de conseguir uma melhor eficiência na seleção das técnicas mais utilizadas, iniciando pela etapa de fundamentação teórica.

Além das aplicações, foram obtidos números de resultados de pesquisas que contribuem para um entendimento visual do panorama de 20 anos de pesquisas neste tema de estudo. A Busca foi realizada a partir das palavras chaves: “*Big Data*” + “área da engenharia”, obtendo números de publicações ano a ano, de 2000 até a presente data de 27 set. 2019.

### IV. ANÁLISE E DISCUSSÃO

A seguir são apresentados alguns estudos e aplicações dos *Big Data* em 7 áreas da engenharia (civil, elétrica, mecânica, produção, materiais, química e engenharia de software).

#### A. Engenharia Civil

Amir Gandomi [41] relata um novo sistema de detecção e rastreamento de veículos baseados nos dados de imagens coletados pelos veículos aéreos não tripulados (*Unmanned Aerial Vehicles* - UAV). Este sistema utiliza quadros capturados consecutivamente para gerar informações dinâmicas dos veículos, como as suas posições e velocidades. O sistema é baseado em quatro principais módulos: registro de imagens, extração de recursos de imagem, detecção da forma do veículo e o rastreamento do veículo. Também é realizada uma análise dos dados obtidos através da captura de imagens pelos UAVs e com isso, é possível customizar o fluxo de tráfego de veículos em determinado local.

Este sistema pode ser utilizado por “cidades inteligentes” em centrais de gerenciamento de tráfego, para a localização de acidentes, no monitoramento real de uma rodovia, entre outros. Além disso, através da utilização de novas tecnologias para a captura e análise dos dados, os UAVs estão atraindo mais atenção de pesquisadores e engenheiros para novas aplicações. De acordo com os autores seu método também pode ser aplicado na área de construção, aplicações de busca e salvamento, inspeção estrutural e monitoramento de saúde; e com algumas modificações também pode ser utilizado no rastreamento de animais na selva.

Já Liang Wang [42] apresenta um sistema baseado em internet das coisas para o desenvolvimento de cidades inteligentes e um planejamento urbano utilizando análises de *Big Data*. O sistema atual em tempo real usando diversos tipos de sensores (sensores de clima, de água, de estacionamento inteligente), câmeras de vigilância, botões de emergência nas ruas, entre outros dispositivos. Sua implementação estrutura-se na coleta, filtragem de dados, classificação, pré-processamento, processamento e tomada de decisão. Os sensores geram grandes quantidades de dados em alta velocidade, que são processados pelos *frameworks Hadoop* e *MapReduce* e garantem escalabilidade e eficiência ao sistema. Este sistema não beneficia somente os cidadãos, mas também as autoridades fornecendo informações que tornam as tomadas de decisão mais rápidas e eficientes.

#### B. Engenharia Elétrica

M. Mazhar Rathore [43] relata uma estrutura de análise *Big Data* para redes elétricas inteligentes e componentes de um sistema de apoio à decisão para economia de energia. A estrutura analisa, em tempo real, os dados de consumo de energia elétrica, identificar padrões de consumo e faz previsão do consumo de energia; além de fornecer cronogramas de operação ideais para eletrodomésticos. A infraestrutura de medição foi instalada em um edifício residencial nas simulações experimentais.

O sistema integra análise de dados e módulos dinâmicos de otimização para gerar padrões de consumo, soluções alternativas para economia de energia e melhoria na eficiência

energética através da automatização e programações operacionais dos eletrodomésticos e sistemas (pela programação de horários ideais para os aparelhos, sistemas de iluminação, aquecimento, ventilação e ar condicionado) que serão obtidos pelo sistema.

Porém, Jui-Sheng Chou [44] relata o desenvolvimento de uma infraestrutura de carregamento inteligente para carros elétricos, a partir da utilização de ferramentas de *Big Data*. A estrutura está sendo implementada na província de Jeju, na Coreia do Sul, onde os autores apresentam a distribuição dos carregadores e o número de carregadores em operação. O objetivo é combinar diversas técnicas estatísticas e de *machine learning* para identificar o padrão de demanda do carregamento dos veículos elétricos e integrar energia renovável para alimentar as redes de carregamento.

Essencialmente, as instalações de carregamento são monitoradas e controladas automaticamente por uma entidade central por meio de redes de comunicação predominantes, com ou sem fio. Os relatórios periódicos permitem investigar a confiabilidade da infraestrutura de carregamento e obter uma visão da qualidade do serviço; além do fato de que as análises estatísticas possibilitam recomendações para melhorar o provisionamento das instalações, o balanceamento da carga de energia e fazer uma previsão precisa pode permitir aos veículos elétricos selecionarem a fonte de energia apropriada e planejarem o tempo de carregamento.

### C. Engenharia de Produção

Junghoon Lee [45] apresenta uma proposta de arquitetura geral de análise baseada em *Big Data* para o ciclo de vida dos produtos (CVP). Esta arquitetura integra as análises de *Big Data* e padrões orientados a serviços que ajudaram nas tomadas de decisão, coordenação e otimização do processo de Produção Limpa (PL). Pela utilização de tecnologias e conceito internet das coisas em cada estágio do CVP desenvolve-se um ambiente de manufatura e manutenção inteligente.

Baseado nas aplicações integradas dos resultados da mineração dos dados, são fornecidos aos gerentes corporativos um melhor suporte à tomada de decisão, o aprimoramento da gestão de operações e serviços. Os resultados preliminares da pesquisa mostraram que a implantação beneficiou clientes, fabricantes, fornecedores, meio ambiente e todas as etapas do CVP, além de promover efetivamente a implementação da PL. Esta nova estratégia da PL fornece uma base teórica e prática para o desenvolvimento sustentável das empresas de manufatura.

Já Yingfeng Zhang [46] expõe as possibilidades e tendências da integração das análises avançadas de dados em manufatura, produtos e serviços, relacionando o *Big Data*, o avanço das Tecnologias de Informação e Comunicação (TIC) e o fato de que os sistemas físicos cibernéticos (*cyber-physical systems* – CPS) facilitam a transformação sistemática de dados massivos em informações e auxiliam na tomada de decisão. Além disso, o grupo ainda apresenta uma arquitetura sistemática denominada 5C que engloba etapas necessárias para integrar totalmente os CPS na manufatura.

Esta arquitetura 5C é capaz de automatizar e centralizar o processamento de dados e engloba as etapas de aquisição de dados, processamento de informações, visualização e apoio à tomada de decisão. Além disso, as informações de integridade geradas pelo sistema podem ser usadas para programação de manutenção e controle otimizado, afim de obter uma maior produtividade e confiabilidade.

### D. Engenharia Mecânica

Jay Lee [47] relata um sistema de apoio à decisão de manutenção inteligente (*Smart Maintenance Decision Support System* – SMDSS) com base nos dados corporativos de uma empresa. O sistema é capaz de fornecer ao usuário recomendações para melhorar os ciclos de vida dos ativos. Os autores ilustram como as ferramentas analíticas *Big Data* podem ser usadas para priorizar a manutenção do equipamento e definem a arquitetura na qual um proprietário de ativo físico pode inserir parâmetros de utilização para equipamentos interconectados e receber uma proposta abrangente para que o serviço cumpra as recomendações geradas pelo modelo analítico.

O SMDSS fornece um modelo analítico preditivo que pode ser aplicado em indústrias de manufatura e serviços. As descobertas e resultados mostram que algoritmos de solução e modelos de otimização existentes podem ser aplicados em grandes conjuntos de dados para definir e recomendar decisões aos gerentes.

Um dos grandes desafios da manutenção preditiva é lidar com grandes quantidades de dados heterogêneos disponíveis; e com proatividade impactar diretamente na competitividade do setor através de um monitoramento contínuo dos equipamentos industriais.

Não obstante, Daniel Bumblauskas [48] apresenta os resultados de um estudo numa empresa de metalurgia onde foram empregados métodos de análise de dados e de seleção de recursos a partir do monitoramento contínuo de equipamentos. As técnicas de aprendizado de máquina e mineração de dados foram usadas para extrair informações e auxiliar na tomada de decisão. As informações obtidas ajudarão no desenvolvimento de modelos de aprendizado adaptável, capazes de lidar com informações complexas que podem ser aplicadas em uma linha completa de produtos e equipamentos industriais.

### E. Engenharia de Materiais

Marta Fernandes [49] destaca, através de estudos de casos, as possibilidades de inovação e otimização de materiais auxiliados pela mineração de dados em exemplos como: a descoberta de materiais de uma nova bateria de lítio e novos materiais termoeletrônicos. No artigo ressalta-se também que os métodos quantitativos e qualitativos de aprendizado de máquina podem ser utilizados para realizar diferentes tarefas na descoberta, projeto e otimização de materiais, principalmente pelo bom desempenho, velocidade e simplicidade da mineração de dados aplicável, por exemplo, na obtenção de diagramas de classificação e na construção de modelos.

A descoberta de novos materiais é um processo sem fim e tem sido um marco no progresso humano.

Wencong Lu [50] apresenta um panorama e cita exemplos da evolução de pesquisas para a descoberta de novos materiais e classifica esta área de pesquisa como Materiais 4.0. O trabalho ressalta também o desenvolvimento de protocolos de aprendizado de máquina e fala dos *softwares* de simulação de propriedades de materiais quânticos, dados de materiais digitalizados, algoritmos inteligentes de aprendizado de máquina, entre outros. Enfatiza a redução considerável no tempo entre o conceito e a comercialização dos novos produtos e fala também da recuperação após a conclusão do ciclo de vida do produto. O desafio ainda é agregar dados de várias fontes, além de gerenciar e analisar dados não estruturados para habilitar os materiais 4.0; recomenda ainda que sejam desenvolvidas plataformas de pesquisas de materiais baseadas na web para explorar oportunidades e identificar lacunas.

#### F. Engenharia Química

Jose Rajan [51] apresenta as possibilidades de separações multidimensionais avançadas em espectrometria de massa utilizando ferramentas *Big Data*; onde a instrumentação analítica híbrida baseada em espectrometria de massa (*mass spectrometry* – MS) tem sido uma técnica poderosa para enfrentar os desafios na ciência e na medicina, inclusive auxiliar na descoberta de medicamentos e biologia sintética.

O artigo ressalta a possibilidade de medições em larga escala para obter informações e também ressalta os desafios do *Big Data* nas análises químicas: da enumeração de isômeros químicos até este campo da análise multidimensional com base na MS.

Enquanto a área de Engenharia Química está coletando cada vez mais dados, de diferentes fontes, essas análises se tornam mais difíceis em termos de uso dos dados certos e das ferramentas certas, para tomar melhores decisões em tempo real.

Já Jody May [52] exibe os avanços recentes em algumas áreas como: indústria química e farmacêutica, entre outras. Os objetivos deste artigo são educar a comunidade de Engenharia Química sobre a capacidade do *Big Data* e melhorar a confiabilidade e a eficiência operacional em vários setores da indústria, inclusive direcionar futuras pesquisas. O estudo ressalta ainda que 88% dos executivos da indústria química reconhecem que a análise de dados será crucial para manter a vantagem competitiva num intervalo de 5 anos.

Historicamente, a indústria de processos químicos foi uma das primeiras a adotar o controle de processos computadorizado. Por isso a coleta de dados do processo para monitoramento e controle tornou-se rotina e forneceu uma plataforma para exploração e desenvolvimento de métodos e aplicativos orientados a dados. Os servidores de banco de dados projetados especificamente para armazenar dados do processo, permitiram acesso fácil e confiável aos dados do processo por engenheiros e pesquisadores. A indústria de processos químicos foi pioneira na adoção de ferramentas orientadas a dados devido a essa vantagem.

#### G. Engenharia de Software

Leo Chiang [53] explora-se como uma tecnologia de Engenharia de Software (*Software Engineering* – SE) pode apoiar o desenvolvimento de projetos de *Big Data* e como as técnicas de *Big Data* podem ser utilizadas para desenvolver novos processos e evoluir as técnicas de SE. O artigo destaca que muitas pesquisas foram feitas nos últimos anos para melhorar a produção de sistemas de *Big Data* e garantir escalonamentos rápidos e elásticos, sempre e quando necessário. Um dos principais focos de pesquisa na aplicação de métodos de SE à produção de sistemas de *Big Data* está na área de desenvolvimento de arquiteturas de software.

Não obstante, as tecnologias da informação e comunicação e os softwares estão cada vez mais difundidos e devem começar a ganhar ainda mais destaque, principalmente em áreas como: nuvem (*cloud*), *Big Data*, *Internet of Things* e *Cyber-Physical Systems*.

Segundo Timothy Arndt [54] a engenharia de software como disciplina deve ser consolidada, pois apesar de apresentar impressionantes conquistas na área, é uma das mais recentes disciplinas científicas.

Atualmente pesquisadores em Engenharia de Software tem se destacado pela elevada qualidade dos softwares e pela adoção de práticas controladas no desenvolvimento e operação.

A Engenharia de Software fornece ferramentas e métodos corretos que auxiliam os usuários em todas as atividades relacionadas ao software e aos serviços. Nos últimos anos, os avanços levaram a uma automação crescente em aspectos como testes, implantação, gerenciamento de novas versões de softwares, além de permitir aos pesquisadores e profissionais identificar novas abordagens para a criação e operação de softwares e serviços.

Como os conjuntos de dados manipulados pelos softwares estão aumentando cada vez mais, além de fornecer novas arquiteturas de sistemas e infraestruturas de softwares que conseguem lidar com o *Big Data*, agora os próprios softwares devem se beneficiar da inteligência extraída de grandes conjuntos de informações, *bugs*, avisos e notificações, confirmações demográficas, padrões de codificação, comportamentos e perfis dos usuários, entre outros.

Os desafios para futuras pesquisas nesta área incluem ferramentas inovadoras que utilizam técnicas de aprendizado de máquina e mineração de dados para extrair informações de arquiteturas baseadas em sensores, minerando conhecimentos que são praticamente impossíveis de serem descobertos pelos seres humanos, mas que necessitam atenção, para melhorar as qualidades de software, estudando estruturas de aplicações, análise de tendências preferencias dos usuários, identificar principais causas de falhas, entre outros.

#### H. Evolução das publicações com o termo *Big Data* nos últimos 20 anos

Para se obter um breve panorama da evolução das publicações contendo o termo “*Big Data*” nos últimos 20 anos foi realizada, no dia 27 set. 2019, uma pesquisa no site Google Acadêmico. A coleta dos dados foi feita ano a ano durante o

período estudado e os dados obtidos do número de publicações em cada ano estão dispostos na Fig. 2.

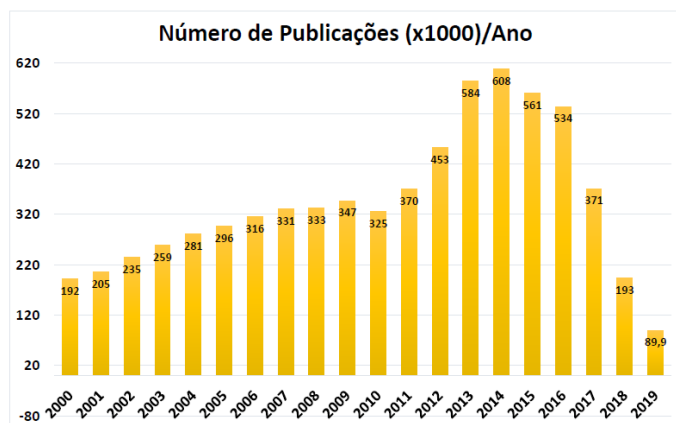


Fig 2- Quantidade de publicações (x1000) por ano

O gráfico da Fig. 2 confronta a quantidade de publicações em cada ano (eixo das ordenadas) com os respectivos anos de publicação, variando de 2000 até a presente data, em 2019 (eixo das abscissas). Pelo gráfico pode-se constatar o pico no número de publicações ocorreu no ano de 2014 (608.000), quando, a partir de então, tem se registrado um declínio quase que linear. Merece destaque o fato de que para o presente ano de 2019 foram consideradas publicações realizadas até a presente data deste artigo (27 set. 19). Portanto, nota-se que houve uma tendência maior de publicações com o termo “*Big Data*” entre os anos 2013 e 2016.

#### I. As publicações com o termo *Big Data* por área

A Busca pelo termo “*Big Data*” + “área da engenharia” estudada no trabalho nos últimos 20 anos, também realizada no dia 27 set. 2019 no Google Acadêmico a partir do ano 2000 até 2019 permitiu a elaboração do gráfico da Fig. 3.

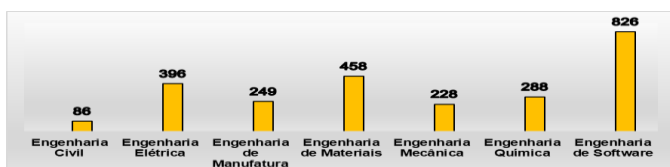


Fig 3- Quantidade de publicações (x 1000) por área da engenharia

O gráfico mostra a tendência de publicações em cada área da engenharia estudada neste trabalho para mesmo período. Pelo gráfico é possível constatar um maior número de publicações para os termos “*Big Data*” e “Engenharia de Software”, com 826.000 publicações no período.

Em seguida vem os termos “*Big Data*” e “Engenharia de Materiais” com 458.000 publicações e “*Big Data*” e “Engenharia Elétrica” com 396.000 publicações. O termo com o menor número de publicações dentre os pesquisados foi “*Big Data*” e “Engenharia Civil”, com 86.000 publicações.

#### V. CONCLUSÕES

As ferramentas e tecnologias relacionadas ao conceito *Big Data* são comprovadamente eficazes para a extração de

informações a partir dos mais diversos tipos de dados; em busca, principalmente, de maior eficiência e vantagem competitiva.

Esta pesquisa não apenas cumpriu o seu papel de, a partir de um levantamento bibliográfico, contextualizar de forma coesa e objetiva o termo *Big Data*, como também apresentou um breve panorama que, além da evolução das publicações num período de 20 anos, destacou algumas aplicações em: engenharia civil, elétrica, mecânica, produção, materiais, química e engenharia de *software*.

A evolução das publicações retrata que o conceito *Big Data* já teve seu pico de interesse (no ano de 2014) e nos últimos 4 anos tem havido um certo desinteresse da comunidade. Por conseguinte, espera-se que nos próximos anos seja atingido um patamar de equilíbrio que indicará ter atingido seu nível de maturidade.

Destaca-se também que globalmente vive-se um momento onde busca-se conectar tudo e todos, onde conceitos como internet das coisas e indústria 4.0 certamente contribuirão para que a população global gere ainda mais dados e informações e numa velocidade cada vez maior, o que necessita uma constante evolução das pesquisas e tecnologias para a captura, processamento, transmissão e análise destes dados; em busca de melhores tomadas de decisão, maior eficiência e também produtividade; já que a população global ainda não parou de crescer, muito menos seu consumo e anseios.

#### AGRADECIMENTOS

Os autores agradecem ao Centro de Ciências e Tecnologia – CCT e ao Fundo Mackenzie de Pesquisas – MackPesquisa pelo apoio aos projetos de pesquisas orientados pelos professores envolvidos neste trabalho.

#### REFERENCIAS

- [1] M. Ge; et al. Big data for internet of things: a survey. Future generation computer systems, 2018.
- [2] J. Gantz, John; D. Reinsel, David. The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east. IDC iView: IDC analyze the future, 2012. Disponível em: <<https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>> acesso em: 28 ago. 2019.
- [3] WE ARE SOCIAL. Digital in 2018: world's internet users pass the 4 billion mark. New York: We Are Social Ltd, 30 jan. 2018. Disponível em: <<https://wearesocial.com/blog/2018/01/global-digital-report-2018>> acesso em: 31 jul. 2019.
- [4] J. Manyika. Big data: the next frontier for innovation, competition, and productivity, 2011. Disponível em: <[http://www.mckinsey.com/insights/mgi/research/technology\\_and\\_innovation/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation)> acesso em: 23, set, 2018.
- [5] S. M. F. S. Massruhá. O papel na agricultura. Agroanalysis, 2015.
- [6] E. Melo. Desafios e oportunidades para a silvicultura de precisão: uma síntese do congresso brasileiro de agricultura de precisão de 2014. Série técnica ipef, piracicaba, 2016.
- [7] U. Srivastava; S. Gopalkrishnan, Santosh. Impact of big data analytics on banking sector: learning for Indian banks. Procedia computer science, 2015.
- [8] W. Lin et al. An ensemble random forest algorithm for insurance big data analysis. IEEE access, 2017.

- [9] P. D. Scaico et al.. O conceito big data na educação. In: anais do workshop de informática na escola, 2014.
- [10] S. Henderson. A few criminal justice big data rules. Ohio st. J. Crim. L., 2017.
- [11] P. Bertens, et al.. Games and big data: a scalable multi-dimensional churn prediction model. In: 2017 ieee conference on computational intelligence and games (cig). Ieee, 2017.
- [12] G. Kim et al. Big-data applications in the government sector. Communications of the acm, 2014.
- [13] M. Victorino et al. Uma proposta de ecossistema de big data para a análise de dados abertos governamentais conectados. Informação & sociedade, 2017.
- [14] Trelewicz. Big data and big money: the role of data in the financial sector. It professional, 2017.
- [15] G. Wang et al. Big data analytics in logistics and supply chain management: certain investigations for research and applications. International journal of production economics, 2016.
- [16] M A. Waller et al.. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. Journal of business logistics, 2013.
- [17] S. Erevelles et al. . Big data consumer analytics and the transformation of marketing. Journal of business research, 2016.
- [18] Z. Xu et al. Effects of big data analytics and traditional marketing analytics on new product success: a knowledge fusion perspective. Journal of business research, 2016.
- [19] J. Beckmann; D. Lew. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. Genome medicine, 2016.
- [20] V. Gligorijević et al. Integrative methods for analyzing big data in precision medicine. Proteomics, 2016.
- [21] T. Murdoch et al. The inevitable application of big data to health care. Jama, 2013.
- [22] W. Raghupathi et al. Big data analytics in healthcare: promise and potential. Health information science and systems, 2014.
- [23] F. Provost; T. Fawcett. Data science and its relationship to big data and data-driven decision making. Big data, 2013.
- [24] M. Janssen et al. Factors influencing big data decision-making quality. Journal of business research, 2017.
- [25] X. Zheng et al. Big data for social transportation. Ieee transactions on intelligent transportation systems, 2015.
- [26] Y. Lv et al. Traffic flow prediction with big data: a deep learning approach. Ieee transactions on intelligent transportation systems, 2014.
- [27] H. Chen et al. Business intelligence and analytics: from big data to big impact. Mis quarterly, 2012.
- [28] H. W. Park; L. Leydesdorff. Decomposing social and semantic networks in emerging “big data” research. Journal of informetrics, 2013.
- [29] T. H. Davenport et al. How 'big data' is different. Mit sloan management review, 2012.
- [30] A. McAfee et al. Big data: the management revolution. Harvard business review, 2012.
- [31] Khan, et al. Seven v's of big data understanding big data to extract value. In: proceedings of the 2014 zone 1 conference of the american society for engineering education. Ieee, 2014.
- [32] D. Barton; D. Court. Get started with big data: tie strategy to performance. Harvard business review, out. 2012. Disponível em: <<https://hbr.org/2012/10/getting-started-on-a-big-data>> acesso em: 15 ago. 2018.
- [33] J. Tabuena. What every internal auditor should know about big data. Compliance week, 2012.
- [34] C. Taurion. Você realmente sabe o que é big data. Blog da ibm, 2012. disponível em: <[https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/voce\\_realmente\\_sabe\\_o\\_que\\_e\\_big\\_data?lang=en](https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/voce_realmente_sabe_o_que_e_big_data?lang=en)> acesso em: 18 jul2018.
- [35] M. Sahoo et al. The emergence of big data: a survey. International journal of computer science and mobile applications, india, jun. 2018.
- [36] D. Laney. “3d data management: controlling data volume, velocity and variety” meta group research note, february 6, 2001.
- [37] P. Russom et al. Big data analytics. Tdwi best practices report, fourth quarter, v. 19, n. 4, p. 1-34, 2011.
- [38] B. Feldman et al. Big data in healthcare hype and hope. Dr. Bonnie, 2012.
- [39] E. G. Ularu et al. Perspectives on big data and big data analytics. Database systems journal, 2012.
- [40] A. Gandomi; M. Haider. Beyond the hype: big data concepts, methods, and analytics. International journal of information management, 2015.
- [41] L. Wang et al. Detecting and tracking vehicles in traffic by unmanned aerial vehicles. Automation in construction, 2016.
- [42] M. Rathore et al. Urban planning and building smart cities based on the internet of things using big data analytics. Computer networks, 2016.
- [43] J. Chou; N. Ngo. Smart grid data analytics framework for increasing energy savings in residential buildings. Automation in construction, 2016.
- [44] J. Lee et al. Big data analysis for an electric vehicle charging infrastructure using open data and software. In: proceedings of the eighth international conference on future energy systems. Acm, 2017.
- [45] Y. Zhang et al. A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. Journal of cleaner production, 2017.
- [46] J. Lee et al. Industrial big data analytics and cyber-physical systems for future maintenance & service innovation. Procedia cirp, 2015.
- [47] D. Bumblauskas et al. Smart maintenance decision support systems (smdss) based on corporate big data analytics. Expert systems with applications, 2017.
- [48] M. Fernandes et al. Data analysis and feature selection for predictive maintenance: a case-study in the metallurgic industry. International journal of information management, 2019.
- [49] W. Lu et al. Data mining-aided materials discovery and optimization. Journal of materiomics, 2017.
- [50] J. Rajan; R. Seeram. materials 4.0: materials big data enabled materials discovery, applied materials today, volume 10, 2018, pages 127-132, issn 2352-9407.
- [51] J. C. May; J. A. Mclean. Advanced multidimensional separations in mass spectrometry: navigating the big data deluge. Annual review of analytical chemistry, 2016.
- [52] L. Chiang et al. Big data analytics in chemical engineering. Annual review of chemical and biomolecular engineering, 2017.
- [53] T. Arndt. Big data and software engineering: prospects for mutual enrichment. Iran journal of computer science, 2018.
- [54] G. Casale et al. Current and future challenges of software engineering for services and applications. Procedia computer science, 2016.