

Big Data: quais os impactos e o que impulsiona

THIAGO MARTINS MATOS
FIAP - SP - Brasil
thiagomatos11@gmail.com

HUMBERTO BRAGA
FIAP - SP - Brasil
hb.braga@gmail.com

LUIZ TERUO KAWAMOTO JÚNIOR
FIAP - SP - Brasil
luizteruo@hotmail.com

Resumo - Com um grande volume de informações, o total de dados existente no mundo vem crescendo em escala exponencial surgindo a necessidade das empresas de analisar estes dados para a tomada de decisões. Big data, é o termo utilizado para definir o conjunto de soluções tecnológicas capaz de coletar, organizar e analisar qualquer tipo de informação em tempo real que pode ser dados estruturados e não estruturados que pode vir de mídias sociais, e-mail, smartphones, sensores, entre outros meios, essa coleção costuma ter 5Vs: Volume, Velocidade, Variedade, Veracidade e Valor, sendo fundamental para tomada de decisões, surgimento de novos produtos e insights, porém seu o funcionamento pode levar a invasão de privacidade.

Palavras-chave: Big Data, Informações, Impactos, Privacidade.

Abstract - *With a large volume of information, the amount of data in the world is growing at exponential emerging need for companies to analyze these data for decision making. Big data is a term used to define the set of technological solutions able to collect, organize and analyze any type of digital information in real time that can be structured and unstructured data that can come from social media, email, smartphones, sensors, transactions, among other means, this collection usually has 5VS: Volume, Velocity, Variety, Value and Veracity is central to decision making, development of new products and insights, but their operation can lead invasion of privacy.*

Keywords: *Big Data, Information, Impacts, Privacy.*

1. Introdução

O moderno mundo da tecnologia está cheio de dispositivos que incluem computadores, sensores, GPS's, antenas de celulares e dispositivos móveis como smartphones e tablets. Essa concentração de recursos computacionais permite detecção,

captura, coleta e processamento de dados em tempo real de bilhões de dispositivos conectados que servem a diferentes aplicações, incluindo redes sociais, monitoramento ambiental, aplicações industriais, comerciais e na medicina. Estes desenvolvimentos nos trouxeram para a era da Internet das Coisas criando grande rede de objetos identificáveis que poderão interoperar uns com os outros e com os data centers e suas nuvens computacionais. A Internet das Coisas, com seus objetos gerando dados a todo instante, é um impulsionador poderoso para Big Data.

Existem diversas definições sobre o Big Data, McKinsey (2011) define o Big Data como um conjunto de dados cujos o crescimento é exponencial devido a intensa utilização de smartphones, sites, redes sociais, transações, sensores e o uso da computação em nuvem vem gerando quantidade de dados incalculáveis cuja a dimensão está além da habilidade das ferramentas de capturar, gerenciar e analisar dados.

A EMC (2011) por sua vez, definição é intencionalmente subjetiva e incorpora de como mover um grande conjunto de dados precisa estar em ordem para serem considerados Big Data, para que a variedade, complexidade ou velocidade de requisitos precisa ser atendida.

Objetivo deste artigo é compreender melhor o fenômeno chamado Big Data, como essa ferramenta pode agregar valor para as organizações, como elas enfrentam a grande quantidade de dados que recebem todos dos dias, como tirar proveito desses dados e quais os desafios encontrados.

2. Cinco V's

O Big Data pode ser resumido com base nas características dos 3 Vs; Velocidade, Volume e Variedade. No entanto muitas vezes são esquecidos 2 Vs adicionais; Valor e Veracidade, que são críticos em uma governança da informação e do contexto legal.

2.1 Volume

André Petry (2013), afirma que 2,5 exabytes de informação são produzidos pela humanidade a cada dia, número que dobra a cada 40 meses. O volume de dados que circulam a cada segundo pela internet é maior do que toda a informação armazenada na rede a 20 anos atrás. Isso dá as empresas oportunidade de trabalhar com vários petabytes de dados em único conjunto de dados e não oriundos da internet. (Andrew McAfee e Erik Brynjolfsson, 2012, p. 40-41).

2.2 Velocidade

Em muitos casos, a velocidade de geração de dados é ainda mais importante do que o volume. A informação em tempo real, ou quase, permite a empresa ser muito mais ágil do que concorrentes. (Andrew McAfee e Erik Brynjolfsson, 2012, p. 41).

2.3 Variedade

Segundo artigo da revista Harvard Business Review, o Big Data inclui mensagens, atualizações e imagens em redes sociais; leituras de sensores; sinais de GPS de celulares e etc. Muitas das fontes mais importantes dessa profusão de dados são novas. A imensa quantidade de informação disponível em redes sociais, por exemplo, tem a mesma idade das redes em si; o Facebook foi lançado em 2004, Twitter em 2006. O mesmo vale para smartphones e outros aparelhos moveis que hoje trazem levadas imensas de dados ligados a pessoas, atividades e lugares. (Andrew McAfee e Erik Brynjolfsson,

2012, p. 41). Portanto existe uma grande variedade de dados, proveniente de várias fontes, podendo ser estruturados ou não. Alguns exemplos são dados gerados por sistemas transacionais, sensores, câmeras, satélites, logs, redes sociais, etc.

2.4 Valor

Segundo as estimativas da consultoria EMC (2011) em um estudo produtivo, descobriu-se que conjunto de dados podem aprimorar a produtividade e competitividade. Por exemplo, uma economia de U\$\$ 300 bilhões anuais somente na saúde, no setor público da Europa uma economia de € 250 Bilhões anuais na administração do setor, mais que o PIB da Grécia. E ajudar varejistas a aumentar margens de lucros em 60%.

2.5 Veracidade

De acordo Taurion (2013) para um dado gerar valor para o negócio é preciso ter veracidade, por que precisamos ter certeza que esses dados são autênticos e fazem sentido, ou seja precisamos nos certificar que os dados fazem sentido e são autênticos.

3. O que alimenta o Big Data

De acordo com a Exame (2013) as informações que alimentam o big data não são somente textos e números dos 640 milhões de sites, essas informações são vindas de sensores de localização em veículos, dos GPS's e das antenas dos 6 bilhões de celulares em uso no mundo e dos 2,7 bilhões de comentários feitos no Facebook diariamente.

Segundo o Oráculo Digital (2013) um estudo realizado pela IDC - International Data Corporation, conclui que o crescimento dos PCs e dos dispositivos móveis, o aumento do acesso à Internet e o “[...] aumento em dados de máquinas, como câmeras de vigilância ou medidores inteligentes têm contribuído para a duplicação do universo digital nos dois últimos anos para um montante de 2,8 ZB (Zetabytes)”.

Com isso, o volume de dados que temos hoje é consequência da diversidade de informações. Temos dados em formato estruturados, isto é, armazenados em bancos de dados, e dados não estruturados oriundos de inúmeras fontes, como documentos, imagens, áudios, vídeos, tuites e assim por diante e com isso temos uma enorme variedade de dados que pode alimentar o grandioso universo big data.

4. Impactos trazidos pelo Big Data

O Big Data pode ocasionar mudanças e impactos na forma como tratamos e protegemos nossas informações e principalmente de nossos clientes e usuários, podemos resumir os impactos tragos pelo Big Data:

4.1 Armazenamento

De acordo com IBM (2012) todos os dias são criados 2,5 quintilhões de bytes de dados, sendo que 90% desses dados foram criados nos últimos dois anos. Diante desse gigantesco volume de dados existe uma preocupação por parte das empresas de como armazenar tanta informação. Segundo o Olhar Digital (2012) um dos desafios para adotar o big data é a capacidade de armazenamento de bilhões de informações, além de oferecer a possibilidade de acesso, recuperação e processamento dessas informações com menor tempo. O site ainda completa que novas tecnologias orientadas a tratar

grandes volumes de dados vindo sendo desenvolvidas para suprir essa crescente demanda.

O TecMundo (2013) afirma que atualmente a questão de armazenamento se tornou simples e barato devido a evolução dos dispositivos de armazenamento. Com isso, existe uma grande demanda de volume de dados que alimentam o big data e graças aos avanços da tecnologia as empresas estão conseguindo acessar as informações através de poderosos dispositivos de armazenamento.

4.2 Segurança

Para os desafios de Big Data, torna-se clara a necessidade de estabelecer sistemas de proteção para toda essa informação. O fenômeno torna variáveis como o volume, a criação de informação em alta velocidade e variedade críticas para a segurança da informação.

A norma NBR ISO/IEC 17799 (2005, p.ix) define segurança da informação como “é a proteção da informação de vários tipos de ameaças para garantir a continuidade do negócio, minimizar o risco ao negócio, maximizar o retorno sobre os investimentos e as oportunidades de negócio”.

“A segurança também deve ser vista com atenção, pois se a empresa consegue acumular muitas informações sobre seus clientes, como comportamento, hábitos etc., devem tomar muito cuidado para que tais informações não sejam acessadas e divulgadas indevidamente. Mesmo quando informações individualmente não contenham explicitamente informação sobre determinado indivíduo, quando triangulada e integrada com diversas outras pode expor segredos industriais ou identificação pessoal íntima. Esta situação pode gerar não apenas processos judiciais, mas arranhar significativamente a imagem da corporação. Portanto, é essencial criar uma política de governança para dados e informações no mundo do Big Data.” (TAURION, 2013 p.59).

Alinhar o Big Data com práticas de segurança da informação pode melhorar visibilidade de ambientes de Tecnologia da Informação, além de reforçar capacidades de distinguir as atividades suspeitas e garantir a confiança dos sistemas contra as ameaças. Essas ameaças podem comprometer informações privadas de pessoas físicas e/ou jurídicas e até expor seus segredos e informações íntimas, levando a organização a sérios problemas como ações judiciais, a perda financeira e de credibilidade de sua marca.

4.3 Privacidade

O surgimento da Internet trouxe à tona questões sobre privacidade que antes não eram relevantes. Questões como até que ponto os dados de uma pessoa podem ser disponibilizados na rede ou até que ponto empresas pode fazer uso de programas capazes de coletar informações sobre seus usuários para alcançar seus próprios objetivos.

De acordo com Taurion (2013 p.33) um desafio que precisa ser debatido é a questão da privacidade. O autor ainda comenta sobre a pegada digital que cada pessoa deixa com o uso da internet “[...] Deixamos nossa pegada digital a todo o momento, seja usando o Internet Banking, comprando pela Internet, acessando um buscador, tuitando, comentando alguma coisa no Facebook, usando o smartphone, ativando serviços de localização”.

Aglutinar todas estas informações permite a uma empresa ou governo ter uma visão bem abrangente daquela pessoa e de seus hábitos e costumes. Taurion (2013 p.33)

afirma que “esta questão nos leva a outro ponto extremamente importante: garantir a segurança deste imenso volume de dados”.

Nesse podemos interpretar que atualmente temos meios de nos “intrrometer” na vida privada de quase qualquer pessoa que use a Internet, ao passo que novas tecnologias de segurança estão cada vez mais sendo usadas para garantir a privacidade e o uso indevido e não autorizado de informações pessoais.

Um bom exemplo de privacidade é suposto caso onde o EUA usou tecnologias de big data para espionar os brasileiros, segundo O GLOBO (2013) documentos secretos revelados por Edward Snowden um ex-técnico da NSA - National Security Agency mostra que em até 2002 funcionou em Brasília uma das estações de espionagem da NSA que trabalhavam em conjunto com a CIA - Central Intelligence Agency dos Estados Unidos. O jornal ainda completa que os documentos revelam que Brasília fez parte das 16 bases instaladas para coleta de informações através de satélites de outros países e ainda afirma que de acordo com os documentos de Edward Snowden a NSA “[...] espionou telefonemas e correspondência eletrônica de pessoas residentes ou em trânsito no Brasil, assim como empresas instaladas no país”.

De acordo com a INFO (2013) a infraestrutura montada pela NSA sugere que a captação e a análise dos dados são feitas num esquema de big data, ou seja, os dados são volumosos, de natureza muito variada e chegam em alta velocidade que caracterizam o big data. A matéria ainda afirma que:

“James Bamford, autor de vários livros sobre a NSA, diz que a rede da agência americana inclui salas de espionagem em operadoras de telecomunicações e um gigantesco data center em Bluffdale, no estado americano de Utah. Com esse aparato, basta um agente digitar o nome de uma pessoa num computador para que todas as conversas dela passem a ser registradas e todos os e-mails e mensagens instantâneas passem a ser interceptados. Depois, os computadores se encarregam de analisar esses dados.”

As matérias das homepage O Globo (2013) e Info (2013) mostram como a tecnologia envolvida no big data pode ser utilizada para quebrar a privacidade de qualquer pessoa ou organização através dos meios de comunicações como dados e voz, no caso mencionado a CIA em conjunto com a agência NSA usaram as tecnologias para espionar os brasileiros, uma vez que os dados foram interceptados pelos espiões das agências, seus poderosos computadores podem facilmente rastrear as pegadas digitais de qualquer pessoa física ou jurídica.

Taurion (2013 p.42) afirma que é “[...] necessário criar uma política de segurança da informação que classifique os dados de acordo com seu nível de privacidade. Existem dados que não podem ser disponibilizados publicamente e outros que podem [...]”. Taurion (2013 p. 58) ainda afirma que “devido à necessidade de integração de tecnologias e acessos a inúmeras bases de dados corporativos, e das questões de segurança e privacidade, TI deve atuar de forma proativa. Caso contrário, terá uma bomba-relógio em suas mãos!”

Diante dos riscos de comprometer a privacidade das pessoas e organizações é muito importante adotar uma política de segurança de informação para proteger os dados evitando que sejam publicados indevidamente afetando a empresa a passar por problemas judiciais e a perda de credibilidade.

4.4 Profissional – Cientista de Dados

A tecnologia de big data vem causando diversificações em carreiras tradicionais, segundo Taurion (2013, p. 78) o novo cargo, chamado de cientista de dados é um bom exemplo de oportunidade para profissionais de TI e de outros setores. De acordo com a

COMPUTERWORLD (2013) com o interesse das organizações pela implementação de projetos de Big Data, a profissão de cientista de dados ganhou importância e se tornou uma carreira promissora para os próximos anos. Para mostrar a importância dessa nova profissão o site alerta que por se tratar de uma profissão nova, achar gente capacitada não é tarefa fácil. Especialmente no Brasil, onde há déficit de talentos qualificados em TI.

A CIO (2013) também alerta o déficit dos profissionais e completa mostrando uma pesquisa realizada por uma grande consultoria onde a “ampliação das iniciativas de Big Data exigirá a contratação de um exército de 4,4 milhões de especialistas nessa área em todo o mundo até 2015”.

Diante das informações apresentadas, existe uma forte demanda para o cientista de dados e sua definição e características serão baseadas na ideia do autor **Taurion** (2013, p.78) que afirma que o cientista de dados “demanda normalmente formação em Ciência da Computação e Matemática, bem como as habilidades analíticas necessárias para encontrar a providencial agulha no palheiro de dados recolhidos pela empresa”.

Davenport e Patil (2012, p. 49) afirmam que um cientista de dados é alguém capaz de achar respostas para grandes dúvidas da empresa em meio a rios de informações desestruturadas. Os autores comentam sobre as dificuldades que as empresas passam para encontrar esse tipo de profissional:

“Na corrida para explorar o potencial do big data, o maior obstáculo enfrentado por empresas é a escassez desses talentos especiais. Como ainda não há cursos de nível superior para formar esse profissional é preciso criatividade para encontrá-lo. Busque gente diligente em qualquer área com forte ênfase em dados e computação - ainda que totalmente distante da gestão, como a física experimental ou a biologia de sistemas. Reconheça, também, que o que atrai e segura um cientista de dados no trabalho pode ser distinto daquilo que deixa satisfeito outros profissionais. Cientista de dados precisa de autonomia, mas querem estar na 'central de comando', dando respostas a problemas de administração com colegas da gestão em tempo real” (DAVENPORT; PARTIL 2012 p. 49).

Como o cientista de dados é uma profissão nova, as empresas tem certa dificuldade em encontrar esse profissional devido à escassez e a falta de formação específica. A melhor forma de superar essa escassez é a usar e explorar a criatividade dos profissionais de TI e/ou de outras áreas que tenham formação em exatas, que seja ligado ao negócio e que tenha capacidade de obter informações relevantes ao negócio através de uma imensidade de dados.

5. O que impulsiona o Big Data?

O mundo está passando por uma grande transformação através da computação em nuvem, interações de pessoas através das mídias sociais e dispositivos que interagem através da internet das coisas, mobilidade que trouxe mudanças do desktop para dispositivos móveis baseados na geolocalização, desencadeando uma avalanche de dados e interações que impulsiona o Big Data.

5.1 Internet das coisas:

Segundo a Gartner, é uma rede de objetos físicos que contenham tecnologia embarcada para interagir com seus ambientes internos ou ambientes externos. Ou seja a internet das coisas é uma relação simbiótica entre o mundo físico e o mundo digital.

A MIT Technology Review (2013), cita que exemplos não faltam, várias cidades instalaram sensores nas suas redes de transporte que transmite a posição dos ônibus, bondes e trens, mais o monitoramento do tráfego, tornando esses dados disponíveis para

o público, com atualizações em tempo real com a possibilidade de verificar o tempo de chegada. Outro exemplo é adoção da tecnologia de sensores para monitorar o desempenho esportivo, no qual estes sensores coletam dados sobre exercícios e enviam para um servidor central que os usuários podem acessar para analisar seu desempenho.

A Frost & Sullivan diz que há uma tendência subjacente que está levando a convergência de produtos, tecnologias e indústrias: Conectividade. Com uma previsão de 80 Bilhões de dispositivos conectados em 2020 ou 500 dispositivos conectados por Km² em 2020, a internet das coisas deve continuar a afetar todos os setores e isso afetará a mobilidade.

5.2 Mobilidade e Conectividade

Um levantamento realizado pela Qualcomm (2012) constatou que: quase seis em cada dez usuários (58%) usam aparelho uma vez a cada 30 minutos e ainda revela que mais da metade dos usuários navega na internet por meio de dispositivos móveis. Outro dado mostra que 91% dos entrevistados a mobilidade faz do País um lugar mais eficiente para fazer negócios e 66% dos brasileiros afirmam que a tecnologia wireless tem melhorado a vida de uma maneira geral.

5.3 Computação em Nuvem:

Segundo o NIST (2011), o Cloud é um modelo para permitir acesso sob demanda onipresente e conveniente via rede a um pool compartilhado de recursos computacionais configuráveis como: (Rede, Servidores, Armazenamento, Aplicativos e Serviços) que podem ser rapidamente provisionados e lançados com o mínimo esforço de gestão ou interação com o provedor de serviço.

Segundo a Intel (2013), a nuvem tem o potencial de desbloquear o Big Data, com isso os modelos de computação de ajudar acelerar o potencial de soluções de análise. Nuvens oferece eficiência e flexibilidade para acessar dados, surgimento de ideias e condução de valor.

6. Metodologia

Conforme Lakatos e Marconi (2005, p.83), o método é o conjunto de atividades que permite alcançar o objetivo, traçando o caminho a ser seguido, detectando erros e auxiliando as decisões na pesquisa.

O método desta pesquisa desse artigo foi de forma qualitativa exploratória, onde o instrumento utilizado foi a coleta de dados em revistas, livros, artigos, vídeos e palestras.

7. Conclusão

Conforme visto no artigo, com o volume de dados crescendo exponencialmente, novas fontes de dados surgem de forma cada vez mais detalhada. A evidência é clara, decisões tomadas com base em dados tendem a ser melhores e confiáveis, com isso o Big Data nas organizações são capazes de obter vantagem competitiva em relação à concorrência e de lidar com oportunidades de negócios com as quais não conseguiam lidar antes.

Porém observa-se muitos obstáculos no caminho como a falta de profissionais como o cientista de dados, os desafios culturais dentro das organizações, além da

questão de privacidade que cada vez mais ganhará importância, tornando necessário criar formas de desenvolver e assegurar a privacidade e segurança de clientes e usuários.

Referências

Associação Brasileira de Normas Técnicas – ABNT. NBR ISO/IEC 17799:2005: *Tecnologia da informação - código de prática para a gestão da segurança da informação*, 2005.

CIO, *Cientista de dados é o cara*. Disponível: <<http://cio.uol.com.br/gestao/2013/03/18/cientista-de-dados-e-o-cara/>>. Acesso em: 09 Jul. 2013.

COMPUTERWORLD, *Cientista de dados é profissão do futuro*. Disponível: <<http://computerworld.uol.com.br/carreira/2013/03/14/cientista-de-dados-e-profissao-do-futuro/>>. Acesso em: 15 Jul. 2013.

DAVENPORT, Thomas H.; PATIL D.J. *Cientista de dados: o profissional mais cobiçado do século 21*. Rev. Harvard Business Review Out. 2012, pag. 49.

EMC, *Big Data as a service: A market and technology Perspective*. Disponível: <<http://www.emc.com/collateral/software/white-papers/h10839-big-data-as-a-service-perspt.pdf>> Acesso em: 22 de Jul. 2013.

EXAME, *Um fenômeno chamado big data*. Disponível: <<http://exame.abril.com.br/revista-exame/edicoes/1025/noticias/para-nao-se-afogar-em-numeros?page=1>>. Acesso em: 25 Fev. 2013.

FROST & SULLIVAN, *Connectivity will change the future of mobility services*. Disponível: <http://www.frost.com/prod/servlet/press-release.pag?docid=261380984> Acesso em: 05 Ago. 2013.

IBM, *What is big data?* Disponível: <<http://www01.ibm.com/software/data/bigdata/>>. Acesso em: 30 Jul. 2013.

INFO, *EUA usaram tecnologias de Big Data para espionar brasileiros*. Disponível: <<http://info.abril.com.br/noticias/seguranca/2013/07/eua-usaram-tecnologias-de-big-data-para-espionar-brasileiros.shtml>>. Acesso em: 09 Jul. 2013.

INTEL, *Big Data in the Cloud: Converging Technologies*. Disponível: <<http://www.intel.com/content/dam/www/public/us/en/documents/product-briefs/big-data-cloud-technologies-brief.pdf>> Acesso em: 09 Jul. 2013.

LAKATOS, E.M.; MARCONI, M.A. *Fundamentos da metodologia científica*. 6.ed. São Paulo: Atlas, 2005.

MCAFEE, A.; BRYNJOLFSSON E. *Big data: a revolução da gestão*. Revista Harvard Business Review Out. 2012, pag. 41.

MCKINSEY, *Big Data: The next frontier for innovation, competition, and productivity*. <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation> Acesso em: 2 mar. 2013.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY, *2013: The year of the internet of Things*. Disponível: <<http://www.technologyreview.com/view/509546/2013-the-year-of-the-internet-of-things/>>. Acesso em: 04 Jul. 2013.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, *The NIST definition of Cloud Computer*. Disponível: <<http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>> Acesso em: 26 Jun. 2013.

O GLOBO, *NSA e CIA mantiveram em Brasília equipe para coleta de dados filtrados de satélite*. Disponível: <<http://oglobo.globo.com/mundo/nsa-cia-mantiveram-em-brasilia-equipe-para-coleta-de-dados-filtrados-de-satelite-8949723>>. Acesso em: 13 Jul. 2013.

OLHAR DIGITAL, *O desafio da TI: como armazenar e processar tantas informações no cenário atual*. Disponível em: <<http://olhardigital.uol.com.br/noticia/o-desafio-da-ti-como-armazenar-e-processar-tantas-informacoes-no-cen-rio-atual/26070>>. Acesso em: 30 Jul. 2013.

ORÁCULO DIGITAL, *Novo estudo do universo digital revela lacuna em big data*. Disponível: <<http://www.oraculodigital.com/?p=interna&id=956>>. Acesso em: 07 abr. 2013.

PETRY, A. *O Berço do Big Data*. Revista Veja, São Paulo Edição 2321, pag. 74, 15 Mai. 2013.

QUALCOMM, *TIME Mobility Poll, in Cooperation with Qualcomm*. Disponível: <<http://www.qualcomm.com/media/documents/files/time-mobility-poll-in-cooperation-with-qualcomm.pdf>>. Acesso em: 16 de Jul. 2013.

TAURION, C. *Big Data*. Editora Brasport Livros e Multimídia Ltda: Rio de Janeiro, 2013 96p.

TECMUNDO, *Big Data: uma grande bola maciça de informações*. Disponível em: <<http://www.tecmundo.com.br/armazenamento/35157-big-data-uma-grande-bola-macica-de-informacoes.htm>>. Acesso em: 30 Jul. 2013.