

Gerência de *Big Data*
e
seu impacto na sociedade

Gustavo Olegário

July 12, 2016

Resumo

Este artigo, traz como principais pontos o tema "Big Data" e seus principais impactos na atual sociedade. Além de explicar alguns conceitos, ligados a área de Banco de Dados, Redes de Computadores e Tecnologia da Informação em geral, o artigo traz consigo citações sólidas sobre o tema para ajudar o leitor a ter uma noção melhor do que se quer tratar. Os resultados obtidos são mostrados ao longo do artigo juntamente com a conclusão a que se chegou, bem como possíveis temas a se discutir sobre tal assunto.

1 Introdução

Não é novidade que a tecnologia e os sistemas embarcados têm sido difundidos pela humanidade de maneira colossal. Essa evolução tecnológica e o avanço na área da Ciência da Computação, permitiu ao ser humano que ele tivesse um computador na palma de sua. Tais fatos permitiram ao ser humano a criação de dados em qualquer lugar e a qualquer momento.

Para se ter uma ideia, em 2012 foram gerados 2.8 trilhões de GB. Essa geração de dados descontrolada foi batizada de "Big Data". Com o passar do tempo, as empresas perceberam que se conseguissem utilizar os dados da forma correta elas poderiam melhorar seus serviços e produtos. Isso fez com que a gerência de "Big Data" ganhasse um destaque importantíssimo. Para que se possa entender melhor tal contexto, far-se-á uma melhor explanação de conceitos e citações mais adiante.

1.1 Motivação

A motivação principal para a confecção deste artigo deve-se a crescente notoriedade que a área de "Big Data" e sua gerência vem ganhando nestes últimos anos. Além disso, como nota-se (Silveira, Marcolin & Freitas 2015) o campo de Big Data tem sido pouquíssimo explorado no Brasil. Por último, o maior conhecimento do "Big Data" pode trazer ganhos significativos para a sociedade, como será visto mais adiante.

1.2 Justificativas

Como já foi comentado anteriormente, o campo de "Big Data" vem ganhando um grande destaque tanto pelas empresas, quanto pelas instituições de ensino. Tal fato ocorre, pois essa quantidade gigantesca de dados combinada com a estatística e outras áreas de ciências exatas tem o poder de melhorar produtos, pesquisas e até mesmo fazer previsões. Além disso, o Brasil tem sido um dos pouquíssimos países

que tem feito pesquisas nesta área. Sendo assim, este artigo procura também estimular pesquisas sobre este tema em nosso País.

1.3 Objetivos (Gerais e Específicos)

Os objetivos principais deste artigo estão em entender como é o funcionamento base do "Big Data" e como essa quantidade massiva de dados é gerenciada. Além disso, o artigo trata, de maneira supérflua, como as empresas tentam usar os dados a seu favor e quais são os desafios relacionados a essa área tão nova e tão complexa.

1.4 Organização do Artigo

O artigo está estruturado da seguinte forma: na seção 2 apresentam-se os conceitos básicos (questões históricas e técnicas). Na seção 3 os trabalhos correlatos (referências). Já na seção 4 os aspectos relevantes, equanto que na seção 5 os problemas existentes são apresentados. Por último na seção 6 temos possíveis soluções, na 7 trabalhos futuros e na 8 conclusões.

2 Conceitos Básicos

O "Big Data", apesar de usualmente ser traduzido apenas como grande quantidade de dados, está muito além disso. O "Big Data" não envolve somente um volume colossal de dados, mas também outras informações.

2.1 Volume

Organizações coletam dados de uma grande variedade de fontes, incluindo transações comerciais, redes sociais e informações de sensores ou dados transmitidos de máquina a máquina. No passado, armazenar tamanha quantidade de informações teria sido um problema – mas novas tecnologias (como o Hadoop) têm aliviado a carga.

2.2 Velocidade

Os dados fluem em uma velocidade sem precedentes e devem ser tratados em tempo hábil. Tags de RFID, sensores, celulares e contadores inteligentes estão impulsionado a necessidade de lidar com imensas quantidades de dados em tempo real, ou quase real.

2.3 Variedade

Os dados são gerados em todos os tipos de formatos - de dados estruturados, dados numéricos em bancos de dados tradicionais, até documentos de texto não estruturados, e-mail, vídeo, áudio, dados de cotações da bolsa e transações financeiras. Mas e para armazenar essa quantidade abusarda de dados? Como isso é feito? Usam-se servidores.

2.4 Servidor

Um servidor é um sistema de computação centralizada que fornece serviços a uma rede de computadores. Esses serviços podem ser de naturezas distintas, como por exemplo, arquivos. Os computadores que acessam os serviços de um servidor são chamados clientes. As redes que utilizam servidores são do tipo cliente-servidor, utilizadas em redes de médio e grande porte (com muitas máquinas) e em redes onde a questão da segurança desempenha um papel de grande importância. O termo servidor é amplamente aplicado a computadores completos, embora um servidor possa equivaler a um software ou a partes de um sistema computacional, ou até mesmo a uma máquina que não seja necessariamente um computador. Os servidores que lidam com banco de dados são chamados de Servidor de banco de dados. Responsáveis por possuírem e manipularem informações contidas em um banco de dados

3 Trabalhos Correlatos

”USO CORPORATIVO DO BIG DATA: UMA REVISÃO DE LITERATURA”
: O Big Data é um advento tecnológico de processamento de grandes volumes de dados que vem ganhando notoriedade por conta de oportunidades e desafios em torno de sua utilidade no apoio aos negócios. Diante disso, este artigo procura, através de uma revisão sistemática da literatura, identificar como estão ligados o Big Data e o mundo corporativo.

”ANÁLISE COMPORTAMENTAL POR INTERMÉDIO DAS TECNOLOGIAS DE BIG DATA”: O mundo está cada vez mais interligado. Novas tecnologias vêm sendo criadas para conectar e diminuir a distância entre as pessoas. Com a ascensão das tecnologias há um crescimento exponencial na geração de dados variados, estruturados ou não. Analisar estes dados em busca de informações úteis para o negócio, visando ainda, a obtenção do comportamento de um grupo de indivíduos sobre uma questão em exame, é o que este artigo aborda. Isso a partir de demonstrações e descrições das tecnologias de Big Data, explorando ainda, os conceitos fundamentais aplicados ao contexto atual.

”Big Data’: Big Gaps of Knowledge in the Field of Internet Science”: Como um membro do conselho editorial e editores da Revista Internacional de Ciência Internet, gostaríamos de aproveitar esta oportunidade para comentar sobre alguns desenvolvimentos interessantes no campo da ciência Internet e Web Ciência. A análise dos chamados ”Big Data” recebeu um impulso notável. Realização de uma pesquisa com ”Big Data” como uma consulta, encontramos 130 entradas no ISI Web of Science, em julho de 12. Destes, 94 publicações têm aparecido desde 2008, o que não é surpresa, porque antes de 2008 não houve consenso terminológico.

”BIG DATA: GETTING A BETTER READ ON PERFORMANCE”: Ao longo dos últimos anos, muitas empresas têm avidamente perseguido o prometido benefícios de big data e avançado analytics. Em uma pesquisa recente da McKinsey de executivos nesta área, quase todos deles disseram que suas organizações tinham feito investimentos significativos, a partir de dados armazéns para o Google Analytics programas.

”What is big data? A consensual definition and a review of key research topics”: Apesar de Big Data ser um chavão de tendências tanto na academia, quanto na indústria, o seu significado ainda está envolta por muito imprecisão conceitual. O termo é usado para descrever uma ampla gama de conceitos: desde a capacidade tecnológica de armazenar, agregar e processar dados, para a mudança cultural que está invadindo as empresas e a sociedade, de tanto afogamento em sobrecarga de informação. A falta de uma definição formal levou a investigação a evoluir para múltiplas e inconsistentes caminhos. Além disso, a ambiguidade existente entre pesquisadores e profissionais fragiliza o desenvolvimento eficiente .

”Why ”Big Data””: Mostra os grandes benefícios do ”Big Data”.

”The Age of Big Data”: O impacto da abundância de dados se estende bem além dos negócios. Justin Grimmer, por exemplo, é uma das nova geração de cientistas políticos. Um professor de 28 anos assistente na Universidade de Stanford, ele combinou matemática com a ciência política em seus estudos de graduação e pós-graduação, vendo ”uma oportunidade, porque a disciplina é cada vez mais intensivo de dados. ”Sua pesquisa envolve a análise automatizada por computador de blogue postagens, discursos do Congresso e comunicados de imprensa e artigos de notícias, à procura de insights sobre como as idéias políticas se espalhar.

4 Aspectos Relevantes

Geram-se milhares de dados por segundo, sejam através de publicações na internet e mídias em geral, ou por transações, dispositivos, conversas etc. Muitas dessas informações são geradas publicamente e de forma voluntária, sendo significativa parte a expressão de opiniões, experiências, localizações e infinitudes de

outras ramificações, isso, principalmente nas redes sociais. Esses procedimentos deixam "rastros" que quando bem analisados se tornam poderosa ferramenta de análise comportamental e social. Em caso de utilização em tempo real, se configuram úteis inclusive para prevenção de tendências e suporte para tomada de decisões críticas, tanto governamentais quanto privadas. O processamento desta enorme quantidade de dados em velocidades aceitáveis exige soluções específicas, que já vêm sendo discutidas amplamente. Uma delas é o Big Data.

As empresas estão tendendo a constituir ou melhorar suas estratégias de negócios com base nessas informações coletadas. Nota-se que entender e utilizar os dados já coletados em conjunto com novas coletas seria essencial para se conhecer melhor as preferências do mercado. Contudo as soluções Big Data não se limitam apenas a isso, também se apresentam extremamente importantes para áreas críticas da sociedade, como por exemplo: a educação, a saúde, a segurança e os avanços científicos.

Implementações de Big Data para análise comportamental tem sido aperfeiçoadas, haja vista o projeto da IBM de análise de "sentimentos" em twitters, utilizado na copa das confederações de 2013 ou as sugestões inteligentes de produtos por lojas virtuais. Porém, mesmo assim, enfrentam-se desafios complexos, ainda a solucionar.

De acordo com Qmee (2014), essa massa de dados que é gerada na internet pode ser visualizada por meio de infográfico (pode ser acessado no sítio), apresentando em 60 segundos o demonstrativo das informações geradas a partir dos principais serviços da internet, representados pelo Google, Facebook, Twitter, dentre outras empresas. O trabalho revela que a cada 60 segundos são realizadas dois milhões de pesquisas no Google, no Facebook se geram 350 GB de dados, sendo distribuídos entre postagens, likes, fotos e vídeos. No Instagram, rede social para compartilhamento de fotografias, publica-se por minuto cerca de 3.600 fotos; no LinkedIn, rede social para profissionais, realizam-se onze mil consultas por minuto em perfis profissionais; e, na loja virtual da Amazon é obtido aproximadamente U\$83.000 em vendas no mesmo tempo.

Mas, não é apenas as redes sociais ou a própria internet em si que são capazes de gerar informações plausíveis de processamento e análise, outros dispositivos físicos, como automóveis, aeronaves, rodovias, robôs, etiquetas de rádio frequência (RFID), dentre outros, também podem ser encarados como fontes de dados.

Estima-se que do início da civilização até o ano de 2013 a humanidade criou em torno de cinco hexabytes de dados, valor no presente se alcança em questão de dias (VILLELA, 2013).

5 Problemas existentes

5.1 Ocupam muito espaço

A lógica sugere que, quanto mais dados as empresas tiverem, mais complexa será a tarefa de geri-los. Será preciso responder perguntas do tipo: Vale comprar equipamentos para guardá-los? Melhor armazenar na nuvem? Com que frequência você precisa acessar esses registros? Como pode lidar com latência?

5.2 Trazem perspectivas difusas

Uma questão relativa à grande quantidade de dados é que isso cria polarização e diferentes perspectivas. Vamos dizer que existam dois silos em uma empresa. Parece simples decifrar o que esses dois recipientes contém de maneira macro. Agora, se forem analisando 100, pode haver muitas nuances dentro desses dados. Certamente, já se ouviu que os números dizem o que as pessoas querem ouvir e que, espremidos, podem dizer qualquer coisa. Pois bem, quanto mais dados tiver, mais espaço de manobra existe para respostas imprecisas.

5.3 Nem sempre trazem aquilo que você estava procurando...

Imagine o Google Maps dando-lhe instruções e sugerindo como alternativa uma rota “mais rápida”. Porém, apenas depois de pegar esse caminho descobrirá que ele é esburacado ou sem asfalto. Às vezes, sistemas de big data oferecem atalhos que, na verdade, não são exatamente aquilo que você estaria procurando.

5.4 Perguntas erradas

A ciência de análise de dados é uma mistura complexa de conhecimento de contextualização (a profunda compreensão do sistema bancário, varejo, ou de outras indústrias), especialidade em matemática e estatística, além de habilidades com programação. Muitas organizações estão para contratar cientistas de dados que são gênios da matemática e da programação, mas que não têm o componente mais importante de um cientista de dados: o conhecimento de contextualização. É por isso que especialistas aconselham que o melhor para as empresas é procurar profissionais a partir dos colaboradores que já estão na empresa e já conhecem o negócio, pois, em teoria, é mais fácil e rápido aprender Hadoop.

5.5 Big Data integrado

O mais interessante do Big Data para qualquer empresa é a possibilidade de implementar a tecnologia de maneira integrada em cada departamento da companhia (e não separá-los), fazendo com que os dados sejam organizados de forma conectada. Isso traz mais valor à análise. No entanto, as políticas de segurança das empresas muitas vezes frustram essa promessa. Ao longo desta lista, um tema comum emerge: por mais que as empresas possam querer se concentrar em dados, as pessoas continuam a ficar no caminho. Tanto quanto podemos querer trabalhar com base nos dados, as pessoas, em última instância, controlam os processos de Big Data, incluindo a tomada de decisões iniciais a respeito de quais dados serão recolhidos e guardados, e quais perguntas serão feitas sobre esses dados.

6 Possíveis soluções

6.1 Armazenamento

Obviamente, não existe quantidade de silício no mundo o suficiente para armanezar todos os bits gerados pela humanidade, em uma janela de tempo de tamanho considerável. Sendo assim, as empresas e instituições de ensino devem armanezar dados específicos e que depois de estudados sejam descartados. Outra possibilidade para a questão do espaço, é a compreensão de dados, de forma similar quando um vídeo é gravado.

6.2 Perspectiva

A medida que a quantidade de dados vai aumentando, a precisão vai ficando cada vez menor e fica mais difícil de computar. Uma possível solução, seria reduzir o universo com que as empresas trabalham. Reduzir a quantidade de dados para um número que ainda represente a realidade, mas que ao mesmo tempo se tenha precisão e facilidade nos cálculos.

6.3 Imprecisão

Este tipo de problema pode decorrer da falta de atualização de dados. Hoje em dia, com o aperfeiçoamento que existe na questão da otimização e da gerência constante de dados tem sido cada vez menos comuns esses problemas. Uma possível solução, seria a participação do usuário. O usuário poderia informar ao aplicativo ou atualizá-lo avisando aos outros usuários que uma estrada está esburacada ou engarrafada (como já acontece no aplicativo Waze por exemplo).

6.4 Otimizar a questão de logística

Soluções como softwares de gestão de logística de transporte, por exemplo, podem ser usados para entender os padrões e as tendências de mercado no que diz respeito a quantidades de pedidos e distâncias de entrega. Estas informações preparam a empresa dando o conhecimento necessário para trabalhar melhor com as tendências de serviço ao comprador, e aplicar os recursos adequados no local para melhor responder melhor às demandas, ganhando assim mais credibilidade com os clientes.

6.5 Economia de combustível e outros insumos

A eficiência dos insumos pode ser melhorada com a utilização do Big Data. Um exemplo disso, é o aproveitamento das informações obtidas através dos computadores de bordo dos veículos para a economia de combustível. Um case de sucesso neste sentido e da empresa de logística UPS, que ao notar que nas rotas que haviam muitas curvas à esquerda, os veículos de transporte gastavam mais combustível, economizaram 10 milhões de litros de gasolina em um ano, apenas alterando os trajetos. Os dados obtidos através do Big Data também podem dar às empresas a vantagem de saber quais são os funcionários mais produtivos neste sentido e, desta forma, definir metas alcançáveis.

6.6 Economia com a otimização dos processos

Através do Big Data, se torna possível a identificação de processos onde ocorrem perdas de produtividade e má alocação de recursos. Desta forma, as estratégias podem ser aperfeiçoadas de forma a gerar economia, e os talentos passam a ter condições de serem muito melhor aproveitados. Uma das formas mais comuns de aproveitamento do Big Data neste sentido, é a avaliação das maneiras mais eficazes de abordagem dos clientes, o que pode aumentar os resultados das campanhas de maneira considerável.

7 Trabalhos futuros

Como possíveis trabalhos futuros sugere-se a relação entre "Big Data" e segurança da vida pessoal. Outro possível tema para trabalhar é com "Big Data" e sua importância no mundo empresarial, mostrando como os dados podem ser utilizados no mundo dos negócios.

8 Conclusão

Sem dúvida o tema de "Big Data" é extremamente amplo e a sociedade está apenas engatinhando nesta área. Ainda há muito para se descobrir e formas de utilizar que facilitarão a vida do ser humano. Todavia, é preciso sagacidade na gerência e uso de tais dados, pois seu uso desenfreado e mal pensado pode acabar sendo mais prejudicial do que benéfico.

Referências

- [1] SILVEIRA, Marcio; MARCOLIN, Carla Bonato; FREITAS, Henrique Mello Rodrigues. *USO CORPORATIVO DO BIG DATA: UMA REVISÃO DE LITERATURA*. Revista FATEC Zona Sul, Fevereiro.2016.
- [2] BERNARDO, Daniel dos Santos; BORLONE, Danilo Romulado. *ANÁLISE COMPORTAMENTAL POR INTERMÉDIO DAS TECNOLOGIAS DE BIG DATA*. Revisão e Gestão de Projetos Volume 6, Nº3. 2015.
- [3] SNIJDERS, Chris; MATZAT, Uwe; REIPS, Ulf-Dietrich. "Big Data": *Big Gaps of Knowledge in the Field of Internet Science*. International Journal of Internet Science, Fevereiro.2012.
- [4] BUGHIN, Jacques. *BIG DATA: GETTING A BETTER READ ON PERFORMANCE*. McKinsey & Company, Fevereiro.2016.
- [5] MAURO, Andrea de; GRECO, Marco; GRIMALDI, Michele. *What is big data? A consensual definition and a review of key research topics*. AIP Proceedings, Setembro.2015
- [6] SHAW, Jonathan. *Why "Big Data"*. Harvard Magazine, Abril.2014
- [7] LOHR, Steve. *The Age of Big Data*. New York Times, 11 de Fevereiro. 2012