



Universidades Lusíada

Alvarez, Isabel Maria Surdinho Borges, 1948-
Viana, Joaquim Mesquita da Cunha, 1954-
Silva, Nuno Sotero Alves da

Big data : suporte consistente das decisões de gestão

<http://hdl.handle.net/11067/3409>

<https://doi.org/10.34628/1w2j-9487>

Metadados

Data de Publicação	2016
Resumo	Este artigo foca a descrição das inovações introduzidas por Big Data, a sua evolução exponencial e as novas abordagens tecnológicas e alguns dos temas éticos associados bem como as possíveis mudanças para a sociedade....
Palavras Chave	Big data, Tomada de decisão
Tipo	article
Revisão de Pares	Não
Coleções	[ULL-FCEE] LEE, n. 21 (2016)

Esta página foi gerada automaticamente em 2024-09-21T10:24:09Z com informação proveniente do Repositório

BIG DATA:
SUPORTE CONSISTENTE DAS DECISÕES DE GESTÃO

Isabel Alvarez
Universidade Lusíada, Lisboa
Universidade Autónoma, Lisboa

Joaquim da Cunha Viana
Universidade Lusíada, Lisboa

Nuno Sotero da Silva
Universidade Lusíada, Lisboa

Resumo: Este artigo foca a descrição das inovações introduzidas por Big Data, a sua evolução exponencial e as novas abordagens tecnológicas e alguns dos temas éticos associados bem como as possíveis mudanças para a sociedade

Palavras-chave: Big Data (megadados), NoSQL, análise de dados, ética de computador, a vigilância de dados.

Abstract: This work will focus on describing the innovations introduced by Big Data, its exponential evolution, the new technological approaches and also some of the associated ethical issues and the possible changes to society.

Keywords: Big Data, NoSQL, data analytics, computer ethics, data surveillance.

1. Introdução

O termo Big Data refere-se, tal como Gerard Broussard [1] o definiu, ao “processamento de um volume de dados demasiado grande para ser gerido por um único servidor de ficheiros”, Os dados em causa poderão mesmo, frequentemente, ser não estruturados e originados em diferentes fontes.

Com a evolução exponencial da disponibilidade das redes de comunicação, dados guardados em servidores dos mais diversos tipos, em redes com as configurações mais díspares e sendo, quaisquer deles e delas, geridos pelas mais variadas organizações, passaram a estar disponíveis para quase todos os utilizadores de equipamentos informáticos.

Tendo esta conjuntura em conta, Mujeeb e Naidu [4] definem como características essenciais “Variedade, Volume e Velocidade”.

Variedade

As novas fontes de dados, principalmente redes sociais e dispositivos móveis disseminam novos tipos de dados de máquina para máquina. Uma grande parte destes dados precisam agora ser adicionados aos dados transaccionais convencionais. Sendo em grande parte não estruturados, nomeadamente sob a forma de texto, áudio e vídeo, estes tipos de dados adicionam um novo grau de complexidade às aplicações que com eles têm de lidar.

Volume

Dia a dia, a quantidade de dados gerados pelas mais diversas aplicações aumenta de uma forma dramática. Esta realidade exige cada vez maior capacidade de armazenamento, bem como torna essencial a necessidade de distribuição dos locais, quer de armazenamento quer de processamento.

Velocidade

A quantidade de dados armazenada e a necessidade de processamento está directamente relacionada com a frequência com que aqueles são gerados e partilhados. A recente evolução da Internet das Coisas (Internet of Things – IoT) faz-nos crer que a situação poderá tornar-se ainda mais exigente à medida que

mais dispositivos vão sendo preparados para recolha e partilha de dados em tempo real. Já hoje, a utilização de bases de dados relacionais clássicas se começa a mostrar insuficiente.

Ainda de acordo com os mesmos autores [4] num ambiente em que os sistemas distribuídos, com a evolução explosiva da World Wide Web, na década de 90 do século passado se tornaram a regra, houve uma primeira tentativa de abordagem a estas questões de “Big Data” através da utilização de tecnologia de bases de dados paralelas. Rapidamente, no entanto, se concluiu que aquele caminho não era adequado, nem do ponto de vista técnico nem do financeiro.

Por outro lado, segundo a IBM [2]:

- 80 % da informação mundial é não estruturada.
- Este tipo de informação está a crescer 15 vezes mais rápido que a estruturada.
- A potência computacional está a crescer de tal maneira que os computadores pessoais hoje, são praticamente tão potentes como os supercomputadores de há uma década atrás.

Por outro lado, um inquérito da mesma companhia concluiu que mais de metade dos gestores, hoje em dia, sentem que não têm acesso à informação necessária para uma tomada de decisão consistente e fundamentada. Neste contexto, a evolução da quantidade de dados espalhados pelo mundo tem despertado muito significativamente a atenção dos académicos, e dos técnicos em geral, que se debruçam sobre a problemática dos sistemas de informação, como pode talvez ser verificado através da análise do gráfico da figura 1, extraído do sítio Web da Google especializado no tema, em que se apresenta a evolução das pesquisas registadas pela companhia no seu motor de busca no decurso dos últimos anos.

Com efeito, à medida que o volume de dados disponível aumenta, a sua utilização (transformação em conhecimento) torna-se de dia para dia uma tarefa cada vez mais complicada. A análise expedita de dados emerge assim como uma área de grande visibilidade relativamente ao futuro dos sistemas de informação. A tal ponto que, Chen et al. [3], citando Hal Varian, Economista Chefe na Google e Professor Emérito da Universidade de Berkeley, na Califórnia, recomenda aos neófitos nestas tecnologias um forte investimento no estudo de bases de dados, aprendizagem máquina, econometria, estatística e visualização dada a grande necessidade que estes sectores do conhecimento apresentarão num futuro próximo - e mesmo já no presente - de técnicos especializados em análise de dados.

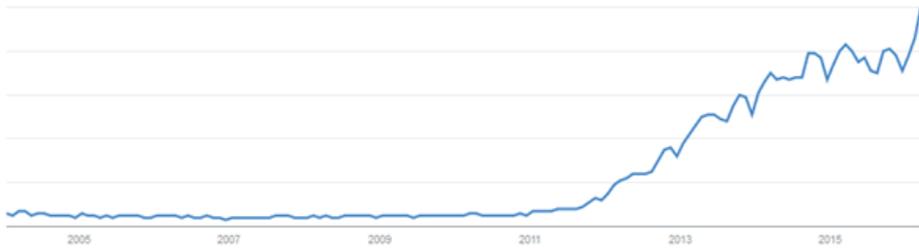


Figura 1 - Interesse ao longo do tempo

2. Novas abordagens tecnológicas

Face a todos os constrangimentos atrás referidos, novas abordagens têm de ser estudadas e implementadas no sentido de garantir processamentos mais rápidos e, conseqüentemente, mais eficientes. Assim sendo, cada vez mais, o processamento de “big data” assenta em técnicas de áreas relacionadas com estatística e inteligência artificial, nomeadamente, redes neuronais, reconhecimento de padrões e processamento de sinais [4].

Todas estas técnicas tiram partido de sistemas de armazenamento e exploração de dados [5].

A especificidade desta área, faz com que, ainda que muitas das plataformas analíticas actuais se suportem em sistemas baseados exclusivamente em SQL, dadas a maior fiabilidade e simplicidade da linguagem de consulta (“query”) e do alto nível de desempenho na análise de grandes fluxos de dados em tempo real, sejam cada vez mais as abordagens conhecidas como NoSQL [6, 7].

O termo “NoSQL” foi criado em 1998 por Carlo Strozzi [8] para o seu sistema de gestão de bases de dados relacional, Strozzi NoSQL. A razão para o nome cunhado por Strozzi prendeu-se com o facto de a sua base de dados não apresentar uma interface SQL com o utilizador. Mais recentemente, o termo NoSQL, cujo significado é ‘não apenas SQL’ (‘not only SQL’) passou a ser usado para descrever bases de dados com propriedades distintas das bases de dados relacionais clássicas e que, geralmente, não são interrogadas recorrendo à linguagem SQL.

Para armazenar e gerir dados não estruturados, o NoSQL, [6] ao contrário das bases de dados relacionais clássicas, separa a função armazenamento da gestão, tornando-as independentes uma da outra.

A arquitectura NoSql sendo projectada para uma utilização eficiente de dados não só não estruturados mas também frequentemente distribuídos procura, quando o repositório de dados é partilhado, que os grupos de equipamentos (‘clusters’) não fiquem dependentes uns dos outros [7]. Quaisquer que sejam as acções de actualização a desencadear, as alterações sofridas por qualquer dos ‘clusters’ não deve afectar nenhum outro.

Para conseguir este objectivo, foi inserido entre o sistema base e o dispositivo de armazenamento de longo prazo, uma nova camada, com elevado potencial de crescimento, sendo que, além disso, os dados a processar são armazenados em memória RAM, muito mais rápida que qualquer disco rígido disponível [9]. Segundo estes autores, as principais características de uma tal arquitectura deverão ser as seguintes:

- Ser capaz de gerir dados como objectos;
- Poder armazenar esses objectos em memória de modo a melhorar os desempenhos e evitar engarrafamentos nos acessos aos discos, tão preocupantes para os gestores de bases de dados;
- Ser capaz de carregar dados em memória principal, a partir da secundária, de modo transparente;
- Conseguir propagar para a memória de longo prazo as alterações efectuadas sobre os dados, possivelmente de forma assíncrona;
- Poder crescer de forma simples.

3. Novas oportunidades e novos desafios

As oportunidades e desafios de Big Data derivam da recolha, agregação e processamento de massivas quantidades de dados, bem como das técnicas usadas para o seu processamento (data analytics). Colocam-se questões sobre como irá alterar a ciência e a natureza do conhecimento, como afectará as áreas de negócio, transportes, saúde, etc. A quem pertencem os dados? Haverá categorias de dados que nunca deverão ser usados ao analisar-se dados pessoais?

Embora Big Data represente grandes oportunidades de benefício para as empresas, educação, saúde, governo e muitas outras áreas, os riscos, contudo, para a privacidade pessoal, a capacidade de gerir a nossa reputação individual e identidade online e o que pode significar perder – ou ganhar – posse sobre os nossos dados pessoais, gera naturalmente questões éticas. Antecipa-se, mesmo, que surja fricção entre as pessoas (que providenciam os dados) e as empresas que agregam e processam os dados. Por exemplo, considere-se a situação em que uma pessoa tem um acidente e é levada inconsciente para o hospital. Muitos dados são gerados neste processo – assumindo mesmo que serão úteis para desenvolver tratamentos mais efectivos – mas parece óbvio que os dados pertencem àquela pessoa. Foram gerados durante o seu tratamento, mas também com equipamento providenciado pelo hospital, baseado em know-how desenvolvido ao longo de décadas em vários negócios, universidades e instituições governamentais, tudo sempre com o objectivo de salvar vidas. Esses dados, além de irem gerar ganhos e lucros nas empresas do sector da saúde, também irão ajudar a salvar vidas. Trata-se, na realidade de um esforço mútuo. Não é óbvio se os dados pertencem ao doente. Mas também não é óbvio se o hospital pode fazer o que quiser com

aqueles dados. Questões como esta terão de ser profundamente consideradas.

É sabido que tanto as empresas privadas como as organizações governamentais procuram ter maior informação sobre os comportamentos e sentimentos das pessoas [9].

As organizações usam várias técnicas analíticas – desde crowdsourcing até algoritmos genéticos, redes neuronais ou análises de sentimentos – para estudar formas de dados estruturados e não estruturados que possam ajudar a produzir e processar descobertas, produtividade e tomadas de decisão.

Desde a introdução da Internet que nos temos movido de comunicações com base em texto para dados de formato mais rico; que inclui imagens, vídeos e mapas interactivos bem como metadados associados tais como informação geolocalizada [9]. Entre os vários tipos de dados usados no processamento de Big Data e atrás referidos, há dados considerados sensíveis (por exemplo, que revelam informação sobre a origem racial de uma pessoa, sua opinião política, religiosa ou de outras crenças, dados sobre saúde ou vida sexual, dados genéticos e biométricos, condenações ou outras medidas penais), muitos destes dados originados em acontecimentos aparentemente inócuos como, por exemplo, nos tipos de filmes ou outros programas de entretenimento que selecciona [12].

Os avanços no armazenamento e pesquisa de dados e nas tecnologias de mineração de dados tornam possível preservar crescentes quantidades de dados gerados directa ou indirectamente pelos utilizadores e proceder à sua análise para obter novas perspectivas muito valiosas. Por exemplo, as empresas podem estudar as tendências de compra dos consumidores para melhor direccionar as suas campanhas de marketing. Os dados em tempo real obtidos a partir dos telefones móveis podem providenciar características detalhadas sobre os compradores que se movem num centro comercial, revelando os seus próprios processos complexos de tomada de decisão de compra.

Casos como o de Snowden que desvendou como os governos monitorizam as populações, recolhendo dados de todas as fontes já referidas, a que se junta o que os utilizadores publicam nas redes sociais – Facebook ou Twitter – ou quando usam os seus telefones, poderão ser dados utilizáveis para a “segurança nacional” e fins políticos [10].

Big Data também apresenta novos desafios éticos. As instituições podem usar a tecnologia para saber mais sobre os seus funcionários, aumentar a produtividade e introduzir processos revolucionários de negócio. Contudo, estes melhoramentos têm um custo: monitorizar os movimentos de cada funcionário medindo o seu desempenho, podendo ser feito no melhor interesse da empresa não será, necessariamente, é no melhor interesse das pessoas que fazem parte dessa instituição [9].

A estrutura sugerida por Davis e Patterson [11] referindo quatro questões - identidade, privacidade, posse e reputação - apresenta um importante primeiro passo para discussão das questões éticas que se colocam com a utilização dos dados pessoais.

4. Conclusões

A questão abordada neste artigo divide-se em duas partes: uma, a descrição do que é Big Data, as características técnicas e seus espectáveis novos desenvolvimentos e em que se analisou a função predominantemente crescente nas acções de informação e orientação em praticamente todos os sectores empresariais e governamentais; a segunda, em que se abordaram as práticas no que se refere às novas oportunidades e desafios que se colocam.

Necessitamos de melhor discernimento para melhor perceber as grandes oportunidades e riscos de Big Data. Sem dúvida que se tornará uma parte integrante da nossa sociedade, com utilização tanto nas aplicações comerciais como governamentais. O nosso desafio será maximizar os seus benefícios e minimizar os seus danos.

Teremos de pensar em que tipo de sociedade queremos viver e enfrentar as questões que a utilização de Big Data coloca no nosso caminho.

Referências

- [1] Broussard, G. (2014) "A Primer for Defining and Implementing Big Data in the Marketing and Advertising Industry", www.researchexcellence.com
- [2] Zikopoulos, P., Eaton, C., Deroos, D., Deutsch, T. & Lapis, G. (2012) "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data", McGraw-Hill
- [3] Chen, H., Chiang, R., Storey, V. (2012) "Business Intelligence and Analytics: From BIG DATA to BIG IMPACT, MIS Quarterly Vol. 36 No. 4, pp. 1165-1188/ December 2012
- [4] Mujeeb, S., Naidua, L. (2015) "Relative Study on Big Data Applications and Techniques", International Journal of Engineering and Innovative Technology (IJEIT), Volume 4, Issue 10, April 2015
- [5] Chen, C.L., Zhang, C. (2014) "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", Information Sciences, www.elsevier.com/locate/ins, January 2014.
- [6] J. Han, E. Haihong, G. Le, and J. Du (2011) "Survey on nosql database", Pervasive Computing and Applications (ICPCA), 2011 6th International Conference on. IEEE, 2011, pp. 363-366.
- [6] Gajendran, S. K., "A Survey on NoSQL Databases"
- [7] Strozzi, C. (2007 - 2010) "NoSQL - A relational database management system, http://www.strozzi.it/cgi-bin/CSA/tw7/I/en_US/nosql/Home%20Page
- [8] Kanwar, R., Trivedi, P., Singh, K. (2013) "NoSQL, a Solution for Distributed Database Management System", International Journal of Computer Applications (0975 - 8887), Volume 67-No.2, April 2013

- [9] Michael, H., Miller K., (2013) "Big Data: New Opportunities and New Challenges", IEEE Computer Society, 0018-9162/13
- [10] Lyon, D., (2014) "Surveillance, Snowden and Big Data: Capacities, Consequences, critique", Big Data & Society, july-september 2014: I-13
- [11] Davis K., Patterson D., Ethics of Big Data: Balancing Risk and Innovation, = Reilly, 2012
- [12] Rouvroy, A., Council of Europe, "Of Data and Men" Fundamental Rights and Freedoms in a World of Big Data, Bureau of the Consultive Committee of the Convention for The Protection of Individuals with Regard to Automatic Processing of Personal Data [ETS 108], [T-PD-BUR], January 2016