

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO (UFRJ)
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS (CCJE)
FACULDADE DE ADMINISTRAÇÃO E CIÊNCIAS CONTÁBEIS (FACC)
CURSO DE BIBLIOTECONOMIA E GESTÃO DE UNIDADE DE INFORMAÇÃO (CBG)

VANESSA ALVES DA SILVA PEREIRA

BIG DATA: UM ESTUDO EM GESTÃO EMPRESARIAL

Rio de Janeiro

2016

VANESSA ALVES DA SILVA PEREIRA

BIG DATA: UM ESTUDO EM GESTÃO EMPRESARIAL

Trabalho de Conclusão de Curso apresentado ao Curso de Biblioteconomia e Gestão de Unidades de Informação da Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de bacharel em Biblioteconomia.

Orientadora: Profa. Dra. Maria Irene da Fonseca e Sá

Rio de Janeiro

2016

Ficha catalográfica

P436v Pereira, Vanessa Alves da Silva.
Big data: um estudo em gestão empresarial / Vanessa Alves da Silva
Pereira. - Rio de Janeiro, 2015.

81f : il.

Orientadora: Maria Irene da Fonseca e Sá.
Monografia (Trabalho de Conclusão de Curso) – Curso de
Biblioteconomia e Gestão de Unidades de Informação, Universidade
Federal do Rio de Janeiro.

1. Big Data. 2. Gestão empresarial. 3. Vantagem competitiva. 4.
Estratégia. 5. Processo decisório. 6. Privacidade. 7. Direito à informação. I.
Sá, Maria Irene da Fonseca e II. Título.

CDD: 658.05

VANESSA ALVES DA SILVA PEREIRA

BIG DATA: UM ESTUDO EM GESTÃO EMPRESARIAL

Trabalho de Conclusão de Curso apresentado ao Curso de Biblioteconomia e Gestão de Unidades de Informação da Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de bacharel em Biblioteconomia.

Rio de Janeiro, 07 de março de 2016.

Profa. Dra. Maria Irene da Fonseca e Sá - UFRJ
Orientadora

Profa. Me. Maria José Veloso da Costa Santos - UFRJ
Membro interno

Prof. Me. Robson Santos Costa - UFRJ
Membro interno

Dedico este trabalho ao meu maior herói, meu pai Volnandes, que se foi e levou consigo uma parte de mim.

AGRADECIMENTOS

Agradeço primeiramente por todas as oportunidades que tive nestes cinco anos de faculdade, usufruídas de forma plena: extensão na Escola de Serviço Social (ESS), Empresa Júnior Ayra Consultoria, estágio na biblioteca Eugênio Gudín do Centro de Ciências Jurídicas e Econômicas (CCJE), estágio na Agência Nacional de Petróleo (ANP) e intercâmbio na École Supérieure de Commerce (ESC) em Rennes, na França, que mudou minha vida para sempre. Tudo isso foi possível graças a Universidade Federal do Rio de Janeiro (UFRJ), uma das mais renomadas instituições do país.

Ao corpo de professores do Curso de Biblioteconomia e Gestão de Unidades de Informação (CBG), cujos ensinamentos eu levarei para a vida.

À banca examinadora, Maria José Veloso da Costa Santos (Mazé) e Robson Santos Costa, vocês são grandes exemplos pra mim e é uma honra ser avaliada por vocês.

À minha orientadora Maria Irene da Fonseca e Sá, que me orienta há quase dois anos, desde que enviei meu primeiro e-mail, ainda na França, pedindo ajuda para escolher um tema e que sempre esteve disposta a me ouvir e ajudar, mesmo nos meus momentos de crise. Este trabalho não seria possível sem você.

À Regina Marteleto, minha orientadora de Iniciação Científica, que me apresentou à Ciência da Informação e me permitiu realizar pesquisas de campo riquíssimas.

À Ayra Consultoria, pelos ensinamentos infinitos, noites em claro e dedicação extrema, que me fizeram crescer e aprender como nunca. E a todos os amigos que fiz nos dois anos de muito trabalho na empresa.

À ESC Rennes Business School, por ter me acolhido durante um ano e pela experiência incrível de fazer amigos do mundo todo, aprendendo a respeitar culturas muito diferentes da minha.

Ao Ágora, por todas as noites que passamos em claro fazendo trabalhos no Facebook e principalmente por todos os momentos divertidos e risadas dos últimos dois anos.

Às CBGGirls por serem o maior presente que a UFRJ me deu. À Rebecca, por ser um grande exemplo pra mim, de dedicação e pé no chão; à Isabelle, que é minha “mãe mais nova”, sempre com conselhos pertinentes e palavras fofas e a Bárbara, minha melhor amiga, que esteve comigo na Ayra, que morou comigo na França e que simplesmente fez toda a diferença na minha vida. Vocês, eu levo pra sempre.

Aos meus amigos, que tiveram toda paciência do mundo, mesmo eu sumindo por quase cinco anos.

Ao meu irmão Volnandes, minha cunhada Vanessa, meu afilhado Lucas e meu sobrinho Miguel, por me ensinarem a ter uma família um pouquinho maior e encherem de alegria minha vida.

À minha avó Alzira (in memoriam), que se foi há muitos anos, mas que está presente no meu coração todos os dias.

À Alban, l'amour de ma vie qui a tout changé. Tu es ma force et si j'ai fini la fac c'est grâce à toi. Merci beaucoup. Tu n'imagines pas à quel point tu es important pour moi. Je t'aime à la folie et pour toujours.

À minha mãe, Margaret, que sempre foi meu maior exemplo. É a pessoa mais forte que conheço. Obrigada por sempre acreditar em mim, mesmo nas minhas ideias mais loucas. Tudo que faço e todo meu esforço é sempre pensando em você e meu pai.

Ao meu pai Volnandes (in memoriam) com quem eu tive a oportunidade de discutir sobre o início deste trabalho, mas que se foi antes de ver sua conclusão. Você está na minha mente em todos os momentos e tudo que eu faço é pensando na sua aprovação. Você se foi, mas transformou-se na minha maior força. A saudade é grande e eterna, mas nada, nem o tempo, irá apagar todos os momentos incríveis que tivemos e tudo que você representa pra mim. Descanse em paz. Eu te amo.

“A tecnologia não é boa, nem má, e também não é neutra.” (KRANZBERG, 1986, p. 544).

RESUMO

Big Data é um novo conceito que chega para revolucionar a relação das empresas com seus consumidores. A produção de dados pela sociedade aumenta em níveis exponenciais, chegando atualmente a um *zettabyte* de dados produzidos diariamente, a grande maioria através de e-mails, em redes sociais - como Twitter, Facebook, LinkedIn -, através do *Global Positioning System* (GPS) dos milhões de aparelhos celulares e da Internet das Coisas, em que as máquinas estão conectadas à Rede produzindo milhares de *bytes* de dados a cada segundo. Analisar, cruzar e produzir informação a partir destes dados não-estruturados, que apresentam-se em volumes exacerbados, velocidade em tempo real e em diversos formatos não é tarefa fácil, porém algumas empresas já estão se aventurando no universo do Big Data para entender melhor seus clientes. O presente trabalho explora através de múltiplos estudos de caso, como cinco diferentes empresas - New South Wales Emergency Services, na Austrália; Nielsen Holdings, nos Estados Unidos; Ramco Cements Limited, na Índia, Dunnhumby, na Inglaterra e Netflix, Inc. nos Estados Unidos - utilizaram o Big Data para revolucionar seu posicionamento de mercado. Desenvolveu-se uma pesquisa bibliográfica exploratória qualitativa. A pesquisa buscou entender com que objetivos as aplicações Big Data estão sendo implementadas nas empresas e os resultados mostraram que, apesar de embrionários, as organizações estudadas já estão se beneficiando com projetos de Big Data para melhorar sua estratégia, ganhar vantagem competitiva sustentável e aprimorar seu processo decisório. No entanto, é preciso ter atenção sobre a questão da privacidade e do direito à informação, pois a tendência atual é que com o Big Data as fronteiras existentes entre os dados públicos e privados diminuam consideravelmente. As empresas que souberem dominar a eminente avalanche de dados que produzimos serão aquelas que dominarão o mercado empresarial no futuro.

Palavras-chave: Big Data. Gestão empresarial. Vantagem competitiva. Estratégia. Processo decisório. Privacidade. Direito à informação. Estudo de caso.

ABSTRACT

Big Data is a new concept that emerges to provoke a revolution in the relationship between companies and their consumers. The data produced by society increases in exponential levels, achieving one zettabyte of data produced daily, the great majority through e-mails, in social networks such as Twitter, Facebook and LinkedIn, through the Global Positioning Systems (GPS) of our smartphones and the Internet of Things, in which machines are constantly connect to the Web producing thousands of bytes of data at every second. Analyzing, data mining and producing information through this non-structured data that appears in enormous volume, real-time velocity and in a variety of formats is not an easy task, nonetheless a few companies had started to enter the world of Big Data to better understand its clients. The present paper explores through multiples case studies how five different companies - New South Wales Emergency Services, in Australia; Nielsen Holdings, in The United States; Ramco Cements Limited, in India, Dunnhumby, in England and Netflix, Inc. in The United States – have explored Big Data to increase their market share. It was developed a qualitative bibliographic research that aimed to understand which are the goals of the companies implementing Big Data projects and the results has shown that even though still in its early state, the firms studied are already harvesting the benefits of Big Data projects to develop their strategies, gain sustainable competitive advantage and to improve their decision making process. However, privacy and the right to information are very important subjects to be considered since the frontiers between public and private data are slowly decreasing. The companies who manage to dominate the eminent flow of data that we produce every day are the ones who will surely dominate the market in the near future.

Keywords: Big Data. Business management. Competitive advantage. Strategy. Decision making process. Privacy. Right to information. Case study.

LISTA DE ILUSTRAÇÕES

| | | |
|-------------------|---|----|
| Quadro 1 - | Dado x Informação x Conhecimento..... | 17 |
| Figura 1 - | Quantidade de dados produzidos em 24 horas..... | 20 |
| Quadro 2 - | Quantidade de dados produzidos na Internet em tempo real..... | 20 |
| Quadro 3 - | Quadro de conceitos Big Data..... | 21 |
| Figura 2 - | Os 5 Vs do Big Data..... | 23 |
| Quadro 4 - | Um cardápio de possibilidades de Big Data..... | 26 |
| Quadro 5 - | Terminologia para o uso e a análise de dados..... | 26 |
| Quadro 6 - | Visão geral das tecnologias do Big Data..... | 29 |
| Figura 3 - | A Internet das coisas surge entre 2008 e 2009..... | 32 |
| Figura 4 - | Familiaridade com o conceito de Big Data..... | 39 |
| Figura 5 - | Uso atual e estratégia para Big Data..... | 40 |
| Figura 6 - | Atitudes das organizações com relação ao Big Data..... | 40 |
| Figura 7 - | Ramco GeoApps..... | 58 |
| Figura 8 - | Ramco Perfmon..... | 59 |
| Figura 9 - | Ramco APO..... | 60 |
| Quadro 7 - | Resumo dos estudos de caso..... | 73 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|---------|---|
| AWS | Amazon Web Service |
| CBG | Curso de Biblioteconomia e Gestão de Unidades de Informação |
| DVD | Digital Versatile Disc |
| ERP | Enterprise Resource Planning |
| EUA | Estados Unidos da América |
| GE | General Electric |
| GPS | Global Positioning System |
| HDFS | Hadoop Distributed File System |
| HMR | Hadoop MapReduce |
| HTML | HyperText Markup Language |
| IBOPE | Instituto Brasileiro de Opinião Pública e Estatística |
| IDC | International Data Corporation |
| ISACA | Information Systems Audit and Control Association |
| KPI | Key Performance Indicators |
| LG | Lucky Goldstar |
| MIT | Massachussets Institute of Technology |
| NO SQL | Not Only Structured Query Language |
| NSW SES | New South Wales State Emergency Service |
| OLAP | OnLine Analytical Processing |
| PNL | Processamento de Linguagem Natural |
| RCL | Ramco Cements Limited |
| SVG | Scalable Vector Graphics |
| TI | Tecnologia da Informação |
| TIC | Tecnologia da Informação e Comunicação |
| UFRJ | Universidade Federal do Rio de Janeiro |

SUMÁRIO

| | | |
|--------------|---|-----------|
| 1 | INTRODUÇÃO..... | 10 |
| 1.1 | OBJETIVOS..... | 12 |
| 1.1.1 | Objetivo geral..... | 12 |
| 1.1.2 | Objetivos específicos..... | 12 |
| 1.2 | JUSTIFICATIVA..... | 12 |
| 2 | PROCEDIMENTOS METODOLÓGICOS..... | 13 |
| 3 | REFERENCIAL TEÓRICO..... | 15 |
| 3.1 | DEFININDO BIG DATA..... | 15 |
| 3.1.1 | O que é Big Data?..... | 16 |
| 3.1.2 | Tecnologias para o Big Data..... | 28 |
| 3.1.3 | Internet das Coisas..... | 31 |
| 3.1.4 | Privacidade e direito à informação..... | 34 |
| 3.2 | BIG DATA NO AMBIENTE EMPRESARIAL..... | 36 |
| 3.2.1 | Big Data e estratégia..... | 36 |
| 3.2.2 | Big Data e vantagem competitiva..... | 41 |
| 3.2.3 | Big Data e processo decisório..... | 43 |
| 4 | ESTUDOS DE CASO..... | 45 |
| 4.1 | CASO NEW SOUTH WALES STATE EMERGENCY SERVICE..... | 45 |
| 4.2 | CASO NIELSEN HOLDINGS..... | 47 |
| 4.3 | CASO RAMCO CEMENTS..... | 54 |
| 4.4 | CASO DUNNHUMBY..... | 62 |
| 4.5 | CASO NETFLIX..... | 67 |
| 4.6 | RESUMO DOS ESTUDOS DE CASO..... | 73 |
| 5 | CONSIDERAÇÕES FINAIS..... | 75 |
| | REFERÊNCIAS..... | 77 |

1 INTRODUÇÃO

As tecnologias vêm revolucionando a maneira como as pessoas lidam e interagem com o mundo a seu redor de forma cada vez mais drástica. Desde o surgimento da Internet até sua popularização, o ser humano depende cada vez mais das máquinas para realizar suas tarefas diárias. Com a criação dos *smartphones* passa-se a estar conectados à Internet 24 horas por dia, 7 dias por semana. Com a Internet das Coisas, muito em breve, vários objetos de nossas casas e trabalho estarão, também, inteiramente conectados. Todos esses aparatos rastreiam cada passo do que é feito na *Web* e cada um desses passos é um potencial gerador de dados. Como passa-se cada vez mais tempo *online*, gera-se progressivamente mais dados, de forma mais rápida. Qualquer tarefa realizada quando se está conectado, desde enviar um e-mail, escrever uma mensagem em uma rede social ou até mesmo deixar o GPS do celular ligado enquanto se movimenta, deixa rastros imperceptíveis que transformam-se em dados.

Gera-se atualmente aproximadamente 1 *zettabyte* de dados por dia mas, até então, a imensa maioria desse volume era ignorada. Porém, percebeu-se em determinado momento que esses dados podem ser valiosíssimos quando bem aproveitados, principalmente para as empresas, que podem obter informações em tempo real, geradas diretamente por seus clientes. Para o governo, significam informações sobre como a sociedade se comporta e formas mais fáceis de monitorar a população. Por outro lado, cria-se uma fina barreira entre os benefícios da exploração de dados e a invasão de privacidade. Neste mundo globalizado e constantemente conectado, surge então uma nova fonte de poder que não pode mais ser ignorada. Os especialistas a chamam de Big Data.

O imenso volume de dados gerados diariamente é impressionante. Esses dados provêm de diversas fontes, como dos 492.480.000 *tuítes* (comentários feitos no Twitter), 199.929.600 horas de vídeos assistidas no Youtube, 398.131.200 buscas no Google, 4.509.734.400 curtidas (forma do internauta dizer que gostou de uma postagem) no Facebook, 59.961.600 fotos postadas no Instagram e 2.940.000.192 e-mails enviados. Esses números geram mais de 1 *zettabyte* de dados todos os dias (PENNY STOCKS LABS, 2015). A grande questão a ser respondida é de que maneira estes dados podem ser utilizados de forma vantajosa pelas empresas. No entanto, sequer existe um consenso sobre o conceito de Big Data e quais tecnologias o sustentam. As empresas ainda possuem muitas dúvidas sobre como tornar o conceito aplicável, trazendo soluções que agreguem valor. Criar estratégias que incorporem o Big Data é a primeira solução a ser pensada pelas organizações. (TAURION, 2013b).

O presente trabalho expõe e discute diversas conceituações sobre o termo Big Data, de forma a entender como essas definições podem ser utilizadas pelas empresas. Apesar das variações relacionadas a definição do que é Big Data, para fins de análise, neste trabalho será adotado o conceito definido por César Taurion (2013a) que utiliza cinco fatores: volume, variedade, velocidade, veracidade e valor.

Big Data não trata apenas da dimensão volume, como parece à primeira vista, mas existe também uma variedade imensa de dados, não estruturados, dentro e fora das empresas (coletados das mídias sociais, por exemplo), que precisam ser validados (terem veracidade para serem usados) e tratados em velocidade adequada para terem valor para o negócio. A fórmula é então, Big Data = volume + variedade + velocidade + veracidade, gerando valor. (TAURION, 2013a, não paginado).

Tendo em vista as incipientes discussões sobre os benefícios que a utilização do Big Data pode trazer às empresas, o problema de pesquisa do presente trabalho é: Com que objetivos as aplicações Big Data estão sendo implementadas nas empresas? Para responder a esta pergunta, foram analisados estudos de caso em cinco empresas diferentes: New South Wales Emergency Services, na Austrália; Nielsen Holdings, nos Estados Unidos; Ramco Cements Limited, na Índia, Dunhumby, na Inglaterra e Netflix, Inc. nos Estados Unidos.

Após breve introdução para contextualizar o assunto e apresentar o problema de pesquisa, são apresentados o objetivo geral e os objetivos específicos do trabalho, além da justificativa para a escolha do tema. O segundo capítulo refere-se aos procedimentos metodológicos utilizados. O terceiro capítulo apresenta o referencial teórico da pesquisa. Em um primeiro momento, propõe-se uma discussão sobre a definição de Big Data. São apresentados e analisados conceitos para entender o que é Big Data, em seguida, discute-se sobre as tecnologias mais utilizadas em projetos de Big Data. É feita uma introdução sobre o conceito de Internet das Coisas, em que aparelhos e máquinas estão constantemente conectados à Internet e conclui-se falando sobre privacidade e direito à informação, tema atualmente em voga tendo em vista as problemáticas que o envolvem e seu caráter atual. O terceiro capítulo pretende relacionar Big Data com o ambiente empresarial. São discutidas, portanto, as relações entre o Big Data e estratégia, vantagem competitiva e processo decisório. O quarto capítulo apresenta os cinco estudos de caso analisados nesta pesquisa, seus resultados e a conexão dos mesmos com referencial teórico previamente estudado. Por fim, apresentam-se as considerações finais, em que o problema de pesquisa é respondido e novas indagações são feitas visando pesquisas futuras.

1.1 OBJETIVOS

Apresenta-se a seguir o objetivo geral e os objetivos específicos do presente trabalho.

1.1.1 Objetivo geral

Identificar a utilização de aplicações Big Data em gestão empresarial.

1.1.2 Objetivos específicos

- a) Conceituar Big Data, de modo a dar subsídios ao desenvolvimento da pesquisa.
- b) Levantar estudos de caso de empresas que implementaram ou planejam implementar projetos de Big Data.
- c) Apresentar as soluções estratégicas utilizadas nos casos estudados.

1.2 JUSTIFICATIVA

Big Data é hoje um tema bastante discutido nas empresas, porém, apesar de seu potencial já ser reconhecido, somente os primeiros passos para explorar este ambicioso conceito foram dados. Ainda há muito o que entender antes que o Big Data seja totalmente incorporado na maioria dos negócios e somente algumas empresas pioneiras se arriscaram e apostaram em explorá-lo. No campo das pesquisas, as discussões também ainda estão em fases preliminares em diversas áreas, como a Ciência da Computação, a Administração, a Ciência da Informação e a Biblioteconomia.

Ainda não foi criada uma profissão específica para o profissional especializado em Big Data, no entanto, algumas características desejáveis para estes profissionais já foram elencadas como “[...] uso básico do computador como ferramenta para busca, avaliação e classificação de informações e cultura de internet” (MACHADO; PALACIOS, 2007, p.79 apud CABRAL; SAID, 2014, não paginado). Cabral e Said (2014) complementam ainda citando o conhecimento teórico sobre redes e seu funcionamento. Todas estas características são inerentes ao bibliotecário, além deste ser o profissional mais bem preparado para extrair

valor dos dados, indexando-os e transformando-os em informação. Possui também as habilidades e conhecimento necessários para realizar o arquivamento dos dados estruturados e não-estruturados de forma a serem recuperados futuramente. (PEREIRA, 2013 apud SILVA, 2014, p.20).

Taurion (2014a) afirma que os cientistas de dados devem conhecer profundamente o negócio e, mais importante, devem saber fazer as perguntas certas. Não necessariamente estes profissionais serão encontrados somente no setor de Tecnologia da Informação. Dificilmente encontraremos uma pessoa que possua todas as características necessárias para trabalhar com o big data, sendo assim, o ideal é que exista uma equipe multidisciplinar, de forma a aproveitar e analisar todo o imenso volume de dados. (DAVENPORT, 2014, p.85).

“Vivemos hoje na sociedade da informação e informação é um produto por si mesmo, além de ser o combustível que impulsiona os negócios da maioria das empresas” (TAURION, 2013a, não paginado). Dessa forma, afirma-se a importância do profissional bibliotecário no contexto do Big Data e justifica-se a necessidade iminente de realizar pesquisas sobre o tema na área de Biblioteconomia.

A justificativa pessoal para a escolha do tema envolve o fato do Curso de Biblioteconomia e Gestão de Unidades de Informação (CBG) oferecido pela Universidade Federal do Rio de Janeiro (UFRJ) focar em três áreas: 37% em disciplinas de Biblioteconomia, 30% em disciplinas de Gestão e 30% em outras disciplinas de cultura geral, entre elas a Tecnologia da Informação e Comunicação (TIC). Falar sobre Big Data é relacionar as três áreas de estudo propostas pelo curso e aproveitar de maneira ampla a multidisciplinaridade que o estudo de Biblioteconomia proporciona.

2 PROCEDIMENTOS METODOLÓGICOS

Para o desenvolvimento da presente pesquisa, seguiu-se as orientações de Létourneau (2011) que afirma ser necessário, em um primeiro momento, delimitar o assunto, sendo este preciso e operacional, seguido de uma estratégia de pesquisa, ou seja, a formulação do problema. O assunto definido foi Big Data, delimitando-o às suas implicações no mundo empresarial. O objetivo geral do trabalho é entender como as aplicações Big Data estão sendo implementadas nas empresas.

Trata-se de uma pesquisa qualitativa, que segundo Minayo (1994, p.21-22) “[...] se preocupa, nas ciências sociais, com um nível de realidade que não pode ser quantificado. Ou seja, ela trabalha com o universo de significados, aspirações, crenças, valores e atitudes [...]

que não podem ser reduzidos à operacionalização de variáveis”. Westbrook (1994 apud BAPTISTA;CUNHA, 2007, p.173) afirma que na pesquisa qualitativa, deve-se realizar movimentos reiterados e cíclicos entre a coleta de dados e sua análise e diz que “[...] essa coleta de dados é vista mais como um processo do que um procedimento, requerendo constantes julgamentos analíticos”. A abordagem qualitativa da pesquisa é tida como a forma adequada de entender fenômenos sociais (RICHARDSON et al., 2012, p.79), sendo esta a principal justificativa para a escolha da utilização da mesma no presente trabalho.

Gil (2008, p.50) afirma que quase todas as pesquisas deparam-se em determinado momento com a pesquisa bibliográfica, porém este trabalho será desenvolvido baseando-se exclusivamente em fontes bibliográficas. Segundo o autor (2008, p.50), “A pesquisa bibliográfica é desenvolvida a partir de material já elaborado, constituído principalmente de livros e artigos científicos”. O lócus de pesquisa foram os livros, artigos científicos e notícias envolvendo o assunto em foco. Foi realizada pesquisa exploratória utilizando o Google e bases de dados como o Portal Capes, o banco de dados da Scielo e da BRAPCI, utilizando as palavras-chave: Big Data, gestão empresarial, internet das coisas, privacidade e direito à informação, processo decisório, vantagem competitiva e estudo de caso. Gil (2008, p. 27) afirma que as pesquisas exploratórias objetivam “[...] desenvolver, esclarecer e modificar conceitos e idéias, tendo em vista a formulação de problemas mais precisos ou hipóteses pesquisáveis para estudos posteriores”. A população do trabalho em questão são os livros, artigos científicos e notícias encontrados no Google e nas bases de dados pesquisadas utilizando as palavras-chave acima descritas, nos idiomas português, inglês e francês. A amostra foi selecionada de acordo com o grau de interesse gerado pela leitura dos resumos apresentados no material encontrado.

Estudos de caso são definidos por Gil (2008, p.57-58) como “[...] um estudo profundo e exaustivo de um ou de poucos objetos, de maneira a permitir o seu conhecimento amplo e detalhado [...]”. Os estudos de caso são interessantes principalmente em pesquisas exploratórias, quando o assunto em questão ainda é embrionário, sem padrões definidos e cujos fenômenos possuem uma grande variedade de fatores que podem ser observados diretamente. Por seu caráter flexível, os estudos de caso estimulam descobertas e “[...] enfatizam a multiplicidade de dimensões de um problema, focalizando-o como um todo e apresentam simplicidade nos procedimentos, além de permitir uma análise em profundidade dos processos e das relações entre eles.” (VENTURA, 2007, p.386).

Em um primeiro momento, realizou-se extensa pesquisa em fontes como livros, artigos e notícias sobre o assunto para delinear o referencial teórico que serviria de base para

o trabalho. Em seguida, foram selecionados os cinco estudos de caso a serem analisados, de acordo com sua pertinência para entender os objetivos das empresas ao implementarem projetos de Big Data e os resultados obtidos à partir da realização destes projetos. Todos os estudos de caso utilizados encontravam-se em língua estrangeira, tendo sido minuciosamente traduzidos pela autora para uma profunda compreensão do contexto e de seus resultados. Foi feita, em seguida, uma análise detalhada dos procedimentos e resultados encontrados pelas empresas, posteriormente expostos no presente trabalho e resumidos em um quadro para melhor compreensão. Por fim, discutiu-se a relação entre os resultados obtidos pelas empresas com o referencial teórico estudado, concluindo com a resposta ao problema da pesquisa.

3 REFERENCIAL TEÓRICO

O presente referencial teórico pretende expor e discutir, baseando-se na literatura estudada, os conceitos de Big Data e seus impactos no ambiente empresarial.

3.1 DEFININDO BIG DATA

“[...] o big data pode ser descrito como um rio de informações em correnteza que nunca para [...]” (DAVENPORT, 2014, p.114). A partir desta afirmação, a próxima seção dedica-se a explicar o que é o Big Data, seus conceitos e expor um panorama geral do crescente número de dados produzidos pela sociedade atual.

Em seguida, disserta-se sobre as tecnologias de Big Data, que surgem com o propósito de analisar os dados não-estruturados e em grandes volumes, diferenciando-se dos modelos anteriores que se deparavam com dados estruturados e internos das organizações.

Discute-se na sequência sobre a Internet das Coisas, conceito em que objetos, e até mesmo pessoas e animais, estão conectados à Internet gerando soluções mais precisas para problemas cotidianos, além de otimizar as tarefas do dia-a-dia.

Para finalizar o capítulo, são levantadas questões relacionadas à privacidade e ao direito à informação, um dos pontos mais polêmicos sobre o Big Data, que gera inúmeras discussões e nos mostra um futuro ainda incerto, que possivelmente trará novos paradigmas ao conceito de privacidade.

3.1.1 O que é Big Data?

Para melhor compreender o significado de Big Data, traduzido literalmente para o português como “Grandes Dados”, é necessário definir o que é Dado. Na Ciência da Informação, a conceituação deste termo está quase sempre ligada aos termos Informação e Conhecimento. Por se tratarem de definições com significados muito similares, não há consonância entre os autores da área para defini-los, considerando o aspecto contextual como essencial para distinção, por vezes tênue, dos termos discutidos. (RUSSO, 2010).

Dado é definido por Setzer (2001, não paginado) como “[...] uma seqüência de símbolos quantificados ou quantificáveis [...] Com essa definição, um dado é necessariamente uma entidade matemática e, desta forma, é puramente sintático”. Já Angeloni (2003, p.18) afirma que “Os dados são elementos brutos, sem significado, desvinculados da realidade.” Davenport (1998, p.18) define dado como uma “Simple observação sobre o estado do mundo”. Sendo ele facilmente estruturado e obtido por máquinas, além de ser quantificado com frequência. O dado é facilmente transferível. Conclui-se à partir das elucidações de Russo (2010, p.15), que “[...] dados são sinais que não foram processados, correlacionados, integrados, avaliados ou interpretados de qualquer forma, e, por sua vez, representam a matéria prima a ser utilizada na produção de informações”.

A informação, segundo Setzer (2001, não paginado) “[...] é uma abstração informal (isto é, não pode ser formalizada através de uma teoria lógica ou matemática), que está na mente de alguém, representando algo significativo para essa pessoa”. Na definição de Angeloni (2003, p.18) “As informações são dados com significado.” A informação é considerada como “[...] dados processados e contextualizados”. Davenport (1998, p.18) diz que informação são “Dados dotados de relevância e propósito”. A informação requer uma unidade de análise e exige consenso com relação ao seu significado e é imprescindível a mediação humana. Para Russo (2010, p.15) “[...] informação pode ser entendida como dados processados e contextualizados”.

Conhecimento é definido por Davenport (1998, p.19 apud RUSSO, 2010, p.17) como “[...] a informação mais valiosa [...] é valiosa precisamente porque alguém deu à informação um contexto, um significado, uma interpretação”. Setzer (2001, não paginado) caracteriza conhecimento como “[...] uma abstração interior, pessoal, de algo que foi experimentado, vivenciado, por alguém”. Angeloni (2003, p.18) define:

O conhecimento pode então ser considerado como a informação processada pelos indivíduos. O valor agregado à informação depende dos conhecimentos anteriores desses indivíduos. Assim sendo, adquirimos conhecimento por meio do uso da informação nas nossas ações. Desta forma, o conhecimento não pode ser desvinculado do indivíduo; ele está estritamente relacionado com a percepção do mesmo, que codifica, decodifica, distorce e usa a informação de acordo com suas características pessoais, ou seja, de acordo com seus modelos mentais. (ANGELONI, 2003, p.18).

O conhecimento de acordo com Davenport (1998, p.18) é uma “Informação valiosa da mente humana. Inclui reflexão, síntese, contexto”. É de difícil estruturação, transferência e captura em máquinas. É frequentemente tácito. O conhecimento tácito é capital dos seres humanos, dificilmente transmitido e capturado, dependendo das experiências e vivências da pessoa (RUSSO, 2010, p.19).

Russo (2010, p.18) distingue os termos analisados como apresentado no quadro a seguir:

Quadro 1 – Dado x Informação x Conhecimento

| Dado | Informação | Conhecimento |
|--|--|--|
| Conjunto de letras, números ou dígitos que não contém significado claro. | Dado trabalhado, útil, tratado, com valor significativo atribuído ou agregado a ele com um sentido natural e lógico. | Informação trabalhada por pessoas e pelos recursos computacionais, possibilitando a geração de cenários, simulações e oportunidades. |

FONTE: Russo (2010, p.18). Quadro produzido pela autora.

Analisando as definições acima, percebe-se que o dado, por si só, não possui significado claro e precisa portanto, ser capturado e analisado, para que se transforme em informação. Uma característica importante a ser analisada sobre os dados é que antes do Big Data, os dados coletados pelas empresas provinham de fontes internas e eram, majoritariamente, dados estruturados, ou seja, “[...] dados formatados em linhas e colunas numéricas organizadas” (DAVENPORT, 2014, p.113). Esse tipo de dado já vem sendo explorado há um tempo pelas organizações, estando presente em bancos de dados, arquivos sequenciais e com relação de importância. (CIO, 2012 apud CANARY, 2013).

No entanto, os dados que as empresas irão trabalhar caso intencionem realizar projetos de Big Data são, em sua maioria, dados não-estruturados provenientes das mais diversas

fontes, como postagens em mídias sociais, vídeos, informações de GPS, fotos, etc. Cada um destes dados possui um formato diferente e todos eles precisam ser lidos e assimilados pelas tecnologias de Big Data.

Segundo uma pesquisa realizada pela International Data Corporation (IDC) em 2011 (apud CANARY, 2013) quase 90% dos dados que provém da Internet são dados não-estruturados e esses números tendem a aumentar. Estes resultados reafirmam a importância das empresas aprenderem a trabalhar com esse tipo de dados.

Os dados não-estruturados podem ser classificados em três tipos, de acordo com a Information Systems Audit and Control Association (ISACA) (2013 apud HENRIQUES;COSTA, 2014, p.5):

- a) Dados voluntários: criados e compartilhados pelos próprios indivíduos, como perfis em redes sociais;
- b) Dados observados: gravações das ações do indivíduo, como dados de localização disponibilizados pelos celulares;
- c) Dados inferidos: dados sobre o indivíduo com base em informações fornecidas pelo mesmo ou observações, como por exemplo, classificações de crédito. (ISACA, 2013 apud HENRIQUES;COSTA, 2014, p.5).

Esses dados provenientes da Internet em números exorbitantes foram os principais responsáveis pelo surgimento do fenômeno do Big Data. Entender o conceito de Big Data não é simplório, tendo em vista que não existe consenso sobre o termo, sobre quais tecnologias o sustentam e sobre como torná-lo tangível, ou seja, como transformar este conceito em soluções de negócio, agregando valor às organizações (TAURION, 2013b). Porém, a compreensão do mesmo faz-se essencial quando se entende a importância da utilização do Big Data como vantagem competitiva.

Segundo Taurion (2013a) entender o Big Data é, primeiramente, pensar em duas palavras-chave: volume e velocidade. A quantidade de dados gerados diariamente é impressionante. Nos últimos anos, produziu-se mais dados que em toda a história da humanidade. (HARVARD BUSINESS REVIEW, 2013).

Estima-se que em 2015 foram gerados 7,9 *zettabytes* (10^{21} bytes) de dados e acredita-se que esses números não parem de crescer. Eric Schmidt, diretor executivo do Google em 2010, afirmou que a sociedade na época produzia em dois dias maior quantidade de dados do que havia produzido desde o surgimento da humanidade até 2003. Isso significa aproximadamente cinco *exabytes* a cada dois dias (TAURION, 2013a). Porém, assume-se que estes números já estejam bastante ultrapassados.

Para melhor compreensão, entende-se que um bit é a abreviação do termo *binary digity* (dígitos binários, em português, tradução nossa) e que representa o “Valor compreendido entre duas possibilidades mutuamente exclusivas”. Isso significa que todas as informações visíveis digitalmente são representadas por apenas dois números, o 0 e o 1. (TENENBAUM et al., 1995, p.2 apud AMARAL;BEZARRA, 2010, p.2).

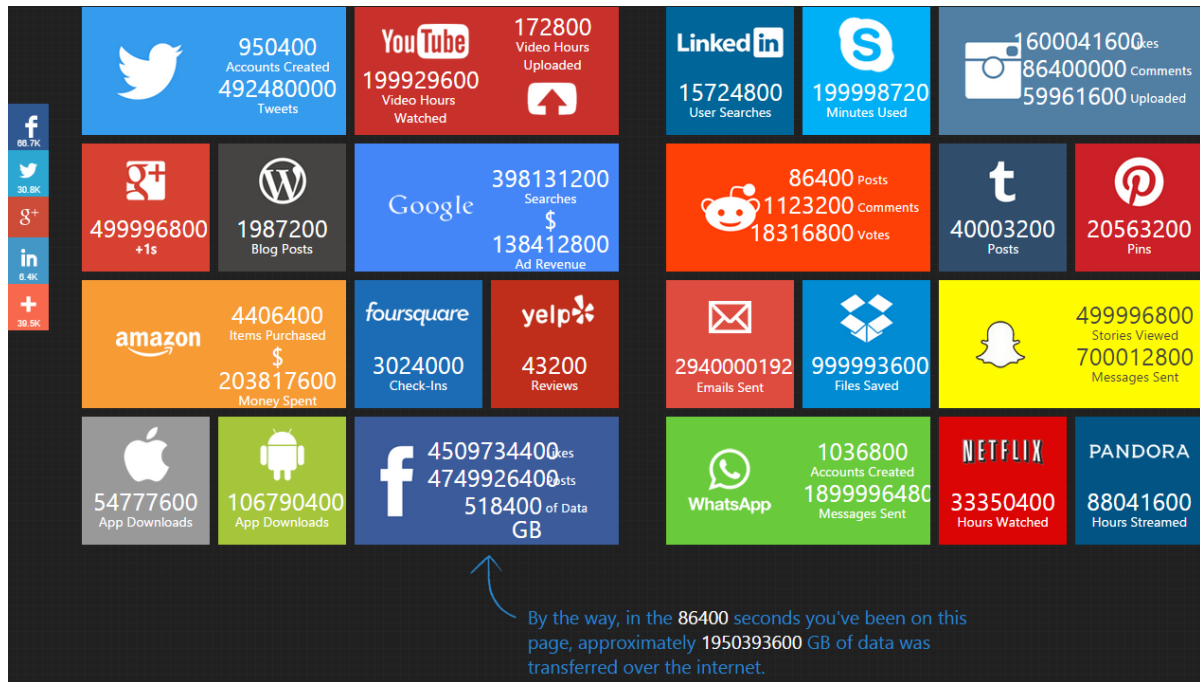
O *byte* é um conjunto de oito bits e, a partir daí, existem várias outras nomenclaturas para determinar múltiplos dos *bytes*. Um *kilobyte* equivale a 1.024 *bytes*, um *megabyte* equivale a 1.024 *kilobytes* e um *gigabyte* equivale a 1.024 *megabytes*. (O QUE..., 2008). No entanto, essas definições não são suficientes quando consideramos a quantidade de dados produzidos atualmente e, assim, fez-se necessário então a criação de novas nomenclaturas como o *terabyte* que equivale a 1.000 *gigabytes*, o *petabyte* que equivale a 1000 *terabytes*, o *exabyte*, que equivale a 1000 *petabytes*. E finalmente o *zettabyte*, que equivale a 1000 *exabytes*. Um *zettabyte* representa em dados o conteúdo equivalente a aproximadamente 250 bilhões de *Digital Versatile Discs* (DVD). (ARTHUR, 2011).

Em 2013, Taurion (2013a) afirmou que os medidores de energia geravam 350 bilhões de medições por ano e apenas uma pequena parcela de casas possuíam esse tipo de medidor. Existiam mais de seiscentos milhões de *websites*, eram escritos mais de cem mil *tuítes* por minutos, em um total de aproximadamente 12 *terabytes* de *tuítes* por dia. Um bilhão de usuários compartilhava algo no Facebook gerando 2,7 bilhões de comentários diariamente. Isso sem contar os dados gerados por sensores e câmeras de segurança pública e que monitoram o trânsito, além dos mais de um bilhão de *smartphones*. (TAURION, 2013a).

A empresa *Penny Stocks Lab* criou um infográfico atualizado em tempo real que nos mostra a quantidade de dados gerados na Internet de maneira geral e por plataforma. Diante destas informações, pode-se visualizar, na imagem abaixo, que em um dia (86.400 segundos) foram produzidos 1.950.393.600 *gigabytes* de dados. O que equivale a aproximadamente 1.959.393 *terabytes*, 1.959 *petabytes* e mais de 1 *zettabyte* de dados em apenas 24 horas. Os dados foram coletados iniciando-se às 16 horas do dia 03 de junho de 2015 até às 16 horas do dia 04 de junho de 2015.

A figura 1, a seguir, mostra a quantidade de dados produzidos em 24 horas na Internet.

Figura 1 – Quantidade de dados produzidos em 24 horas



Fonte: Penny Stocks Labs (2015).

A partir das informações coletadas nesse site, criou-se o quadro a seguir para facilitar a análise da quantidade de dados produzidos por diversas plataformas e mídias sociais nos tempos de um minuto, dez minutos, trinta minutos, uma hora e um dia.

Quadro 2 – Quantidade de dados produzidos na Internet em tempo real

(continua)

| | 1 minuto | 10 minutos | 30 minutos | 1 hora | 1 dia |
|------------------------------|-----------|------------|------------|-------------|---------------|
| Twitter - tuítes | 342.000 | 3.420.000 | 10.260.000 | 20.520.000 | 492.480.000 |
| Youtube - horas assistidas | 138.840 | 1.388.400 | 4.165.200 | 8.330.400 | 199.929.600 |
| Google - buscas | 276.480 | 2.764.800 | 8.294.400 | 16.588.800 | 398.131.200 |
| Amazon - itens comprados | 3.060 | 30.600 | 91.800 | 183.600 | 4.406.400 |
| Facebook - curtidas | 3.131.760 | 31.317.600 | 93.952.800 | 187.905.600 | 4.509.734.400 |
| LinkedIn - busca de usuários | 10.920 | 109.200 | 327.600 | 655.200 | 15.724.800 |
| Skype - minutos utilizados | 1.388.880 | 13.888.800 | 41.666.400 | 83.332.800 | 199.998.720 |
| Instagram - fotos enviadas | 41.640 | 416.400 | 1.249.200 | 2.498.400 | 59.961.600 |

Quadro 2 – Quantidade de dados produzidos na Internet em tempo real

(conclusão)

| | | | | | |
|-------------------------------------|-------------|---------------|---------------|---------------|---------------|
| E-mails - enviados | 204.166.680 | 2.041.666.800 | 6.125.000.400 | 1.225.000.080 | 2.940.000.192 |
| WhatsApp - mensagens enviadas | 13.194.420 | 131.944.200 | 395.832.600 | 791.665.200 | 1.899.996.480 |
| Netflix - horas assistidas | 23.160 | 231.600 | 694.800 | 1.389.600 | 33.350.400 |
| Total de dados gerados em gigabytes | 1.354.440 | 13.544.400 | 40.633.200 | 81.266.400 | 1.950.393.600 |

Fonte: Quadro produzido pela autora com base no site Penny Stock Labs.

Baseando-se nesses números e na diversidade de formatos que a sociedade gera dados atualmente, compreende-se que Big Data refere-se ao volume e a variedade de dados. No entanto, existem outras características que necessitam de análise.

Taurion (2013a, não paginado) diz que “Big Data significa coisas diferentes para pessoas diferentes”. Considerando que ainda não há um consenso sobre a definição de Big Data, apresenta-se, a seguir, um quadro com alguns conceitos levantados de diversos autores e instituições.

Quadro 3 – Quadro de conceitos Big Data

(continua)

| | |
|---------------------------------------|---|
| McKinsey Global Institute (2011, p.1) | <i>“Big data” refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. This definition is intentionally subjective and incorporates a moving definition of how big a dataset needs to be in order to be considered big data [...].¹</i> |
| ISACA (2013 p.5) | Big data refere-se, principalmente, aos conjuntos de dados que são muito grandes ou com rápidas mudanças para serem analisados com técnicas de banco de dados relacionais tradicionais ou multidimensionais ou ferramentas de software comumente usadas para capturar, gerenciar e processar os dados em um tempo razoável. |

¹ “Big Data” refere-se a um conjunto de dados cujo tamanho é superior a habilidade de ferramentas de software de bancos de dados típicos capturarem, estocarem, gerenciarem e analisarem. Essa definição é intencionalmente subjetiva e incorpora uma definição variável do tamanho que um conjunto de dados precisa ter para ser considerado big data [...]. (MCKINSEY GLOBAL INSTITUTE, 2011, p.1, tradução nossa)

Quadro 3 – Quadro de conceitos Big Data

(conclusão)

| | |
|--|--|
| Magalhães et al. ([2014?], não paginado) | [...] Big Data é uma grande massa de dados/metadados aos quais geramos todos os dias, com características estruturadas (armazenados em banco de dados) e não estruturadas (fotos, vídeos, e-mails) e que, na maioria das vezes, é analisada para atender a eventos em tempo real, buscando a partir da autenticidade dos dados dar sentido as informações relevantes passíveis de agregar valor tanto para empresas que buscam estratégias para seu negócio como para governos que buscam entender as demandas e características da população. |
| Dumbill (2012, não paginado) | <i>Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the structures of your database architectures.</i> ² |
| Breternitz e Silva (2013, p. 107). | [...] um conjunto de tendências tecnológicas que permite uma nova abordagem para o tratamento e entendimento de grandes conjuntos de dados para fins de tomada de decisões. |
| Gartner ([201-], não paginado) | <i>Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.</i> ³ |
| Davenport (2014, p.1) | [...] big data é um termo genérico para dados que não podem ser contidos nos repositórios usuais; refere-se a dados volumosos demais para caber em um único servidor; não estruturados demais para se adequar a um banco de dados organizado em linhas e colunas; ou fluidos demais para serem armazenados em um data warehouse estático. Embora o termo enfatize seu tamanho, o aspecto mais complicado do big data, na verdade, envolve sua falta de estrutura. |

FONTE: Quadro produzido pela autora.

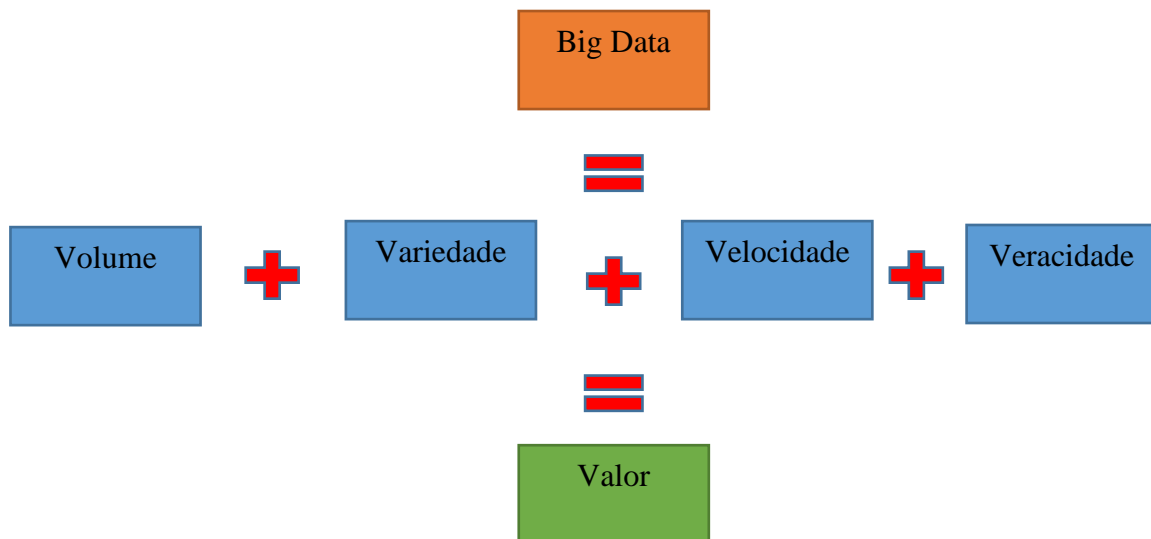
² Big data são dados que excedem a capacidade de processamento de sistemas de bases de dados convencionais. Os dados são muito grandes, movem-se muito rápido ou não se enquadram nas estruturas da arquitetura de sua base de dados. (DUMBILL, 2012, tradução nossa)

³ Big data é um grande volume, uma grande velocidade e uma grande variedade de ativos de informação que demandam formas de processamento de informação custo-efetivas e inovadoras para aprimorar os insights e o processo decisório. (GARTNER, [2011-], tradução nossa)

Taurion (2013a) define o Big Data utilizando cinco fatores: volume, variedade, velocidade, veracidade e valor.

Big Data não trata apenas da dimensão volume, como parece à primeira vista, mas existe também uma variedade imensa de dados, não estruturados, dentro e fora das empresas (coletados das mídias sociais, por exemplo), que precisam ser validados (terem veracidade para serem usados) e tratados em velocidade adequada para terem valor para o negócio. A fórmula é então, Big Data = volume + variedade + velocidade + veracidade, gerando valor. (TAURION, 2013a, não paginado).

Figura 2 – Os 5 Vs do Big Data



Fonte: A autora

O volume é o que mais chama atenção e o sentido mais comum atribuído ao Big Data. Apesar de subjetivo, é inegável que a quantidade de dados gerada diariamente – 2,5 quintilhões de *bytes*, o que equivale a 2,5 seguido de dezoito zeros – (DAVENPORT, 2014, p.10) e que podem ser analisadas pelas empresas chegam a números astronômicos. “Geramos *petabytes* de dados a cada dia. Estima-se que este volume dobre a cada dezoito meses”. (TAURION, 2013a, não paginado). Esses dados provêm das mais variadas fontes, como a Internet das Coisas, que será melhor exemplificada posteriormente neste trabalho, sensores e câmeras, além do que se gera nas mídias sociais, através de diferentes dispositivos. “Integram o chamado Big Data o conteúdo de 640 milhões de sites, dados de seis bilhões de celulares e os três bilhões de comentários feitos diariamente no Facebook” (TAURION, 2013a, não

paginado). Vale lembrar que esta afirmação foi feita há mais de dois anos e que estes números são, certamente, muito maiores no momento. E que continuarão a crescer exponencialmente.

A variedade refere-se às diferentes fontes em que os dados podem ser encontrados e processados (CABRAL; SAID, 2014). Estes dados podem ser estruturados ou não-estruturados, sendo este último sua imensa maioria, gerados por e-mails, mídias sociais, câmeras de segurança, buscas na Internet, etc. Este fator é de extrema importância, pois estas fontes são, em geral, não relacionadas, porém, quando combinadas podem derivar em informações de extrema importância para os gestores tomarem suas decisões e fazerem previsões mais eficientes. (TAURION, 2013a).

A velocidade está relacionada à rapidez com que podemos capturar e analisar os dados gerados em tempo real (CABRAL; SAID, 2014). O Big Data existe em um fluxo contínuo de informações que não param de fluir. Um dado coletado ontem, já pode ser ultrapassado hoje, no momento da análise. Por isso, é importante que as análises sejam feitas em tempo real, atualizadas quase constantemente, adaptando-se aos novos dados inseridos no sistema. Ao contrário das análises de dados tradicionais, com o Big Data deve-se obedecer um fluxo contínuo. (DAVENPORT, 2014, p.16).

“Veracity is an indication of data integrity and the ability for an organization to trust the data and be able to confidently use it to make crucial decisions”.⁴ (UNIVERSITY ALLIANCE, [20--], não paginado). É uma característica de extrema importância, tendo em vista que dados sem sentido ou que não sejam autênticos podem levar a decisões equivocadas e atitudes que gerarão consequências negativas para a organização. Para certificar-se da validade dos dados encontrados é preciso ter confiança nas fontes pesquisadas e nos programas de análise de dados, além de minuciosa análise dos cientistas de dados de que as informações geradas são confiáveis.

E, por fim, o valor gerado, que são resultados das ações da empresa ao coletar, armazenar, processar e analisar o Big Data.

O conceito de que o Big Data pode gerar valor para empresa é constatado, principalmente, quando se percebe a oportunidade de reutilização destes dados. Utilizados primeiramente em um único aspecto, eles podem ser analisados sobre uma diferente perspectiva ou mesmo poderá ser combinado com os outros dados, de forma a indicar novas

⁴ “A veracidade é uma indicação da integridade dos dados e a habilidade de uma organização de confiar nos dados e ser capaz usá-los com confiança para tomar decisões cruciais.” (UNIVERSITY ALLIANCE, [20--], tradução nossa).

informações à empresa, que podem transformar-se em valiosas decisões. (TAURION, 2014b).

Taurion (2013a, não paginado) conclui sua explicação dizendo que “Big Data não é apenas um produto de software e hardware, mas um conjunto de tecnologias, processos e práticas que permitem às empresas analisarem dados a que antes não tinham acesso e tomar decisões ou mesmo gerenciar atividades de forma muito mais eficiente”.

Em contrapartida, Davenport (2014, p.6) acredita que a maioria das definições do termo são equivocadas, apesar do grande potencial do conceito.

Percebe-se, pela análise destes conceitos apresentados, que boa parte dos especialistas definem o Big Data pelas características dos três Vs (volume, variedade e velocidade) e, em alguns casos, até mesmo acrescentando mais dois Vs (veracidade e valor). Davenport (2014, p.7) discorda destas definições por acreditar que pode existir Big Data mesmo sem a presença dos três ou cinco Vs. Ele afirma que “É verdade que essas são características importantes, mas e se você só tiver um ou dois Vs? Será que isso significa que você só tem um terço ou dois quintos do big data?”

Davenport (2014, p.6) também critica o erro mais comum relacionado ao entendimento do termo Big Data, que seria acreditar que se trata somente de um grande volume de dados. De acordo com o autor, se não se pode tomar esta afirmação como verdade, considerando que o que definimos como um grande volume de dados na atualidade poderá ser insignificante no futuro.

Para algumas empresas, o volume de dados analisados não chega a ser exorbitante, sendo a falta de estrutura dos mesmos o grande problema a ser considerado com relação ao Big Data (DAVENPORT, 2014, p.6-7).

Davenport (2014, p.8) aconselha então que, ao invés de prender-se a uma definição única do termo, deve-se analisar todas as muitas variações e possibilidades do Big Data e explorá-la de acordo com as necessidades da organização. (DAVENPORT, 2014, p.8).

Apresenta-se, a seguir, um quadro que deve ser analisado de forma a escolher um tipo de possibilidade de cada coluna, para assim ter um melhor aproveitamento do Big Data. (DAVENPORT, 2014, p.8).

Quadro 4 – Um cardápio de possibilidades de Big Data

| Estilo de dados | Fonte de dados | Setor afetado | Área afetada |
|------------------------|-----------------------|----------------------|-----------------------|
| Grande volume | On-line | Serviços financeiros | Marketing |
| Não estruturados | Vídeo | Saúde | Cadeia de suprimentos |
| Fluxo contínuo | Sensores | Manufatura | Recursos humanos |
| Formatos variados | Genômicos | Viagens/transportes | Finanças |

Fonte: DAVENPORT (2014, p.8).

Apesar de Big Data ser uma novidade, a noção de analisar dados de forma a gerar vantagem competitiva aos negócios já existe há algum tempo sob nomes diferentes. A grande inovação com relação ao Big Data é, não só seu volume exacerbado e a velocidade quase em tempo real com que surgem, mas também o fato destes dados serem, em sua maioria, não-estruturados. (DAVENPORT, 2014, p.9).

Davenport (2014, p.10) exemplifica os diversos termos que já definiram estas explorações de dados e seus diferentes enfoques, sendo o Big Data o conceito mais atual. O quadro 5, mostra a terminologia desse campo.

Quadro 5 – Terminologia para o uso e a análise de dados

| Termo | Período | Significado específico |
|--|-----------------|---|
| Suporte à decisão | 1970-1985 | Uso da análise de dados para dar suporte à tomada de decisões. |
| Suporte aos executivos | 1980-1990 | Foco na análise de dados para dar suporte ao processo decisório dos altos executivos. |
| Processamento Analítico On-Line (OLAP) | 1990-2000 | Software para a análise de tabelas de dados multidimensionais. |
| Business Intelligence | 1989-2005 | Ferramentas para dar suporte a decisões orientadas por dados, com ênfase em relatórios. |
| Analytics | 2005-2010 | Foco em análises estatísticas e matemáticas para a tomada de decisões. |
| Big Data | 2010-atualmente | Foco em um grande volume de dados não estruturados e em rápido movimento. |

Fonte: Davenport (2014, p.10).

Uma grande diferença no que tange o Big Data é que antes os dados eram, em sua maioria, encontrados internamente nas empresas e agora eles estão em toda parte. Podem “[...] incluir sequências do genoma humano, sensores instalados em poços de petróleo, comportamentos de células cancerígenas, localização de produtos em paletes, interações nas mídias sociais ou sinais vitais do paciente”. (DAVENPORT, 2014, p.119).

Percebe-se que as empresas acreditam no potencial do conceito, apesar de encontrarem-se ainda em fase de planejamento. Segundo Taurion (2013a, não paginado) “[...] pelo menos 2/3 das empresas pesquisadas sentem que Big Data oferece um potencial muito grande para a criação de vantagens competitivas”, porém somente 6% já possui algum projeto dessa natureza em andamento. 47% estão ainda na fase de estudos e 28% possuem algum projeto piloto. No entanto, as empresas não devem somente se impressionar com o volume de dados, mas sim fazer algo a respeito para analisá-los e compreendê-los, de forma a transformá-los em conhecimento, inovações e valor para o negócio. Deve-se saber que nem tudo será utilizável - estima-se que somente 25% destes dados possuem algum tipo de valor – porém, até então, somente 1% dos *zettabytes* gerados são analisados. (HARVARD BUSINESS REVIEW, 2013, p.4).

Para tal, é necessário que estas revoluções na análise dos dados extrapolem os processos de gestão e tecnologia. Precisam existir mudanças nas orientações básicas e mesmo na cultura organizacional. A forma de pensar no negócio irá mudar com o advento do Big Data. Os produtos e serviços serão agora totalmente focados no cliente, no que ele quer, pensa e espera. “As empresas que conseguirem analisar e se adaptar rapidamente, usando dados a partir de fontes internas e externas, serão claramente as vencedoras.” (HARVARD BUSINESS REVIEW, 2013, p.5-6).

Portanto, Big Data não é teoria ou futurologia. Geramos um imenso volume de dados a cada dia e análises de padrões e correlações nesta massa de dados podem produzir informações valiosíssimas em todos os setores da sociedade humana, de governos buscando entender demandas da população a empresas buscando se posicionar mais competitivamente no mercado. (TAURION, 2013a, não paginado).

Entende-se que os conceitos e tecnologias que envolvem o Big Data influenciarão as empresas por muitas décadas vindouras. Davenport (2014, p.14) afirma que “Para que tais conceitos e tecnologias desaparecessem, as empresas e organizações teriam de perder subitamente o interesse em maneiras de poupar dinheiro, vender mais produtos e serviços e encantar os clientes – um cenário bastante improvável”.

Ainda que exista muita discordância sobre uma definição concreta do que é Big Data, é interessante ressaltar que mais importante que o termo, são os benefícios que o Big Data pode trazer às organizações.

3.1.2 Tecnologias para o Big Data

Como previamente discutido, ainda existem divergências sobre o termo Big Data, apesar de sua crescente popularidade. Sendo assim, tornar este conceito aplicável, ou seja, transformar os dados e “[...]criar soluções de negócios que agreguem valor para as companhias” (TAURION, 2013a, não paginado) ainda apresenta-se como uma tarefa complexa e controversa.

Criam-se então, novas tecnologias capazes de administrar o imenso volume de dados gerados pelo Big Data, que excedem a capacidade das antigas soluções, preocupadas em analisar dados estruturados gerados pelos sistemas internos das empresas.

Uma tecnologia, para decifrar o fluxo veloz do Big Data, deve ser capaz de criar rapidamente modelos estatísticos, que se encaixem aos dados, otimizando-os e prevendo-os. Para tal, são interessantes tecnologias com aprendizado de máquina, que são flexíveis e adaptáveis. (DAVENPORT, 2014, p.111).

O aprendizado de máquina, também chamado de modelagem automatizada, significa que o software consegue adaptar diferentes modelos de dados a fim de obter o melhor ajuste possível. Esse processo é muito rápido, porém pode gerar resultados difíceis de interpretar e explicar. (DAVENPORT, 2014, p.116).

Apesar do Big Data usualmente gerar uma discussão em torno do grande volume de dados não-estruturados, Davenport (2014, p.113) acredita que a grande novidade relacionada à nova tecnologia é como ela irá gerar valor às organizações, seja reduzindo os custos, aumentando a velocidade de processamento de dados, desenvolvendo novos produtos ou serviços e proporcionando um melhor suporte ao processo decisório, informando novos dados e modelos.

A grande vantagem das tecnologias de Big Data é que a maioria delas é gratuita ou barata, não sendo necessário grandes investimentos por parte das empresas neste sentido. No entanto, para arquitetar e programar esses softwares faz-se necessário uma equipe extremamente especializada e bem treinada, o que demanda tempo e um custo elevado. (DAVENPORT, 2014, p.115).

O quadro 6, a seguir, apresenta uma visão geral das tecnologias do Big Data, segundo Davenport (2014, p.112).

Quadro 6 – Visão geral das tecnologias do Big Data

| Tecnologia | Definição |
|--|---|
| Hadoop | Software de código aberto para o processamento de big data em uma série de servidores paralelos |
| MapReduce | Um framework arquitetônico no qual o Hadoop se baseia |
| Linguagens de script | Linguagens de programação adequadas ao big data (por exemplo, Python, Pig Hive) |
| Aprendizado de máquina | Software para identificar rapidamente o modelo mais adequado ao conjunto de dados |
| Visual Analytics | Apresentação dos resultados analíticos em formatos visuais ou gráficos |
| Processamento de Linguagem Natural (PLN) | Software para análise de texto – frequências, sentido, etc. |
| In-memory analytics | Processamento de big data na memória do computador para obter maior velocidade |

Fonte: Davenport (2014, p.112).

As tecnologias de Big Data são divididas sob duas óticas:

- a) As tecnologias envolvidas com análise de dados, cujos principais representantes são o Hadoop MapReduce (HMR);
- b) As tecnologias de infraestrutura, responsáveis pelo armazenamento e processamento dos dados. São os bancos de dados *Not Only Structured Query Language* (NoSQL).

Alguns exemplos de softwares banco de dados NoSQL que tratam dados estruturados e não-estruturados são o Big Table, usado pela Google, o DynamoDB, usado pela Amazon, o Neo4j, entre outros.

O Hadoop é uma das principais tecnologias relacionadas ao Big Data. Criado em 2005 pelo Yahoo, o Hadoop é um projeto da comunidade Apache, dividido em duas partes: a primeira, o HMR é baseado no MapReduce, software utilizado pelo Google para acelerar as pesquisas realizadas em seu buscador, a segunda é o Hadoop Distributed File System

(HDFS), que é “[...] um sistema de arquivos distribuídos otimizados para atuar em dados não-estruturados” (TAURION, 2013a, não paginado). Na prática, os dados precisam primeiramente estar armazenados no HDFS para então serem processados pelo HMR.

Além destas duas tecnologias, existe também o Hadoop Common que é um “[...] conjunto de bibliotecas e utilitários que suportam os projetos Hadoop” (TAURION, 2013a, não paginado).

Davenport (2014, p.113) define o Hadoop como “[...] um conjunto de ferramentas de software de código aberto e um framework para distribuir os dados em vários computadores; trata-se de um ambiente unificado de armazenamento e processamento altamente escalonável para grandes e complexos volumes de dados”.

Outra tecnologia de extrema importância para o Big Data é o Stream Process. Sua principal diferença com relação ao Hadoop é que este último analisa os dados que são primeiramente armazenados em seu sistema, enquanto o Stream Process faz a análise de dados em tempo real. A vantagem desta tecnologia é que a análise dos dados no momento que os mesmos estão sendo criados permite a realização de ações imediatas em resposta às questões levantadas. (TAURION, 2013a).

O grande desafio para os gestores que pretendem se arriscar a trabalhar com Big Data é escolher a opção de banco de dados mais adequada, dentre tantas e, especialmente, qual o melhor componente de análise de dados a ser utilizado, sendo este último o principal responsável por transformar estes dados em valor para as empresas.

Outros aspectos a serem considerados, quando se analisa as tecnologias de Big Data são o conceito de *stream process*, que permite o tratamento de dados em tempo real e as tecnologias que permitem a visualização destes dados. Alguns exemplos são as nuvens de *tags*, *clustergramas*, *History Flow* e *Spatial Information Flow*.

A utilização da computação em nuvem é outro fator tecnológico interessante, tendo em vista sua capacidade de suportar imensos volumes, sua elasticidade e menor custo se comparado ao armazenamento em servidores físicos.

Para obter sucesso tirando proveito do Big Data, faz-se essencial para as empresas começarem a investir em tecnologias que o comportem, tendo sempre em vista que um dos atributos essenciais quando se fala deste novo conceito é a resiliência. O volume de dados, sua natureza, a velocidade com que são criados novos dados, está em constante mudança e somente as organizações que souberem se adaptar prontamente e tirar proveito dessas variações, é que serão bem sucedidas. (TAURION, 2013a).

3.1.3 Internet das coisas

“A Internet das Coisas vai criar uma rede de centenas de bilhões de objetos identificáveis e que poderão interoperar uns com os outros e com os data centers e suas nuvens computacionais. A Internet das Coisas vai aglutinar o mundo digital com o mundo físico, permitindo que os objetos façam parte dos sistemas de informação. Com a Internet das Coisas podemos adicionar inteligência à infraestrutura física que molda nossa sociedade. A Internet das Coisas, com seus objetos gerando dados a todo instante, é um impulsionador poderoso para o Big Data.” (TAURION, 2013a, não paginado).

Na década de 1980, o filósofo francês Pierre Levy (apud DINIZ, 2006, p.59) criou o termo "ecologia coletiva" para designar a articulação de "coletividades pensantes homens-coisas". Foi este pensamento que inspirou a criação da chamada "Internet das Coisas". A Internet foi criada com o propósito de proporcionar a comunicação a qualquer tempo e em qualquer lugar, entre pessoas, porém a Internet das coisas extrapola esta definição, adicionando a comunicação entre quaisquer coisas.

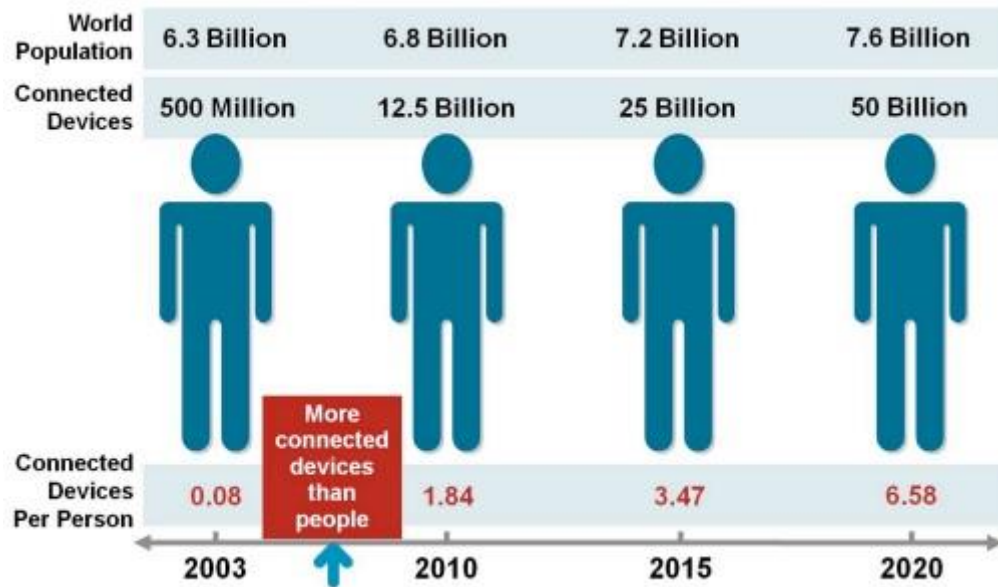
Segundo Singer (2012, não paginado), o termo “[...] é usado para designar processos que envolvam objetos conectados em rede e que produzam e/ou processem informação em tempo real e de forma autônoma.”

Benghozi, Bureau e Massit-Folea (2008, p.10) definiram a Internet das Coisas como “[...] un réseau de réseaux qui permet, via des systèmes d’identification électronique normalisés et unifiés, et des dispositifs mobiles sans fil, d’identifier directement et sans ambiguïté des entités numériques et des objets physiques et ainsi de pouvoir récupérer, stocker, transférer et traiter, sans discontinuité entre les mondes physiques et virtuels, les données s’y rattachant”.⁵

Já Evans (2001, p.2) da Cisco Internet Business Solutions Group, considera que a Internet das Coisas só passou a existir a partir do momento que haviam mais objetos ou coisas conectados do que pessoas. Isso aconteceu oficialmente em 2010, quando o número de aparelhos conectados à Internet chegou a 12,5 bilhões, enquanto a população mundial era de 6,8 bilhões, um total de 1,84 aparelhos conectados à Internet por pessoa. Porém, refinando estes números, Evans (2011, p.3) acredita que a Internet das coisas surgiu entre 2008 e 2009. A figura 3, a seguir, representa a Internet das Coisas segundo Evans (2011, p.2).

⁵ “[...] uma rede de redes que permite, via sistemas de informações eletrônicos normalizados e unificados, e dispositivos móveis sem fio, identificar diretamente e sem ambiguidade as entidades numéricas e os objetos físicos, podendo assim recuperar, estocar, transferir e tratar, sem discontinuidade entre os mundos físicos e virtuais, os dados que se relacionam entre si”. (BENGHOZI;BUREAU;MASSIT-FOLEA, 2008, p.10, tradução nossa)

Figura 3 – A Internet das coisas surge entre 2008 e 2009



Fonte: Evans (2011, p.2).

Os exemplos de objetos conectados à Internet são inúmeros. O primeiro eletrodoméstico “inteligente” foi uma geladeira fabricada pela Lucky Goldstar (LG) em junho de 2000. O objetivo era que a geladeira pudesse se conectar automaticamente com outros aparelhos eletrônicos como a televisão, o rádio, a câmera digital etc.

Em 2005, foi criado o Nabaztag17, um objeto em forma de coelho, que recebia via Internet a previsão do tempo, lia *e-mails* e notícias, entre outras coisas. Este foi o primeiro objeto inteligente comercializado em escala, que não foi produzido para indústrias. (SINGER, 2012).

Atualmente, as televisões, sistemas de segurança e de ar-condicionado estão cada vez mais em rede, otimizando seu uso, consumo e manutenção. A previsão é que em breve os dispositivos de consumo, como a geladeira, por exemplo, estejam ainda mais conectados, possibilitando facilidades como encomendar automaticamente algum produto cujo estoque esteja acabando.

Apesar do nome estar relacionado às coisas, os sensores não são exclusivos aos objetos. Já existem experimentos sendo feitos em animais e humanos. A empresa J.R. Simplot realiza experiências nas quais implementa sensores no estômago de vacas, que determinam sua temperatura e a presença de bactérias E.Coli. Caso ocorra alguma alteração, os veterinários podem prontamente tratar qualquer doença.

Relacionado aos seres humanos, as atividades diárias estão sendo cada vez mais monitoradas por meio de *smartphones* e relógios digitais. Existem centenas de aplicativos

capazes de coletar e analisar o condicionamento físico, produtividade e saúde pessoal. Os dispositivos da Apple, como o iPhone e o Apple Watch, possuem integrados ao seu sistema um aplicativo de controle de saúde que indica o número de passos diários, resultados de exames médicos e mesmo análise do sono. Outros aplicativos como o Nike+ e o RunKeeper são dedicados a esportistas que desejam monitorar seus exercícios físicos, com controle de tempo, velocidade e percurso, através do GPS de seu celular ou relógio. Um aplicativo chamado Me-trics possui a ferramenta de analisar e monitorar todo e qualquer âmbito da vida, o que inclui a saúde, estado de espírito, finanças, condicionamento físico, atividades on-line e mesmo consumo diário de água.

Ainda que esteja acontecendo uma grande evolução dos sensores na vida cotidiana, a maioria dos dados da Internet das Coisas provém da “Internet industrial”, ou seja, as empresas estão conectando suas fábricas, malhas de transporte, redes de energia etc. A General Electric (GE) declarou que o monitoramento de lâminas de turbinas a gás de produção de energia pode produzir até 588 gigabytes por dia. (DAVENPORT, 2014, p.11-12).

Os exemplos são inúmeros e prediz-se que, em breve, será possível inclusive saber em que local as pessoas se encontram (casa, trabalho, escola, etc.) com uma simples procura por seu nome no Google.

Apesar de todos os benefícios que a Internet das Coisas pode proporcionar, levantam-se questões como a privacidade e riscos ambientais e biomédicos, que possam vir a ser desenvolvidos através do uso contínuo destas tecnologias. (DINIZ, 2006, p.59).

Outros autores são mais otimistas, como Evans (2011, p.2) que acredita que a Internet foi uma das mais importantes e poderosas criações da história humana e que a Internet das coisas representa a próxima evolução da Internet, “[...] *taking a huge leap in its ability to gather, analyze, and distribute data that we can turn into information, knowledge, and, ultimately, wisdom*”⁶ (EVANS, 2011, p.2), permitindo à humanidade conquistar o conhecimento e sabedoria para triunfar nos anos vindouros (EVANS, 2011, p.7). “*La perspective est celle d’un monde de connexion encore plus dense, entre les hommes mais aussi avec les objets – une connexion permanente et de plus en plus invisible, qui engendre autant de cranter qu’elle est porteuse de promesses*”⁷ (BENGHOZI; BUREAU; MASSIT-FOLEA, 2008, p.7).

⁶ [...] dando um enorme salto em sua habilidade de juntar, analisar e distribuir dados que podemos transformar em informação, conhecimento e, finalmente, sabedoria”. (EVANS, 2011, p.2, tradução nossa)

⁷ “A perspectiva é de um mundo de conexões ainda mais densas, entre os homens, mas também entre os objetos – uma conexão permanente e cada vez mais invisível, que gera tanto medo quanto traz promessas”. (BENGHOZI; BUREAU; MASSIT-FOLEA, 2008, p.7, tradução nossa)

3.1.4 Privacidade e direito à informação

Um dos aspectos bastante discutido entre os teóricos que analisam o Big Data é a questão da privacidade e dos limites éticos que as empresas devem levar em consideração com relação aos dados de seus clientes antes de coletá-los e analisá-los sem restrições. (TAURION, 2013a).

Privacidade é definida por Ferreira (2010 apud MAGALHÃES et al., [2014?], não paginado) como “[...] a condição de quem tem a vida íntima, os assuntos e os afazeres pessoais preservados, ou ainda, os assuntos considerados privados”. O dicionário online Michaelis (DICIONÁRIO, 2012, não paginado) define privacidade como “vida privada, intimidade”. Magalhães et al. ([2014?]) conclui então que, preservar a privacidade relaciona-se com o controle do indivíduo sobre as informações que lhe concernem, protegendo sua vida íntima de forma que seu espaço privado não seja utilizado por terceiros.

Sendo assim, as empresas e seus clientes devem tomar precauções especiais quando consideram a privacidade relacionada ao Big Data. Essa preocupação é legítima pois a empresa só poderá ter acesso aos dados de seus clientes se os mesmos estiverem cientes e dispostos a compartilhá-los. Caso contrário, “[...] a promessa do Big Data é vazia”. (HARVARD BUSINESS REVIEW, 2013, p.8).

As questões de privacidade versus direito à informação ainda causam diversas dúvidas relacionadas à governança dos dados. A maioria dos dados que compõem o Big Data provêm de fora da organização e ainda não é claro quem detém a propriedade dos mesmos. Davenport (2014, p.120) relata que em uma pesquisa em que foi perguntada a questão “[...] a quem os dados dos nossos clientes pertencem?”, a maioria das empresas teve dificuldade em responder.

Magalhães et al. ([2014?]) atenta-se ainda, em questionar se usuários possuem realmente consciência da forma que os dados que fornecem na Internet (em sites de relacionamento, aplicativos, sites de comércio eletrônico, etc.) podem ser utilizados e ao grau de exposição a que estão sujeitos. Para se afiliar a estes sites, o usuário concorda com os Termos de Usos, associados a uma Política de Privacidade, que garante à empresa diversos direitos sobre os dados de seus clientes. No entanto, grande parte dos usuários destes sites não possuem o hábito de ler estes termos e concordam com os mesmos, sem saberem exatamente ao que estão se sujeitando. (MAGALHÃES et al., [2014?]).

Em sites de relacionamento, como o Facebook estas questões podem ir além. Quando o usuário concorda com os Termos de Uso e adiciona suas fotos e dados no *site*, o Facebook

armazena os mesmos por tempo indeterminado, mesmo que o usuário decida apagar sua conta, sob o pretexto que um dia possa mudar de ideia e decidir reabri-la. Sendo assim, tudo que um dia foi publicado em sua conta torna-se, automaticamente, propriedade da empresa. (MAGALHÃES et al., [2014?]).

Taurion (2013a) afirma que possuindo a data de nascimento, o gênero e o código postal de uma pessoa é possível identificá-la em bases de dados abertas e públicas em 87% dos casos, o que pode vir a ser preocupante. Em alguns casos, esses dados podem ser usados de forma maliciosa, pois informações como “estou de férias” em mídias sociais podem, por exemplo, aumentar o risco de roubo à propriedade, dentre outros problemas.

As empresas devem sempre atentar-se as questões legais relacionadas ao uso de dados, para não gerar constrangimentos e inconvenientes que podem levar, não só a perda de clientes, mas também gerar processos legais. Para evitar tais consequências indesejadas é interessante “[...] criar uma política de segurança da informação que classifique os dados de acordo com seu nível de privacidade”. (TAURION, 2013a, não paginado).

É necessário atentar-se também, ao fato de, quando tratados, os dados do Big Data poderem ser facilmente triangulados e integrados com diversas outras fontes e, por vezes, um dado que individualmente não exporia um cliente (como comportamentos, hábitos de compra etc.), combinados podem revelar identificação pessoal íntima ou mesmo segredos industriais que podem não só gerar processos judiciais às organizações, mas também prejudicar sua imagem. (TAURION, 2013a).

Deixamos nossa pegada digital a todo momento, seja usando o Internet Banking, comprando pela Internet, acessando um buscador, tuitando, comentando alguma coisa no Facebook, usando o smarphone, 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. Onde esteve a cada dia e o que viu na Internet. Se tem alguma doença ou se tem propensão a sofrer de uma. Esta questão nos leva a outro ponto extremamente importante: garantir a segurança deste imenso volume de dados. (TAURION, 2013a, não paginado).

Com o acesso das empresas a esses dados, Taurion (2013a) afirma que pode-se, inclusive, dar um novo sentido à palavra anonimato. Tendo em ao vista o acesso cada vez mais frequente e livre de barreiras a informações pessoais, a fronteira entre o público e o privado torna-se menos evidente e a privacidade cada vez menos respeitada. Os clientes não se importam mais em oferecer seus dados às empresas, desde que recebam em troca serviços e produtos melhores e personalizados. Porém, para esta troca, acredita-se na responsabilidade da empresa em salvaguardar estas informações. Violações destes dados podem estragar

relações entre o cliente e a empresa e gerar custos irreparáveis às últimas. “As empresas que podem ser transparentes com relação à coleta e uso de informações ao mesmo tempo em que demonstram que a gestão de risco cibernético tem alta prioridade, terão uma vantagem competitiva.” (HARVARD BUSINESS REVIEW, 2013, p.11).

3.2 BIG DATA NO AMBIENTE EMPRESARIAL

Diante desta revolução digital, as empresas podem escolher não tomar nenhuma medida de forma a conquistar seu espaço no novo mercado que se anuncia ou identificar as oportunidades que o Big Data apresenta para se sobressair sobre seus concorrentes. “Com o mundo hiperconectado podem redesenhar sua proposição de valor para seus clientes”. (TAURION, 2013a, não paginado).

No presente capítulo, apresenta-se o potencial do Big Data para auxiliar as empresas a conquistarem vantagem competitiva. É necessário então compreender o papel do Big Data na estratégia das organizações e seu poder no auxílio à tomada de decisão.

3.2.1 Big Data e estratégia

Segundo a Enciclopédia Barsa (1979, v.6, p.43), a palavra estratégia tem origem no grego *strategus*, uma designação dos generais na antiga Grécia, e significa, literalmente, a arte do general. Porém, afirma-se que, não é possível encontrar uma definição única para o termo atualmente, pois estratégia “[...] passou a representar uma interdependência entre os objetos, planos e potenciais, de ordem política, social, econômica e militar de uma nação”.

Clausewitz (apud ENCICLOPÉDIA, v.6, p.43) definiu estratégia no século XIX como “[...] a combinação entre si de vários combates isolados... Em outras palavras, a estratégia elabora o plano da guerra, delinea o rumo proposto para as várias campanhas que a compõem e prevê as batalhas a serem travadas em cada campanha”.

Sun Tzu foi um general chinês que viveu no século I a.C. e é considerado, até hoje, um dos maiores estrategistas da História. Em seu legado, deixou anotações que se transformaram no célebre livro *A Arte da Guerra*. Todo seu discurso discorre sobre estratégias militares ditas infalíveis para a vitória e, apesar de terem sido escritas especificamente para os generais conduzirem seus exércitos nas guerras, as mesmas estratégias são amplamente utilizadas na atualidade pelas organizações, adaptando-as para obter vantagens na acirrada concorrência empresarial.

Sobre estratégia, Sun Tzu (2007, p.26) dizia que “Nos planos, jamais um deslocamento inútil; na estratégia, jamais um passo em vão. O comandante hábil assume uma posição tal que não pode sofrer nenhuma derrota; não negligencia nenhuma circunstância que lhe garanta o controle do inimigo.”

Conceituações mais recentes entendem a estratégia como “*The pattern of major objectives, goals and purposes and the fundamentals, plans, policies and philosophies for achieving those goals, that are declared in such a way as to define what business the firm is engaged in, or wishes to be engaged in, and the kind of organization it is or would like to be*”.⁸ (LOWSON, 2002, p.42).

Lowson (2002, p.40) entende a estratégia empresarial como “[...] *an organization’s sense of purpose – a guiding purpose or policy, a focus statement, even a philosophy, for the achievement of an objective. It is the mapping out of future directions that need to be adopted using the resources possessed*”.⁹

Todas as organizações que intencionam o sucesso, necessitam adotar uma estratégia detalhada, com objetivos de curto e longo prazo que sejam desafiadores, porém realizáveis. Para tal, é necessário que os funcionários da empresa entendam esse plano com clareza e o cumpram com maestria. Somente assim a empresa irá se destacar perante as outras.

Diante desse cenário competitivo, é preciso que as organizações entendam o grande potencial que possuem ao implementar estratégias de Big Data. É de extrema importância que as empresas, neste momento inicial, deem a devida relevância ao termo e aproveitem-se das oportunidades criadas neste ambiente atual em que o grande volume de dados pode ser utilizado a seu favor, permitindo que conheçam melhor seus clientes, direcionem as vendas, reduzam os custos e, até mesmo, criem um novo produto baseados nos dados coletados.

No entanto, é essencial aos gestores adotarem uma estratégia adequada capaz de permitir a utilização ideal destes dados, evitando assim, gastos inúteis com projetos infrutíferos. Entende-se, portanto, que o maior desafio não é tecnológico, mas sim humano. (FEIJÓ, 2013 apud HENRIQUES; COSTA, 2014, p.4). Ignorar o Big Data é perder espaço no mercado, ao mesmo tempo que, investir excessivamente sem antes entender quais são os potenciais e limitações do Big Data é extremamente prejudicial ao negócio. (TAURION,

⁸ “O padrão dos grandes objetivos, metas e propósitos e os fundamentos, políticas e filosofias para atingir essas metas, são declarados de forma a definir em que negócio a empresa está engajada, ou deseja se engajar, e o tipo de organização que é ou deseja ser”. (LOWSON, 2002, p.42, tradução nossa)

⁹ “[...] o senso de propósito de uma organização – seu propósito ou política de orientação, seu foco, ou mesmo uma filosofia para alcance de um objetivo. É o mapeamento das direções futuras que precisam ser adotados usando os recursos existentes”. (LOWSON, 2002, p.40, tradução nossa)

2013b). A empresa precisa, antes de traçar sua estratégia, entender qual seu objetivo com o Big Data e quais perguntas pretende responder com sua utilização.

Taurion (2013a) afirma que muitas empresas implementam iniciativas de Big Data sem, primeiramente, definir uma estratégia que os oriente. O autor alerta que Big Data vai além de adquirir pacotes de tecnologias, mas, mais importante, significa uma nova forma de explorar e utilizar o exorbitante volume de dados existentes dentro e fora das empresas. “Big Data embute transformações em processos de negócios, fontes de dados, infraestrutura de tecnologia, capacitações e mesmo mudanças organizacionais na empresa e em TI”. (TAURION, 2013a, não paginado).

As organizações devem aproveitar-se do Big Data para coletar dados relevantes que sejam úteis a estratégia de seu negócio, utilizando estas informações para melhorar um produto, melhorar uma estratégia de marketing, diminuir gastos e o desperdício de recursos e destacar-se perante seus concorrentes ao disponibilizar um serviço que satisfaça seu cliente. (MAGALHÃES et al., [2014?]).

Entretanto, as empresas devem focar-se em explorar somente o que precisam. “Há muita água no oceano, mas você não consegue beber tudo isso. E também reconheça o que é “bom saber” versus o que você “precisa saber”. (HARVARD BUSINESS REVIEW, 2013, p.9). Lowson (2002, p.40) afirma que “*Competitive strategy is about difference, the choice of certain activities to deliver a unique value-mix to a selected market. Strategy, then, is first about choices. Choices concerting markets, product and service combinations, resources in their widest sense and directions for the future*”.¹⁰

Grande parte das empresas ainda não tomou conhecimento sobre as imensas vantagens que uma exploração eficiente do Big Data pode acarretar, não implementando assim projetos voltados à exploração de dados em suas estratégias ou não dispendo a devida atenção que deveriam à potencialidade do Big Data.

A Harvard Business Review realizou uma pesquisa em 2013, com 951 leitores em que a maioria relatou conhecer o conceito de Big Data, mas que apenas 28% o utilizava em suas organizações “[...] para tomar decisões melhores ou criar novas oportunidades de negócios”. Somente 23% afirmaram utilizar uma estratégia para Big Data, 6% concordaram com “Minha organização considerou o impacto do big data nas funções-chave da empresa” e apenas 3% disseram que “Minha organização sabe como utilizar big data em nosso negócio”.

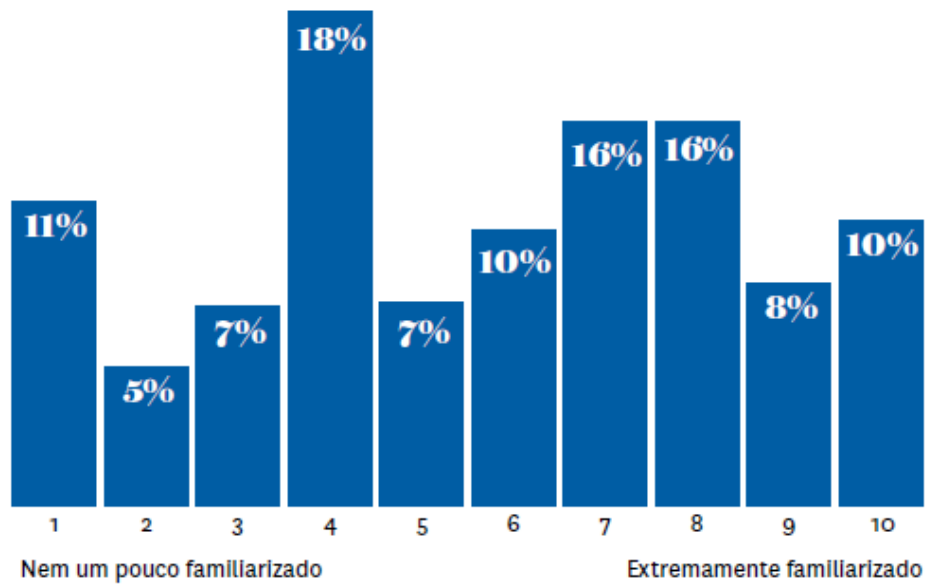
¹⁰ “Estratégia competitiva é sobre diferença, sobre a escolha de certas atividades para demonstrar uma mistura de valores únicos para um mercado selecionado. Estratégia é, então, primeiramente sobre escolhas. Escolhas sobre mercados, combinações de produtos e serviços, recursos em seu sentido mais amplo e direções para o futuro”. (LOWSON, 2002, p.40, tradução nossa)

(HARVARD BUSINESS REVIEW, 2013, p.3). As figuras 4, 5 e 6, a seguir, ilustram essas afirmações.

Figura 4 – Familiaridade com o conceito de Big Data

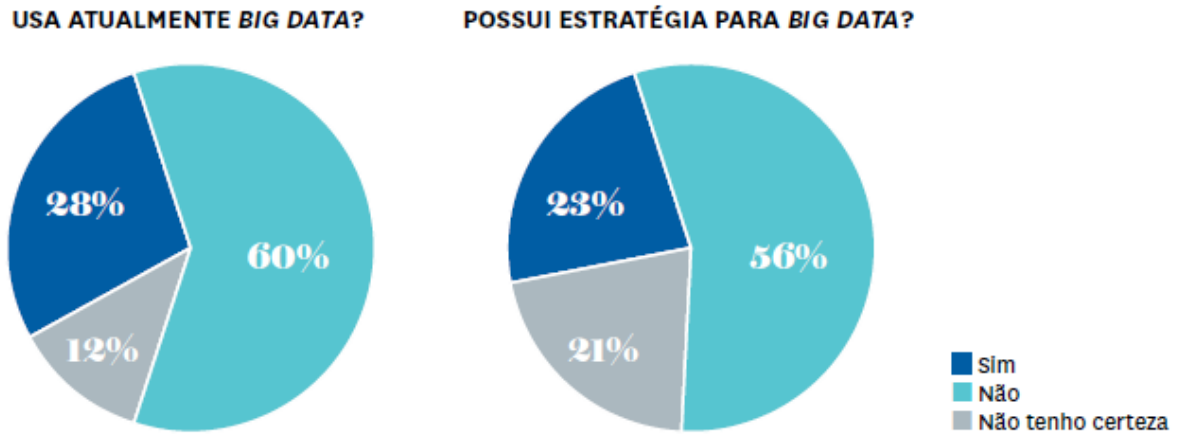
Familiaridade com o Conceito de *Big Data*

O quão familiarizado você está com o Conceito de *Big Data*?



Fonte: HARVARD BUSINESS REVIEW (2013, p.3).

Figura 5 – Uso atual e estratégia para Big Data

Uso Atual e Estratégia para Big Data

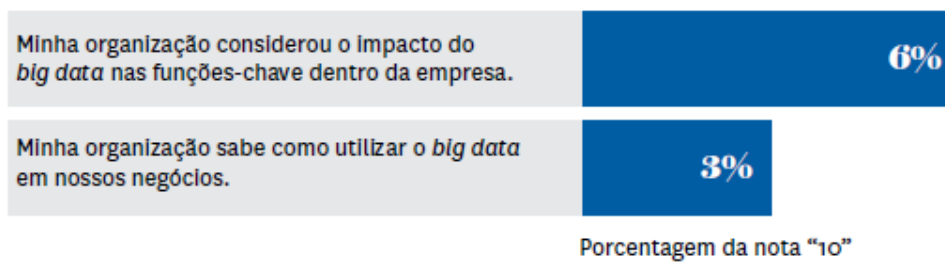
Fonte: HARVARD BUSINESS REVIEW (2013, p.3).

Figura 6 – Atitudes das organizações com relação ao Big Data

Atitudes das Organizações com Relação ao Big Data

Classifique seu nível de concordância com cada uma das seguintes afirmações.

1 = discordo totalmente; 10 = concordo totalmente



Fonte: HARVARD BUSINESS REVIEW (2013, p.3).

Essa pesquisa indica que a maioria das empresas ainda está na fase inicial de percepção do Big Data. “O ponto não é ficar deslumbrado com o volume de dados, mas sim analisa-los – convertê-los em conhecimento, inovações e valor de negócio. O mesmo estudo sugere que apenas 1% dos 2,8 zettabytes é analisada de alguma maneira”. (HARVARD

BUSINESS REVIEW, 2013, p.4). Sendo assim, a análise do Big Data ainda passará por muitas transformações até que todo seu potencial seja utilizado pelas organizações.

Davenport (2014, p.57) afirma:

Antes de mais nada, vocês precisam parar de pensar em como o big data se encaixa no seu negócio [...] No entanto, o passo mais importante é decidir-se por uma estratégia específica para o big data. A sua equipe de administração sênior precisa se reunir para começar a falar sobre o que big data pode fazer pela empresa e quais das vastas possibilidades vocês pretendem explorar. Esse processo deve começar com uma reflexão ponderada sobre os objetivos que vocês visam atingir com o big data. (Davenport, 2014, p.57).

A revolução nas estratégias empresariais está apenas no começo e ela se dará não somente nos processos de gestão e tecnologia, mas na empresa como um todo, desde suas orientações básicas à cultura das organizações. “Nós simplesmente não podemos pensar nos negócios da mesma forma agora que temos esse novo recurso.” (HARVARD BUSINESS REVIEW, 2013, p.5).

Taurion (2013a, não paginado) alerta em seu livro que “A questão não é mais se vou ou não adotar Big Data, mas quando e com que estratégia adotarei.”

3.2.2 Big Data e vantagem competitiva

Vantagem competitiva está atrelada diretamente com estratégia e pode ser entendida como o desempenho superior das empresas sobre seus concorrentes, porém seu conceito ainda é bastante amplo e discutido. (VASCONCELOS; BRITO, 2004, p.52).

A vantagem competitiva pode ser vista como o objetivo das ações da empresa, pode ser usada para explicar a diversidade entre as empresas, pode ser vista como o objetivo final da função corporativa e, finalmente, pode explicar o sucesso ou fracasso na competição internacional. (VASCONCELOS; BRITO, 2004, p.52).

Existem duas teorias principais que embasam as definições de vantagem competitiva, a visão baseada em recursos, que afirma serem os recursos internos da empresa os fatores determinantes para gerar lucro econômico; e as teorias de posicionamento estratégicos, que combinam as vantagens de possuir recursos específicos com posições privilegiadas de mercado. (VASCONCELOS; BRITO, 2004, p.52).

Vasconcelos e Cyrino (2000, p.20) definem vantagem competitiva como “[...] a ocorrência de níveis de performance econômica acima da média de mercado em função das estratégias adotadas pelas firmas”.

Um dos primeiros autores a definir o termo foi Ansoff (1965, p.188-194 apud VASCONCELOS; BRITO, 2004, p.52) afirmando que vantagem competitiva trata-se da “[...] vantagem de perceber, de forma pró-ativa, tendências de mercado à frente dos concorrentes e de ajustar a oferta em função dessa antecipação”.

Porter (1985) conceitua como “[...] a medida de sucesso da estratégia”. O autor afirma que “Uma empresa conquistaria uma vantagem competitiva executando as atividades estrategicamente mais importantes da cadeia de valor de forma mais barata ou melhor do que a concorrência”. (PORTER apud VASCONCELOS; BRITO, 2004, p.53).

Barney (1991) amplia a conceituação de vantagem competitiva e diz que as empresas devem buscar a vantagem competitiva sustentável, ou seja, duradoura. Para tal, ele define que a empresa deve possuir quatro atributos:

1. Valioso: Para ser considerado um recurso é necessário possuir valor para a empresa e para o mercado;
2. Raro: Muitas empresas podem ter recursos valiosos, mas eles só significarão vantagem competitiva caso sejam raros entre seus competidores;
3. Inimitável: O recurso deve ser difícil de imitar por parte de seus competidores, para que garanta a vantagem competitiva sustentável;
4. Sem substitutos: Os competidores não devem possuir substitutos estratégicos equivalentes para este recurso que é valioso, raro e não imitável.

Quando as empresas optam por utilizar o Big Data para compreender melhor o que pensam e o que fazem seus clientes, direcionando desta forma seus produtos e serviços para atender suas necessidades, ela está criando valor de forma diferenciada de seus concorrentes.

“Com a tecnologia de Big Data, as organizações podem melhorar suas operações, oferecer melhores produtos, desenvolver relacionamentos mais profundos com os clientes e se transformar em organizações mais ágeis e preditivas, distanciando-se dos concorrentes”. (HENRIQUES; COSTA, 2014, p.10).

Deve-se entender, no entanto, que o diferencial não está em investir em novas tecnologias, já que as mesmas estarão disponíveis a todos, mas sim “[...] na sofisticação e maturidade da gestão da empresa”. (TAURION, 2013b, p.6).

As empresas também precisam adaptar-se em relação a mitigação de riscos. Com o Big Data, minimizar os riscos de forma rápida e ativa, assim que identificados, torna-se

essencial para alcançar vantagem competitiva. Investir em soluções de Big Data é uma maneira de minimizar consideravelmente os riscos (HARVARD BUSINESS REVIEW, 2013, p.10-11). Sabendo o valor das informações armazenadas e analisadas, as empresas podem prevenir-se e evitar o desperdício. (TAURION, 2013a).

A informação sempre foi, e sempre será, um dos ativos de maior importância para uma organização. No entanto, até então, a informação não possuía diretamente valor monetário. Pensando nisso, a Gartner, uma das maiores empresas de pesquisa em tecnologia da informação e consultoria, criou um modelo econômico chamado Infonomics que tem como objetivo “[...] mensurar e avaliar a significância econômica das informações que uma empresa possui, de modo que estas informações possam ser valorizadas monetária e contabilmente”. (TAURION, 2013a, não paginado).

Para obter vantagem competitiva, as organizações devem investir em descoberta e experimentação com dados. É preciso ser ágil e flexível, explorar fontes de dados novas e existentes, procurando padrões, eventos e oportunidades em um ritmo constante. “As empresas que conseguirem analisar e se adaptar rapidamente, usando dados a partir de fontes internas e externas, serão claramente as vencedoras”. (HARVARD BUSINESS REVIEW, 2013, p.5-6).

3.2.3 Big Data e processo decisório

“Decisões são a moeda corrente no mundo dos negócios. Cada êxito, cada infortúnio, cada oportunidade aproveitada ou perdida é o resultado de uma decisão que alguém tomou – ou deixou de tomar”. (HARVARD BUSINESS REVIEW, 2008, p.2).

Decisão é a escolha de uma dentre várias alternativas propostas para a resolução de um problema cujo objetivo é transformar a situação atual real em uma situação futura desejada. “Assim, a tomada de decisão, pode ser representada pelo processo de escolha entre os diversos cursos de ação para resolver um problema”. (PORTO, 2004, p.33).

Tomar as decisões corretas, de maneira rápida e eficaz, e executá-las com primor é o fator primordial para o sucesso de qualquer organização. “Não importa em qual setor você opere, quão grande e bem conhecida seja a sua empresa nem quão engenhosa a estratégia pareça”, se houver falhas no processo decisório, todo o esforço feito no planejamento será em vão. (HARVARD BUSINESS REVIEW, 2008, p.2).

De acordo com Bazerman (2004, p.4-5), em um processo racional de tomada de decisão, deve-se seguir seis etapas: Definir o problema, identificar os critérios, ponderar os

critérios, gerar alternativas, classificar cada alternativa segundo cada critério e identificar a solução ótima. Esse processo racional indicaria aos gestores qual a melhor solução possível a ser seguida. No entanto, um dos fatores mais importantes para que este modelo seja implementado com sucesso é o acesso ao maior número possível de informações relativas aos critérios identificados.

Podemos então inferir a importância do Big Data no processo decisório das empresas. O volume de dados gerados, coletados, em sua maioria, externamente à organização, é fonte quase infindável de informações futuras que podem auxiliar os gestores a identificarem as soluções ótimas.

De acordo com uma pesquisa realizada por Erik Brynjolfsson, economista da *Sloan School of Management* do *Massachusetts Institute of Technology* (MIT) nos Estados Unidos, as empresas que optam pela tomada de decisão com base em dados obtêm um aumento de cinco a seis por cento em produtividade. (HENRIQUES; COSTA, 2014, p.7).

Ressalta-se a importância da qualidade e veracidade dos dados coletados que serão utilizados, posteriormente, na tomada de decisão, pois a informação só é eficaz quando atende às necessidades do consumidor desta informação. “Quanto melhor a qualidade do dado, melhores serão as decisões com base no dado, criando valores para a empresa” (HENRIQUES; COSTA, 2014, p.6).

O Big Data deve ser aproveitado pelas empresas por sua capacidade de interconectar à informação que aparece dispersa em diferentes sistemas em um único sistema central. Agregando estes dados presentes em diferentes formatos, as organizações podem melhorar sua capacidade de decisão e, conseqüentemente, a performance da empresa como um todo. (UNIVERSITY ALLIANCE, [20--]).

Breternitz e Silva (2013, p.20) afirmam que a ideia central do Big Data é sua utilização na tomada de decisões, sendo assim, as organizações devem preparar-se, implementando sistemas automatizados que permitam extrair respostas rápidas e ágeis com relação às informações dos clientes (HARVARD BUSINESS REVIEW, 2013, p.1), pois cada vez mais as decisões deverão ser tomadas “[...] baseadas em dados e rápidas experimentações e não mais em intuição e estimativas”. (HARVARD BUSINESS REVIEW, 2013, p.3).

Há algum tempo, as empresas vêm utilizando análises de dados para auxiliar no processo decisório. No entanto, com o surgimento do Big Data, algumas estruturas internas precisarão passar por mudanças. Tradicionalmente, um analista separaria alguns dados e o analisaria, criando um modelo que oriente o gestor para tomada de decisão baseada nos resultados obtidos. Porém, com o Big Data, os dados possuem um fluxo contínuo e rápido que

necessitam uma análise constante com resultados sempre atualizados, para que, desta forma, possam auxiliar nas decisões em tempo real. (DAVENPORT, 2014, p.16).

Os gestores e tomadores de decisão podem obter grandes benefícios com a utilização do Big Data, tornando as decisões cada vez menos intuitivas e cada vez mais baseadas em fatos. Davenport (2014, p.67) afirma “A inteligência competitiva e de mercado costumava ser um exercício bastante intuitivo, mas o big data está começando a mudar essa abordagem. Se vocês puderem obter dados mais detalhados e fazer uma análise mais sistemática desses dados, o resultado provavelmente será decisões estratégicas”.

4 ESTUDOS DE CASO

Apresenta-se a seguir estudos de caso de cinco empresas que implementaram projetos de Big Data, obtendo resultados significativos em sua gestão estratégica. As empresas são: New South Wales State Emergency Service, Nielsen Holding, Ramco Cements Limited, Dunhumby e Netflix, Inc.

4.1 CASO NEW SOUTH WALES STATE EMERGENCY SERVICE

A New South Wales State Emergency Service (NSW SES), caso relatado por Wamba et al. (2015), é uma organização governamental australiana responsável por prestar serviços de emergência em casos de crises e desastres naturais ou causados pelo homem. Criada em 1955 pelo governo do estado de New South Wales, a organização atua em uma área de aproximadamente 800.000 km², o que equivale a quatro vezes o tamanho do Reino Unido. A NSW SES possui uma particularidade que a torna única, sendo formada por cerca de 280 funcionários que gerenciam quase 9000 voluntários. Sua sede localiza-se na região de Wollongong e possui 229 unidades voluntárias que atuam em 17 regiões. Outra característica importante da organização é sua rede de relacionamento com empresas parceiras, como a *Fire and Rescue New South Wales* e a *New South Wales Rural Fire Services*, que atuam como corpo de bombeiros, e o *Bureau of Meteorology*, responsável por fornecer informações meteorológicas. Essas redes de relacionamentos são essenciais para auxiliar com informações sobre possíveis emergências.

Visando melhorar os serviços de emergência, a NSW SES decidiu explorar a imensa quantidade de dados (estruturados e não-estruturados) que recebe diariamente para aprimorar a gestão em casos de emergência através da criação de uma plataforma colaborativa baseada

no *Microsoft SharePoint*, conectando em uma única instância todos os dados recebidos das agências da NSW SES e organizações parceiras. Realiza três funções principais: coleta, armazenamento e difusão dos dados, além de apoiar os processos importantes de forma inter e intra organizacional. Outra função importante realizada pela plataforma é o armazenamento de dados históricos (por exemplo, dados sobre inundações nos últimos 200 anos). A difusão da informação é feita através da *web 2.0* por inúmeros canais de comunicação, incluindo o site da NSW SES, Twitter, Facebook, rádio e telefones celulares dos *stakeholders*¹¹.

O sucesso da plataforma só foi atingido por conta da colaboração entre todos os *stakeholders*, desde a alta-gerência, peça fundamental para o desenvolvimento do projeto, até os voluntários, que ajudaram com conselhos e recomendações para o comitê de implementação do projeto. Durante todo o projeto, houve comunicação bilateral entre todas as partes envolvidas.

O caso da NSW SES (WAMBA et al., 2015) mostrou que quando se lida com Big Data é uma boa estratégia a criação de uma plataforma colaborativa que reúna todos os dados produzidos por diversas fontes, sejam eles estruturados ou não-estruturados, em uma única instância, que seja capaz de receber estes dados, decodificá-los, cruzá-los e disseminá-los da forma mais eficiente possível para todas as partes interessadas. O armazenamento de dados-chave para consultas futuras também é essencial. Outro fator de extrema importância para o sucesso na implementação de um projeto de Big Data é o envolvimento de todos os funcionários da organização. O apoio e envolvimento da alta gerência é o ponto nevrálgico do projeto. Além disso, todos devem entender a relevância das mudanças a serem implementadas e caso tenham liberdade para fornecer sugestões e conselhos, em um ambiente em que a comunicação entre todos é estimulada, as chances do projeto ser bem sucedido aumentam bastante.

Como resultado da implementação da plataforma na NSW SES, houve uma melhoria significativa no processo decisório da organização em casos de crise ou emergência. Através do cruzamento de dados, fornecidos por diversas fontes, é possível prever com maior exatidão a ocorrência de desastres, permitindo a execução de medidas preventivas. Além disso, em empresas que trabalham com serviços de emergência, saber exatamente “quando” e “onde” irá acontecer um desastre é peça fundamental para decidir sobre “quando”, “aonde” e “em que quantidade” transportar equipamentos, alimentos e pessoal especializado. Através da

¹¹ "*Stakeholder* é uma pessoa ou grupo que possui participação, investimento ou ações e que possui interesse em determinada empresa ou negócio. [...] *Stakeholder* também pode significar partes interessadas, sendo pessoas ou organizações que podem ser afetadas pelos projetos e processos de uma empresa." (BEZERRA, 2014)

plataforma, os gerentes tiveram acesso não só as informações prévias sobre possíveis desastres, mas também sobre a disponibilidade de voluntários e outros ativos em determinadas regiões, tornando mais simples e eficaz a tomada de decisões sobre a movimentação e alocação de ativos. O processo decisório é de extrema importância nas organizações, pois mesmo com estratégias impecáveis, qualquer falha na tomada de decisões pode destruir todo seu planejamento e ser fatal para o sucesso da empresa. (HARVARD BUSINESS REVIEW, 2008, p.2).

4.2 CASO DA NIELSEN HOLDINGS

A Nielsen Holdings, caso relatado por Prescott (2014), é uma empresa que coleta, cruza, armazena e analisa dados de audiência de canais de televisão, computadores e dispositivos móveis – conhecidos como três telas – vendendo-os para redes de televisão e agências publicitárias. Esses dados de audiência são cruzados com dados demográficos (idade, gênero, raça, classe econômica e região) fornecendo um perfil acurado sobre o público que assistiu um programa. Essas informações são essenciais para uma rede de televisão determinar o preço a ser pago por um comercial de 30 segundos no intervalo de certo programa televisivo. O cruzamento dos dados demográficos que ditam o perfil de quem estará assistindo determinado programa, permite que as empresas de propaganda saibam qual o nicho de mercado que assistirá um programa de televisão em determinado dia e horário, direcionando assim os comerciais para o momento mais propício de acordo com o perfil dos consumidores de seus produtos ou serviços. Além disso, essas informações também são úteis para a decisão sobre quais programas de televisão continuarão a serem produzidos e quais serão cancelados, baseando-se em sua audiência.

A Nielsen Holdings é uma empresa americana com sedes em Nova Iorque e na Holanda e filiais em diversos países do mundo, incluindo o Brasil, com quem trabalhou em parceria com o Instituto Brasileiro de Opinião Pública e Estatística (IBOPE), coletando dados de navegação na Internet até 31 de dezembro de 2015 (EQUIPE NIELSEN IBOPE, 2015). A empresa foi fundada em 1923 por Arthur C. Nielsen e desde então tornou-se a principal distribuidora de informações sobre audiência televisiva no mundo. Durante muitos anos, a Nielsen utilizou dois métodos principais para medir a audiência televisiva: o diário e o medidor de pessoas.

O diário é um método em que famílias selecionadas recebem um caderno em formato de diário nos meses de novembro, fevereiro, março e julho (chamados meses de varredura)

onde devem anotar, no período de uma semana, todos os programas televisivos que assistiram, incluídos aqueles que foram gravados para serem assistidos posteriormente. Estes diários são então, enviados para a Nielsen que prepara manualmente um painel, cruzando os dados informados pelas famílias com suas informações demográficas. Os resultados são então, disponibilizados aos clientes da Nielsen, aproximadamente, um mês após o recebimento dos diários das famílias participantes. Esse método é bastante criticado por conta de sua pouca precisão e confiabilidade, entendendo que as pessoas responsáveis por preencher os diários podem postergar o preenchimento dos mesmos e esquecer o que assistiram ou mesmo colocar informações sobre programas que pensam que deveriam assistir por pressão social, ao invés do que realmente assistem. Além disso, os dados demoram um longo período (um mês) para serem disponibilizados, estando de certa forma obsoletos para os clientes, em uma época em que dados em tempo real já são analisados e entregues aos clientes em curtíssimo período de tempo. Apesar dos esforços da Nielsen para melhorar este método, como por exemplo, com a criação de um site que auxilia as famílias no preenchimento dos diários (NIELSEN, 2015), com o advento de novas formas de medição muito mais precisas, o diário está caindo em desuso e o método possivelmente será descontinuado no futuro.

O método medidor de pessoas trata-se de um conversor enviado a famílias selecionadas que deve ser anexado no aparelho de televisão de forma a coletar dados sobre todos os programas assistidos, seja ao vivo ou gravados para visualização posterior. É analisado também quando o telespectador está assistindo uma gravação e decide avançar, deixando de assistir alguma parte do programa. Além disso, cada membro da família deve criar um perfil com seus dados demográficos, que são escolhidos antes de assistir um programa. Visitantes também são convidados a informar seus dados quando assistem televisão na casa das famílias. Criado nos anos de 1970, o medidor de pessoas em seus primórdios era passivo e não possuía nenhuma interação com as famílias. Com a evolução da tecnologia, incorporou-se um sensor no conversor que emite luzes quando a televisão é ligada para que o telespectador informe qual membro da família está assistindo televisão naquele momento. As luzes também são emitidas para lembrar os membros da família a manterem-se ativos durante toda a emissão. Esta é uma forma de interação que torna a análise de dados mais precisa e acurada. Os dados gravados minuto a minuto, são enviados então à base de dados da Nielsen, que os analisa diariamente, criando categorias a partir do cruzamento das informações que serão posteriormente oferecidas aos clientes da empresa.

Durante muitos anos, a Nielsen exerceu verdadeiro monopólio¹² no mercado de audiência televisiva. Uma empresa que desejasse competir com a Nielsen precisaria de um investimento altíssimo, além do desenvolvimento de tecnologias que superassem a do medidor de pessoas. No entanto, há poucas décadas essa vantagem foi seriamente ameaçada por dois fatores eminentes: a digitalização e a criação das empresas de televisão a cabo e a satélite.

Com a popularização da Internet, a participação do telespectador com o programa vai muito além de apenas assisti-lo passivamente em seu aparelho televisivo. Ele agora pode utilizar seu computador ou dispositivo móvel (*tablets*, celulares, etc.) para interagir. Além do acesso no momento que lhe for mais conveniente, o consumidor ainda conta com serviços como a Netflix, que permite assistir filmes e programas online. Outro fator diferencial é a interação do telespectador, que possui agora a possibilidade de debater sobre o programa em fóruns e redes sociais, além de procurar mais informações sobre os personagens, o enredo, os próximos episódios, etc.

As televisões a cabo e a satélite, em sua maioria oferecidas em sistemas de assinatura no qual o cliente paga um valor para ter acesso a determinado número de canais de televisão, trouxe uma inovação revolucionária: para ter acesso a esse serviço, os clientes recebem um conversor que é acoplado ao seu televisor e que pode ser facilmente monitorado. Esses conversores geram as empresas de televisão a cabo/satélite quantidade exorbitantes de dados granulares segundo a segundo, fornecendo informações sobre o que está sendo assistido. Pela imensa complexidade de análise desses dados, eles são, normalmente, vendidos para outras empresas que irão, portanto, possuir imenso potencial competitivo para com a Nielsen. Com a popularização das televisões a cabo/satélite (em 2007/2008, nos Estados Unidos, já existiam aproximadamente 39 milhões de casas com conversores) cria-se diariamente uma infinidade de dados que superam largamente os dados criados pelo medidor de pessoas da Nielsen. Os pontos negativos destes conversores com relação ao medidor de pessoas da Nielsen é a complexidade de análise do mesmos além, principalmente, da falta dos componentes demográficos presentes no medidor de pessoas, que permitem avaliar quem está assistindo o quê, em que momento.

Os principais competidores da Nielsen são as empresas de TV a cabo ou a satélite. Como quase todas oferecem também acesso à internet, elas são capazes de coletar também

¹² Monopólio diz respeito ao mercado em que apenas uma organização oferece determinado produto ou serviço, não existindo substitutos para o mesmo, ou quando apenas uma organização possui completo domínio de mercado, sendo capaz de ditar os preços dos serviços/produtos, sem se preocupar com concorrentes. (PINDYCK; RUBINFELD, 2002)

dados de uso da Internet e vendê-los a empresas que irão analisar quais dados são relevantes. Um exemplo é a TiVo. A TiVo é uma empresa que produz aparelhos de gravador de vídeos digitais, no qual o cliente escolhe os programas que possui interesse em assistir na televisão e os mesmos são gravados para serem assistidos posteriormente, havendo a possibilidade de excluir os comerciais publicitários. Além disso, oferece acesso à Internet. A TiVo trabalha em parceria com uma empresa que mede dados de uso da Internet e está presente em mais de 30.000 domicílios.

A Rentrak é uma empresa criada em 2007 que compra dados de provedores de televisão a cabo e a satélite. Esses dados são analisados através de um *software* proprietário desenvolvido pela própria organização que apresenta depois as informações relevantes para seus clientes em um portal na Internet. A Rentrak compra dados de mais de 500.000 conversores de televisão a cabo/satélite e dados de vídeos por demanda de mais de 45 milhões de conversores, o que significa que eles processam trilhões de linhas de dados. Através de seu robusto software de análise de dados, eles são capazes de dizer quando um telespectador termina de ver um programa de televisão e começa a ver um vídeo na Internet, por exemplo.

Outro grande competidor da Nielsen é a TRA Inc. que inovou por comprar dados de empresas de televisão a cabo e a satélite e cruzá-los com dados de compras realizadas pelos telespectadores. A TRA possui uma base de dados com 1.5 milhões de dados de audiência de televisão coletados através dos conversores de televisão a cabo/satélite e 55 milhões de dados coletados de compras. Juntos totalizam 370.000 residências com dados cruzados. A TRA faz análises em mais de 50 *terabytes* de dados, baseados nos pedidos dos consumidores. Através dessas análises é possível fornecer informações sobre qual programa de televisão ou vídeo na Internet um cliente que possui o hábito de comprar determinado produto ou serviço costuma assistir. As empresas de publicidade podem assim direcionar com ainda mais eficiência e eficácia os horários a serem transmitidos seus comerciais a seu público-alvo.

A Nielsen Holdings (PRESCOTT, 2014) monopolizou o mercado de informações sobre audiência televisiva por muitos anos, até o advento da digitalização e dos dados coletados através dos conversores de empresas de televisão a cabo e a satélite. Para competir com a Nielsen, era necessário às outras empresas investirem um alto valor financeiro de forma a equiparar sua tecnologia com aquela disponibilizada pela Nielsen para coleta de dados, através de diários e de medidores de pessoas, e seu sistema de cruzamento dos dados coletados com os dados infográficos dos telespectadores analisados.

Como proposto por Barney (1991), a Nielsen reunia todos os critérios necessários para sustentar sólida vantagem competitiva. Seus recursos eram valiosos, pois a tecnologia de

extração de dados da Nielsen (medidor de pessoas) e o cruzamento de dados relevantes com dados demográficos, fornecia a seus clientes valiosas informações; eram raros, pois a tecnologia de extração de dados, o cruzamento e a análise de dados oferecidas a seus clientes foram criados pela Nielsen e eram exclusivos da empresa; eram inimitáveis, pois a Nielsen possuía patentes sobre sua tecnologia, não sendo possível adquiri-las; e, por fim, eram sem substitutos, pois os competidores não possuíam recursos para fornecer aos clientes os mesmos dados e informações providos pela Nielsen.

No entanto, com o advento da digitalização e o big data produzido pelos conversores das empresas de televisão a cabo e a satélite, esse quadro mudou drasticamente e a Nielsen viu-se, pela primeira vez, severamente ameaçada. Em 2006, a empresa foi vendida para um consórcio de investidores e possuía, em 2008, dívidas equivalentes a cerca de oito bilhões de dólares. Para solucionar o problema, a primeira medida tomada pela empresa foi a contratação de um ex-diretor executivo da General Electric (GE), empresa conhecida por suas inovações na exploração de Big Data. Seu objetivo era não só sanar as dívidas da Nielsen e trazer lucros, mas também revitalizá-la, utilizando avançada tecnologia que traria novamente vantagem competitiva sustentada a empresa.

O novo diretor executivo teve como objetivo primário entender que a Nielsen precisava expandir seus serviços para além da coleta de dados de audiência televisiva e começasse a explorar também o uso da Internet em computadores e dispositivos móveis. Equipou-se então os medidores de pessoas com extratores de dados que capturassem também o uso da Internet pelos membros da família. Com relação aos dados de conversores de empresas de televisão a cabo/satélite, a Nielsen resolveu testar este sistema tão popular entre seus concorrentes e comprou e analisou dados de cerca de 64.000 famílias que possuem estes conversores, comparando-os com os dados do medidor de pessoas. A conclusão foi que os dados obtidos através dos conversores, apesar de serem mais granulares para análise e em maior quantidade, não se equiparavam ao medidor de pessoas por não possuir os dados demográficos. Além disso, apresentava o mesmo problema do medidor de pessoas no que tange a incerteza de que alguém está realmente assistindo o programa quando a televisão está ligada. É ainda mais grave por não possuir o sistema de luzes emitidas para lembrar os telespectadores a manterem-se ativos enquanto assistem o programa. A solução encontrada pela Nielsen foi continuar aprimorando seu medidor de pessoas e não seguir a estratégia de seus competidores em comprar dados de televisões a cabo/satélite. Com isso em mente, o setor de TI da Nielsen passou a ser central para aprimorar as formas de coletas de dados da empresa.

A Nielsen investiu também na compra e parceria com empresas que a ajudariam na extração e análise de dados. Um exemplo é a BuzzMetrics que captura, cruza e analisa dados em redes sociais, blogs, sites, etc. sobre o que está sendo falado a respeito de uma marca, com o objetivo de responder o que os consumidores pensam de determinada propaganda, qual a reputação da marca, como os consumidores reagem a determinados eventos de divulgação da marca, etc. A compra da empresa NetRating permitiu a Nielsen ter acesso a dados de visitação de *websites* e vídeos *online*. A empresa Telephia, outro investimento da Nielsen, oferece dados sobre o mercado de telefonia celular, o uso dos aparelhos móveis e, inclusive, a eficácia de propagandas visualizadas nestes dispositivos, permitindo um melhor entendimento dos hábitos destes consumidores e um perfil de sua experiência na utilização dos aparelhos celulares.

Em 2008, a Nielsen adquiriu a empresa IAG research, que mede a eficácia do marketing indireto, ou seja, reúne dados sobre a quantidade de vezes que determinada marca ou produto aparece em um programa. Marketing indireto ainda não é explorado por nenhuma outra empresa e ainda não existe um padrão para testar sua eficiência, logo, é uma grande vantagem para a Nielsen investir neste setor. Um dos maiores desafios para as empresas que medem audiência televisiva é saber quais programas os telespectadores assistem fora de casa – no trabalho, na academia, no bar, etc. A Nielsen saiu na frente de seus competidores, adquirindo a empresa Integrated Media Measurement Inc. que desenvolveu um sistema capaz de captar, através de aparelhos celulares, uma espécie de assinatura sonora emitida pelos programas de televisão, de forma a gerar dados sobre o que as pessoas assistem fora de casa. Desta forma, a Nielsen distribuiu aparelhos celulares às famílias que possuem os medidores de pessoas e pôde, portanto, ter acesso a esses dados de tão difícil captura.

Um dos maiores desafios da Nielsen é a grande quantidade de dados produzidos diariamente. Só com os medidores de pessoas, gera-se cerca de 300.000 linhas de dados a cada segundo. Pensando em melhorar a coleta, cruzamento e análise deste Big Data, a Nielsen comprou a empresa Audience Analytics Technology Inc., que permitiu a melhora na análise de dados tão robustos. Outra grande revolução da empresa foi a sociedade com a Catalina Solutions em 2009, que permitiu a Nielsen obter dados sobre compras que são cruzados com os dados dos medidores de pessoas, apresentando um melhor entendimento e uma profunda compreensão das ligações entre a exposição de um consumidor ao produto ou serviço através da propaganda e a sua compra efetiva. Isso permitirá que os clientes da Nielsen possam melhor direcionar suas propagandas, utilizando os programas e momentos em que há maior chance da exposição reverter-se em compras. Também relacionado a compras, a aliança com

a empresa 4INFO permitiu acesso a dados que medem o impacto das propagandas em dispositivos celulares que convertiam-se em compras. Obtém-se esses dados através do cruzamento das informações que a Nielsen possuía de 60.000 compradores com os 70.000 usuários de dispositivos móveis da 4INFO.

A Nielsen criou, em 2010, uma plataforma de sincronização de mídias que utiliza as assinaturas sonoras presentes nos programas televisivos, sincronizando-as com aplicações da *web*. Quando conectadas através de computadores e/ou dispositivos móveis, permite ao telespectador encontrar conteúdo relacionado ao programa que está assistindo, como roteiro, informações sobre os atores, sobre os próximos episódios, responder questionários, etc. Esta tecnologia não só permite a Nielsen coletar dados sobre o tempo/horário em que o programa foi assistido, mas também permite que propagandas específicas sejam disparadas nas plataformas. Outra aliança estratégica da Nielsen foi com a empresa Clarabridge Inc. que oferece análise de conteúdos escritos extraídos de mídias sociais, desta forma, a Nielsen pode entender melhor o que está sendo dito sobre determinada marca ou propaganda no Facebook ou Twitter, por exemplo, e oferecer este serviço a seus clientes. Por fim, e não menos inovador, a Nielsen adquiriu a empresa NeuroFocus INC. que usa a neurociência para entender a atenção, o envolvimento e a memória dos telespectadores quando expostos a anúncios, propagandas, marcas, etc., medindo suas atividades cerebrais e condutas elétricas da pele. A NeuroFocus desenvolveu uma patente de tecnologias e técnicas que fornecem pesquisas de marketing muito mais precisas e direcionadas ao público-alvo.

Após as estratégias adotadas pela Nielsen, os resultados foram favoráveis e paulatinamente a Nielsen está reobtendo sua vantagem competitiva. Analisando de acordo com os preceitos expostos por Barney (1991), os serviços da Nielsen voltaram a ser valiosos, pois percebeu-se através da compra e análise de dados obtidos por empresas de televisão a cabo/satélite que os medidores de pessoa da Nielsen fornecem análises mais acuradas, por cruzarem dados de audiência com dados demográficos; são raros pois com a compra da empresa Integrated Media Measurement Inc., a Nielsen é capaz de obter dados de audiência fora de casa através de assinatura sonora presente nos programas televisivos, sendo este fator raramente observado por seus concorrentes; a aquisição da empresa NeuroFocus permitiu a Nielsen ter serviços inimitáveis, por conta da tecnologia patenteada da empresa, que obtém dados da reação dos telespectadores ao serem expostos a uma propaganda utilizando-se da Neurociência; por fim, mas não menos importante, os serviços da Nielsen são sem substitutos quando falamos de análise da eficácia do Marketing indireto, realizado pela empresa IAG research, pois este mercado não é explorado por nenhuma outra empresa.

Através desta análise é possível concluir que, acompanhando as mudanças nas tecnologias, a Nielsen precisou entender e expandir seu alcance às três telas (televisão, computador e dispositivos móveis), adaptando-se à digitalização, além de perceber que o mercado passava por uma grande revolução por conta do enorme volume de dados produzidos por conversores de televisão a cabo/satélite que são explorados por empresas concorrentes. Com o uso e análise apropriados do Big Data, a Nielsen Holdings caminha para readquirir a vantagem competitiva sustentável que transformou, no passado, o mercado de análise e medição de audiência televisiva em um verdadeiro monopólio.

4.3 CASO RAMCO CEMENTS

A Ramco Cements Limited (RCL), caso relatado por Dutta e Bose (2015), é uma organização de manufatura de cimento fundada em 1961 que atua no sul da Índia. É a principal empresa do Ramco Group que opera em diversas áreas, incluindo indústrias têxteis, cimento, produtos de fibras de cimento, software, algodão cirúrgico e biotecnologia. O faturamento anual do grupo Ramco Group é de US\$800 milhões, o que os torna uma das maiores organizações do sul da Índia, com planos de expansão. A RCL é a quinta maior produtora de cimento do país e seu principal produto é o cimento Portland, que se destaca por suas características e propriedades (ASSOCIAÇÃO BRASILEIRA DE CIMENTO PORTLAND, 2009). Tendo sua sede em Chennai, quase todas as fábricas e unidades da RCL estão situadas no sul da Índia, com exceção de uma unidade de moagem em Kolaghat, localizada no leste do país. Atualmente, a empresa possui cinco fábricas de cimento, três unidades de moagem, duas fábricas de embalagem, uma fábrica de processamento de cimento a seco e uma fábrica de processamento de concreto. A capacidade combinada de suas fábricas é de 16 milhões de tonéis de cimento. Para uso interno de energia em suas instalações, a RCL também é grande produtora de energia eólica, possuindo seis parques eólicos com moinhos de vento que geram um total de aproximadamente 159 megawatts.

A RCL sempre procurou aplicar em seus processos ferramentas de alta tecnologia, destacando-se por ser uma das poucas empresas de manufatura de cimento que está engajada em tornar a tecnologia da informação (TI) como uma das principais preocupações da empresa. Ao longo dos anos, a empresa implementou diversos projetos de TI como sistemas de softwares baseados em lógica difusa para controle de processos, tecnologia de pré-calcinador, controladores lógicos programáveis, tecnologia de mineração à céu aberto, moinhos verticais para moagem de cimento, entre outros. Recentemente, a empresa implementou um sistema

centralizado, com uma base de dados única, de forma a controlar os custos da empresa com transparência e confiabilidade nos dados analisados.

Pensando em padronizar os processos e relatórios de todas as fábricas e unidades da RCL e auxiliar no processo decisório através da análise de dados, criou-se um sistema baseado em *Entreprise Resource Planning*¹³ (ERP – em português, planejamento de recurso corporativo, tradução nossa). Porém, a implementação deste sistema inicial não foi bem sucedida por diversos fatores. Em primeiro lugar, não houve um estudo aprofundado para definir quais requisitos eram necessários para o desenvolvimento do sistema, levando em consideração o contexto e os problemas da empresa. Sendo assim, quando o sistema foi implementado, ele apresentou falhas graves, como a falta de padronização. Além disso, não foi oferecido aos funcionários treinamento compatível, o que resultou na falta de conhecimento para a utilização dos dados fornecidos pelo sistema como auxílio ao processo decisório.

Percebeu-se, então, que era necessário a criação de um novo sistema que pudesse analisar de forma clara o Big Data produzido pela empresa e, assim, fornecer subsídios aos tomadores de decisão. Os dados produzidos pela empresa podem ser classificados em duas categorias: O primeiro trata-se de dados operacionais na forma de parâmetro de processos-chaves, que medem a eficácia e eficiência dos processos da empresa; o segundo são dados de ERP concernentes a diversos processos, como a logística de saída, com detalhes como: Pedidos de clientes, seus registros passados, histórico de crédito, pagamentos, aquisições, status do envio de pedidos, etc. Os dados operacionais são úteis apenas para decisões restritas a cada uma das fábricas, não sendo relevantes quando trata-se de decisões estratégicas mais amplas. Já os dados de ERP são essenciais para extrair informações úteis como o padrão de comportamento de seus clientes, o rastreamento dos *Key Performance Indicators* (KPI – em português, indicadores-chave de performance, tradução nossa), etc. Os dados eram inseridos automaticamente no sistema, no momento exato que o processo estava acontecendo e, para prevenir informações errôneas ou fraudulentas e para garantir a transparência, não era permitido inserir dados manualmente para fins de relatórios. Todos os dados eram coletados diretamente da base de dados unificada.

Apesar deste sistema produzir subsídios mais acurados para o processo decisório que anteriormente, ainda haviam algumas questões a serem resolvidas. Primeiramente, a empresa

¹³ *Entreprise Resource Planning* (ERP) “[...] é um sistema que facilita o fluxo de informações dentro de uma empresa, integrando as diferentes funções: manufatura, logística, finanças, recursos humanos e engenharia, entre outras. Ainda apresenta uma base de dados que opera em uma única plataforma e que consolida todas as informações em um único ambiente computacional.” (LAUGENI; MARTINS, 2005, p. 388).

precisava de um sistema que avaliasse as KPIs, com uma visualização fácil das metas e do que foi conquistado até o momento. Outra questão importante era desenvolver um sistema responsivo, ou seja, que se ajustasse a qualquer dispositivo, sejam computadores, celulares, *tablets*, etc. Outro ponto importante a ser tratado era a redução de custos no que se refere a logística de saída, não era mais viável a empresa fazer esse controle através de planilhas estáticas. Com esses problemas apresentados, a RCL resolveu executar um projeto para implementação de um sistema que trouxesse soluções.

A primeira etapa do projeto consistiu em extensiva pesquisa interna visando compreender as falhas do sistema existente. As informações foram coletadas através de entrevistas e observação dos funcionários usuários do sistema, além de análise do histórico de dados e dos registros de chamadas do *help desk*. Outro fator importantíssimo desta fase foi a padronização dos KPIs que poderiam ser utilizados para avaliação de performance. Antes, existiam 800 KPIs e a equipe reduziu-os para 200, desenvolvendo processos de cálculo para cada um deles, customizando-os se necessário.

Na sequência, realizou-se uma pesquisa externa em que a equipe de Tecnologia da Informação (TI) da RCL pesquisou quais sistemas de análise de dados já existiam no mercado, suas tecnologias e se elas poderiam ser customizadas para atender os requisitos da empresa. Estudou-se sobre diversas tecnologias, como geração de relatórios em tempo real, interface de programação de aplicação do Google Maps, ferramentas de gráficos, gráficos Scalable Vector Graphics (SVG), HyperText Markup Language (HTML) 5 Canvas, etc. Outro fator importantíssimo analisado nesta etapa, relaciona-se à privacidade e a segurança da informação. O objetivo da RCL era que os dados coletados por seu sistema pudessem ser acessados por pessoas autorizadas em qualquer dispositivo, a qualquer dia ou momento, em qualquer lugar. Taurion (2013a) nos alerta sobre os perigos da triangulação de dados extraídos do Big Data, pois mesmo quando se trata primariamente de dados que não exporiam um cliente ou a empresa, quando combinados podem revelar identificação pessoal íntima ou mesmo segredos industriais. A conclusão das pesquisas encontradas pela equipe de TI é que a melhor solução seria a RCL produzir seu próprio sistema ERP. No entanto, várias informações técnicas e ideias sobre tecnologias existentes foram extraídas através das consultas e entrevistas com vendedores de sistemas pré-existentes.

Formou-se então a equipe do projeto. É essencial que em um projeto de Big Data, as equipes sejam multidisciplinares, ou seja, formadas por pessoas de diversas áreas. O projeto da RCL foi chamado de Análise e Otimização de Dados e contou com doze pessoas: Dois funcionários da equipe de TI, dois funcionários da Ramco Systems, empresa de criação e

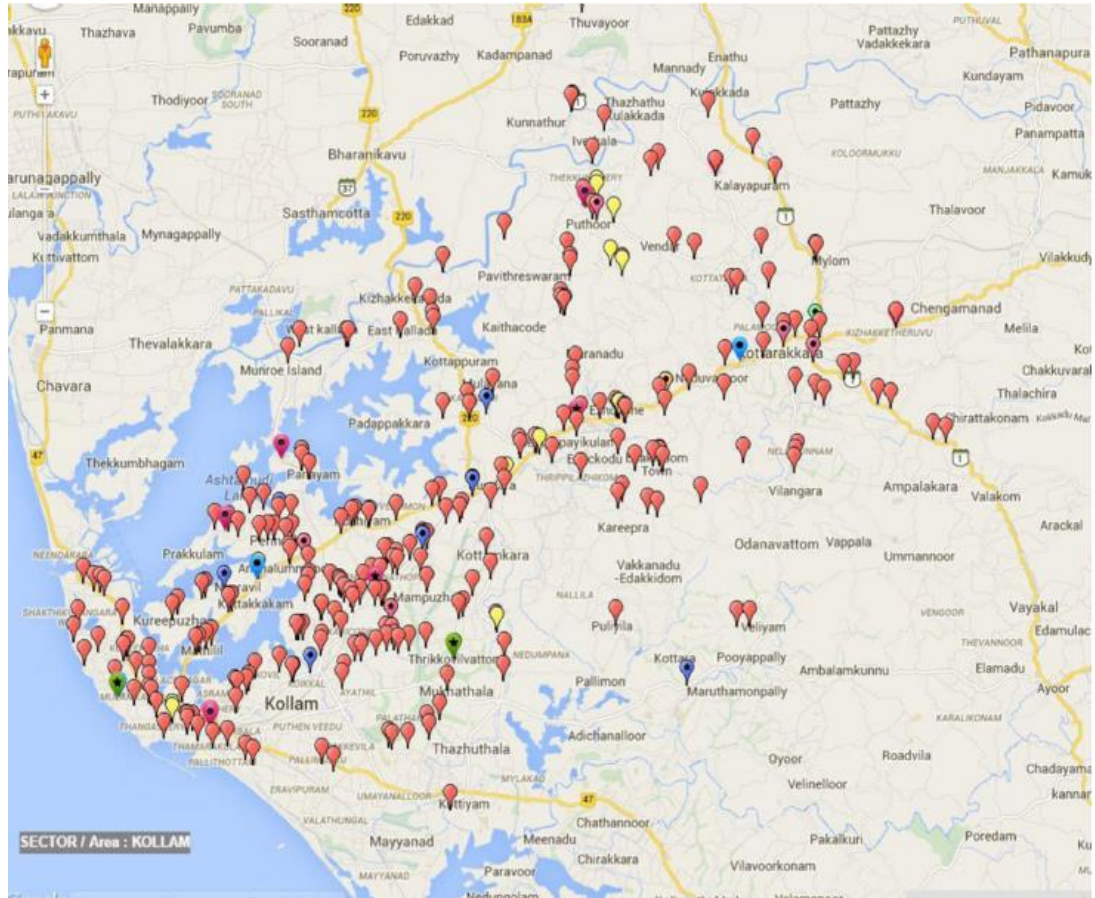
implementação de softwares do Ramco Groups; e seis funcionários de áreas funcionais da empresa, como *Marketing*, por exemplo. A grande vantagem de possuir pessoas das áreas funcionais da empresa é que no momento da implementação do novo *software*, elas serviriam de guia e motivador interno para a melhor aceitação por parte do setor.

O mapa do projeto foi dividido em nove etapas subsequentes, sendo elas: Desenvolver os requisitos-chave do projeto, criar o escopo do projeto com informações sobre *stakeholders*, dividir as responsabilidades entre os membros da equipe, modelar os dados que vão suprir os requisitos, desenvolver a lista de entregas previstas, conduzir treinamentos e *workshops* com a equipe para que todos estejam alinhados com os objetivos do projeto, criar cronograma viável e delinear as etapas do projeto, conseguir aprovação do projeto, conseguir orçamento previsto, decidir a periodicidade dos relatórios de andamento do projeto.

O projeto criou três aplicativos que foram incorporados ao novo sistema da RCL, todos centralizados e utilizando dados extraídos de uma base de dados unificada. São eles: Ramco GeoApps: Ferramenta de análise de dados baseada na interface do Google Maps que captura e disponibiliza a geolocalização de toda a logística de saída, ou seja, mostra a localização dos clientes que solicitaram uma encomenda e a localização dos caminhões que saíram para entrega. Esses dados são cruzados com os existentes na base de dados e através de análise em tempo real permite visualização de possíveis problemas como o não recebimento da entrega, a performance do funcionário de vendas, análise da concorrência na área, etc. Outros ativos da RCL, como dados sobre os parques eólicos, também são passíveis de captura, cruzamento e análise através do aplicativo. O grande diferencial é a visualização da análise de dados que é mapeada e diferenciada por cores, de acordo com seu KPI (por exemplo, volume de vendas, dívidas ativas, crescimento, etc.). As cores diferenciadas mostram o crescimento, estagnação ou declínio do indicador.

A figura 7, a seguir, demonstração a visualização da análise de dados.

Figura 7 – Ramco GeoApps

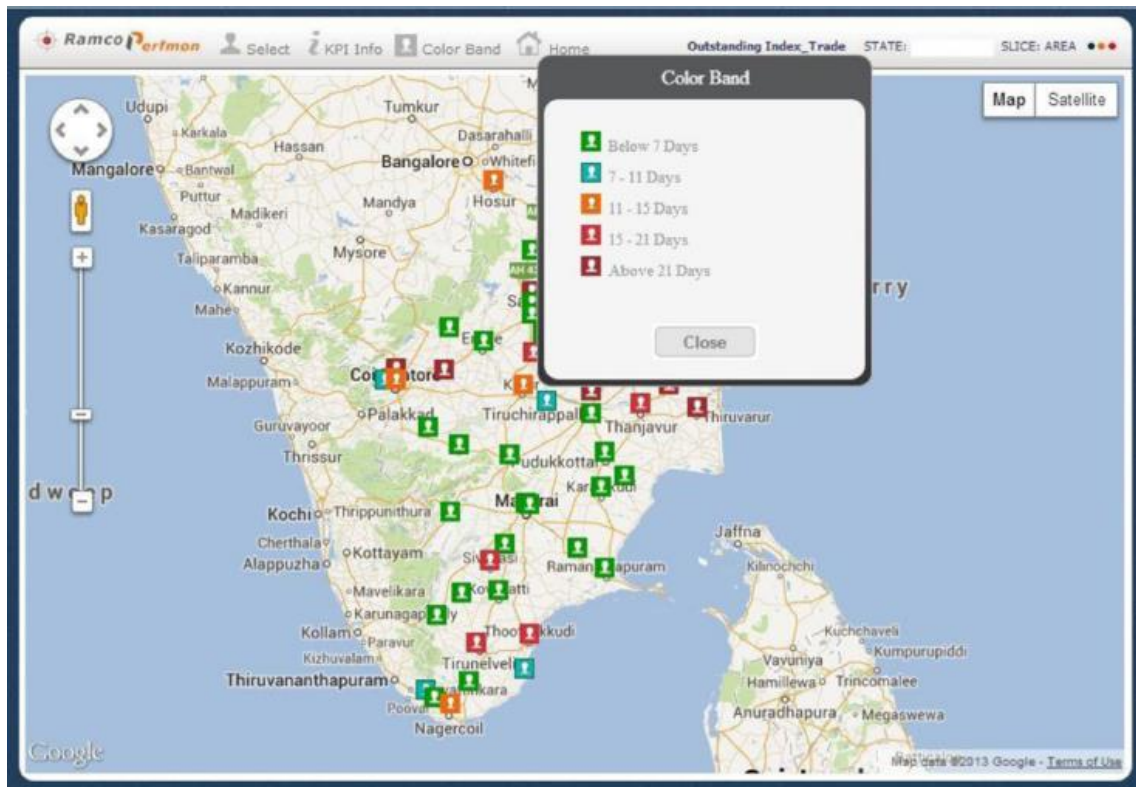


Fonte: (DUTTA; BOSE, 2015, p.301).

Ramco Perfmon: Painel que apresenta os KPIs (por exemplo, horário das entregas, aumento das vendas, etc.) de cada funcionário, com suas conquistas e falhas. A interpretação dos dados é feita através de gráficos, cores, mapas de calor, mapas temáticos, entre outros. O principal objetivo é tornar complexos dados de Big Data em informações simples e compreensíveis. As cores são utilizadas para mostrar o nível de performance dos funcionários. Os pontos de cores escuras apontam rapidamente os pontos com problemas ou anomalias, facilitando aos gestores a decisão de promover uma ação de melhoria específica.

A figura 8, a seguir, apresenta os KPIs de cada funcionário.

Figura 8 – Ramco Perfmon

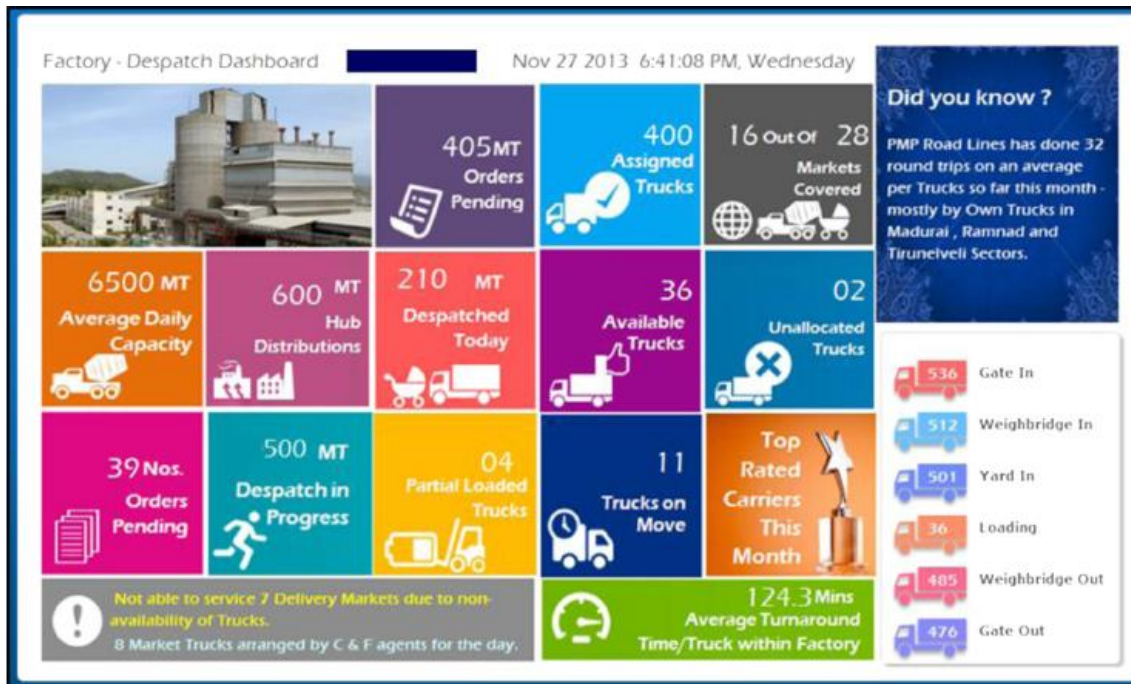


Fonte: (DUTTA; BOSE, 2015, p.302).

Ramco APO: Aplicativo de planejamento e otimização da análise e visualização de dados criado para melhorar a logística de saída, baseando-se em dados em tempo real. Cada encomenda recebida é analisada pelo aplicativo que define a melhor fonte de entrega (que podem ser fábricas, unidades de moagens, armazéns, etc.) de acordo com sua disponibilidade de estoque, custo de logística e outras restrições. Todo o processo é realizado em tempo real. Através da interface do Ramco APO é possível visualizar detalhes do planejamento, uma visão geral das operações de despacho e resumo detalhado analisando as rotas a serem seguidas. Pelo menu de detalhamento, o gestor pode ter mais informações relevantes, como por exemplo, detalhes sobre o caminhão de entrega, localização atual, horário de chegada previsto, peso sendo transportado, tipo de caminhão, dados sobre o proprietário e possíveis razões para atrasos, quando houverem. Esse sistema permitiu a RCL identificar as razões exatas dos atrasos nas entregas e aumentou sua eficiência, reduziu os custos de logística e melhorando sua relação com os clientes.

A figura 9, a seguir, apresenta o aplicativo de planejamento e otimização da logística de saída da empresa.

Figura 9 – Ramco APO



Fonte: (DUTTA; BOSE, 2015, p.302).

O sistema criado pela RCL foi implementado integrando-se totalmente ao antigo sistema e foram realizados diversos treinamentos com os funcionários de forma que o novo sistema fosse bem aceito pelos mesmos. Durante os treinamentos, o foco foi demonstrar aos gestores as facilidades que o sistema trazia para auxiliar no processo decisório. A confiabilidade dos dados é um fator crucial para que a tomada de decisões seja feita da forma mais correta.

Os resultados alcançados pela implementação deste sistema pela RCL foram de muito sucesso, facilmente visualizados por números. A redução de moratórias, penalidades, perdas e prejuízos decorrentes dos processos de logística de saída foi de até 70% e a melhoria na apuração de encomendas foi de até 40%. Além disso, após a implementação dos aplicativos, os gestores puderam melhor visualizar os locais com mercado em potencial, criando planos de *marketing* proativos que melhoraram as vendas e aumentaram sua fatia de mercado em 20% a 30%. Outra melhoria significativa foi a padronização do peso de sacos de cimento, que agora são monitorados pelo sistema para checar variações de peso. Através dessa medida, a variação de peso nos sacos de cimento caiu de 40% para apenas 2%, isso significa uma economia de custos de quase 2 milhões de dólares por ano para a RCL.

Outro ponto de melhoria que satisfaz os objetivos iniciais da implementação do novo sistema diz respeito ao auxílio nas tomadas de decisão. Os gestores agora podem focar no que

é realmente importante, que estará assinalado em cores de fácil visualização, permitindo uma ação quase imediata e com base em dados de confiança. Isso economiza tempo para dispensar na discussão de planos de ação ao invés de verificar a veracidade dos dados. O sistema também alerta o gestor quando há mudança de cor, ou seja, quando há algum problema ou anomalia, permitindo que ações reparatórias sejam tomadas imediatamente. A comunicação com a equipe de vendas também é facilitada pelo sistema, pois ao clicar em um vendedor é possível enviar um e-mail ou SMS diretamente.

Os setores de Marketing e Vendas também foram beneficiados, de forma que agora os vendedores podem visualizar dados sobre seu desempenho e de seus colegas e até mesmo sobre a rede de distribuição de seus competidores. Desta forma, eles podem criar estratégias de vendas mais eficazes e focar em um mercado em especial, baseando-se em sua competitividade. Além de ter também informações sobre armazéns de cimento, entregas e clientes-chave, permitindo criar estratégias logísticas, melhorando o serviço para os clientes.

Por fim, o sistema aumentou o nível de satisfação de seus clientes, fator comprovado comparado as taxas de abandono de clientes (53% a menos que no ano anterior), a redução do tempo de entrega (51% mais rápido que no ano anterior), além do feedback geral ser bastante positivo.

Conclui-se através deste estudo de caso que projetos de big data demandam muito esforço e investimento de toda a empresa. É necessário um planejamento minucioso que indique principalmente qual objetivo a organização pretende atingir com este projeto. Na implementação do projeto é necessário atentar-se para os treinamentos e para a motivação dos funcionários em usarem um novo sistema. A RCL conseguiu, através de seu projeto de Big Data, melhorar seu processo decisório, pois conta agora com uma quantidade muito maior de dados, facilmente visualizados e analisados pelos gestores. Além disso, aumentou sua vantagem competitiva, pois seus concorrentes não possuem a tradição de investir em grandes projetos tecnológicos. Adquirir vantagem competitiva no ramo de manufatura de cimentos não é tarefa fácil. Considerado um *commodity*¹⁴, para destacar-se no mercado, a empresa precisa ir muito além de melhorar seus produtos ou ajustar seu preços. O principal fator competitivo, neste caso, refere-se primordialmente aos serviços prestados. Com a melhora de seus serviços, a Ramcon Cements adquiriu larga vantagem com relação a seus concorrentes.

¹⁴ “Commodity pode ser definida como um ativo físico que possui características padronizadas, de ampla negociação em diversas localidades, que pode ser transportado e armazenado por um longo período de tempo.” (KALDOR, 1939, p.3; COPELAND; WESTON, 1988, p.302; GERMAN 2005, p.1 apud PEREIRA, 2009, p. 16)

4.4 CASO DUNNHUMBY

A empresa Dunnhumby, fundada na Inglaterra em 1989, é a empresa líder mundial em ciência dos consumidores, ou seja, a organização é especializada em entender os consumidores através de análise de dados. O objetivo da empresa é prever as necessidades dos clientes, atuais e futuros, de forma a fidelizá-los. A organização conta atualmente com mais de 2.000 funcionários e possui filiais em 23 países, incluindo o Brasil. A Dunnhumby analisa o Big Data de quase um bilhão de consumidores em todo o mundo, cruzando os dados para obter informações sobre o que os clientes querem, onde querem e quanto pagariam pelo que querem. De posse da análise dessas informações é possível personalizar suas experiências, oferecendo o que precisam, quando precisam, onde precisam e pelo preço que querem pagar, o que leva a clientes satisfeitos. Essa satisfação faz com que os consumidores confiem na empresa e voltem a comprar nela, tornando-se clientes fiéis. A Dunnhumby é a empresa responsável pelo sucesso do Tesco Clubcard, programa de lealdade da Tesco, maior rede de varejo do Reino Unido. (DUNNHUMBY, 2016).

A Dunnhumby, caso relatado por Martinez e Walton (2014) é uma empresa de análise de dados de consumidores e possui, entre seus funcionários, diversos cientistas de dados com ótimas qualificações que são capazes de criar modelos de análise preditiva de dados excelentes. No entanto, a empresa não estava satisfeita. Estudos apontam que utilizando os recursos internos da empresa é possível chegar a um modelo preditivo de dados referência, porém esse modelo ainda não é o ideal ou o ótimo. Para encontrar o modelo ideal seria necessário utilizar-se de recursos externos, ou seja, da ideia de mentes que não necessariamente trabalham na empresa. Não importa se os cientistas da Dunnhumby são excepcionais, sempre existirão mentes ainda mais brilhantes fora da organização. Para solucionar este problema, a empresa buscou soluções de inovação aberta.

Inovação aberta é um termo cunhado por Henry Chesbrough que significa “[...] um paradigma que assume que as empresas podem e devem usar ideias externas, além de ideias internas, [...] para criar valor.” (CHESBROUGH, 2006, p.24, tradução nossa). O recurso escolhido pela Dunnhumby foi o *crowdsourcing*, cuja tradução literal significa uma junção das palavras *crowd* (multidão, tradução nossa) e *outsourcing* (terceirização, tradução nossa). Segundo Martinez e Walton (2014, p.203, tradução nossa), *crowdsourcing* é “[...] uma nova forma de tecnologia aberta de entrada em que empresas e instituições selecionam um processo de busca de ideia ou soluções tradicionalmente feito por funcionários internos e terceirizam na

Internet, para um grupo geralmente grande e indefinido de pessoas, chamadas de multidão [...]”.

Para arriscar-se a realizar uma competição de *crowdsourcing*, no entanto, a Dunnhumby, com receio de cometer erros que comprometessem o nome da organização, optou por contratar uma empresa que possuísse mais experiência no assunto: a Kaggle, líder mundial em competições de modelagem preditiva. Fundada em 2010 por Anthony Goldbloom, um analista de dados australiano, seu objetivo é mediar conhecimento entre empresas, governos e pesquisadores que pretendem resolver seus problemas acessando uma rede de contribuintes externos à sua empresa, competindo para encontrar a melhor solução. A rede de contribuintes da Kaggle possui mais de 100.000 cientistas de dados e estatísticos, que buscam dados reais de empresas com os quais possam desenvolver e refinar suas técnicas de análise.

A Kaggle oferece três tipos de competições a seus clientes: abertas, privadas e em sala de aula. A competição aberta caracteriza-se na definição de um problema por parte da empresa participante, que disponibiliza seus dados à Kaggle, encarregada de torná-los anônimos e, em seguida, disponibilizá-los à sua rede de cientistas em formato de competição. Os participantes devem então enviar suas respostas através de um portal online. Ao final da competição, o vencedor ganhará um prêmio pela propriedade intelectual de sua produção. A competição privada se realiza com a participação de convidados que devem assinar acordos de não-divulgação dos dados. Os antecedentes dos convidados também são verificados. Os participantes são selecionados baseados em performances anteriores em outros concursos da Kaggle. Nesta modalidade, todos os participantes ganham um valor em dinheiro, sendo que os vencedores ganham um prêmio de maior valor. Por fim, a competição em sala de aula refere-se a competições entre estudantes que desafiam seus colegas com base nas matérias ministradas no curso de estratégia e mineração de dados. A vantagem é que não há custos com prêmios, pois os estudantes participam gratuitamente.

A Dunnhumby optou por testar uma competição aberta de *crowdsourcing* em parceria com a Kaggle com três objetivos principais: testar as adequações de análise de dados nesses concursos, explorar a complexidade de promover uma competição e medir o interesse dos analistas de dados participantes da competição. Para tal, optou por escolher um problema não-crítico da organização e lançou a competição desafio do comprador, que consistia em prever a data e os gastos de determinado cliente em sua próxima visita em uma loja de varejo. Os participantes (individuais ou em grupo) recebiam um conjunto de dados com detalhes de cada visita ao mercado de 100.000 consumidores no período de um ano (01 de abril de 2010 à 31

de março de 2011). Em cada visita, haviam dados com a data e o valor gasto pelo consumidor. A tarefa então consistia em prever a data e o valor gasto na próxima visita de cada um dos consumidores. Com relação ao valor, era considerado uma variação de até 10 dólares da despesa real. O vencedor da competição seria o participante que previsse o maior número de visitas (data e valor) corretas.

As respostas deveriam ser postadas em um portal da Kaggle somente com as informações sobre as visitas, os modelos e códigos só deveriam ser submetidos após o término da competição. As respostas postadas eram validadas e pontuadas e os resultados eram expostos em um painel de classificação *online*, atualizado em tempo real. Os participantes foram limitados ao máximo de duas respostas por dia. Os funcionários da Dunnhumby também foram encorajados a participar, porém, como não poderiam receber recompensas financeiras, o prêmio para o funcionário melhor colocado na competição foi um iPad.

Um fator motivacional muito importante utilizado pela Kaggle foi a tabela de classificação *online* dos participantes. É possível acompanhar os desempenhos em tempo real e conforme respostas corretas eram enviadas, um participante poderia ultrapassar outro, o que gerava um senso de competição e motivava a multidão a trabalhar arduamente para superar seus oponentes. Além disso, eles recebiam feedback sobre seu trabalho imediatamente, o que raramente acontece em suas empresas e entre pares.

A competição ocorreu durante dois meses, entre o início de agosto e o final de setembro de 2011. O prêmio total oferecido pela Dunnhumby foi de 10.000,00 dólares, divididos entre o primeiro lugar, que ganharia 6.000,00 dólares, o segundo lugar, que ganharia 3.000,00 dólares e o terceiro lugar, que ganharia 1.000,00 dólares. O valor do prêmio foi pensado levando em consideração que os competidores sentem-se desmotivados quando uma grande empresa, com um alto valor de receita anual, oferece um prêmio não compatível. Essa informação foi obtida com base em antigas competições da Kaggle, em que outras empresas tiveram problemas com relação a isso. A repartição do prêmio entre os três primeiros lugares também foi sugestão da Kaggle, tendo em vista que não diminuiria a qualidade dos participantes e daria a Dunnhumby o direito sobre três modelos de análise de dados, ao invés de um. Os participantes sentiam-se motivados não só pelos prêmios monetários, mas também pelo status alcançado entre a rede de cientistas de dados da Kaggle. Os participantes internos da Dunnhumby aspiravam provar que poderiam alcançar o mesmo nível de cientistas de várias áreas diferentes, além de impressionar seus gestores sêniores obtendo bom desempenho.

A rede de cientistas de dados da Kaggle conta com mais de 100.000 pessoas, dos mais variados perfis, de diferentes áreas e diferentes partes do mundo. Isso é um grande diferencial na criação de modelos de análise de dados, levando em consideração que normalmente a Dunnhumby não teria acesso a expertise desses cientistas de tão variadas origens, muitas vezes com técnicas e ferramentas desconhecidas pelos funcionários da empresa.

A competição desafio do comprador da Dunnhumby (MARTINEZ; WALTON, 2014) realizada no período de dois meses no ano de 2011, encerrou-se com a participação de 537 pessoas que submeteram 2029 respostas diferentes. Foi a competição da Kaggle de maior sucesso até então. O vencedor foi um professor de estatística de Moscou, Rússia, enquanto o segundo lugar foi um estatístico de Indiana, nos Estados Unidos. O funcionário da Dunnhumby que obteve a melhor colocação ficou em 50º lugar, o que prova o sucesso da competição de *crowdsourcing*. Os modelos e as abordagens vencedoras foram apresentadas para o conselho da empresa junto com os resultados e modelos preditivos. O principal objetivo da Dunnhumby, como discutido anteriormente, não era solucionar o problema e obter modelos preditivos ideais, mas sim testar o *crowdsourcing*, o que se provou extremamente vantajoso. Através desta experiência, a Dunnhumby pôde experimentar como atrair as pessoas certas e motivá-las, impulsionando sua criatividade, que tipo de problemas de inovação podem trazer para a multidão e como criar uma competição atraente, protegendo os dados sensíveis e obedecendo as leis de proteção de dados.

Porém, através da execução da competição, a Dunnhumby percebeu que para organizar e produzir uma competição de *crowdsourcing* a participação de uma empresa como a Kaggle foi essencial. As dificuldades técnicas e logísticas da produção da competição poderiam perfeitamente ser construídas internamente pela empresa, no entanto, as dificuldades de reunir uma rede de analistas de dados tão numerosos e de variadas áreas como a da Kaggle, seria tarefa deveras trabalhosa. Portanto, grande parte do sucesso para obter tantas boas respostas deu-se ao fato da utilização da rede de cientistas da Kaggle.

A motivação dos participantes também é fator chave para o sucesso na utilização do *crowdsourcing*. Além da recompensa financeira, as tabelas de classificação atualizadas em tempo real eram essenciais para manter os participantes sempre motivados a melhorarem seus modelos, recebendo *feedbacks* por seu trabalho e esforçando-se para superar seus competidores. Um valor considerado apropriado para o prêmio também é importante. Não importa se será compartilhado entre várias pessoas, contanto que o montante seja atrativo e motive os participantes a dedicarem-se ao máximo.

Apesar do enorme sucesso do concurso desafio do comprador promovido pela Dunnhumby que comprovou a eficácia do *crowdsourcing*, um grande problema deveria apresentar-se como a maior preocupação das empresas que pretendem adotar as competições abertas: a privacidade dos dados. É impreterível que as informações sensíveis que possam ser extraídas através dos dados sejam apagadas ou mascaradas, o que é foi feito pela Kaggle, no entanto, é importante atentar-se que os participantes destas competições são cientistas de dados extremamente talentosos e curiosos, capazes de cruzar os dados disponibilizados com outros conjuntos de dados disponíveis na Internet ou em bases de dados, conseguindo, em alguns casos, ter acesso a dados sensíveis que identifiquem os indivíduos e suas ações. No exemplo da Dunnhumby, o país e a moeda em que foram feitas as compras disponibilizadas para os participantes foram ocultadas, no entanto, logo no início da competição, um dos participantes cruzou as datas com certos feriados em determinado país e, sendo assim, descobriu a procedência dos mesmos.

Como discutido anteriormente, a privacidade e o acesso à informação deve ser uma das maiores preocupações das empresas que pretendem realizar projetos de Big Data. Apesar das barreiras que garantam a privacidade estarem se rompendo e as empresas terem acesso a nossos dados pessoais de formas cada vez mais fáceis, ainda existem diversas leis que protegem os dados e restringem o movimento das empresas. Sendo assim, é mister ter muita atenção ao tornar públicos dados que podem ser associados a um indivíduo. Casos como esse podem levar a empresa a processos judiciais e prejudicar a marca de forma profunda.

No caso da Dunnhumby, os dados dos consumidores eram internos a empresa, não estando disponíveis de forma alguma *online*, portanto era quase impossível fazer os cruzamentos necessários para associações com indivíduos, no entanto, a empresa considera que no futuro, caso deseje trabalhar novamente com *crowdsourcing*, a privacidade dos dados será uma de suas maiores preocupações.

Algumas vezes, as empresas enfrentam um problema cuja solução não faz parte do conjunto de técnicas e *expertises* de seus funcionários. Nesses casos, utilizar o *crowdsourcing* pode ser a solução ideal. No estudo de caso analisado, a Dunnhumby optou por utilizar a competição desafio do comprador para resolver um problema não crítico, no entanto, considerando a velocidade com que as empresas podem conseguir respostas favoráveis, e mesmo a solução ideal, essa modalidade de inovação aberta pode ser usada também para problemas críticos, estratégicos ou urgentes, desde que tomadas as devidas precauções com relação a proteção de dados. As competições de *crowdsourcing* podem ser ideais quando um problema precisa ser resolvido através de experimentações e pode possuir diversas soluções.

Prova-se uma modalidade interessante para empresas que enfrentam os desafios de analisar Big Data. Além disso, os custos de uma competição de *crowdsourcing* restringem-se aos prêmios e ao contrato de empresas como a Kaggle.

Conclui-se, portanto, que adotar concursos de *crowdsourcing*, quando disponibilizado um grande grupo de cientistas de dados, constantemente motivados e cujos dados sensíveis são protegidos, pode ser de grande utilidade para as empresas que pretendem se arriscar na análise de Big Data. Através de competições de inovação aberta é possível receber um número incontável de soluções de modelagem e mineração de dados em um curto período de tempo, por um valor não elevado.

4.5 CASO NETFLIX

A Netflix, Inc., caso relatado por Kumar (2013), é uma conhecida empresa de aluguel de DVDs e transmissão de vídeos através da Internet pelo sistema de assinatura, é um sucesso sem precedentes neste setor. Fundada na Califórnia, Estados Unidos da América (EUA), em 1997, a Netflix começou seus serviços como uma empresa de aluguel de DVDs através dos correios. Em 1999, a empresa criou um sistema de assinaturas em que os clientes pagavam um valor mensal para receber em casa número ilimitado de DVDs. Em 2000, a empresa já se mostrava voltada para análise de dados, prevendo recomendações a seus assinantes baseadas nas notas enviadas por eles para cada DVD assistido. Em 2007, veio a grande revolução: a Netflix criou um sistema de assinatura para transmissão de vídeos em qualquer dispositivo que possa ser conectado a um televisor (computador, *notebook*, *tablet*, telefone celular, etc.). No início, este serviço só estava disponível nos EUA, mas em 2010, seus serviços de transmissão de vídeos começaram a se expandir globalmente. Hoje, a Netflix está presente em todos os países do mundo, exceto a China, e conta com quase 75 milhões de assinantes. Sua receita líquida no ano de 2015, tanto para o serviço de aluguel de DVDs, quanto para o serviço de transmissão de vídeos, foi de 122 milhões de dólares. Em 2013, a Netflix começou a investir na produção de conteúdo original, lançando a série de sucesso *House of Cards*. Em 2015, a empresa produziu 450 horas de conteúdo original e os planos são de aumentar ainda mais seus investimentos neste setor, produzindo séries novas e *remakes* de antigas séries de sucesso. (NETFLIX, 2016)

O grande diferencial da Netflix, que a torna líder no setor de transmissão de vídeos por Internet em todo mundo, é a sua análise contínua de Big Data. Conhecida por ser uma empresa voltada para dados, praticamente todas suas decisões estratégicas são tomadas após

criteriosa análise dos mais diferentes tipos de dados produzidos por seus assinantes e por dados fornecidos por outras empresas, como a Nielsen, por exemplo. Os dados coletados são os mais variados: histórico de transmissão de vídeos, notas atribuídas pelos assinantes, recomendações de terceiros, etc. Porém, a Netflix se atenta a todos os detalhes para conhecer seus assinantes de forma bastante acurada; a empresa consegue, através da análise de seus dados, descobrir, por exemplo, o que pessoas em uma determinada localidade assistiram no sábado à noite utilizando seus *tablets*. Existem tabelas de dados para entender quais filmes foram interrompidos antes dos créditos finais. Através dessa análise é possível conhecer cada um de seus assinantes analisando seus padrões de comportamento ao assistir vídeos: quantas vezes a pessoa pausa, se ela troca um vídeo por outro ou se interrompe a transmissão na metade, se é um *binge viewer*¹⁵, quais dispositivos costuma utilizar para acessar a Netflix, quais os dias e horários que costuma assistir vídeos e, principalmente, quais são suas preferências, analisadas através de seu histórico e de suas classificações, ou seja, as notas atribuídas pelo assinante para cada filme ou série. A combinação desses dados conferiu ao Netflix um entendimento de seus assinantes como nenhuma outra empresa de transmissão de vídeos ou rede de televisão jamais obteve.

Todos esses dados coletados pela Netflix, além de outros como dados demográficos sobre seus assinantes, contratos e licença de transmissão de filmes e séries, todos os vídeos disponíveis em todos os países e suas devidas legendas, são armazenados por pelo menos dez anos, o que significa uma base de dados crescente e com um volume imenso, armazenada nas nuvens. Esses dados são inseridos em algoritmos sofisticados que são capazes de realizar operações muito complexas, por exemplo, quando um assinante atribui uma nota a um filme ou série que não assistiu através do Netflix, essa nota terá um peso inferior àquela atribuída logo após o final da exibição do programa. O mesmo acontece com notas atribuídas antes do início dos créditos finais de um filme que terão um peso proporcional ao tempo de filme assistido.

A partir da análise dos dados coletados surge o grande diferencial da Netflix: suas recomendações. Em 2012, três quartos de todos os dados coletados foram utilizados em algoritmos para gerar recomendações personalizadas a cada um dos usuários. A empresa acredita que cruzando todos os dados como histórico de visualização, notas atribuídas, tempo de visualização, etc. é possível ter um perfil adequado do gosto do assinante e, a partir daí,

¹⁵ *Binge viewer* ou *binge watcher* diz respeito a um espectador que costuma assistir vários episódios de uma série ou programa seguidos.

cruza-se com a descrição dos programas e dados de outros usuários para chegar a recomendação adequada. Quando um assinante assiste um dos programas recomendados e gosta, a tendência é que ele assista mais programas e se sinta satisfeito com o serviço. A página inicial da Netflix também é altamente personalizada de acordo com o usuário. Os filmes são dispostos em linhas horizontais de acordo com sua legenda (lançamento, grandes sucessos, drama, comédia, ação, etc.) e a seleção dos gêneros é feita de acordo com os dados do usuário (histórico, notas, pausas, desistências, feedbacks, etc.). Além disso, também utiliza-se a similaridade, que é encontrada através das buscas feitas pelo usuário e os programas que selecionou para assistir depois.

A arquitetura de software que a Netflix utiliza é uma das bases para oferecer uma experiência rica a seus assinantes em termos de interação e recomendação. O modelo de classificação básico da empresa é feito através da classificação prevista x popularidade do filme/série, afinando o resultado através de testes A/B¹⁶. Um dos recursos de grande sucesso da Netflix é o “assista depois”, que caracteriza-se por uma tela que surge no momento que os créditos finais do filme começam a passar, recomendando três novos filmes a serem assistidos. No caso de uma série, o próximo episódio começa a ser exibido automaticamente em alguns segundos, caso o assinante não pressione o botão de parar.

Em 2009, a Netflix assinou um contrato com a Amazon Web Services (AWS), subsidiária de computação nas nuvens da Amazon e paulatinamente migrou todos seus dados para o serviço em nuvens, o que alavancou seus recursos, de forma que sua equipe de Tecnologia da Informação (TI) pôde, então, focar nas operações mais significativas, ao invés de se preocupar em verificar se os servidores estão funcionando corretamente. Além disso, a criação de um servidor em nuvens próprio custaria muito tempo e dinheiro para a Netflix que, terceirizando, paga apenas pelos recursos utilizados. Alguns dos recursos utilizados são o Amazon Elastic MapReduce, que avalia as transmissões de vídeo e cria métricas no que tange a performance e padrões de transmissões, melhorando a qualidade de transmissões de vídeos da Netflix; e o Hadoop, que permite análise do Big Data com a criação de representações visuais de padrões de tráfego de dados para cada tipo de dispositivo utilizado na visualização de vídeos, aumentando consideravelmente as competências de processamento, o que permite a equipe de TI da Netflix visualizar onde o tráfego de dados na rede está mais lento, permitindo

¹⁶ “Um teste A/B [...] consiste em mostrar diferentes versões de uma mesma página para fatias diferentes da audiência do site – e assim conseguir testar qual página tem melhor performance.” (TEIXEIRA, 2016, não paginado).

o aumento da capacidade de rede. Em 2012, a AWS gerenciava 95% dos requisitos de Big Data da Netflix.

Até 2011, a Netflix era uma empresa que reproduzia filmes e séries que já haviam sido exibidos nos cinemas ou em algum canal de televisão. No entanto, isso a deixava em desvantagem, pois muitas vezes os canais que produziam as séries, temendo a concorrência, não negociavam a licença de exibição de seus programas. A conhecida rede de televisão HBO, que produz séries de grande sucesso como *Game of Thrones*, por exemplo, não permite a exibição de nenhuma de suas séries na Netflix. Já a Showtime só permite a exibição de temporadas muito antigas. Além disso, a concorrência de diversas empresas de renome, como a Hulu, Amazon.com, Comcast, Apple, Google e outras, começou a surgir e ameaçar a supremacia da Netflix na transmissão de vídeos através da Internet. A solução para diferenciar-se foi começar a produzir conteúdo original.

Porém, sua estratégia era bem diferente da tradicional, utilizada pelas grandes redes de televisão na escolha de quais séries produzir. Normalmente, um episódio piloto é exibido para testar a aceitação do público sobre uma nova série e, quando bem aceito, a mesma é então financiada. A Netflix inovou e utilizou sua imensa rede de assinantes (em 2011 eram 29 milhões) e pediu que rotulassem as séries e filmes, criando metadados para descrevê-los, comentando sobre atores, diretores, roteiro, categoria e apelo global. Quando a série *House of Cards*, um *remake* de um famoso drama político da rede de televisão BBC, foi leiloada, a Netflix acessou seu banco de dados para prever o sucesso da mesma entre seus assinantes. Sabendo que a produção contava com David Fincher e Kevin Spacey, muito populares segundo os dados coletados, e levando em consideração o roteiro, a Netflix não hesitou em comprar duas temporadas completas (26 episódios) da série por 100 milhões de dólares. A estratégia de *Marketing* utilizada também foi diferenciada, foram criados dez *trailers* diferentes que eram disponibilizados na página inicial dos assinantes, baseados em suas preferências. Por exemplo, um assinante que costumasse assistir programas de ação, veria uma versão energética do trailer, já alguém que gostasse de romances e dramas, veria uma versão focada nos personagens. A primeira temporada completa de *House of Cards* foi disponibilizada no dia 1 de fevereiro de 2013 para todos os assinantes.

A Netflix (KUMAR, 2013) é líder absoluta do mercado de transmissão de filmes e séries pela Internet através de assinaturas, graças a bem-sucedida exploração do Big Data. Com o lançamento de sua primeira série original, *House of Cards*, a Netflix aumentou ainda mais sua vantagem competitiva, por utilizar uma estratégia única, analisando os dados de seus assinantes para saber, antes mesmo que eles assistam a série, que a mesma seria um sucesso.

Somente no dia de estreia de *House of Cards*, mais de 100 mil comentários foram postados em redes sociais, como o Facebook e o Twitter. A empresa de análise de comentários em redes sociais Fizziology declarou que 62% destes comentários eram positivos. Na IMDb, um dos mais renomados *websites* sobre cinemas e séries, *House of Cards* possui nota 9 e é a trigésima terceira série mais popular de acordo com pesquisa do site (IMDB, 2013). A empresa Unmetric, que também analisa redes sociais, declarou que o trailer da série teve 1.2 milhões de visualizações até a sua estreia, sendo o mais assistido do canal do Youtube da Netflix, o que é excelente para o *Marketing* da empresa. Até 2013, *House of Cards* era a série mais assistida do Netflix. A série também ganhou, até o momento, 19 prêmios, além de ter tido 123 indicações, nas premiações mais renomadas no que tangem séries, como o *Golden Globes* (ganhou em 2015 com o prêmio de melhor ator em série para Kevin Spacey e em 2014 melhor atriz em série para Robin Wright) e o *Emmy Awards*, que ganhou em seis categorias nos anos de 2013, 2014 e 2015. (IMDB, 2013)

O enorme sucesso de *House of Cards* foi estrategicamente planejado pela Netflix. Através de sua base de dados NoSQL Apache Cassandra, eles coletaram dados de seus clientes e analisaram, chegando ao número exato de assinantes que gostavam das séries de David Fincher e eram fãs do ator Kevin Spacey, além do número de pessoas que assistiram e gostaram da primeira versão britânica da série, transmitida pela BBC em 1990. Cruzando as informações, percebeu-se que muitos dos fãs de David Fincher e Kevin Spacey assistiram também *House of Cards* original. Esses dados provaram à Netflix que a série seria um sucesso antes mesmo dela estrear.

A Netflix apresenta, então, um caso de grande sucesso do uso de Big Data. Através dos dados minuciosamente coletados de todos seus 75 milhões de assinantes em todos os países do mundo (com exceção da China) e algoritmos de análise muito sofisticados, a empresa é capaz de prever, quase com precisão, o gosto de seus assinantes, personalizar seus conteúdos, recomendando filmes e séries, e produzir material original de grande sucesso. Essa estratégia dá clara vantagem competitiva sobre seus concorrentes, pois nenhuma outra rede de televisão possui análise de dados tão eficientes. Por exemplo, em 2011, 43 novas séries que estrearam em grandes redes de televisão foram canceladas após a primeira temporada devido à baixa audiência ou avaliação negativa dos espectadores. Uma pesquisa de um economista revelou que apenas 22% dos filmes lançados por Hollywood são realmente lucrativos. Além disso, a Netflix não precisa gastar muito dinheiro com *Marketing* para divulgar suas séries, já que ela sabe que o público irá assisti-las.

Os esforços da Netflix em lançar novas séries originais continua. Também em 2013, foram lançadas três novas séries: *Hemlock Grove*, um remake da série *Arrested Development* e *Orange is The New Black*, que se tornaram grandes sucessos, ganhando diversas premiações. Em 2014 e 2015, foram lançadas outras séries originais como *Narcos*, *Sense8*, *Bloodline*, *Master of None* e séries em parceria com a Marvel como *Demolidor* e *Jessica Jones*. Na lista das dez melhores novas séries do IMDB, as seis produções da Netflix aparecem, provando que sua estratégia é muito bem sucedida. Recentemente, a empresa também começou a investir na produção de filmes. Algumas das séries esperadas para o ano de 2016 que estão sendo muito comentadas em redes sociais, são novos episódios das clássicas séries *Full House* (agora chamada de *Fuller House*) e *Gilmore Girls*. Além disso, um drama político francês chamado *Marseille* com Gerard Depardieu e uma série brasileira chamada *3%*, dirigida por Pedro Aguilera, estão em fase de produção (NETFLIX, 2016).

No entanto, apesar das vantagens na análise de Big Data da Netflix, alguns problemas resultam do uso excessivo de dados. Por exemplo, sua grande dependência nos serviços da AWS, que esporadicamente podem não funcionar por uma sobrecarga nos servidores, o que impede os assinantes da Netflix de acessarem o serviço. Um caso crítico foi na véspera de natal do ano de 2012, época em que muitas famílias estão reunidas e utilizam o serviço para assistir filmes e séries juntos. Isso causou grandes prejuízos a empresa, pois seus clientes buscaram soluções na concorrência.

Outro fato de extrema importância a ser observado pela Netflix é a privacidade dos dados de seus clientes. Como relatado por Magalhães et al. ([2014?]), quando os usuários se afiliam a um site (no caso, assinam a Netflix), deve-se concordar com os termos de uso que são associados a uma política de privacidade, promovendo à empresa alguns direitos e deveres com relação aos dados. O grande problema é que a maioria das pessoas não lê detalhadamente esses termos de uso ao qual concordam e, portanto, não possuem ideia do que pode vir a ser feito com seus dados. A Netflix enfrentou um processo judicial em 2010 por infringir a privacidade de alguns de seus usuários ao promover um concurso de *crowdsourcing*, chamado Prêmio Netflix, em que o objetivo era criar um algoritmo de análise de dados. As queixas diziam que eram revelados dados pessoais de assinantes do Netflix, o que poderia revelar suas identidades.

A preocupação relacionada com a criatividade também é levantada por especialistas. A tendência é que, cada vez mais, séries e filmes sejam criadas baseadas no gosto de quem possivelmente irá assistir, para garantir o sucesso da mesma, no entanto isso restringe a criação de séries que as pessoas possivelmente iriam gostar, mas não sabem disso ainda por

nunca terem assistido nada do gênero. Essa estratégia restringe também a criatividade dos escritores, que serão sempre obrigados a escrever um tipo específico de roteiro, demandado pelas empresas.

No entanto, as experiências com Big Data na Netflix estão proporcionando à empresa, até o momento, sucesso crescente. Praticamente todo seu processo decisório é baseado no Big Data e os resultados mostram larga vantagem competitiva sobre seus concorrentes. Para manter-se no topo, a Netflix precisa ter cuidado com a questão do uso inadequado de dados, que pode levar a diversos problemas judiciais e que comprometam a marca, caso os mesmos não sejam utilizado de forma clara e preservando a privacidade de seus assinantes. Em 2016, por R\$19,90 mensais, é possível ter acesso a um catálogo incomparável de filmes e séries, além de produções originais, acessados de qualquer lugar do mundo (exceto a China) e com diversos tipos de dispositivos. A Netflix revolucionou, portanto, a maneira que consumimos entretenimento.

4.6 RESUMO DOS ESTUDOS DE CASO

O quadro 7, a seguir, apresenta um resumo dos estudos de caso analisados, divididos em cinco categorias: nome da empresa, contexto, objetivo, as tecnologias utilizadas e os resultados obtidos à partir da implementação de projetos de Big Data.

Quadro 7 – Resumo dos estudos de caso

(continua)

| Empresa | Contexto | Objetivo | Tecnologia(s) utilizada(s) | Resultado(s) |
|--|--|---|--|-----------------------------------|
| New South Wales State Emergency Service (WAMBA et al., 2015) | Organização governamental que presta serviços de emergência em caso de crises e desastres no estado de New South Wales, Austrália. Composta por 280 funcionários e 9000 voluntários. | Coletar, analisar, armazenar e difundir dados de diversas fontes, centralizados em uma única instância. | Plataforma colaborativa baseada no Microsoft SharePoint. | Otimização do processo decisório. |

Quadro 7 – Resumo dos estudos de caso

(conclusão)

| | | | | |
|---|--|---|--|---|
| Nielsen Holdings (PRESCOTT, 2014) | Organização americana de medição e análise de audiência televisiva criada em 1923, que monopolizava o mercado até o surgimento da digitalização e das televisões a cabo/satélite com seus conversores. | Reobter vantagem competitiva sustentável explorando o big data. | Medidor de pessoas. | Obtenção de vantagem competitiva. |
| Ramco Cements Limited (DUTTA; BOSE, 2015) | Empresa de manufatura de cimento criada em 1961 que atua no sul da Índia. É a principal empresa do Ramco Group. Quinta maior empresa do setor na Índia. | Criar um sistema integrado que melhorasse seu processo decisório, sua logística de saída e que diminua custos operacionais. | Tecnologia de ERP. Interface baseada no Google Maps. | Melhoria no processo decisório, redução de custos operacionais, melhoria na satisfação dos clientes e vantagem competitiva. |
| Dunnhumby (MARTINEZ; WALTON, 2014) | Empresa líder mundial em entender os consumidores através da análise de dados, criada em 1989 na Inglaterra. | Testar as vantagens de promover um concurso de <i>crowdsourcing</i> para análise de Big Data. | Plataforma online da empresa Kaggle. | As soluções encontradas através do concurso de <i>crowdsourcing</i> não poderiam ter sido desenvolvidas internamente. O concurso foi um sucesso, porém é preciso atenção com relação à privacidade dos dados. |
| Netflix, Inc. (KUMAR, 2013) | Empresa de aluguel de DVDs e transmissão de vídeos através da Internet pelo sistema de assinatura, criada em 2007 com sede nos Estados Unidos. É líder de mercado. | Criar conteúdo original com base no gosto de seus clientes, deduzido através do uso de Big Data. | Amazon Elastic MapReduce, Hadoop e NoSQL Apache Cassandra, | Produção da série de sucesso <i>House of Cards</i> , entre outras. Larga vantagem competitiva. Problemas com privacidade de dados. |

Fonte: quadro produzido pela autora.

5 CONSIDERAÇÕES FINAIS

Através da análise dos estudos de caso apresentados foi possível concluir que a implementação de projetos de Big Data em empresas está em fase embrionária. O próprio conceito de Big Data ainda não está claramente definido e os benefícios que projetos de análise e cruzamentos destes grandes volumes de dados podem trazer às empresas estão sendo testados. No entanto, baseando-se nos casos analisados, os principais objetivos que as empresas buscam ao implementar aplicações de Big Data relacionam-se, principalmente, às questões estratégicas, como a melhoria no processo de tomada de decisões e criação de vantagem competitiva sustentável com relação a seus concorrentes. Uma das principais preocupações das empresas no que tange esses ousados projetos relaciona-se à privacidade e ao direito à informação.

O conceito de Big Data geralmente está relacionado ao grande volume, velocidade e variedade de dados, porém são levantados questionamentos no que diz respeito a estes valores, pois com a evolução exponencial na produção de dados, o que é volumoso hoje não será amanhã. Em 2016, a produção de dados já é contada em *zettabytes*, no entanto, talvez daqui a 10 anos esse valor seja insignificante. Apesar das divergências em sua definição, é consenso pensar que o Big Data “[...] é um tsunami em alto mar, pouco perceptível da costa, mas que vai chegar com muita força, arrastando negócios estabelecidos e criando oportunidades para outros novos e inovadores modelos de negócios.” (TAURION, 2013a, não paginado).

Durante a pesquisa, foram analisados cinco estudos de caso em empresas localizadas em diferentes partes do mundo: New South Wales Emergency Services, na Austrália; Nielsen Holdings, nos Estados Unidos; Ramco Cements Limited, na Índia; Dunnhumby, na Inglaterra; e, Netflix, Inc. nos Estados Unidos. Apesar de ramos, mercados e culturas muito diferentes, todas as empresas possuem algo em comum: foram beneficiadas por projetos de Big Data. Os projetos apresentados são, em sua maioria, projetos-pilotos, porém já é possível perceber o grande potencial de trabalhar com um grande volume de dados, em uma velocidade constante e nos mais variados formatos. Algumas destas empresas analisadas, como por exemplo a Ramco Cements Limited, uma empresa de produção de cimento, faz parte de um ramo que não possui muita tradição na análise de dados, enquanto outras empresas, como a Dunnhumby, trabalha justamente com predição de mercado através da análise de dados e a Netflix, que é conhecida por ser totalmente voltada para dados, tendo toda sua base estratégica fundamentada no cruzamento das informações que coleta diariamente de seus

assinantes. Apesar das diferenças nas estruturas organizacionais, foi possível concluir que todas elas obtiveram muitos benefícios estratégicos com a utilização do Big Data.

Falar sobre Big Data é discursar sobre um conceito novo. No entanto, percebe-se que ainda é necessário muitas pesquisas sobre o assunto, de forma não só a beneficiar empresas, mas também à sociedade como um todo. Novas pesquisas podem ser realizadas para auxiliar as empresas na implementação de aplicações de Big Data, além de exemplificar vantagens estratégicas, que esses recursos podem vir a trazer. No Brasil, este assunto ainda é praticamente desconhecido, sendo necessário profundos debates para o entender o conceito de Big Data e suas aplicações nas empresas nacionais. Outro ponto de interesse para trabalhos futuros relaciona-se à questão da privacidade e o direito à informação. Privacidade de dados está se tornando um assunto cada vez mais recorrente, já que as barreiras do que é público ou privado diminuem paulatinamente.

Conclui-se, portanto, que o Big Data veio para ficar. As empresas que souberem aproveitar-se desse imenso universo de dados que se apresenta, certamente criarão vantagens competitivas significativas. Produz-se todos os dias *zettabytes* de dados que expressam nossos hábitos, gostos, opiniões e vontades. Ter acesso a esses dados, cruzá-los, aproveitar o que é útil, criar padrões e analisá-los é a chave para entender os consumidores. Há muito tempo sabe-se que possuir informação é possuir poder. O Big Data traz uma fonte infindável de informações. Acredita-se que as empresas que melhor se apropriarem desses dados serão aquelas que, no futuro, irão se destacar neste mundo multipolarizado, dominado pelas grandes organizações multinacionais.

REFERÊNCIAS

- AMARAL, F.M.; BEZERRA, D.S. Termos técnicos em inglês das disciplinas introdução a sistemas de banco de dados, algoritmos, estruturas de dados e tecnologia em orientação a objetos para o curso de tecnologia em sistemas de informação. In: SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DO INSTITUTO FEDERAL DE EDUCAÇÃO, Ciência e Tecnologia de Goiás, 4., 2010, Inhumas. **Anais...** Goiânia: IFG, 2010. p.1-4.
- ANGELONI, M.T. Elementos intervenientes na tomada de decisão. **Ci. Inf.**, Brasília, DF, v.32, n.1, p.17-22, abr. 2003. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652003000100002&lang=pt>. Acesso em: 23 maio 2015.
- ARTHUR, C. What's a zettabyte?: By 2015, the internet will know, says Cisco. **The Guardian**, 29 June 2011. Não paginado. Disponível em: <<http://www.theguardian.com/technology/blog/2011/jun/29/zettabyte-data-internet-cisco>>. Acesso em: 2 jun. 2015.
- ASSOCIAÇÃO BRASILEIRA DE CIMENTO PORTLAND. **A versatilidade do cimento brasileiro**. São Paulo: ABCP, 2009. Disponível em: <<http://www.abcp.org.br/conteudo/basico-sobre-cimento/tipos/a-versatilidade-do-cimento-brasileiro>>. Acesso em: 6 fev. 2016.
- BAPTISTA, S. G.; CUNHA, M. B. Estudo de usuários: visão global dos métodos de coleta de dados. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 2, p. 168-184, maio/ago. 2007.
- BARNEY, J. Firm resources and sustained competitive advantage. **Journal of Management**, v.17, n.1, p. 99-120, Mar. 1991.
- BAZERMAN, M.H. **Processo decisório**: para cursos de administração e economia. Tradução de Arlete Simille Marques. Rio de Janeiro: Elsevier, 2004. 232 p.
- BENGHOZI, P.; BUREAU, S.; MASSIT-FOLEA, F. **L'Internet des objets: quels enjeux pour les Européens?** Paris: École Polytechnique; Telecom Paris Tech, 2008. Disponível em: <<https://hal.archives-ouvertes.fr/hal-00405070/document>>. Acesso em: 14 set. 2015.
- BEZERRA, F. **Stakeholders**: do significado à classificação. [S.l.]: Portal Administração, 2014. Disponível em: <<http://www.portal-administracao.com/2014/07/stakeholders-significado-classificacao.html>>. Acesso em: 16 jan. 2016.
- BRETERNITZ, V.J; SILVA, L.A. Big Data: um novo conceito gerando oportunidades e desafios. **RECT**, Jundiaí, n.13, p.106-113, out. 2013. Disponível em: <<http://www.fatecjd.edu.br/retc/index.php/RETC/issue/view/34>>. Acesso em: 30 maio 2015.
- CABRAL, P.A.; SAID, G. A sociedade na era do big data: Dados demais, filtros de menos. In: SIMPÓSIO INTERNACIONAL DE CIBERJORNALISMO, 5., 2014, Campo Grande, MS. **Anais...** Campo Grande, MS: UFMS, 2014. Não paginado. Disponível em: <<http://www.ciberjor.ufms.br/ciberjor5/files/2014/07/pedroalexandre.pdf>>. Acesso em: 24 maio 2015.

CANARY, V. P. **A tomada de decisão no contexto Big Data**: estudo de caso único. 2013. 73 f. Monografia (Graduação em Administração) – Departamento de Ciências Administrativas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013. Orientador: Antônio Carlos Gastaud Maçada. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/87757/000911900.pdf?sequence=1>>. Acesso em: 23 maio 2015.

CHESBROUGH, H.W. **Open innovation**: the new imperative for creating and profiting from technology. Boston: Harvard Business School Press, 2006.

DAVENPORT, T.H. **Big data no trabalho**: derrubando mitos e descobrindo oportunidades. Tradução Cristina Yamagami. Rio de Janeiro: Elsevier, 2014. 221 p.

_____. **Ecologia da Informação**: por que só a tecnologia não basta para o sucesso na era da informação. Tradução Bernadette Siqueira Abrão. São Paulo: Futura, 1998. 316 p.

DICIONÁRIO Michaelis. [São Paulo]: Melhoramentos, 2012. Não paginado. Disponível em: <<http://michaelis.uol.com.br/moderno/portugues/index.php?lingua=portugues-portugues&palavra=privacidade>>. Acesso em: 05 jul. 2015.

DINIZ, E.H. Internet das coisas. **GV Executivo**, [S.l.], v.5, n.1, p.59, 2006. Disponível em: <<http://bibliotecadigital.fgv.br/ojs/index.php/gvexecutivo/article/view/34372/33170>>. Acesso em: 14 set. 2015.

DUMBILL, E. **What is big data?** an introduction to the big data landscape. O'Reilly Media, Inc., 2012. Não paginado. Disponível em: <<http://strata.oreilly.com/2012/01/what-is-big-data.html>>. Acesso em: 2 jun. 2015.

DUNNHUMBY. **What we do**. Londres: Dunnhumby, 2016. Disponível em: <<https://www.dunnhumby.com/what-we-do>>. Acesso em: 14 fev. 2016.

DUTTA, D.; BOSE, I. Managing a big data Project: the case of Ramco Cements Limited. **International Journal of Production Economics**, [S.l.], v.165, p.293-306, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.ijpe.2014.12.032>>. Acesso em: 5 nov. 2015.

ENCICLOPÉDIA Barsa. Rio de Janeiro: Encyclopaedia Britannica, 1979. v.6.

EQUIPE NIELSEN IBOPE. **Nielsen NetRatings**. Nielsen Ibope, 2015. Disponível em: <https://br.nielsennetpanel.com/images/ServiceDown_Page_BR.html>. Acesso em: 28 jan. 2016.

EVANS, D. **The Internet of things**: how the next evolution of the Internet is changing everything. [São Francisco]: Cisco IBSG, 2011.

GARTNER. Big Data. In: _____. **IT Glossary**. Disponível em: <<http://www.gartner.com/it-glossary/big-data>>. Acesso em: 1 jun. 2015.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

HARVARD BUSINESS REVIEW. **Big Data**: o futuro da informação e dos negócios. [S.l.]: Serasa Experian, 2013. 24 p.

_____. **Decisões mais inteligentes**. Tradução Tom Venetianer. Rio de Janeiro: Elsevier, 2008. 199 p.

HENRIQUES, D.A; COSTA, H.R. Big Data: como utilizar a extraordinária quantidade de informações coletadas por novas tecnologias para obter vantagens competitivas. **Revista Pensar Tecnologia**, Belo Horizonte, v.3, n.1, p.1-11, jan. 2014. Disponível em: <<http://revistapensar.com.br/tecnologia/artigo/no=a72.pdf>>. Acesso em: 26 maio 2015.

IMDB. **House of cards**. IMDb, 2013. Disponível em: <<http://www.imdb.com/title/tt1856010/>>. Acesso em: 10 fev 2016.

ISACA. **Big data**: impactos e benefícios. 2013. Disponível em: <http://www.isaca.org/Knowledge-Center/Research/Documents/Big-Data_whp_Por_0413.pdf/>. Acesso em: 3 jun. 2015.

KRANZBERG, M. Technology and history: “Kranzberg’s Laws”. **Tech Culture**, v. 27, n. 3, nov. 1986.

KUMAR, T.V. **Netflix**: leveraging big data to predict entertainment hits. ICMR, 2013. Disponível em: <http://poslis.fon.bg.ac.rs/index.php?option=com_docman&task=doc_download&gid=497&Itemid=28>. Acesso em 7 fevereiro 2016.

LAUGENI, F.P.; MARTINS, P.G. **Administração da produção**. São Paulo: Saraiva, 2005.

LÉTOURNEAU, J. **Ferramentas para o pesquisador iniciante**. Trad. Ivone C. Benedetti. São Paulo: M. Fontes, 2011. Cap. 11, p. 247-256.

LOWSON, R.H. **Strategic Operation Management**: the new competitive advantage. London: Routledge, 2002. 352 p.

MAGALHÃES V.R.V. et al. **O uso do Big Data na violação da privacidade dos usuários para estratégias de negócios**. Canindé: IFCE, [2014?]. Não paginado.

MARTINEZ, M.G; WALTON, B. The wisdom of crowds: the potential of online communities as a tool for data analysis. **Technovation**, v.34, p.203-214, 2014. Disponível em: <<http://dx.doi.org/10.1016/j.technovation.2014.01.011>>. Acesso em: 8 nov. 2015.

MINAYO, M. C. S. Ciência, técnica e arte: o desafio da pesquisa social. In: _____ (org.). **Pesquisa social**: teoria, método e criatividade. 21. ed. Petrópolis: Vozes, 2002. Cap. 1. p. 9-29.

NETFLIX. **Q4 15 Letter to shareholders**. Los Gatos: Netflix, 2016. Disponível em: <http://files.shareholder.com/downloads/NFLX/1489790719x0x870685/C6213FF9-5498-4084-A0FF-74363CEE35A1/Q4_15_Letter_to_Shareholders_-_COMBINED.pdf>. Acesso em: 10 fev. 2016.

NIELSEN. **Nielsen TV Diary Support**. [S.l.]: Nielsen, 2015. Disponível em: <<http://tvdiary.nielsen.com/>>. Acesso em: 25 jan. 2016.

O QUE é: bit e byte. **G1**: o portal de notícias da Globo, 9 abr. 2008. São Paulo, Internet. Não paginado. Disponível em: <<http://g1.globo.com/Noticias/0,,MUL394860-15524,00O+QUE+E+BIT+E+BYTE.html>>. Acesso em: 2 jun. 2015.

PENNY STOCKS LABS. **The Internet in real-time**. 2015. Não paginado. Disponível em: <<http://pennystocks.la/internet-in-real-time/>>. Acesso em: 3 jun. 2015.

PEREIRA, L.M. Modelo de formação de preços de commodities agrícolas aplicado ao mercado de açúcar e café. São Paulo: USP, 2009.

PINDYCK, R. S.; RUBINFELD, D. L. **Microeconomia**. 5. Ed. São Paulo: Prentice Hall, 2002.

PORTO, G.S. Características do processo decisório na cooperação empresa-universidade. **Revista Administração Contemporânea**, Curitiba, v.8, n.3, p.29-52, set. 2004. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-65552004000300003&lng=en&nrm=iso>. Acesso em: 09 jun 2015.

PRESCOTT, M.E. Big data and competitive advantage at Nielsen. **Management Decision**, [S.l.], v.52, p.573-601, 2014. Disponível em: <<http://dx.doi.org/10.1108/MD-09-2013-0437>>. Acesso em: 5 nov. 2015.

RICHARDSON, R. J. et al. **Métodos quantitativos e qualitativos**. In: _____. Pesquisa social: métodos e técnicas. 3. ed. São Paulo: Atlas, 2012. Cap. 5, p. 70-89.

RUSSO, M. **Fundamentos de Biblioteconomia e Ciência da Informação**. Rio de Janeiro: E-papers, 2010. 177 p.

SETZER, V.W. **Meios eletrônicos e educação: uma visão alternativa**. São Paulo: Escrituras, 2001. Não paginado. Disponível em: <<https://www.ime.usp.br/~vwsetzer/livro-meios.htm>>. Acesso em: 23 maio 2015.

SILVA, L. C. **O bibliotecário no contexto big data**. 25f. Monografia (Graduação em Biblioteconomia e Gestão de Unidades de Informação) – Centro de Ciências Jurídicas e Econômicas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014. Orientadora: Maria Irene da Fonseca e Sá.

SINGER, T. Tudo conectado: conceitos e representações da internet das coisas. In: SIMPÓSIO EM TECNOLOGIAS DIGITAIS E SOCIABILIDADE, 2012, Salvador. **Anais eletrônicos...** Salvador: UFBA, 2012. Não paginado. Disponível em: <<http://www.simsocial2012.ufba.br/modulos/submissao/Upload/44965.pdf>>. Acesso em: 14 set. 2015.

TAURION, C. Big Data. Rio de Janeiro: **Brasport**, 2013a. Não paginado.

_____. Data scientist. **Litteris Consulting**, 2014a. Não paginado. Disponível em: <<http://litterisconsulting.com.br/data-scientist/>>. Acesso em 27 maio 2015.

TAURION, C. O estágio atual do big data no Brasil. **Revista Power channel**, São Paulo, n.20, p.5-7, jun. 2013b. Disponível em: <<http://www.powerchannel.com.br/2013/06/03/cezar-taurion-o-estagio-atual-do-big-data-no-brasil/>>. Acesso em: 30 maio 2015.

_____. Volume, variedade, velocidade, veracidade e valor em big data. **Litteris Consulting**, 2014b. Não paginado. Disponível em: <<http://litterisconsulting.com.br/v-de-valor-em-big-data/>>. Acesso em 27 maio 2015.

TEIXEIRA, F. **Testes A/B agora podem ser feitos via Google Analytics**. [S.l.]: Blog de AI, 2012. Disponível em: <<http://arquiteturadeinformacao.com/user-experience/testes-ab-agora-podem-ser-feitos-via-google-analytics/>>. Acesso em: 10 fev 2016.

TZU, S. **A arte da guerra**. Tradução de Sueli Barros Cassal. Porto Alegre: L&PM, 2007. 152p.

UNIVERSITY ALLIANCE. **What is big data?** Villanova, [201-]. Não paginado. Disponível em: <<http://www.villanovau.com/resources/bi/what-is-big-data/#.VTmUaSFViko>>. Acesso em: 23 abr. 2015.

VASCONCELOS, F.C.; BRITO, L.A.L. Vantagem competitiva: o construto e a métrica. **RAE**, São Paulo, v.44, n.2, p. 51-63, abr./jun. 2004.

VASCONCELOS, F.C.; CYRINO, A.B. Vantagem competitiva: os modelos teóricos atuais e a convergência entre estratégia e teoria organizacional. **RAE**, São Paulo, v.40, n.4, p.20-37, out./dez. 2000.

VENTURA, M.M. O estudo de caso como modalidade de pesquisa. **Rev SOCERJ**, Rio de Janeiro, v.20, n.5, p.383-386, set./out. 2007.

WAMBA, S.F. et al. How big data can make big impact : finding from a systematic review and a longitudinal case study. **International Journal of Production Economics**, [S.l.], v.165, p. 234-246, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.ijpe.2014.12.031>>. Acesso em: 5 nov. 2015.