

CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA
UNIDADE DE PÓS-GRADUAÇÃO, EXTENSÃO E PESQUISA
MESTRADO PROFISSIONAL EM GESTÃO E TECNOLOGIA
EM SISTEMAS PRODUTIVOS

RAFAEL FERNANDES NOVO

MELHORIA NAS DIMENSÕES COMPETITIVAS DAS EMPRESAS POR MEIO DO USO
DO BIG DATA

São Paulo

Dez/2014

RAFAEL FERNANDES NOVO

MELHORIA NAS DIMENSÕES COMPETITIVAS DAS EMPRESAS POR MEIO DO USO
DO BIG DATA

Dissertação apresentada como exigência parcial para a obtenção do título de Mestre em Gestão e Tecnologia em Sistemas Produtivos do Centro Estadual de Educação Tecnológica Paula Souza, no Programa de Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos, sob a orientação do Prof. Dr. José Manoel Souza das Neves

São Paulo

Dez/2014

N945m Novo, Rafael Fernandes
Melhoria nas dimensões competitivas das empresas por meio do uso do Big Data. / Rafael Fernandes Novo. – São Paulo : CEETPS, 2014.
100 f. : il.

Orientador: Prof. Dr. José Manoel Souza das Neves
Dissertação (Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos) – Centro Estadual de Educação Tecnológica Paula Souza, 2014.

1. Big Data. 2. Dimensões competitivas. 3. Inteligência de negócios. 4. Inovação tecnológica. I. Neves, José Manoel Souza das. II Centro Estadual de Educação Tecnológica Paula Souza. III. Título.

RAFAEL FERNANDES NOVO

MELHORIA NAS DIMENSÕES COMPETITIVAS DAS EMPRESAS POR MEIO DO USO
DO BIG DATA

Prof. Dr. José Manoel Souza das Neves

Profª. Dra. Marília Macorin Azevedo

Prof. Dr. Joshua Onome Imoniana

São Paulo, 17 de dezembro de 2014

Dedico esse trabalho à Patrícia, meu amor
nessa vida, que com sua paixão pelos estudos
me apoiou e me motivou a chegar até aqui.

AGRADECIMENTOS

Ao meu mestre e orientador Prof. Dr. José Manoel Souza das Neves, que me ajudou a encontrar meu caminho no mundo acadêmico e me guiou na elaboração deste trabalho.

Aos meus companheiros, que compartilharam das angústias e desesperos das aulas, trabalhos e prazos.

A todos da Unidade de Pós-Graduação, Extensão e Pesquisa do Centro Paula Souza, pelo carinho, competência e dedicação aos alunos.

Ao Joel Brawerman, meu chefe e amigo, que me permitiu conciliar a vida profissional e a acadêmica durante 2 anos.

Às empresas e pessoas que emprestaram seu tempo, tão escasso na vida moderna, para colaborarem com a pesquisa aqui apresentada.

A todos que, ao longo dessa jornada, me incentivaram. Mesmo que apenas com pequenas palavras de apoio, ajudaram a vencer o desânimo que, por muitas vezes, me abateu.

O Universo é uma harmonia de contrários
(Pitágoras)

Aquele que não é corajoso o suficiente para
assumir riscos não realizará nada na vida
(Muhammad Ali)

RESUMO

NOVO, R. F. **Melhoria nas dimensões competitivas das empresas por meio do uso do Big Data**. 100– f. Dissertação (Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos). Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2014.

O presente trabalho tem por objetivo examinar como as empresas podem melhorar suas dimensões competitivas por meio do uso do Big Data. A metodologia usada foi de objetivo descritivo com abordagem qualitativa, delineamento bibliográfico (revisão de literatura) e estudos de casos múltiplos. Big Data ainda é um tema em desenvolvimento dentro das empresas, porém já é possível constatar que existem diferenças entre Big Data e soluções tradicionais de análise de dados e que as empresas estudadas obtiveram ganhos nas suas dimensões competitivas por meio do uso do Big Data. Essas empresas, mesmo possuindo sistemas de análise de dados, estão construindo sistemas de Big Data para analisar maiores volumes de dados de forma mais rápida e eficiente, obtendo, assim, melhorias em algumas dimensões competitivas, como a redução de custos operacionais em manutenção de equipamentos, reduções de custo de infraestrutura, aumento da receita por meio da medição mais precisa do uso dos serviços e melhoria na qualidade do serviço prestado.

Palavras-chave: Big Data; dimensões competitivas; inteligência de negócios; inovação tecnológica.

ABSTRACT

NOVO, R. F. **Melhoria nas dimensões competitivas das empresas por meio do uso do Big Data**. 100 f. Dissertação (Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos). Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2014.

This study aims to examine how businesses can improve their competitive dimensions using Big Data. The methodology used was descriptive with qualitative research and multiple case studies. Although Big Data is still under evaluation by companies, it was possible to identify that there are differences between Big Data and traditional data analytic tools; also, that the studied companies improved their competitive dimensions using Big Data. Even though there are using traditional data analytic tools, these same companies are building Big Data systems to analyze larger amounts of data faster and more efficiently. Those companies are achieving improvements in some competitive dimensions, such as reducing operating costs in equipment maintenance, reductions in infrastructure costs, increase revenue through more accurate measurement of their service usage and improved quality of service delivered to customers.

Keywords: Big Data; competitive dimensions; business intelligence; technological innovation.

LISTA DE QUADROS

Quadro 1: Principais dimensões competitivas.....	55
Quadro 2: Caracterização das empresas.....	64
Quadro 3: Questões do estudo de caso	66
Quadro 4: Dinâmica do processo de coleta de dados	68

LISTA DE TABELAS

Tabela 1: Principais ganhos com a adoção do Big Data.....	94
---	----

LISTA DE FIGURAS

Figura 1: Artigos classificados por ano de publicação.	17
Figura 2: Estrutura do trabalho.....	18
Figura 3: A lacuna de dados sem classificação e análise.....	24
Figura 4: O crescimento do volume de dados do universo digital	24
Figura 5: Oportunidade para o Big Data.....	25
Figura 6: O impacto dos consumidores na geração de dados	29
Figura 7: Facilidade de uso <i>versus</i> potencial de benefício.....	30
Figura 8: % de usuários de <i>smartphones</i> (set/2012).....	36
Figura 9: Análise de palavras relacionadas a Big Data em postagens no Twitter.....	38
Figura 10: % de adultos usuários do Twitter classificados por etnia e idade.....	39
Figura 11: Amostragem de local de origem de <i>tweets</i> com a <i>hashtag</i> #bigdata	41
Figura 12: O crescimento dos mercados emergentes na geração de dados.....	43
Figura 13: Percepção dos indivíduos sobre o rastreamento da navegação na Internet.....	44
Figura 14: Informações disponíveis em 13 redes sociais investigadas.....	45
Figura 15: Forças que determinam a competição de uma indústria	52
Figura 16: Cadeia de valor de Porter	53
Figura 17: Abordagem de pesquisa em artigos de periódicos QUALIS A entre 2003 e 2006	62
Figura 18: Estudo de caso em artigos do ENEGEP.....	63
Figura 19: Fontes de dados em pesquisas de Estudo de Caso.....	67
Figura 20: As três plataformas tecnológicas	71
Figura 21: Arquitetura tradicional de processamento de dados <i>versus</i> Big Data.....	72
Figura 22: Inteligência Potencial das soluções de análises de dados	74
Figura 23: Curva de oportunidade	78
Figura 24: Resumo do estudo da Empresa “B”	79
Figura 25: Estrutura do piloto para complemento do BI	80
Figura 26: Identificação dos clientes que acessam o WhatsApp.....	83
Figura 27: Resumo do estudo da Empresa “C”	85
Figura 28: Interligação dos sistemas de atendimento e engenharia.....	86
Figura 29: Resumo do estudo da Empresa “D”	88
Figura 30: Estrutura do piloto para monitoração dos ATMs	90
Figura 31: Resumo do estudo da Empresa “E”	91

LISTA DE SIGLAS

ANATEL	Agência Nacional de Telecomunicações
ATM	<i>Automated Teller Machine</i> (caixa automático)
BI	<i>Business Intelligence</i> (inteligência de negócio)
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CDR	<i>Call Detail Record</i> (registro de detalhes da chamada)
CEO	<i>Chief Executive Officer</i> (principal executivo da empresa)
CIO	<i>Chief Information Officer</i> (principal executivo da área de TI)
CRM	Customer Relationship Management (gerenciamento do relacionamento com o cliente)
DPI	<i>Deep Packet Inspection</i> (inspeção detalhada do pacote)
EB	<i>Exabyte</i>
ENEGEP	Encontro Nacional de Engenharia de Produção
ETL	<i>Extract Transform Load</i> (extração, transformação e carga)
EUA	Estados Unidos da América
GPS	<i>Global Positioning System</i> (sistema de posicionamento global)
HDFS	<i>Hadoop File System</i> (sistema de arquivo do Hadoop)
IP	<i>Internet Protocol</i> (protocolo da internet)
IPI	Informação Pessoal Identificável
MIT	<i>Massachusetts Institute of Technology</i> (Instituto de Tecnologia de Massachusetts)
MPP	<i>Massively Parallel Processing</i> (Processamento Paralelo Massivo)
noSQL	<i>Not Only Structured Query Language</i> (não apenas linguagem de busca estruturada)
OMC	Organização Mundial de Comércio
RFID	<i>Radio-Frequency Identification</i> (identificação por rádio frequência)
SMP	<i>Symmetric MultiProcessing</i> (Multi-Processamento Simétrico)
SMS	<i>Short Message Service</i> (serviço de mensagem curta)
TI	Tecnologia da Informação

SUMÁRIO

1 INTRODUÇÃO	16
1.1 Questão de pesquisa.....	17
1.2 Objetivo da pesquisa.....	17
1.3 Relevância do tema	17
1.4 Estrutura do trabalho.....	18
2 FUNDAMENTAÇÃO TEÓRICA	19
2.1 Origem do termo Big Data	20
2.2 Conceituação e principais fatores de diferenciação do do Big Data	22
2.2.1 Volume	23
2.2.2 Velocidade.....	25
2.2.3 Variedade.....	27
2.3 Utilização do Big Data por empresas	30
2.4 A relevância das redes sociais como fonte de dados	35
2.5 A ética no uso das informações do consumidor	39
2.6 O cientista de dados	48
2.7 Possíveis obstáculos ao Big Data	49
2.8 Dimensões competitivas.....	51
2.8.1 Estratégia analítica	58
3 MÉTODO	61
3.1 Escolha do método de pesquisa	61
3.2 Protocolo do estudo de caso.....	64
3.2.1 Coleta de dados.....	67
3.2.2 Análise dos dados coletados	69
4 RESULTADOS E DISCUSSÕES	70
4.1 Estudo das empresas.....	70
4.1.1 Empresa “A”.....	70
4.1.2 Empresa “B”.....	75
4.1.3 Empresa “C”	79
4.1.4 Empresa “D”	85

4.1.5 Empresa “E”.....	88
5 CONSIDERAÇÕES FINAIS.....	92
REFERÊNCIAS.....	96

1 INTRODUÇÃO

O termo Big Data foi cunhado por fornecedores de sistemas tradicionais de armazenamentos de dados e de *business intelligence*¹ para definir sistemas capazes de armazenar grandes quantidades de informação e de processar tais informações de forma mais eficiente. Porém, como explicado por Davenport, Barth e Bean (2012), Big Data é muito mais do que isso, entregando às empresas que fizerem uso dessa tecnologia a capacidade de entender o seu ambiente de negócio em um nível muito mais granular, possibilitando, entre outras vantagens, criar novos produtos e serviços, responder às mudanças de forma mais eficiente e explorar melhor o relacionamento entre empresas e seus clientes.

Os sistemas de análise de dados estão presentes em empresas dos mais diversos segmentos, seja no cerne do negócio de multinacionais, como Google ou Walmart, seja oferecendo *insights* para áreas específicas, como marketing e finanças de empresas de diversos segmentos. A inteligência analítica permite melhorar o poder de competição das empresas, sendo uma importante adição aos sistemas de BI (DAVENPORT; HARRIS, 2007).

Brynjolfsson e McAfee (2012) afirmam que soluções de Big Data possuem um potencial ainda maior de beneficiar as empresas do que as soluções analíticas tradicionais, as quais já são utilizadas e conhecidas por abrir novas oportunidades e oferecer vantagens competitivas para aquelas empresas que as utilizam. Os autores, por considerarem que Big Data trará uma revolução na gestão, afirmam que os executivos precisam encarar esse novo modelo o quanto antes. É descrito nesta pesquisa um conjunto de estudos técnicos e acadêmicos apontando a crescente importância do Big Data como uma tecnologia estratégica para as empresas, pontuando com exemplos de instituições que obtiveram melhorias nas suas dimensões competitivas.

A revista *Veja* (2013) dedicou a capa de sua edição 2.321 ao Big Data e, audaciosamente, o comparou a outras inovações tecnológicas que causaram grande revolução ao cotidiano das pessoas, como o computador pessoal e a internet. Apesar de não constituir fonte científica, isso denota que o tema, ainda relativamente novo e inexplorado, vem ganhando atenção e relevância, devendo ser estudado com mais atenção.

¹ Sistemas de inteligência de negócio são comumente chamados pelo nome em inglês de *business intelligence* ou simplesmente BI.

1.1 Questão de pesquisa

Como o uso do Big Data pode melhorar as dimensões competitivas das empresas?

1.2 Objetivo da pesquisa

Examinar como as empresas podem melhorar suas dimensões competitivas por meio do uso do Big Data.

1.3 Relevância do tema

Por tratar-se de um tema novo, torna-se pertinente um breve levantamento sobre a relevância do tema. Como demonstrado na Figura 1, a quantidade de artigos publicados em periódicos CAPES sobre Big Data vem crescendo de forma acentuada nos últimos anos, demonstrando tratar de um assunto novo, porém com crescente interesse no seu estudo.

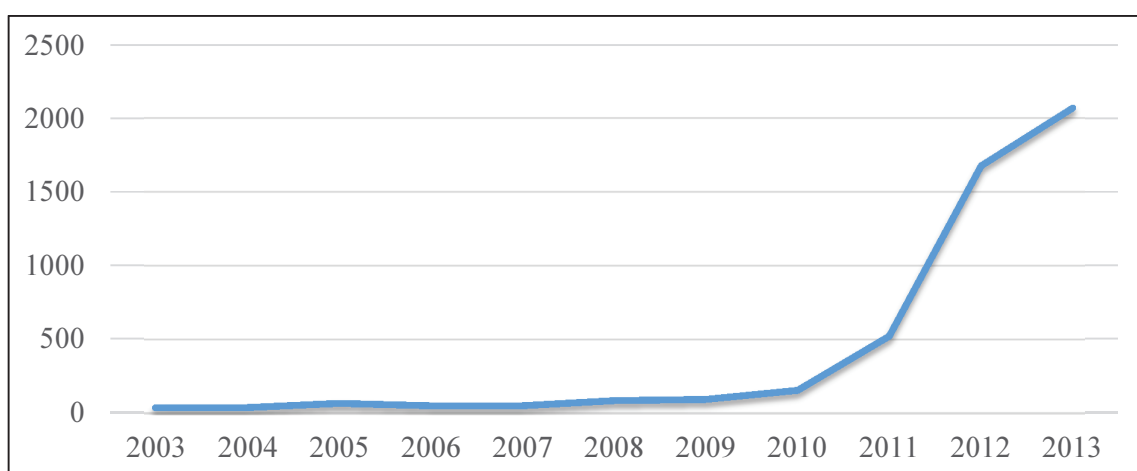


Figura 1: Artigos classificados por ano de publicação.

Fonte: Novo e Neves (2013, p. 8).

Possivelmente por ser este um tema recente, quase a totalidade desses artigos publicados nos periódicos CAPES até o ano de 2013 está na língua inglesa. Dentre os 3819 artigos identificados, 3791 foram escritos em inglês, restando apenas 28 artigos escritos em outros idiomas. (NOVO; NEVES 2013)

1.4 Estrutura do trabalho

Este trabalho está estruturado conforme demonstrado na Figura 2:

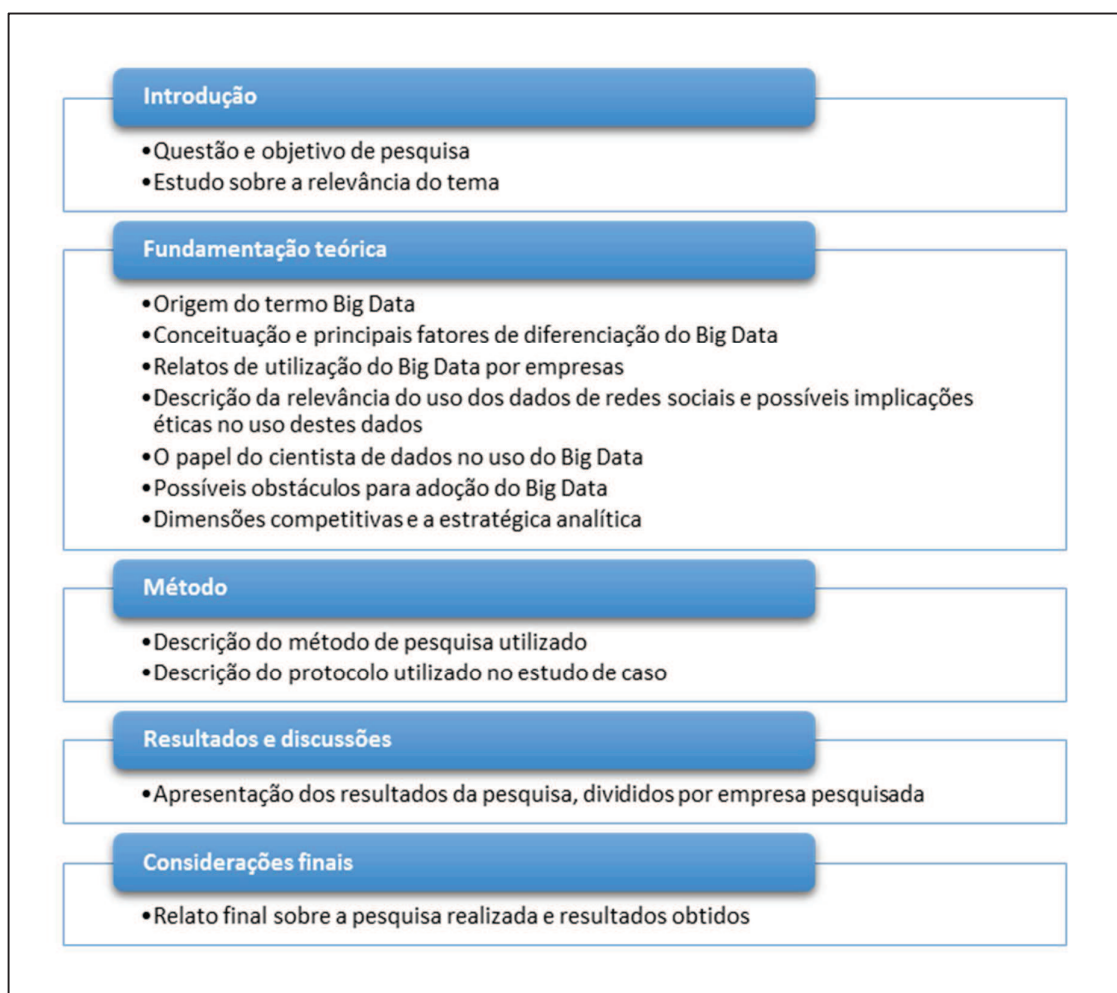


Figura 2: Estrutura do trabalho.

Fonte: Elaboração do autor.

2 FUNDAMENTAÇÃO TEÓRICA

Já foi demonstrado por Davenport e Harris (2007) que a inteligência analítica pode melhorar a capacidade de uma empresa competir dentro dos mais diversos segmentos, gerando melhorias no processo produtivo, nos sistemas de gestão da cadeia de suprimentos, no relacionamento com clientes e no fornecimento de produtos novos e diferenciados.

Brynjolfsson e McAfee (2012) afirmam que soluções de Big Data possuem um potencial maior do que as soluções analíticas tradicionais, que já são utilizadas e conhecidas por abrir novas oportunidades e por melhorarem a competitividade das empresas que as utilizam. Alertam, ainda, que, embora organizações que já nasceram na forma de empresas digitais, tais como Google e Amazon, conhecem há muito tempo os benefícios da análise de dados e hoje dominam o Big Data, as organizações tradicionais ainda não têm as ferramentas e filosofias do Big Data devidamente difundidas em suas camadas de gestão. Os autores, por considerarem que Big Data trará uma revolução na gestão, afirmam que os executivos dessas empresas precisam encarar esse novo modelo o quanto antes.

Big Data foi listado no estudo dirigido por Cearley e Claunch (2013) como uma das 10 tendências de tecnologias estratégicas com o potencial de afetar a vida das pessoas e das empresas. Essa lista foi montada com base no potencial de determinada tecnologia ser altamente disruptiva, ou por estar na eminência de sofrer uma aceleração, mudança ou ponto de inflexão, tornando a tecnologia estratégica e aplicável a um vasto mercado nos próximos 2 anos.

Cearley e Claunch (2013) avisam que, embora esse rol tenha sido elaborado considerando o potencial de impacto dessas tecnologias em uma ampla gama de empresas, cada companhia deve individualmente criar sua própria lista, considerando fatores particulares como necessidades de negócio e modelo de adoção de novas tecnologias.

2.1 Origem do termo Big Data

Um dos maiores desafios ao buscar a origem do termo Big Data² é o fato de serem palavras comuns, que podem ser utilizadas de forma ampla, diferentemente de termos novos, cunhados com objetivo específico, como, por exemplo, *software*, que, de acordo com Lohr (2013), apareceu pela primeira vez em 1958, no artigo “The American Mathematical Monthly”, escrito por John Tukey.

Outro fator que dificulta traçar a origem do termo é a mudança na forma de expressão humana. O surgimento de sítios de internet, *blogs*, redes sociais, entre outros, torna mais fácil a criação e disseminação de novos termos, sem que seja possível identificar exatamente sua origem.

Oxford (2013, tradução nossa) incluiu em sua base o termo Big Data e definiu-o como:

Dados de computação de um tamanho muito grande, tipicamente na medida em que a sua manipulação e gestão apresentam desafios logísticos significativos, também o ramo de computação envolvendo tais dados.

A inclusão do termo nesse relevante dicionário constitui indício do crescimento e importância do Big Data. Essa mesma publicação sugere a origem do termo em artigo de Tilly (1980, p. 10, tradução nossa), que diz: “nenhuma das grandes questões realmente renderam-se ao espancamento das pessoas de big-data”. Essa definição, bem como seu uso no artigo indicado, soa incompleta frente às definições de Brynjolfsson e McAfee (2012), Davenport, Barth e Bean (2012), Gantz e Reinsel (2012), Cearley e Claunch (2013) e Casonato et al (2013), colocando o Big Data como um modelo completo, envolvendo pessoas, processos e tecnologias, com o objetivo de analisar massas de dados, considerando três vetores: volume, velocidade e variedade.

Uma das primeiras referências a utilizar o termo Big Data como sua conotação atual é de Diebold (2003), abordando-o como uma explosão da quantidade de dados disponível e suas possibilidades de criar modelos de dados melhores e mais precisos, utilizando essa massa de dados.

² A tradução literal de Big Data seria grandes dados, entretanto utiliza-se sempre o termo em inglês, ao contrário de outros termos tecnológicos como *cloud computing*, que é aceito também traduzido para computação em nuvem.

Outra referência que figura entre as primeiras com o conceito atual de Big Data, mesmo que sem usar esse nome, está presente no artigo de Laney (2001), “3D Data Management: Controlling Data Volume, Velocity, and Variety”. O autor aborda o crescimento da quantidade de dados armazenada e disponível para análise, no mais puro sentido literal do Big Data; discute a crescente variedade das fontes de dados e dos tipos de dados, incluindo dados estruturados e não estruturados; e também aborda o aumento da demanda por soluções que permitam gerar *insights* quase em tempo real, melhorando os pontos de interação com os clientes. O autor trouxe a definição aplicada apenas às empresas de comércio eletrônico, entretanto não ficou explícito no artigo se ele acreditava que tais conceitos poderiam ser aplicados para quaisquer verticais empresariais.

Outro indício da origem do termo vem dos estudos de Diebold (2012) e Lohr (2013), que apontam John Mashey, cientista chefe da empresa Silicon Graphics na década de 90, como um dos primeiros a usar o termo Big Data. Mashey usou o termo em palestras e apresentações de negócio. Diebold (2012) destaca uma apresentação de Mashey intitulada “Big Data and the Next Wave of InfraStress Problems, Solutions, Opportunities”, que demonstra claramente os princípios do Big Data.

Obviamente, ao analisar esses artigos e suas respectivas concepções de grandes massas de dados, é necessária a contemporização deles. Por exemplo, Diebold (2003), ao falar de Big Data, comenta sobre conjuntos de dados de 200 *gigabytes*³, tamanho considerado pequeno para os padrões atuais, porém um desafio para a tecnologia daquela época. Um dos fatores que determina o crescimento da importância do Big Data é justamente o crescimento acelerado no volume de dados existente e disponível para análise, como descrito em pesquisa de Gantz e Reinsel (2012).

Diebold (2012) destaca que o termo de Big Data está evoluindo de apenas um termo e um fenômeno momentâneo para uma nova disciplina, pois a necessidade de lidar com grandes volumes de dados e de extrair informação destes dados vem, tornando-se um tema-chave para a maioria das ciências. Talvez justamente por tanger várias disciplinas surja a discussão se Big Data poderia ser apenas um pedaço do que hoje é conhecido como ciência da computação ou estatística. Entretanto um dos diferenciais do Big Data é justamente a capacidade de ser multidisciplinar, requerendo conhecimentos específicos de diversas áreas, como tecnologia, estatística e negócios, e fazendo com que a soma das partes seja maior do que o todo. O autor

³ 1 *gigabyte* = 10^9 *bytes* (considerando base decimal).

coloca-se otimista no crescimento do Big Data, pois, apesar de o termo poder parecer estranho à primeira vista e ainda haver debate sobre o nome em si, o fenômeno por detrás é real.

Palavras e frases são blocos de construção fundamentais da língua e da cultura, assim como genes e células são para a biologia da vida. E as palavras são como nós expressamos ideias, então traçar a sua origem, desenvolvimento e propagação não é apenas uma busca acadêmica, mas uma janela para a evolução intelectual da sociedade (LOHR, 2013, tradução nossa).

Para Davenport, Barth e Bean (2012), o termo Big Data é utilizado de forma distorcida por provedores de soluções para classificar sistemas de análise de dados, quando na realidade o Big Data é muito mais do que isso, permitindo o surgimento de novas ofertas e serviços, possibilitando às empresas responderem às novas demandas de forma ágil e assertiva, competindo de forma mais eficiente no mercado.

Cearley e Claunch (2013) mostram-se céticos com relação ao termo, por acreditarem que, com o crescimento do Big Data, sua adoção pelas instituições passará de projetos isolados para uma abordagem estratégica, e nesse ponto a abordagem diferenciada dos dados com relação ao volume, variedade e velocidade será incorporada em uma estratégia mais completa de gerenciamento da informação pelas empresas, o que faria o nome Big Data desaparecer com o tempo. Porém, é fundamental ressaltar que a argumentação dos autores gira apenas ao redor do nome. Os próprios autores destacam que empresas que não adotarem uma visão estratégica com esses conceitos diferenciados que hoje fazem parte do termo Big Data poderão estar em desvantagem frente às outras empresas do mercado que assim o fizerem.

2.2 Conceituação e principais fatores de diferenciação do do Big Data

É comum a comparação – e, algumas vezes, a falta de uma definição mais clara e objetiva – entre Big Data e a inteligência analítica tradicional, pois ambos têm a capacidade de analisar dados e convertê-los em informação para gerar melhorias para as dimensões competitivas das empresas. Há, contudo, três grandes diferenças listadas por Brynjolfsson e McAfee (2012) e corroboradas por Casonato et al (2013): volume; velocidade e variedade. Nota-se que a existência de apenas uma dessas características já pode constituir um problema

de Big Data, entretanto, como exemplificado adiante, são constantes os exemplos de empresas abordando problemas com dois ou até mesmo os três fatores em conjunto.

O Big Data abre ainda a possibilidade de um modelo de decisão diferente, pois permite às empresas criação de experimentos controlados para testar hipóteses que guiarão a tomada de decisão em, por exemplo, novos investimentos ou mudanças operacionais. Possibilitando centenas ou milhares de experimentações, é possível determinar com maior precisão a correlação de eventos que possuam real ligação de causa e efeito (BROWN; CHUI; MANYIKA, 2011).

2.2.1 Volume

A quantidade de dados existente tem crescido de forma acelerada. Segundo Brynjolfsson e McAfee (2012), em 2012 foram criados diariamente cerca de 2,5 *exabytes*⁴ de dados, e o volume de dados trafegado a cada segundo na internet nesse mesmo ano foi maior do que toda a informação que estava armazenada na rede 20 anos atrás. Isso cria uma nova oportunidade para as empresas analisarem esse crescente volume de dados que podem conter informações relevantes para gerar uma percepção mais apurada do mercado, melhorando o modelo de relacionamento com clientes e outros aspectos corporativos que permitam a essas empresas melhorarem suas dimensões competitivas.

A transformação de dados crus em informação relevante ainda é um processo extremamente complexo, especialmente para grandes volumes de dados, e, por esse motivo, muitas empresas restringem o escopo dessa transformação em uma parcela muito pequena da massa de dados disponível. Como visto no estudo de Gantz e Reinsel (2012), existe uma lacuna muito grande entre a quantidade de dados considerada útil para as empresas e o volume atual de dados efetivamente classificados e selecionados para análise. Existem duas lacunas aqui que podem ser preenchidas ao longo dos próximos anos com adoção de sistemas de Big Data pelas empresas. A primeira é identificar as informações relevantes, dentro de toda a massa de dados disponível. A segunda consiste em classificar e analisar estes dados, visando obter mais *insights* e uma melhor assertividade. Existe um potencial para obtenção de benefício na exploração desta

⁴ 1 *exabyte* = 10^9 *gigabytes* ou 10^{18} *bytes* (considerando base decimal).

massa de dados ainda intocada, como levantado por Brown, Chui e Manyika (2011) e demonstrado na Figura 3.

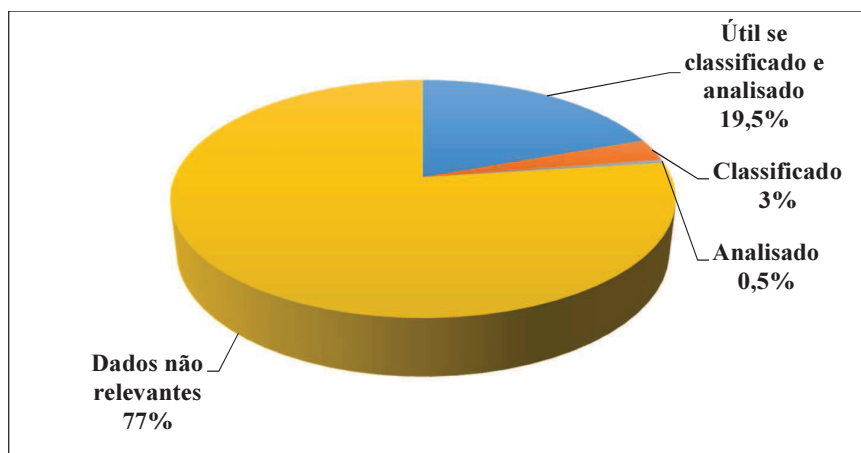


Figura 3: A lacuna de dados sem classificação e análise

Fonte: Adaptado de Gantz e Reinsel (2012).

Essa necessidade de analisar grandes volumes de dados irá aumentar nos próximos anos, posto que a quantidade de informações existente, chamado de universo digital, segundo estudo de Gantz e Reinsel (2012), irá crescer em até 50 vezes entre os anos de 2010 e 2020 e totalizará um crescimento de até 300 vezes, se considerados os anos de 2005 a 2020, saindo de 130 *exabytes* para aproximadamente 40 *zettabytes*⁵, como demonstrado na Figura 4.

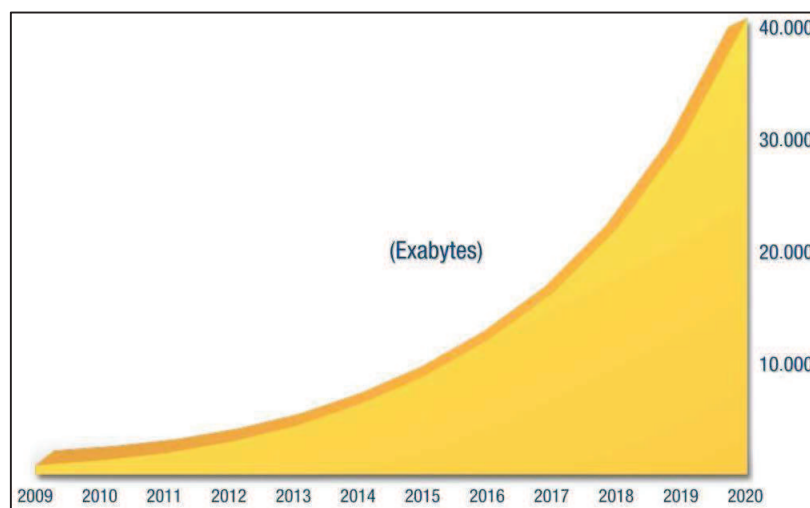


Figura 4: O crescimento do volume de dados do universo digital

Fonte: Traduzido de Gantz e Reinsel (2012, p. 3).

⁵ 1 *zettabyte* = 10^3 *exabytes* ou 10^{21} bytes (considerando base decimal).

Esse crescimento deve-se a diversos fatores, explicados por Brynjolfsson e McAfee (2012, p. 2), como o surgimento de novas fontes de informação oriundas de equipamentos capazes de gerar grandes quantidades de dados que podem ter relevância para as empresas, tais como “telefones celulares, compras na internet, redes sociais, comunicação eletrônica, GPS e maquinário computadorizado”, que “produzem torrentes de dados como subproduto de sua operação normal”.

Agregadas as informações demonstradas na Figura 3 e na Figura 4, Gantz e Reinsel (2012) preveem que a quantidade de dados potencialmente útil, entretanto desprezados pelas empresas, irá crescer significativamente até o ano de 2020, como demonstrado de forma consolidada na Figura 5:

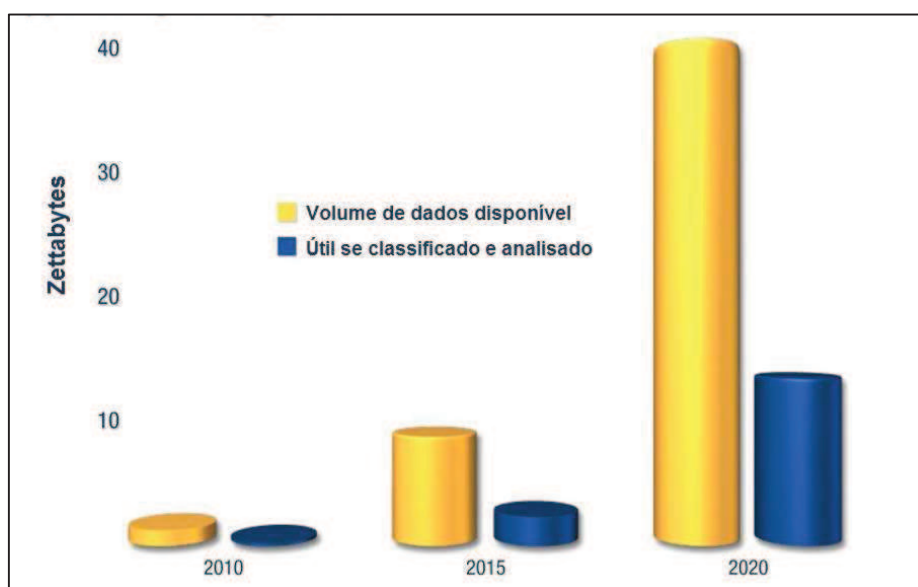


Figura 5: Oportunidade para o Big Data

Fonte: Traduzido de Gantz e Reinsel (2012, p. 9).

2.2.2 Velocidade

O segundo aspecto de diferenciação do Big Data é a velocidade de processamento das informações, que permite uma análise quase instantânea dos dados, possibilitando à empresa ser muito mais ágil, trazendo perspectivas rápidas as quais podem, por exemplo, permitir que

uma rede de supermercados crie ofertas de marketing em tempo real para seus consumidores (BRYNJOLFSSON; MCAFEE, 2012).

Brown, Chui e Manyika (2011) afirmam que as lojas líderes do varejo estão constantemente executando experimentos para identificar fatores que aumentem a interação com seus clientes, gerando maiores vendas. Embora muitas empresas já façam esses experimentos há algum tempo, o Big Data pode levá-los para um nível mais avançado. A futura geração de lojas de varejo será capaz de identificar comportamentos individualizados de cada cliente, utilizando, por exemplo, o *clickstream*, que é a sequência de cliques de mouse em uma página da internet, determinando o padrão de navegação e se adaptando às preferências do cliente em tempo real. Nesse modelo, é possível identificar quando um cliente está próximo de efetuar uma compra, permitindo à empresa influenciar na decisão do cliente, por meio de uma promoção relâmpago, oferecendo um pacote com outro produto em que o cliente esteja interessado, ganhos extras em programas de milhagem ou outra oferta que tenha sido identificada como relevante para o cliente e que seja vantajosa à empresa. Dessa forma, é possível não apenas garantir a venda como agregar produtos com maiores margens para a empresa.

Torna-se necessário que as empresas passem a pensar e considerar o fluxo contínuo de informação e processos, analisando os dados à medida que esses são gerados, ao invés de apenas armazená-los em grandes bancos de dados para análise posterior. Tom Deutsch, diretor do programa de tecnologias de Big Data e analítica aplicada na IBM, ressalta: “Análise durante a transmissão permite o processamento dos dados durante um evento, melhorando o resultado” (DAVENPORT; BARTH; BEAN, 2012, p. 2, tradução nossa).

Gallant (2011) pontua a evolução do modelo tradicional de análise de dados do século XX para um novo modelo do século XXI, comparando-o com o próprio comportamento humano. O ser humano consegue processar uma quantidade grande de informações e tomar decisões em tempo real. Quando uma pessoa está dirigindo, ela precisa estar atenta às diversas fontes de informação, processando todas em tempo real para tomar a melhor decisão. Se identificar um obstáculo na via, precisa responder imediatamente, pois de nada adianta querer desviar depois de já ter atingido o obstáculo. Essa metáfora pode ser aplicada para o uso corporativo de soluções de Big Data, que consegue analisar em tempo real quantidades massivas de informações, advindas de diversas fontes.

2.2.3 Variedade

O terceiro item listado, a variedade, refere-se às diversas fontes de dados, tais como: mensagens; leituras de sensores; câmeras de segurança; aparelhos de GPS; telefones celulares, entre outros, que podem ser usados para análise e posterior extração de informação pertinente ao negócio. Fontes para o Big Data podem englobar tudo, desde dados de voz de *call center* até dados de genoma de pesquisa biológica e medicinal (DAVENPORT; BARTH; BEAN, 2012), e podem ser provenientes das mais diversas origens, como uma copiadora ou até mesmo um motor de jato (BROWN; CHUI; MANYIKA, 2011). Nota-se que algumas das fontes de dados de grande relevância para análise por sistemas de Big Data são relativamente novas, como as redes sociais, dentre as quais se destacam o Facebook, lançado em 2004, e o Twitter, em 2006, que, aliadas às também recentes proliferações de smartphones, tablets, ultrabooks, redes wireless e redes de dados 3G e 4G, permitem às pessoas estarem sempre conectadas, gerando conteúdo, ideias e opiniões.

O alcance dessas redes sociais merece um destaque especial, pois são uma poderosa fonte de dados a ser utilizada pelas empresas nos seus projetos de Big Data, como no estudo mencionado por Brynjolfsson e McAfee (2012), no qual pesquisadores, analisando o fluxo de mensagens do Twitter, conseguiram antever um surto de cólera no Haiti após o terremoto em 2010, com duas semanas de antecedência aos boletins oficiais. Os autores mencionam ainda o estudo realizado por Erik e Lynn Wu, que usaram dados obtidos livremente em buscas na internet para prever a variação média do mercado imobiliário dos EUA, conseguindo obter estimativas mais precisas do que a oficial, da Associação Americana dos Corretores de Imóveis⁶, a qual utilizava um modelo complexo, porém alimentado por dados históricos que não conseguiam acompanhar a nova dinâmica do mercado.

Todas essas fontes de dados podem ser transformadas em informações valiosas, tais como gostos, hábitos, modo de consumo e satisfação dos clientes, a serem utilizadas pelas empresas como diferencial competitivo, como descrito por Davenport, Barth e Bean (2012).

O exposto pode ainda ser exemplificado comparando uma empresa de seguros automotivos tradicional que define o valor de sua apólice somente com base em uma lista limitada de dados sobre seus clientes: idade, região onde mora, local de trabalho, se já sofreu algum sinistro, entre outras informações, que, apesar de úteis, são incompletas, passíveis de

⁶ Tradução nossa.

respostas inverídicas e podem, muitas vezes, levar a resultados equivocados para o valor da apólice, causando a perda do cliente para a concorrência ou um prejuízo financeiro com as desventuras dos segurados.

Davenport e Harris (2007) descrevem o caso da Progressive, empresa de seguros baseada nos Estados Unidos que usa informações do score de crédito do segurado e informações obtidas por rastreadores instalados nos veículos para ajudar a identificar se o segurado respeita as velocidades das vias e pratica condução defensiva, qualificando o condutor para determinar com maior precisão a probabilidade que esse se envolva em um acidente. Com o advento do Big Data, é possível acrescer a esta análise: *tweets*⁷ enviados, posts no Facebook ou uso de outras informações de fontes diversas espalhadas na internet que indiquem um bom ou mal comportamento como condutor e consequente grau do risco do segurado ser responsável por um acidente de trânsito.

Obviamente, esse simples exemplo esconde uma complexidade técnica elevada, pelo volume de dados a serem manipulados e por serem, em sua imensa maioria, dados não estruturados.

A maior parte desses novos dados – que estão se tornando diferencial competitivo na análise de dados – são do tipo não estruturado, ou seja, não estão organizados em tabelas e colunas de um banco de dados tradicional. Por exemplo, em 2012 o Google processava diariamente aproximadamente 24.000 *terabytes*⁸ de dados e apenas uma ínfima porção desses dados está armazenado em bases de dados tradicionais (DAVENPORT; BARTH; BEAN, 2012). Mesmo nos dados provenientes de dentro das empresas, Hea, Zhab e Li (2013) colocam que 80% deles também não são estruturados, como, *e-mails* e outros documentos diversos.

Além de não estruturadas, essas novas fontes de dados possuem outra característica marcante: a volatilidade. São informações que, conforme descritas por Davenport, Barth e Bean (2012), não estão guardadas em um banco, mas em movimento constante, sendo alteradas a cada momento, tais como o fluxo de cliques dos consumidores em um site, ou os relacionamentos de potenciais clientes em sites como Facebook ou LinkedIn.

Gantz e Reinsel (2012) mostram que os consumidores terão grande impacto na geração dos dados que poderão ser analisados por sistemas de Big Data. Os autores determinam a origem dos dados conforme exposto na Figura 6, classificando os dados como: gerados pelos

⁷ *Tweets* são as postagens dos usuários na rede social Twitter, em geral frases com até 140 caracteres.

⁸ 1 *terabyte* = 10^3 *gigabytes* ou 10^{12} *bytes* (considerando base decimal).

consumidores, pertencentes às empresas e uma grande área de sobreposição, ou seja, dados gerados pelos consumidores, mas com alguma responsabilidade das empresas sobre esses dados.

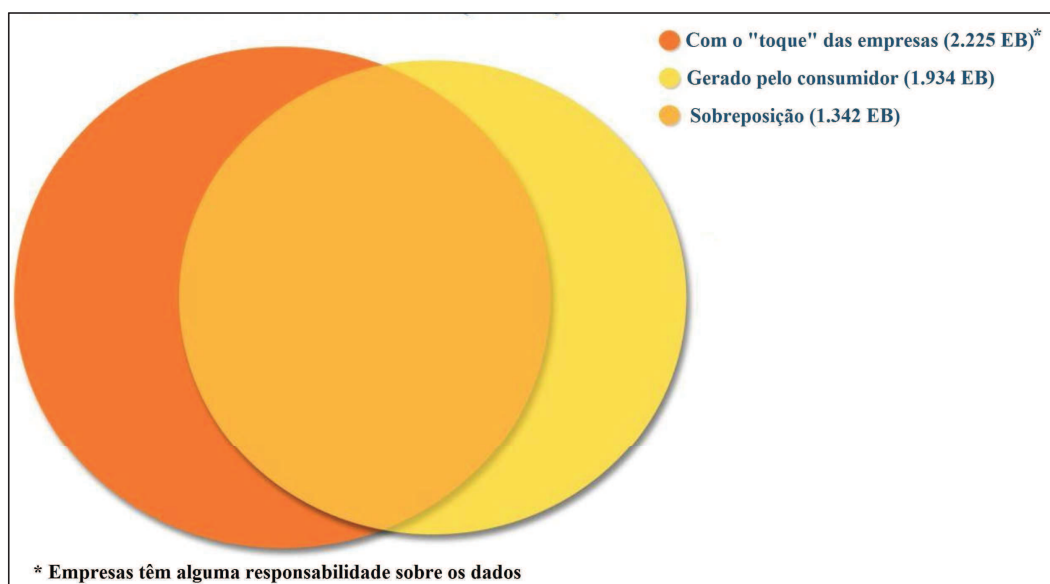


Figura 6: O impacto dos consumidores na geração de dados

Fonte: Traduzido de Gantz e Reinsel (2012, p. 4).

Dentre as diversas tecnologias que compõem o Big Data, a chamada *text mining*⁹ ou mineração de texto é responsável pela análise desses dados não estruturados. A análise de textos possui uma complexidade superior à análise de dados estruturados, e o seu principal objetivo é encontrar padrões para conseguir extrair informações e inteligência analítica desses textos (HEA; ZHAB; LI 2013).

No entanto, os exemplos acima trazem à discussão um outro aspecto importante a ser considerado: o lado ético e as possíveis implicações relacionadas à segurança da informação e privacidade no uso desses dados, como na análise das próprias redes sociais comentadas aqui – que, apesar de serem gerados e disponibilizados por indivíduos que em certo modo optaram por tornar essas dados públicos, certamente não o fizeram com o intuito de serem usados por corporações, como no exemplo citado, para calcular o seguro do seu carro.

⁹ Text mining ou mineração de texto é uma das tecnologias que compõem soluções de Big Data. Trata-se da capacidade de analisar de forma automatizada palavras e frases para extrair correlações e informações relevantes destas.

2.3 Utilização do Big Data por empresas

Davenport e Harris (2007) já destacavam que qualquer organização poderia competir utilizando a inteligência analítica nos mais diversos aspectos, como melhoria dos processos de produção, criação de novos produtos ajustados às necessidades do consumidor ou campanhas de marketing mais assertivas.

Segundo levantamento realizado por Brown, Chui e Manyika (2011), a maioria dos segmentos empresariais pode obter benefícios para as dimensões competitivas com o uso do Big Data. Entretanto, é preciso notar que alguns segmentos podem ser mais beneficiados do que outros. A Figura 7 analisa o benefício potencial para cada segmento e compara com a dificuldade em viabilizar projetos de Big Data. Para determinar o potencial de benefício, foram considerados fatores como: turbulência do mercado, estruturação do mercado, quantidade de transações, número de clientes potenciais, parceiros de negócios e quantidade disponível de dados. Enquanto isso, para calcular a facilidade de captura dos dados, foi avaliado o número de empregados com conhecimento analítico no segmento de indústria, investimento médio em TI, facilidade de acesso às fontes de dados e o grau de uso de dados nas decisões gerenciais.

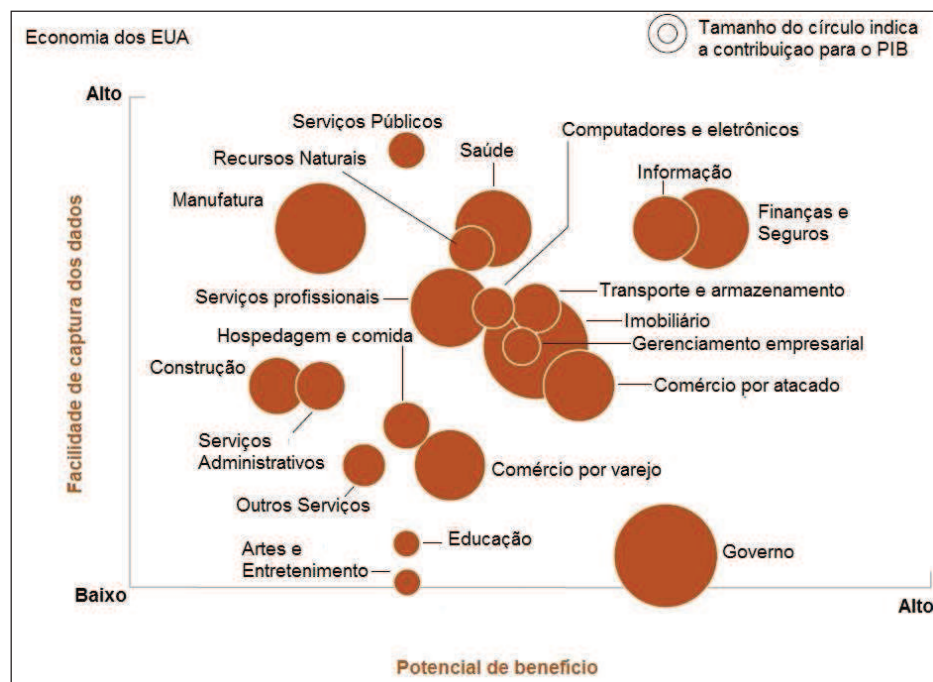


Figura 7: Facilidade de uso *versus* potencial de benefício

Fonte: Adaptado de Brown, Chui e Manyika (2011).

Já existe uma série de exemplos de empresas dos mais diversos segmentos utilizando o Big Data para melhorar seus processos produtivos, desenvolver novos produtos ou conhecer melhor o seu mercado consumidor.

O Walmart é um conhecido exemplo de empresa que faz uso intensivo da análise de dados, detalhado por Hays (2004) e posteriormente comentado por Davenport e Harris (2007). Esse gigante do varejo utiliza técnicas de análise de dados para competir de forma mais eficiente:

- a) Entendendo melhor o padrão de consumo dos seus clientes, correlacionando o acontecimento de eventos à compra de produtos alimentícios, exemplificado por Linda M. Dillman, CIO do Walmart, em entrevista a Hays (2004, tradução nossa): “No passado nós não sabíamos que as vendas de *Pop-Tarts*¹⁰ de morango aumentavam em até 7 vezes antes de um furacão”, mas afirma que “o item mais vendido antes de um furacão é cerveja”. Tal *insight* permite ao Walmart garantir o abastecimento correto das lojas, a satisfação dos clientes e um maior lucro.
- b) Munir os seus fornecedores de informação, melhorando o gerenciamento de seu *supply chain*¹¹, garantindo um melhor abastecimento das lojas, reduzindo os seus custos e competindo de forma mais eficiente.

Quando Hays (2004) escreveu seu artigo para o jornal *New York Times*, os volumes estimados de dados armazenados pelo Walmart referentes aos seus clientes e padrões de consumo era de aproximadamente 460 *terabytes* de dados, armazenados em um sistema da empresa Teradata, que atualmente alega, em seu sítio de internet, ser o líder em soluções de Big Data. Davenport e Harris (2007) apresentaram uma estimativa de 583 *terabytes* armazenados pelo Walmart, e, recentemente, Brynjolfsson e McAfee (2012) trouxeram a estimativa que esse gigante do varejo coleta 2,5 *petabytes* de dados por hora, com informações diversas sobre os seus clientes, histórico financeiro e padrões de compras.

Bughin e Livingston (2011, tradução nossa) trazem o exemplo de uma empresa europeia de telecomunicações, na qual foi criado um comitê com membros de diversas áreas da empresa. Esse comitê supervisionou o projeto de Big Data com o objetivo de direcionar os experimentos para responder perguntas como: “Quão competitiva as nossas marcas aparecem nas mentes dos

¹⁰ Pop-Tart é uma espécie de bolo industrializado.

¹¹ Cadeia de suprimentos é comumente referenciada por seu equivalente em inglês.

nossos clientes no momento que eles fazem uma decisão de compra?” e “Quais os fatores que mais importam para os usuários e quão bem posicionados nós estamos para comunicar estes fatores aos nossos clientes?”. Com base nessas perguntas, foram analisadas diversas fontes de dados, obtendo algumas revelações interessantes, como, por exemplo, se os clientes seriam melhor influenciados a fechar um pacote de serviços (TV, Internet e telefonia) de acordo com a programação do serviço de TV e menos influenciados por características dos serviços de voz. Essa foi uma informação contraditória da obtida por meio de questionários tradicionais e permitiu que a empresa mudasse sua campanha de marketing. Os autores ressaltam que, para os dados serem úteis, eles precisam atravessar as barreiras organizacionais, o que é, geralmente, difícil e causa brigas e disputas internas por poder corporativo.

Brown, Chui e Manyika (2011) trazem 5 exemplos de empresas que usaram soluções de Big Data para melhorar suas dimensões competitivas de diversas formas, otimizando seus processos produtivos, diminuindo estoques, reduzindo os custos operacionais e aumentando suas vendas por meio da melhora do conhecimento dos hábitos de seus clientes ou criando novos produtos:

O caso de uma grande empresa de petróleo que conseguiu reduzir os custos com pessoal em quase 25% e, ao mesmo tempo, aumentar a produtividade em 5%. Esse ganho de eficiência operacional foi possível utilizando sensores ao longo da linha de produção, enviando dados que eram analisados em tempo real, permitindo ajustes imediatos na linha e evitando, assim, intervenções manuais, que são, em geral, demoradas, caras e perigosas.

Uma empresa de bebidas europeia acrescentou informações sobre a previsão do tempo ao seu sistema de estimativa de demanda e, analisando dados como previsão de chuva, temperatura e número de horas de sol em um determinado dia, conseguiu melhorar a assertividade nas estimativas de consumo em aproximadamente 5%, possibilitando uma diminuição do estoque e melhorando o processo de abastecimento das lojas.

O McDonalds equipou algumas lojas com dispositivos que rastreiam dados operacionais, interações com os clientes, tráfego nas lojas e padrão de ordens, permitindo a análise dos impactos na produtividade e nas vendas, gerados por mudanças em itens como cardápio, atendimento e desenho da loja.

Uma empresa líder do varejo norte-americano, a qual conduziu experimentos utilizando dados de movimentação das lojas em conjunto com registros de vendas para criar um modelo que permitisse encontrar o melhor posicionamento dos produtos na loja e o preço ideal para esses produtos. Como resultado, essa empresa conseguiu reduzir o estoque em 17% e, ao mesmo tempo, maximizar a venda de produtos de sua própria marca, aumentando seu lucro.

Uma empresa global de manufatura desenvolveu um novo produto, após utilizar o Big Data para melhorar o seu processo de manufatura. Eles aprenderam tanto sobre os próprios processos que decidiram prestar consultoria para outras empresas em situações semelhantes. Atualmente, tal estabelecimento agrega dados de chão de fábrica e cadeia de suprimentos de diversas empresas e vende ferramentas de *software* para melhorar a *performance* dessas empresas. Esse novo produto/serviço baseado no Big Data atualmente representa uma renda maior do que sua manufatura.

Gallant (2011) traz o exemplo de uma empresa de telecomunicações da Índia, que estava adicionando à sua base de cliente uma média de 5 milhões de novos assinantes por mês; entretanto, estavam perdendo 1,5 milhões de clientes para a concorrência todos os meses. Após uma análise minuciosa, identificaram que os clientes tinham uma alta tendência de mudar de operadora depois de 6 chamadas interrompidas. Dessa forma, a empresa, após a quinta chamada interrompida, disparava um torpedo com uma promoção para o cliente, diminuindo, assim, a sua fuga para a concorrência. Apenas com um modelo de Big Data foi possível a análise de bilhões ou até mesmo trilhões de registros ou eventos, que permitiram a empresa criar um modelo individualizado de relacionamento com seus clientes.

Mais uma evidência da melhoria nas dimensões competitivas causada pelo uso do Big Data nos mais diversos segmentos é o estudo de caso realizado por Bughin e Livingston (2011) com a AstraZeneca, uma gigante do ramo farmacêutico. Mark Lelinski, vice-presidente da empresa, relata que antigamente as principais preocupações da área de pesquisa e desenvolvimento era garantir que um novo medicamento era eficiente, seguro e aprovado pelas regulamentações. Recentemente, uma nova preocupação surgiu, tão importante quanto as demais: como esse produto pode ser comparado com os concorrentes e por qual motivo ele deve ser escolhido frente à concorrência cada vez mais acirrada.

Esse problema levou à criação de um projeto de Big Data concatenando diversas fontes de dados para medir o custo total para tratar uma determinada doença, incluindo idas ao médico, exames de diagnóstico, atendimentos de emergência, entre outros procedimentos médicos. Os medicamentos representavam apenas 11% desse custo total; por isso, se uma droga conseguisse reduzir a fatia remanescente de custos de 89%, poderia ser um diferencial e uma forte justificativa financeira para sua aquisição. Importante lembrar que os clientes da AstraZeneca não são somente os consumidores finais, mas também planos de saúde e hospitais. Os autores destacam que o apoio da alta administração foi fundamental para viabilizar esse projeto.

Com esse novo modelo em funcionamento, é possível determinar, ainda no começo do ciclo de desenvolvimento, a viabilidade comercial dessa nova droga, possibilitando interromper o que seria um grande investimento com baixo retorno, e focar o investimento em pesquisas e desenvolvimento de drogas que sejam lucrativas para a empresa e que melhorem a vida dos pacientes, reduzindo o custo total de atendimento.

Brynjolfsson e McAfee (2012) descrevem um estudo feito no MIT Center for Digital Business, objetivando comprovar a hipótese de melhor desempenho das empresas que utilizam a análise de dados no cerne de sua estratégia corporativa. Para isso, foram aplicadas entrevistas estruturadas com executivos de 330 empresas norte-americanas de capital aberto sobre suas práticas de gestão organizacional e TI (Tecnologia da Informação), agregando o resultado desses dados de desempenho disponíveis em balanços anuais e fontes independentes. Averiguou-se neste estudo que as empresas cujos gestores tomavam decisões fortemente embasadas em dados possuíam, em média, 5% mais produtividade e uma rentabilidade 6% maior do que as demais.

A Sears¹² era uma das empresas que sofriam com o crescimento da massa de dados. Necessitando desenvolver promoções personalizadas para seus clientes, o processo de analisar o gigantesco volume de informações que possuía levava cerca de 8 semanas, tornando as promoções, muitas vezes, obsoletas e ineficientes. Utilizando tecnologias de Big Data e promovendo a unificação das bases de dados, o tempo exigido para análise caiu para apenas 1 semana (BRYNJOLFSSON; MCAFEE, 2012).

¹² Gigante do setor do comércio de varejo norte-americana, com uma rede de lojas de departamentos espalhadas por todo o país.

Davenport, Barth e Bean (2012), por sua vez, citam empresas de cartão de crédito que desejavam encontrar os melhores prospectos de clientes em grandes bases de dados, porém demoravam semanas para extrair, preparar e elaborar ofertas. Atualmente, com a monitoração das interações no seu sítio de internet e no *call center*, as empresas conseguem criar ofertas personalizadas em menos de um segundo, além de otimizá-las constantemente por meio da mensuração dos resultados.

Big Data torna possível para empresas a criação de sistemas preditivos. Gallant (2011) ressaltou que a importância de conseguir inferir o que pode acontecer com um mínimo de antecedência é mais relevante do que poder analisar toda a informação global, porém 6 meses depois do ocorrido. Prever a intenção de um cliente de mudar para a concorrência tem muito mais valor do que analisar por que o cliente deixou a empresa. Não significa que entender os motivos que levaram o cliente a tomar essa decisão não tenha seu valor, mas poder tomar ações de contenção, impedindo a saída do cliente, é mais importante do que agir reativamente, melhorando o que estava errado ou investindo em campanhas caras para tentar trazer o cliente perdido de volta. Essa mesma visão é compartilhada por Hea, Zhab e Li (2013), que adicionam a importância do uso de dados disponíveis em redes sociais para a capacidade de realizar essas predições.

2.4 A relevância das redes sociais como fonte de dados

As redes sociais são definidas por Boyd e Ellison (2008) como sistemas nos quais o usuário pode criar um perfil público ou semipúblico, manter uma lista de contatos com os quais ele estabelece diferentes tipos de conexão ou relação e acessar ou visualizar os contatos dos seus contatos, criando, assim, a sua rede de relacionamento. Ainda segundo Boyd e Ellison (2008), a história dessas redes é relativamente nova, datando de 1997 o primeiro site, denominado sixdegrees.com, segundo o qual duas pessoas em qualquer parte do mundo estavam separadas por, no máximo, 6 níveis, graus ou conexões.

Cresce o número de indivíduos que faz uso das redes sociais para expressar suas opiniões, comunicar e postar nesses sites da internet os mais diversos tipos de informações. Um dos fatores de aumento significativo no uso da internet e redes sociais é a proliferação de dispositivos móveis, como *smartphones*, que permitem o acesso a partir de qualquer lugar. A

Figura 8 mostra que, em alguns casos, mais da metade dos usuários dos principais sites da internet realizaram acessos via *smartphones*. Krishnamurthy e Wills (2010) lembram que dentro deste novo contexto, algumas redes sociais foram criadas especificamente para serem acessadas de dispositivos móveis, como o Foursquare.

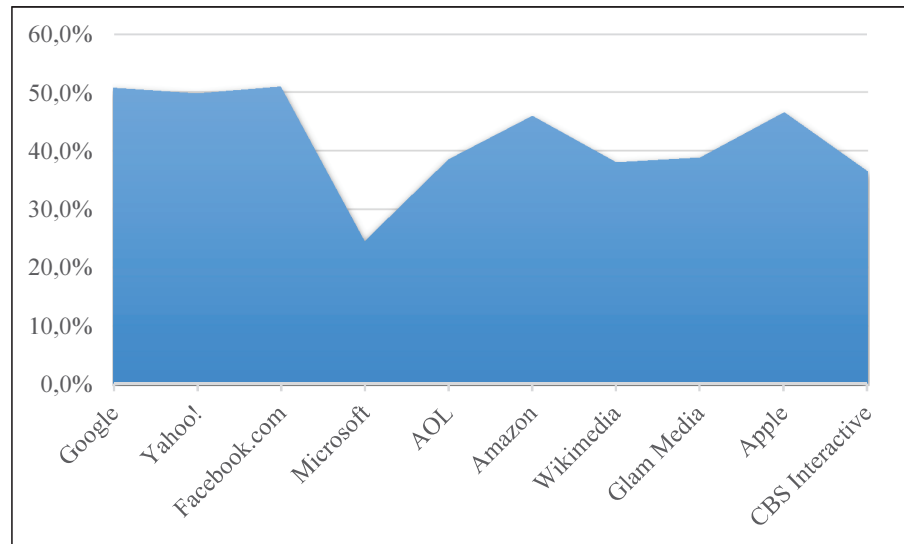


Figura 8: % de usuários de *smartphones* (set/2012)

Fonte: Adaptado de COMSCORE (2013).

Essas informações tornam-se, então, disponíveis para que outros indivíduos ou empresas extraiam os dados desses sites e por meio do uso de tecnologias, como o Big Data, analisem quantidades massivas de dados, gerando informações relevantes sobre os usuários das redes sociais.

O Big Data, em conjunto com as redes sociais, permite o que Brown, Chui e Manyika (2011) chamaram de técnicas de análise de sentimento, isto é, fazendo uso da análise de quantidades imensas de dados criados pelos clientes em redes sociais, blogs e outras fontes na internet, é possível obter um *feedback* real do cliente com relação ao atendimento prestado, qualidade do produto ou impacto de uma campanha de marketing. A análise e obtenção deste *feedback* imediato permite que a empresa resolva o problema ou que a área de marketing ajuste sua campanha rapidamente, maximizando o resultado.

Torna-se possível a individualização do consumidor pela empresa, gerando uma potencial vantagem para a empresa ao conhecer melhor seu consumidor. Esse processo pode trazer vantagens também ao próprio consumidor, o qual pode receber ofertas direcionadas às suas necessidades. Porém, esse novo cenário abre espaço para uma importante discussão ética sobre o uso destes dados e sobre o risco ao qual as pessoas estão se expondo, ao perderem o controle ou influência sobre o uso desses.

Dentro deste contexto, as redes sociais vêm crescendo e hoje possuem elevada significância no cotidiano das pessoas, como ferramenta de compartilhamento de ideias, opiniões e sentimentos. Por exemplo, no ano de 2011 constatou-se que 13% dos norte-americanos adultos usaram o Twitter regularmente (SMITH, 2011). Outra evidência da crescente importância, não apenas das redes sociais, mas do conteúdo que circula por essas redes, foi a decisão em 2010 da Biblioteca do Congresso Norte Americano em arquivar e disponibilizar todos os *tweets* trafegados pelos usuários dessa rede (LOC, 2013).

Esse é um fenômeno global e em franca expansão. O LinkedIn alegou ter em março de 2012 mais de 160 milhões de usuários ativos, crescendo a uma taxa de 2 novos membros por segundo, e esse ainda é um número pequeno perto dos 900 milhões de usuários ativos do Facebook (AQUINO, 2012).

Como relatado por Hea, Zhab e Li (2013), as redes sociais já são utilizadas por diversas empresas, em especial para propaganda, campanhas de marketing e vendas, monitoração de satisfação dos clientes e até mesmo para desenvolvimento ou inovação de produtos. Exemplifica-se o exposto com a cadeia de hotéis Starwood¹³, que há alguns anos usam as redes sociais para estarem mais próximos de seus hóspedes, resolvendo rapidamente qualquer percalço que possa ter ocorrido em suas estadas nos hotéis ou até mesmo para fazer sugestões de novas viagens e enviar aos seus clientes ofertas direcionadas.

Utilizar as redes sociais como repositório de dados para análise torna-se mais simples com o avanço da tecnologia e até mesmo pessoas sem experiência ou conhecimentos técnicos avançados podem utilizar ferramentas como o ScraperWiki¹⁴ para fazer extração de dados de redes sociais como o Twitter e, automaticamente, realizar processos básicos de *text mining*

¹³ Cadeia norte-americana com milhares de hotéis espalhados pelo mundo. No Brasil sua marca mais conhecida é a rede de hotéis Sheraton.

¹⁴ Disponível em: <<https://scraperwiki.com/>>. Acesso em: 22/jan/2014

Adicionalmente, caso a empresa deseje usar dados de redes sociais como fontes de informações para geração de estatísticas de, por exemplo, aceitação de um produto, repercussão de uma ação de marketing, percepção do consumidor, entre outros, é de extrema relevância entender o perfil dos usuários das redes sociais. Gayo-Avello (2011) relata a experiência falha de tentar prever o resultado das eleições presidenciais americanas de 2008 utilizando modelos de Big Data e análise de *tweets*. Nesse estudo, Obama ganharia com uma margem imensamente maior do que foi alcançada na realidade. Essa falha pode ser explicada analisando a pesquisa de Smith (2011), a qual relata que em média 13% dos adultos norte-americanos fazem uso do Twitter, entretanto esse número está alocado de forma díspar entre os diferentes perfis de indivíduos adultos. A Figura 10 mostra que o uso do Twitter nos Estados Unidos é maior entre pessoas de etnia negra que possuem entre 18 e 29 anos, ou seja, um perfil mais inclinado a votar no Obama.

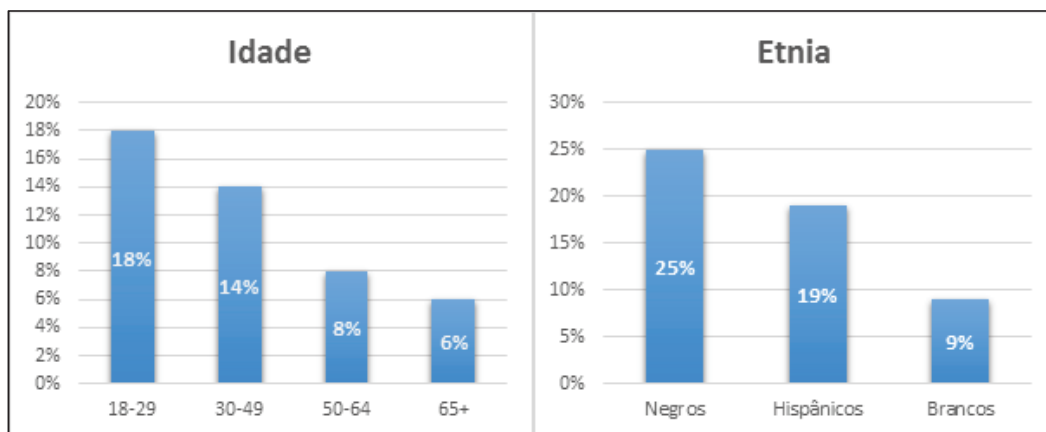


Figura 10: % de adultos usuários do Twitter classificados por etnia e idade

Fonte: Adaptado de Smith (2011)

2.5 A ética no uso das informações do consumidor

A palavra ética vem do grego – *ethos* e, nesse idioma, significa caráter, modo de ser. Ética está ligada às diferentes formas como os indivíduos reagem ou respondem ao tradicionalmente estabelecido, depois de julgarem as ideias de acordo com seus valores individuais sobre o bem e o mal. Segundo Murgel, Silva e Neve (2006), ética existe dentro de um conceito de moral, absorvido por cada indivíduo, que é influenciado de forma distinta por diversos fatores existentes na sociedade. Podem existir situações cuja ética seja “cinza”, ou seja,

situações nas quais a classificação ética pode divergir entre pessoas, dependendo da sociedade em que vivem e influências sofridas ao longo da vida. Motivo esse que ocasiona, em diversas instituições, a criação de conselhos de ética, responsáveis por ponderar essas situações.

No âmbito empresarial, é claro que o objetivo final de qualquer empresa é dar lucro, porém Murgel, Silva e Neve (2006) lembram que o caminho para o lucro passa pela satisfação do cliente, sendo necessário, para isso, oferecer produtos e serviços relevantes ao cliente e ouvir a sua opinião de forma a fechar o ciclo com um processo contínuo de melhoria empresarial. Entretanto, a própria forma de ouvir o cliente também está sujeita a uma avaliação ética. Ética pode ser usada para definir ações certas ou erradas, podendo ser aplicada nas relações das empresas com os seus clientes. Ética existe em um plano distinto da legislação. Uma empresa pode atuar dentro da lei, porém tomar decisões que firam a ética.

Nesse novo paradigma, com informações criadas de maneira pulverizada pelos consumidores e com os novos sistemas para armazenamento e análise dos dados, os chamados sistemas de Big Data, surge a necessidade de observar novas disciplinas da informação: privacidade, segurança, retenção, processamento, propriedade e integridade (THE ECONOMIST, 2010).

Não se discute aqui a relevância das redes sociais como forma de comunicação entre empresa e cliente, já utilizado por várias instituições. Entretanto, novas formas complementares de uso, como a análise dos dados, abre novas possibilidades como ações de marketing direcionais e previsão comportamental.

Um dos exemplos do risco e exposição por meio do compartilhamento de informações nas redes sociais podia ser observado no site <http://icanstalku.com/>. O site atualmente não está mais em operação, mas deixa um relato de como, por intermédio da análise dos metadados de fotos postadas no Twitter, permitiu inferir informações sobre os usuários do Twitter, tais como:

- a) onde vivem;
- b) quem mais vive no mesmo lugar;
- c) padrões de deslocamento;
- d) onde almoçam todos os dias;
- e) com quem almoçam;
- f) opções de entretenimento.

Dados extraídos de redes sociais tradicionais como LinkedIn ou Twitter podem trazer metadados¹⁶ que incluam diversas informações adicionais sobre os usuários, como a geolocalização desses. A pesquisa apresentada na Figura 11 aponta a geolocalização dos usuários que fizeram *tweets* com a *hashtag* #bigdata, durante um período de tempo específico. Esses dados podem ser utilizados como mais um fator de decisão em uma campanha de marketing, permitindo uma abordagem mais assertiva para as empresas, direcionando uma propaganda para determinado mercado ou medindo a aceitação de uma campanha em determinada região.



Figura 11: Amostragem de local de origem de *tweets* com a *hashtag* #bigdata

Fonte: Elaboração do autor.

Rastrear informações geradas por indivíduos não é algo novo. O uso de *cookies*¹⁷ para rastrear padrões de navegação dos usuários na internet existe há vários anos. Alguns sites chegam a requerer, para o seu funcionamento, que o *software* de navegação (Internet Explorer, Firefox, Chrome etc.) aceite armazenar *cookies*. No entanto, conforme explicado por Krishnamurthy e Wills (2009), tornam-se mais frequentes entidades (empresas) que executam

¹⁶ São dados sobre os dados. Exemplificando, um texto postado em uma rede social constitui o dado principal, e dados como nome autor, localidade, dispositivo de acesso constituem os metadados sobre aquela postagem.

¹⁷ *Cookies* são pequenos arquivos de dados, enviados pelos sites da internet acessados pelos usuários e armazenados nos computadores desses. Esses arquivos podem ser posteriormente consultados por eles, ou por outros sítios, para obter informações sobre sítios navegados, pesquisas realizadas, entre outras informações.

a função de agregadores. Tais empresas coletam informações dos *cookies* armazenadas nos computadores dos usuários, executam processos de análise de dados e vendem essas informações para empresas de propaganda, marketing ou qualquer outra empresa que queira contratar seus serviços. Geralmente, a única informação do *cookie* vinculada ao usuário é o IP ou o navegador que esse está utilizando. Com isso, após uma pessoa buscar na internet por um novo produto, desde um refrigerador até uma viagem de férias, qualquer pessoa que utilize o mesmo computador passará a receber em outros sites que visitar anúncios direcionados sobre o referido tema. Porém, essa informação não é individualizada, ou seja, não existe uma relação entre o *cookie* e o nome ou endereço residencial do usuário.

Entretanto, como ressaltado por Krishnamurthy e Wills (2009), o uso de *cookies* de navegação, juntamente com as redes sociais e o poder computacional de soluções de Big Data, pode trazer novas consequências para os usuários, como a geração de informações pessoais identificáveis (IPIs), ou seja, o estabelecimento de vínculos entre informações coletadas pelos *cookies* (como sites visitados, padrão de navegação e interesses) com as informações da rede social (como nome, *e-mail*, amigos, endereço).

Quanto aos agregadores de informação, apesar de esses alegarem não armazenar ou gerar essas IPIs, Krishnamurthy e Wills (2009) dizem que não há como garantir ou controlar que essas empresas, de fato, não o façam.

Embora este estudo tenha foco nas redes sociais, o risco para o usuário e a questão ética na manipulação dos dados está presente de forma geral em toda a internet, como comentado por Krishnamurthy e Wills (2009), que citam os ambientes de comércio eletrônico como Amazon, eBay, entre outros, que armazenam informações pessoais dos clientes e podem ser agregadores dessas informações como *cookies* ou outras fontes de dados.

Atualmente, não existe um entendimento consolidado de que os dados gerados por um usuário, desde informações postadas em redes sociais, informações quanto ao uso de cartão bancário, do telefone ou até *cookies* de internet, sejam de propriedade do indivíduo. Isso cria abertura para as empresas usarem esses dados de forma ampla. Dentro desse contexto, o *The Economist* (2010) alerta para o fato de que cada vez mais as pessoas estão disponibilizando informações potencialmente confidenciais nas redes sociais e isso irá ocasionar uma tensão entre empresas que querem utilizar esses dados em benefícios próprio e os indivíduos que querem proteger as suas informações pessoais.

Diante da natureza dinâmica da informação digital em um mundo altamente conectado, dar-se-á a necessidade de regulamentações globais que norteiem o uso da internet entre países. Por exemplo, um usuário brasileiro que usa o Facebook tem os seus dados armazenados em servidores localizados nos Estados Unidos e, portanto, sujeitos a uma série de legislações locais. Uma empresa europeia que utilize esses dados obedece a um outro grupo de normas que podem ser dissonantes das leis brasileiras e estadunidenses. Modelos de regulamentações como os dos oceanos podem servir como base de inspiração para a criação de normas globais para a internet. Além disso, é necessário um órgão mundial responsável por este controle, como a OMC (THE ECONOMIST 2010).

A participação dos países emergentes na geração de dados irá crescer significativamente nos próximos anos, sobrepujando os mercados maduros, como demonstrado na Figura 12. Essa mudança dá-se devido ao volume populacional desses países, a crescente facilidade na aquisição de instrumentos de comunicação, como smartphones e tablets e também a crescente facilidade de conexão à internet, potencialmente agravando os problemas já mencionados de legislação e cultura, dada a menor maturidade destes países no uso e controle de informações na internet.

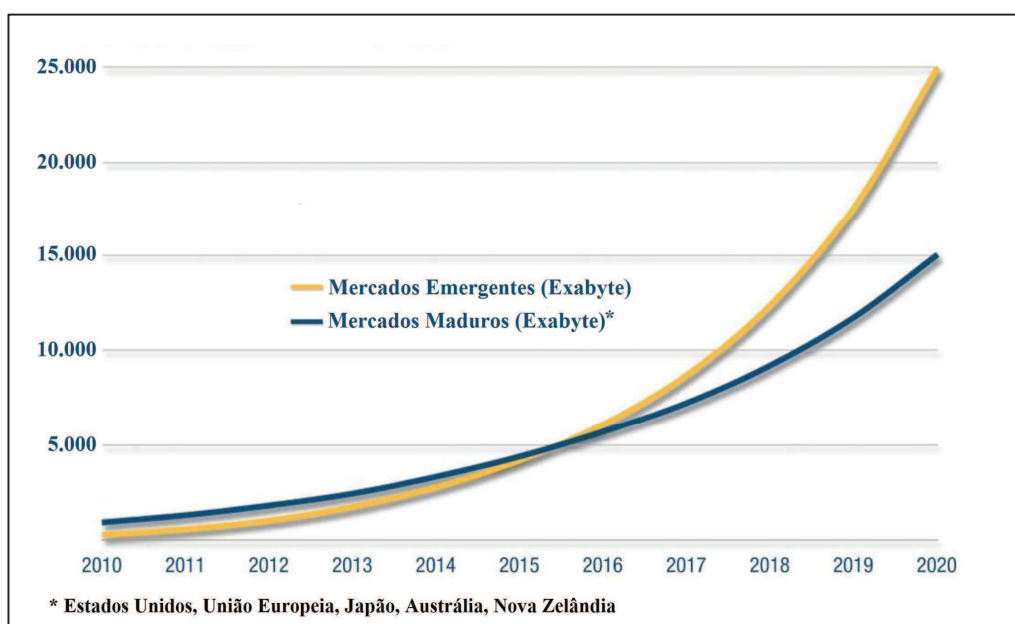


Figura 12: O crescimento dos mercados emergentes na geração de dados

Fonte: Traduzido de Gantz e Reinsel (2012, p.7).

Aquino (2012) revela que o uso de *cookies* ou redes sociais para direcionar propagandas divide a opinião dos consumidores. Esse conflito fica evidenciado na Figura 13, com 43% considerando uma invasão de privacidade, e 45% estando dispostos a receber anúncios relevantes durante o uso das redes sociais, entendendo ser essa uma contrapartida pelo uso gratuito dos serviços desses sites. Tucker e Thaler (2013) relatam que, embora possa ser interessante para o consumidor ter anúncios feitos sobre medida para a sua necessidade atual, existe um receio destes usuários em serem rastreados por empresas de propagandas, sem a compreensão completa de quão extensa é essa monitoração ou como esses dados são utilizados. Essa divisão sugere que é necessário cautela por parte das empresas ao usarem dados de clientes provenientes das redes sociais.

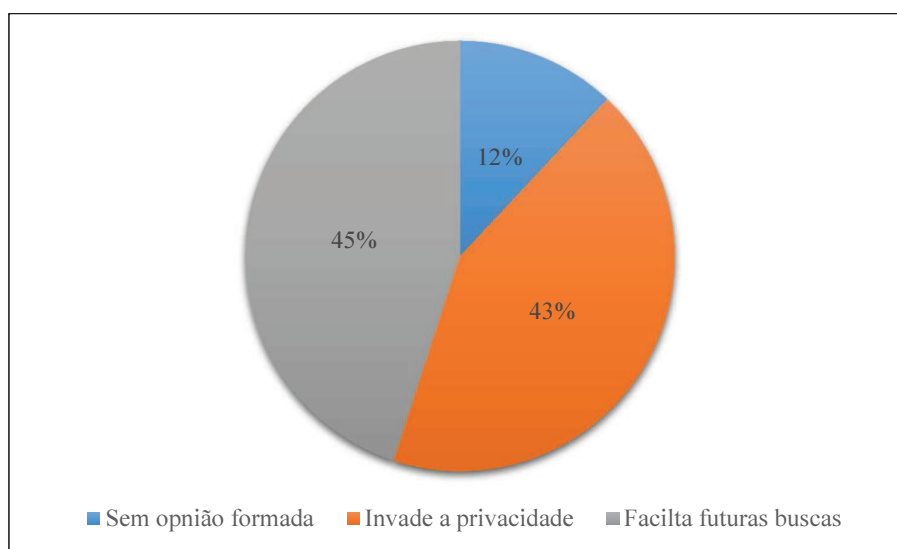


Figura 13: Percepção dos indivíduos sobre o rastreamento da navegação na Internet

Fonte: Adaptado de Borba, Santos e Kawamoto (2013).

Para Kosinski, Stillwell e Graepel (2013), a grande quantidade de dados disponíveis *on-line* oriundos desse novo modelo de interação social trará discussões no tangente à privacidade e propriedade dos dados, na medida em que torna possível para governos, empresas ou até mesmo outras pessoas inferir informações privadas e não divulgadas dos indivíduos, usando análises de dados.

Tene e Polonetsky (2012) alertam que, apesar de existirem políticas de privacidade nas redes sociais, inclusive exigindo a “aceite” dos usuários no momento da criação de uma conta para acesso, essas, muitas vezes, servem apenas como isenção de responsabilidade para as

empresas, ao invés de balizadores das garantias de privacidade para os usuários. Adicionalmente, as redes sociais deixam disponíveis uma série de dados pessoais, os quais podem ser livremente acessados, como demonstrado na Figura 14.

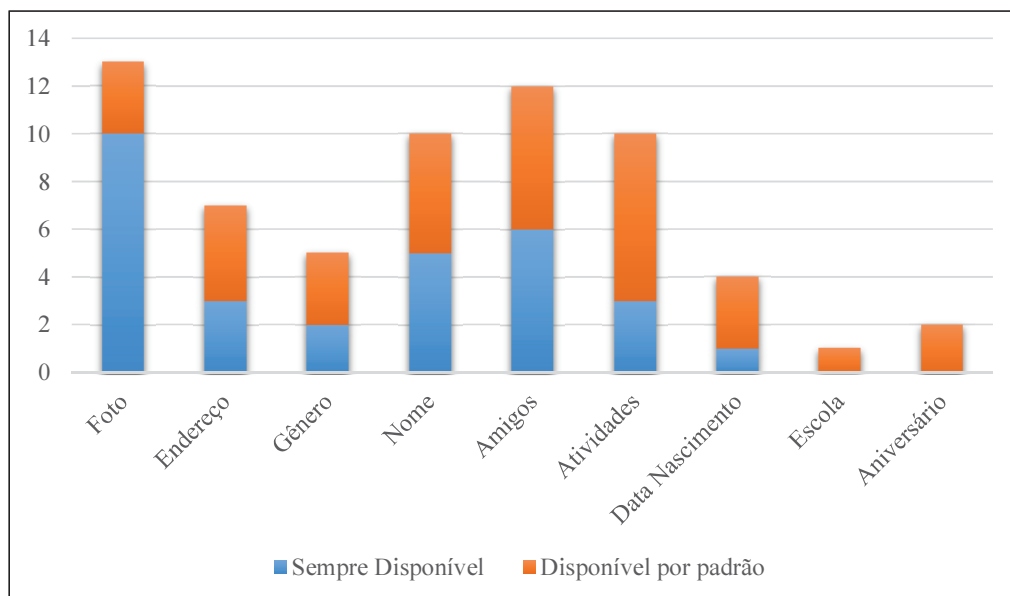


Figura 14: Informações disponíveis em 13 redes sociais investigadas

Fonte: Adaptado de Krishnamurthy e Wills (2010).

Torna-se cada vez mais difícil ocultar ou restringir totalmente os rastros digitais de cada indivíduo nas redes sociais. Mesmo quando é feita a opção por não informar dados pessoais, como sexo ou situação social, Kosinski, Stillwell e Graepel (2012) demonstraram que é possível inferir com elevado grau de assertividade estas e outras informações privadas dos indivíduos, tais como: orientação sexual; etnia; religião; orientação política; inteligência; estado de espírito; uso de drogas; divórcio dos pais; idade e sexo. Para levantar esses dados, não foram analisados nenhuma informação postada pelo próprio indivíduo no Facebook. Utilizou-se apenas os *likes*¹⁸ desses indivíduos e análise dos textos que outras pessoas postaram, e que esses indivíduos “curtiram”. Este estudo foi feito com mais de 58 mil voluntários que consentiram na análise dos seus *likes* e preencheram um questionário para comparação dos resultados. Algumas das características analisadas, como etnia, obtiveram mais de 95% de taxa de acerto. Atualmente, não existe opção de tornar privado ou ocultar os *likes* do Facebook.

¹⁸ Um usuário do Facebook pode marcar postagens (texto, foto etc.) de outros usuários como *Like* ou “curtido”, significando que esse usuário gostou ou concorda com o conteúdo postado pelo outro.

Segundo Tene e Polonetsky (2012), é imperativa a criação de regulamentação pertinente que defina o conceito de *personally identifiable information* ou informações pessoais identificáveis (IPI), com o objetivo de estabelecer princípios e limitações quanto ao seu uso. Porém, essa é uma tarefa complexa pelo fato de as empresas poderem armazenar os dados em locais remotos, com influência de leis distintas daquelas em que vivem os usuários.

Uma vez que esses dados são capturados pelas empresas, surgem dois novos pontos relevantes para debate: a segurança que as empresas oferecem sobre esses dados e por quanto tempo esses ficarão retidos em poder dessa corporação. Isso significa que, muito depois do usuário ter removido a informação do seu perfil público, esses dados podem ainda estar armazenados em bancos de dados ao redor do globo, sem que haja controle do usuário sobre isso (THE ECONOMIST, 2010).

Por esse motivo, Stein (2011) relata um projeto de lei desenhado pelo Senador norte-americano John Kerry, responsável pelo subcomitê de Comunicação, Tecnologia e Internet, que obriga as empresas a oferecer proteção contra hackers para os dados armazenados sobre os consumidores, e também as obriga a permitir que os consumidores tenham acesso a esses dados e possam alterar informações que não condigam com a realidade.

Tankard (2012) alerta para o fato que, ao obter dados dos usuários, como redes sociais, padrão de navegação com *cookies*, padrões de compra, entre outros, muitas empresas criam bancos de dados com informações sensíveis dos clientes. Essas informações são armazenadas sem que exista uma regulamentação consistente no mercado, podendo expor esses clientes ao vazamento de informações pessoais no caso de falhas de segurança. Deibert e Crete-Nishinata (2011) advertem que o furto ou perda desses dados pode ocorrer sem que o usuário saiba disso, aumentando ainda mais o risco para esse.

Vale ressaltar que os indivíduos estão sujeitos a tais problemas com quaisquer dados coletados pelas empresas, como, por exemplo, o seu histórico de compra em uma rede de varejo ou cadastro no próprio site da empresa. Caso prático ocorreu em 2011, quando a rede de jogos *on-line* da Sony, chamada PlayStation Network, foi invadida e clientes tiveram seus dados pessoais, inclusive número de cartão de crédito, furtados por hackers¹⁹.

¹⁹ Fato divulgado pela mídia mundial, como, por exemplo, na matéria “Dados pessoais de usuários da PSN foram roubados, admite Sony”. Disponível em: <http://g1.globo.com/tecnologia/noticia/2011/04/dados-pessoais-de-usuarios-da-psn-foram-roubados-admite-sony.html>. Acesso em: 20/nov/2013

Existem ainda outras questões éticas não abordadas aqui, como a discriminação baseada na análise dos dados. Uma instituição financeira não pode negar um empréstimo com base na cor da pele do cliente. Mas, no exemplo trazido pelo *The Economist* (2010), com a decisão de aprovação ou não do empréstimo sendo tomada por sistemas automatizados, como garantir que esses sistemas não levem em consideração fatores como educação ou bairro de residência, fatores esses ainda fortemente ligados à raça? Poderia ser essa uma forma velada de discriminação.

Por outro lado, a restrição total aos dados tampouco é uma solução. Caso o uso de dados gerados por indivíduos seja demasiadamente restringido ou regulado, alguns dos benefícios do Big Data também serão severamente afetados, como nos exemplos trazidos por Tene e Polonetsky (2012):

- a) o remédio Vioxx foi banido do mercado depois que seu uso foi relacionado com 27.000 mortes por parada cardíaca entre os anos de 1999 e 2003;
- b) os padrões de propagação do vírus da influenza foram mapeados e foi possível prover uma resposta mais rápida de amparo à população.

É imperativo haver um balanço entre a privacidade à qual todos têm direito e os benefícios que podem ser provenientes do uso dessas informações, desde questões como saúde e segurança até a coleta de dados estatísticos para entender padrões de serviços, permitindo a entrega de ofertas mais relevantes ao próprio consumidor (TENE; POLONETSKY 2012).

Existe uma luta tecnológica descrita por Tene e Polonetsky (2012) entre os processos de remoção dos traços de identificação da informação, tais como: anonimização; pseudo-dados; criptografia; chave de codificação; fragmentação de dados e os processos de re-identificação dos dados, que permite a pesquisadores, empresas ou instituições do governo identificarem os indivíduos geradores dos dados.

Levará algum tempo para que as organizações entendam completamente o que é viável em relação a capturar, gerenciar, analisar e, acima de tudo, criar inteligência social, a fim de agir sobre o conteúdo gerado pelo usuário (LESLIE, 2012, p. 9, tradução nossa).

2.6 O cientista de dados

Davenport e Harris (2007) já colocavam ser essencial que haja um balanço entre as competências envolvidas em um projeto de análise de dados, ou seja, investir demais em tecnologia sem ter uma equipa capacitada para análise dos dados pode resultar na falha do projeto ou em resultados sub-ótimos.

Os autores afirmam que empresas que se destacam pela análise de dados montaram times de profissionais analíticos, que transitam na estrutura da organização possuindo vínculos tanto para a TI quanto para o presidente da empresa.

Surge, então, um novo profissional, o cientista de dados, com uma formação que agrega, além dos conhecimentos sobre as ferramentas analíticas, conhecimentos de negócio, tecnologia e principalmente estatística. Esse profissional será, para Brynjolfsson e McAfee (2012), peça fundamental para viabilizar o Big Data dentro das empresas, ajudando durante todo o ciclo, desde a elaboração das perguntas de negócio relevantes até a extração das respostas da massa de dados.

Esse profissional terá um papel diferente na organização comparado ao antigo analista de dados. Davenport, Barth e Bean (2012) colocam o cientista de dados mais próximo às áreas de produção, marketing e estratégia organizacional, ressaltando que não basta a esse profissional conhecer tecnicamente a solução, sendo necessário também um conhecimento do negócio e a capacidade de se comunicar com os tomadores de decisão da companhia. Complementa dizendo que esse novo profissional pode fazer as empresas repensarem o relacionamento entre TI e o negócio.

Ainda não existe um currículo oficial ou padronizado para o cientista de dados, que, para Brynjolfsson e McAfee (2012), pode incluir, além das disciplinas comentadas acima, conhecimentos de física ou ciências sociais. Torna-se, assim, uma tarefa complexa a formação ou contratação de profissionais com formação e conhecimentos tão ecléticos, o que, para Davenport, Barth e Bean (2012), pode tornar-se uma restrição para o sucesso de um projeto de Big Data.

2.7 Possíveis obstáculos ao Big Data

É preciso ter cuidado com as promessas de ganhos fáceis, ou, como exposto por Davenport e Harris (2007), as lendas analíticas, como o caso das fraldas e cervejas do Walmart: supostamente, pessoas que comprem fraldas também comprariam cervejas. Sendo assim, bastava colocar os dois produtos juntos para obter um aumento no número de vendas. Embora o Walmart seja um dos pioneiros na análise de dados e mencionado em estudos como um grande entusiasta de soluções de Big Data, tal relação simples nunca foi evidenciada em nenhum estudo.

Adicionalmente, Davenport e Harris (2007, p. 152) avisam que existem as “mentiras, malditas mentiras, e estatísticas”, em alusão ao problema do cientista/analista de dados, mesmo que acidentalmente, mascarar ou distorcer os resultados estatísticos na esperança de obter correlações forçadas ou confirmar suas impressões pessoais.

Outro obstáculo está na aquisição dos dados, os quais, obviamente, constitui subsídio fundamental nesse tipo de projeto. Davenport e Harris (2007) relatam que obtê-los pode ser uma tarefa extremamente complexa e traz uma lista de questões que precisam ser tratadas no processo de obtenção dos dados:

- a) Quais são os dados necessários? A quantidade de dados existente tem crescido nos últimos anos, porém isso traz um problema de quais dados são realmente necessários e quais dados são apenas “ruído”, ou seja, apenas informações sem relevância que tornam a análise dos dados relevantes mais complexa;
- b) Onde estes dados podem ser obtidos? As fontes de dados são as mais diversas possíveis, desde bancos de dados espalhados pela organização até mesmo fontes externas como redes sociais;
- c) Qual a quantidade de dados? Coletar dados apenas para o caso de talvez precisar deles no futuro pode levar a um volume de dados muito grande, porém de baixa relevância;

- d) Qual a qualidade desses dados? Principalmente no caso de uso de dados externos à organização, garantir a qualidade e veracidade dos dados é fundamental para que o resultado da análise seja verossímil;
- e) Como gerenciar o ciclo de vida dos dados, desde a sua obtenção, retenção, até o descarte? Após obter os dados e até o seu descarte, é importante que esses dados sejam regidos por uma política de governança rígida e estruturada, garantindo a integridade e segurança dessa informação.

É grande o desafio enfrentado pelas empresas no compartilhamento de informação entre unidades de negócio, criando, algumas vezes, verdadeiros feudos de dados. Brown, Chui e Manyika (2011) citam as empresas financeiras como exemplos desse modelo em que a falha de compartilhamento de informação entre setores, como gerenciamento das finanças e empréstimos, impede a criação de uma visão completa de seus clientes e o entendimento das possíveis ligações entre os diversos segmentos.

Bughin e Livingston (2011) relatam que muitos executivos só procuram por soluções de Big Data quando têm uma necessidade urgente de responder a uma ameaça ou identificam uma oportunidade de atacar um determinado mercado. No entanto, identificar essas oportunidades e ameaças pode ser difícil. É necessário que a estratégia de uso do Big Data esteja intimamente conectada à estratégia da empresa e, para isso, é imperativo um extenso planejamento dos executivos a fim de decidir como os recursos analíticos devem ser aplicados para obter o resultado esperado. Porém, os autores ressaltam que, em alguns casos a melhor decisão pode ser esperar antes de embarcar em um projeto analítico. A empresa pode estar se reestruturando, redefinindo o foco estratégico ou passando por dificuldades financeiras. Entretanto, é importante que os executivos tenham uma visão estratégica de longo prazo e acompanhem de perto a concorrência, a qual pode estar desenvolvendo avanços na competição analítica.

O Big Data é capaz de trazer *insights* importantíssimos para a empresa, mas a decisão e visão do executivo sempre será essencial para o negócio, porém, agora, muito mais embasada em dados. Como no exemplo trazido por Vivek Ranadive, CEO da Tibco²⁰ e um autor de livros sobre o tema: uma determinada loja varejista de material de jardinagem sabe que compradores de semente têm 90% de probabilidade de comprar fertilizante e 40% de chance de comprar móveis de jardim. Em uma primeira análise, pode-se imaginar que a melhor estratégia esteja

²⁰ Empresa americana de capital aberto, que fornece soluções de análise de dados

em incentivar a compra de fertilizante, colocando-o perto das sementes. Porém, analisando mais a fundo, os executivos determinaram que a melhor estratégia era colocar os móveis próximos às sementes, para desenvolver a venda desses, afinal há uma margem maior de crescimento para esse produto (GALLANT, 2011).

Junto com o desenvolvimento do Big Data, a arquitetura irá se desenvolver em um ecossistema de informação: uma rede de serviços internos e externos continuamente compartilhando informação, otimizando decisões, comunicando resultados e gerando novas perspectivas para o negócio (DAVENPORT; BARTH; BEAN, 2012, p. 4, tradução nossa)

Apesar do impacto positivo do Big Data, Cearley e Claunch (2013) advertem que novas tecnologias possuem diversos vetores de impacto, podendo causar a disrupção de outras tecnologias tradicionais como *data warehousing*²¹ e gerenciamento da informação. Contudo, essa disrupção não precisa, necessariamente, ser algo ruim, pois se explorada corretamente pode trazer benefícios para a empresa.

2.8 Dimensões competitivas

Segundo Porter (1990), o estudo das forças que se aplicam a determinada indústria são subsídios importantes na elaboração da estratégia de uma empresa. Tal estratégia deve ter como foco a melhoria na capacidade da empresa de competir no mercado, analisando os fatores externos e utilizando as competências internas de uma empresa para se colocar em um lugar de destaque no mercado.

O modelo de 5 forças de Porter (1991), descreve as forças que, particulares à cada indústria, determinam seu nível de concorrência:

- a) Potenciais entrantes: novas empresas que desejem entrar nesse segmento e que frequentemente trazem substanciais aportes financeiros, em geral causando uma batalha de preços e diminuindo as margens de lucro das empresas já instaladas;

²¹ Repositório central de dados, criado pelas empresas para armazenar dados oriundos de diversas bases de dados espalhadas pela organização, com o objetivo de permitir a análise deles de forma mais fácil e eficiente.

- b) Compradores: poder dos compradores em negociar preço e condições para venda dos produtos, em geral quando existem poucos compradores ou que adquirem em volume substancial. Pode ser devido também à alta padronização e falta de diferenciação no produto;
- c) Fornecedores: quando os fornecedores de matéria-prima para a indústria em questão são em menor número ou mais concentrados do que a indústria para a qual eles fornecem, limitando, assim, as opções da empresa e corroendo as margens de lucro e a flexibilidade da empresa;
- d) Produtos substitutos: quando o produto da indústria pode ser substituído por outro, causando, assim, um cenário de competição indireta entre indústrias;
- e) Concorrentes existentes: disputa entre as diversas empresas já constituídas por busca de uma maior parcela do mercado ou melhores margens de lucro. Essa disputa ocorre usando as mais diversas táticas, como concorrência de preços, batalha de publicidade, diferenciação de produtos ou criação de novos produtos.

Essas forças são apresentadas na Figura 15:

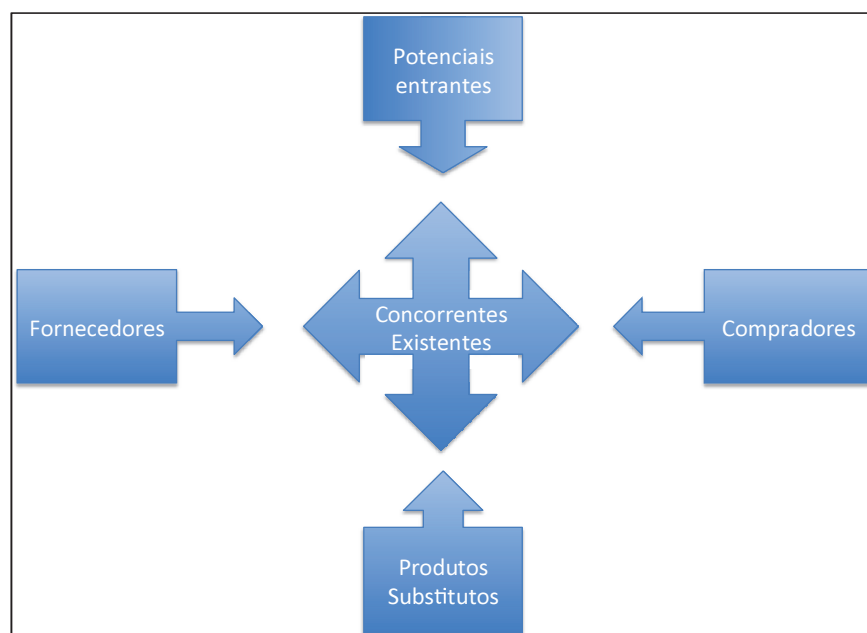


Figura 15: Forças que determinam a competição de uma indústria

Fonte: Adaptado de Porter (1991, p. 23).

Para Porter (1990), a vantagem competitiva de uma empresa vem da soma do resultado da implementação das dimensões competitivas de sua estratégia nas diversas áreas de uma empresa, como produção, marketing, entrega e suporte aos seus produtos. Nesse modelo competitivo, a tecnologia possui um papel significativo:

a tecnologia está contida em toda atividade de valor e está envolvida na obtenção de elos entre atividades, ela pode ter um efeito poderoso sobre o custo e sobre a diferenciação (PORTER, 1990, p. 157).

Porter (1990) coloca que as atividades desempenhadas pela empresa criam uma cadeia de valor, conforme demonstrado na Figura 16:

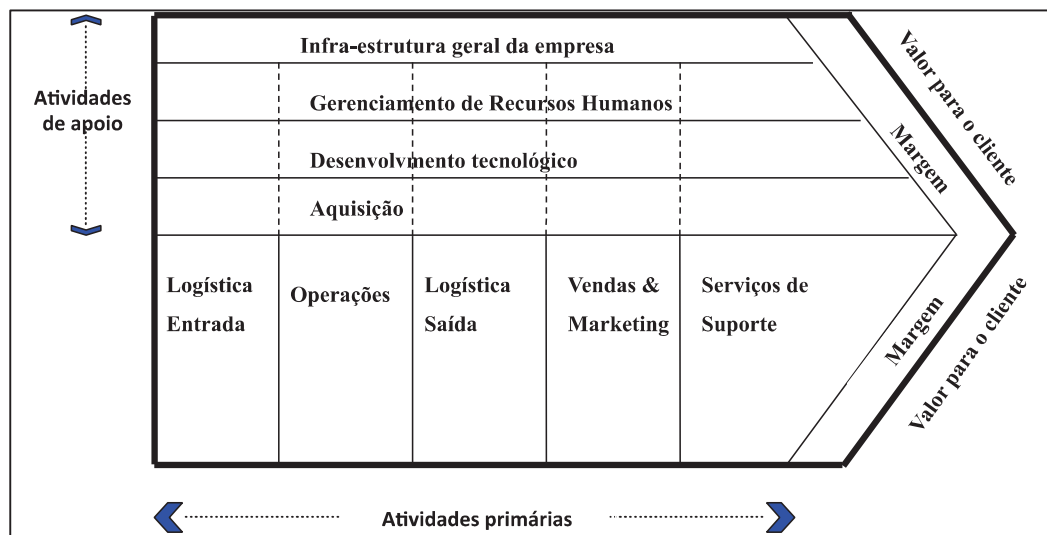


Figura 16: Cadeia de valor de Porter

Fonte: Porter (1990, p. 35)

Vasconcelos e Cyrino (2000) destacam 4 fontes diferentes de vantagem competitiva:

- acesso privilegiado ou exclusivo a recursos únicos e valiosos, como, por exemplo, detenção de patentes ou artigos de alto luxo;
- capacidade de produção ou detenção de processos produtivos superiores, capazes de produzir em menor tempo ou com menor custo do que os concorrentes;

- c) melhor utilização de recursos e capacidades, como otimização de estoque, criação de novos produtos, adaptando recursos e produtos existentes;
- d) criação de fluxo contínuo de inovação por meio do desenvolvimento de competências que permitam o melhoramento contínuo da empresa.

Vasconcelos e Cyrino (2000) reforçam que a estratégia deve ser sempre revista e atualizada, visto que as forças do mercado estão em constante evolução e adaptação. A vantagem competitiva de uma empresa dura, apenas, até o próximo movimento de seus concorrentes. É condição fundamental de perenidade de uma empresa a análise constante do mercado onde essa está inserida para a criação e atualização contínua da estratégia competitiva dessa organização.

Já as dimensões competitivas ou prioridades competitivas podem ser entendidas, segundo Porter (1990) e Diehl, Gonçalves e Martins (2001), como a forma na qual as empresas implementam sua estratégia de negócio.

Neto, Fensterseifer e Formoso (2003, p. 1) acrescentam que as dimensões competitivas:

são objetivos que a empresa deve buscar mais fortemente com o intuito de aumentar a sua competitividade e, por conseguinte, a sua participação no mercado e a sua lucratividade

Complementa definindo que essas são um conjunto de critérios individualizados para cada empresa, dependendo do mercado onde está inserida, seus competidores, seus produtos e suas competências internas, como recursos humanos, físicos e tecnológicos.

As dimensões competitivas são relativas ao segmento de mercado em que a empresa está inserida, como posto por Porter (1990 e 1991) e demonstrado pelos estudos de Santos e Pires (1999) e Diehl, Gonçalves e Martins (2001). Existe, entretanto, um conjunto de dimensões reconhecidas como as principais e mais abrangentes, mencionadas em grande quantidade de estudos sobre o tema e listadas no Quadro 1.

Quadro 1: Principais dimensões competitivas

Dimensão Competitiva	Principais componentes
Custo	<p>Busca por menores custos de produção.</p> <p>Redução e otimização dos custos de operação da empresa, incluindo as diversas áreas de apoio (financeiro, recursos humanos, entre outros).</p> <p>Oferecer produtos e serviços com menor custo que os concorrentes.</p> <p>Aumentar a lucratividade da empresa por meio do aumento da margem de contribuição de cada produto.</p> <p>No passado, foi a principal prioridade para as empresas, tendo perdido o posto nos dias modernos para a flexibilidade. Entretanto, o custo permanece como uma parte fundamental da estratégia competitiva das organizações.</p>
Qualidade	<p>Oferta de produtos com qualidade reconhecida pelos clientes.</p> <p>Prestar serviços adequados.</p> <p>Diferenciar produtos de similares do mercado.</p> <p>Construir uma boa imagem do produto, serviço e da própria empresa.</p> <p>A qualidade deixou de ser apenas uma norma interna a ser seguida e passou a buscar atender e superar as expectativas dos clientes, diferenciando o produto ou serviço da concorrência.</p>
Desempenho	<p>Eficiência operacional na produção ou entrega dos serviços e produtos, permitindo atender a demanda com maior agilidade e menor custo.</p> <p>Permitir a entrega do produto ou serviço no tempo desejado pelo cliente, ou mais rapidamente do que a concorrência.</p> <p>O desempenho pode ser medido de diferentes formas, dependendo do setor empresarial analisado, mas sempre tendo como principal objetivo conquistar a confiança do cliente.</p>
Flexibilidade	<p>Adaptar-se rapidamente às mudanças de mercado ou à demanda dos consumidores.</p> <p>Criação de novos produtos e serviços.</p> <p>Ter um <i>mix</i> adequado de produtos.</p> <p>Mudar rapidamente volume de produção e estrutura de distribuição.</p> <p>Adaptação de seus processos e estrutura organizacional.</p> <p>Adaptar seus produtos de forma rápida para atender a um grupo seletivo de clientes.</p> <p>Tem sido a dimensão mais estudada nos últimos anos, tornando-se um diferencial competitivo de grande relevância.</p>

Fonte: Alves Filho, Pires e Vanelle (1995), Santos e Pires (1999), Diehl, Gonçalo e Martins (2001), Neto, Fensterseifer e Formoso (2003)

Coloca-se ainda a inovação como uma dimensão competitiva em separado, estando, na maioria das vezes, ligada ao uso de novas tecnologias. Entretanto, a inovação não precisa existir como uma nova dimensão, ficando disseminada entre as demais prioridades da estratégia competitiva.

Alves Filho, Pires e Vanelle (1995) destacam que é necessário considerar a incompatibilidade entre algumas dimensões competitivas, como, por exemplo, a incapacidade de reduzir custos, ao mesmo tempo que se busca melhorias de qualidade ou desempenho, e relata que na estratégia empresarial cada dimensão pode ter um peso diferente, dependendo do momento da empresa e do mercado, podendo a empresa alterar suas prioridades de acordo com a necessidade ou o momento do ciclo de vida do produto.

Entretanto, Alves Filho, Pires e Vanelle (1995) expõem que a teoria da incompatibilidade das dimensões é questionada por diversos estudos de casos que relatam empresas cuja busca por melhoria na qualidade acabou levando a uma redução de custos, ou cuja melhoria do desempenho permitiu às empresas serem mais flexíveis. Existem, então, ao menos, faixas em que as dimensões podem coexistir, podendo essas faixas serem ampliadas pela inovação e uso de novas tecnologias.

Teixeira et al (2005) destacam que a competitividade de uma empresa vem de sua eficiência em montar uma estratégia de negócio considerando suas dimensões competitivas, sendo essas provenientes do mercado em que a empresa está inserida.

Schmarzo (2013) lista alguns dos aspectos dos processos de negócio que possuem impacto direto nas dimensões competitivas de uma empresa:

- a) identificar os fornecedores com a melhor relação custo-benefício que possam entregar os produtos no prazo e sem avarias;
- b) agilizar o desenvolvimento de novos produtos e tornar as respectivas campanhas de lançamento mais eficientes;
- c) identificar de forma automatizada possíveis problemas na linha de produção que possam levar à diminuição na qualidade do produto;
- d) quantificar níveis ótimos de estoque e otimizar a cadeia de suprimentos, considerando, inclusive, fatores externos, como clima, feriados e fatores econômicos;

- e) identificar pela análise do comportamento do consumidor e do canal de vendas a efetividade das campanhas de marketing em atingir seus objetivos, capturar a atenção dos clientes, conectar esses com a marca ou produto e gerar vendas;
- f) determinar o preço ideal para um produto, especialmente produtos perecíveis, desde laticínios até artigos de moda e assentos de voo;
- g) otimizar as ações promocionais, considerando os números de venda, níveis de estoque e interesse dos consumidores pelo referido produto;
- h) otimizar os recursos de vendas, incentivo ao *mix* de produtos, modelo de comissionamento e distribuição de contas;
- i) contratar e reter as melhores pessoas dada análise de características e comportamento dos profissionais mais bem-sucedidos da organização.

Por fim, Schmarzo (2013) exemplifica como uma empresa de varejo pode usar os 3 “Vs” do Big Data (Velocidade, Variedade e Volume) em conjunto com a cadeia de valor de Porter (1990), para, na prática, criar valor para si, aprimorando suas dimensões competitivas e obtendo vantagem competitiva em seu mercado. Destaca-se, ao final de cada atividade, as principais dimensões competitivas afetadas:

- a) **logística de entrada:** utilizar dados obtidos dos pontos de vendas, analisados em tempo real (velocidade) para identificar e notificar fornecedores sobre os produtos mais ou menos vendidos, evitando a falta de determinados produtos em estoque, enquanto outros ficam encalhados. Permite melhoras nas dimensões de desempenho e custo;
- b) **operações:** utilizar dados obtidos dos pontos de vendas em conjunto com dados de RFID que, analisados em tempo real (velocidade e variedade), podem ajudar a gerar campanhas relâmpagos de vendas para produtos com baixa saída, ajudando a otimizar o estoque da loja. Permite melhoras nas dimensões de desempenho e qualidade;
- c) **logística de saída:** usar dados de mídias sociais (variedade e volume) para prever efeito de campanhas de marketing, preparando-se melhor para a demanda. Permite melhoras nas dimensões de desempenho e flexibilidade;
- d) **vendas & marketing:** usar dados de pontos de vendas em conjunto com informações dos clientes oriundas de um programa de fidelidade (variedade, volume) para criar microssegmentos de clientes e realizar campanhas de marketing mais assertivas, vinculadas a eventos e datas significantes do mercado, usando dados de mídias

sociais para rapidamente identificar a efetividade da campanha, fazendo ajustes se necessário (velocidade). Permite melhoras na dimensão de custo;

- e) **serviços de suporte:** capturar o sentimento dos clientes, por meio de análise das redes sociais (variedade e volume), objetivando identificar possíveis problemas e insatisfação dos clientes. Permite melhoras nas dimensões de qualidade e custo;
- f) **infraestrutura:** utilizar análise preditiva, agrupando fontes de dados, como números de vendas e de mercado (variedade, velocidade) para melhorar o gerenciamento das áreas. Permite melhoras na dimensão de custo;
- g) **recursos humanos:** utilizar dados obtidos de números de vendas, números da indústria e microsegmentação dos clientes (variedade, volume) para prever demanda e dimensionar corretamente o efetivo necessário de funcionários. Permite melhoras na dimensão de custo;
- h) **desenvolvimento tecnológico:** uso de novas tecnologias como *in-memory database*²² ou *text mining* para permitir novas formas de análise de dados por meio da experimentação de modelos, gerando novas possibilidades de insights para o negócio. Permite melhoras nas dimensões de custo, qualidade, desempenho e flexibilidade;
- i) **aquisição:** uso de informações dos pontos de vendas, aliadas à previsão de demanda obtidas de redes sociais (variedade, volume) para negociar de forma mais assertiva com os fornecedores. Permite melhoras nas dimensões de custo e flexibilidade.

2.8.1 Estratégia analítica

Davenport e Harris (2007, p. 27) caracterizam a competitividade empresarial à luz da análise de dados, conceituando essas empresas como “organização que utiliza extensiva e sistematicamente a inteligência analítica para pensar melhor e ter um desempenho melhor do que a concorrência”.

²² Sistema de banco de dados que carrega todos os dados na memória dos computadores, permitindo a execução muito mais rápida de pesquisas e atualizações dos dados.

Davenport e Harris (2007) ressaltam que todo projeto de análise de dados deve ser vinculado a métricas financeiras, como redução de custos, aumento do volume de vendas ou do lucro da empresa ou departamento.

Importante ressaltar que Davenport e Harris (2007) já previam que o uso de análise dos dados pelas empresas seria uma disciplina em constante evolução, sendo essa condição fundamental para que as organizações continuassem obtendo vantagem competitiva.

À medida que as empresas evoluem no modelo de análise de dados, tornam-se empresas analíticas, como já descritas por Davenport e Harris (2007). Um dos principais fundamentos dessas empresas é a tomada de decisão embasada em números e informações, ao invés de contar somente com a opinião ou o “palpite” do executivo, ou, como descrito por Brown, Chui e Manyika (2011) e Brynjolfsson e McAfee (2012), adotar a decisão baseada em dados no lugar da decisão baseada na “opinião da pessoa com maior salário”²³. Claro que sempre haverá ocasiões em que a decisão, mesmo baseada em dados, poderá mostrar-se equivocada, entretanto Brynjolfsson e McAfee (2012) destacam que essa será sempre a aposta mais segura.

É importante que a alta gestão da empresa seja patrocinadora do projeto de Big Data. As características fundamentais de um executivo analítico foram listadas por Davenport e Harris (2007):

- a) adepto do processo decisório baseado em fatos;
- b) agir sobre os resultados da análise;
- c) administrar a meritocracia.

Davenport e Harris (2007, p. 130) obtiveram a seguinte afirmação de uma alta executiva do Bank of America:

Comece com uma avaliação – qual é o nível de diferenciação de sua oferta em comparação com o que você e seus clientes querem que ela seja? Se ela precisar ser mais diferenciada, você deve encontrar maneiras de criar mais valor com o conhecimento que possui. Você precisa perceber isso antes; de outra forma, nem se dê ao trabalho de desenvolver *data warehouses* e sistemas de CRM. Você precisa ter a visão do valor se quiser descobrir como concretizar essa visão.

Isso posto, não existe uma fórmula mágica ou receita a ser seguida de forma idêntica por todas as empresas que queiram melhorar suas dimensões competitiva utilizando o Big Data e, infelizmente, nem todos os projetos de implementação de sistemas de análise de dados

²³ Do inglês: *the highest-paid person's opinion*.

serão bem-sucedidos, com descrito por Davenport e Harris (2007). Entretanto, mesmo projetos parcialmente executados podem promover melhorias incrementais em pontos específicos das empresas. Os autores citam a United e a American Airlines como exemplos de empresas que, apesar de fazerem uso extensivo da análise de dados, não conseguiram criar vantagem competitiva significativa e sofrem sérios problemas financeiros.

Assim como ocorreu com outras tecnologias, os ganhos obtidos provêm não somente do uso da tecnologia, mas também da adoção de novas práticas de gestão e mudanças organizacionais para maximizar os benefícios obtidos (BROWN; CHUI; MANYIKA, 2011).

Bughin e Livingston (2011) ressaltam que poucos líderes compreendem o verdadeiro potencial do Big Data para os seus negócios ou veem os seus dados como ativos valiosos. As organizações que definirem uma estratégia para utilização correta do Big Data podem se posicionar na liderança da competição analítica.

Brown, Chui e Manyika (2011) admitem que o Big Data ainda está no seu começo, evoluindo junto com as tecnologias que o compõem, mas acredita que com o passar do tempo Big Data tornar-se-á um novo ativo das corporações, representando um fator fundamental na competitividade, tanto quanto a própria marca da empresa.

Brynjolfsson e McAfee (2012) e Cearley e Claunch (2013) destacam o impacto na gestão corporativa por meio da análise de padrões e geração de insights, permitindo a tomada de decisões de forma mais rápida e eficiente, baseadas na análise de um conjunto expandido de dados, incluindo dados da empresa, do mercado e até mesmo dos consumidores.

Pelo descrito, as empresas devem colocar o Big Data dentro do seu planejamento estratégico e deliberar sobre seu uso para os próximos anos, como, por exemplo, começar um plano de testes e implementação piloto. Cearley e Claunch (2013) recomendam que as empresas criem um modelo para pesquisar e acompanhar as tendências tecnológicas como o Big Data. Dessa forma, seria possível determinar o melhor modelo, e o momento ideal para a adoção da tecnologia. Isso é particularmente importante em indústrias sujeitas às rápidas mudanças. Os autores complementam afirmando que tecnologias como Big Data têm o potencial de melhorar a eficiência operacional, reduzindo custos e aumentando a rentabilidade, inclusive com a criação de novos modelos de negócio, produtos e serviços, o que resulta em um aumento na lucratividade da empresa.

3 MÉTODO

É apresentado aqui o método e procedimentos utilizados para atingir o objetivo geral, especificado no item 1.2 Objetivo da pesquisa.

3.1 Escolha do método de pesquisa

O método de pesquisa foi definido considerando o delineamento, objetivo e abordagem descritos a seguir.

A pesquisa é descritiva, uma vez que, como posto por Andrade (2010, p. 112), “neste tipo de pesquisa, os fatos são observados, registrados, analisados, classificados e interpretados, sem que o pesquisador interfira neles”. A autora também destaca que a pesquisa descritiva pode eventualmente se aproximar da pesquisa exploratória, que, segundo Malhotra (2004, p. 109):

é adequada quando é preciso definir o problema com maior precisão, identificar os cursos alternativos de ação, desenvolver as perguntas ou hipóteses da pesquisa e isolar e classificar as variáveis-chave como dependentes ou independentes.

Quanto à abordagem, trata-se de pesquisa qualitativa. Godoy (1995) aponta que a pesquisa qualitativa não procura enumerar ou medir os eventos estudados. Partindo do contato direto do pesquisador com o objeto de pesquisa, permite desenvolver questões de interesse amplo, que vão ganhando foco e definição à medida que a pesquisa se desenvolve.

Flick (2009) aponta a pesquisa qualitativa como um termo usado no passado para descrever de forma crítica uma alternativa à pesquisa quantitativa. Entretanto, com o passar do tempo, a pesquisa qualitativa foi ganhando características próprias, utilizando o texto ao invés de números para a construção da realidade em estudo, com observância da perspectiva dos participantes, considerando suas práticas cotidianas e conhecimentos adquiridos. Narcizo e Silva (2010) complementam dizendo que a abordagem qualitativa permite capturar com maior exatidão a complexidade de cenários, situações e usos que podem ser atribuídos ao tema de pesquisa.

A pesquisa qualitativa e quantitativa possui semelhante relevância na Engenharia da Produção, como demonstrado na Figura 17.

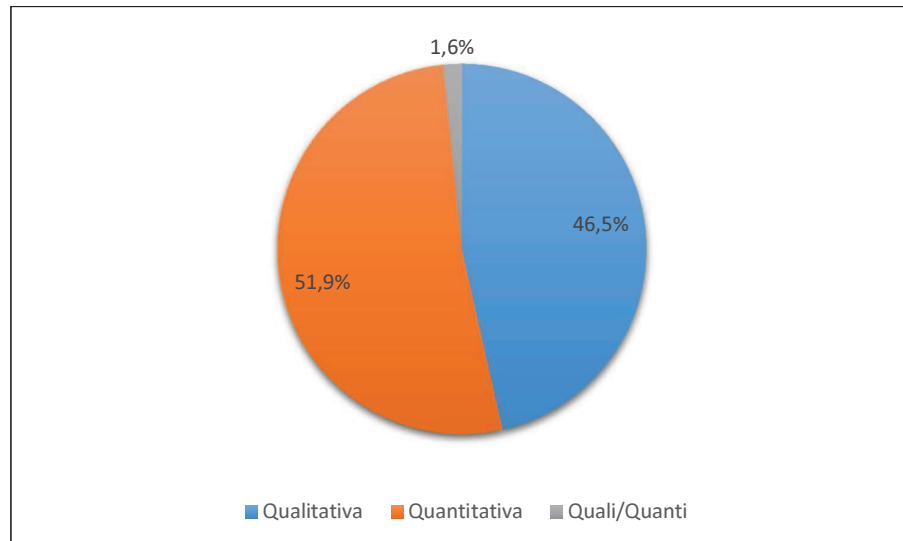


Figura 17: Abordagem de pesquisa em artigos de periódicos QUALIS A entre 2003 e 2006

Fonte: Adaptado de Lacerda et al (2007).

Segundo Teixeira e Pacheco (2005), o método qualitativo possui 5 características básicas:

- a) supõe o contato direto do pesquisador com a situação que está sendo investigada, tendo o ambiente natural como fonte direta de dados;
- b) o material obtido é rico em descrições de pessoas, situações e acontecimentos;
- c) a preocupação com o processo é muito maior do que com o produto;
- d) tenta-se capturar a “perspectiva dos participantes”, ou seja, como esses encaram as questões em estudo;
- e) a análise dos dados tende a seguir um processo indutivo.

O delineamento da pesquisa é bibliográfico (revisão de literatura) e estudos de casos múltiplos, visto que esses irão se complementar, e ainda, de acordo com Vergara (2011), os tipos de pesquisa se complementam.

A pesquisa descreve o uso do Big Data pelo mercado, lançando mão de fontes de dados secundárias com o objetivo de aprofundar o conhecimento sobre o tema. Buscou-se utilizar,

principalmente, estudos e pesquisas de institutos renomados, como Gartner e McKinsey, além de artigos científicos contendo fundamentação teórica sobre o tema. Complementa-se a pesquisa com estudos de casos de empresas ou segmentos empresariais que estão utilizando o Big Data. Realizando a agregação dessas fontes de dados, foi possível discorrer sobre o tema com o embasamento de casos práticos.

A pesquisa bibliográfica, ou revisão de literatura, é para Roesch (2007) uma das partes mais longas e trabalhosas, pois requer seleção, leitura e análise para determinar os textos relevantes ao tema abordado.

Para Miguel (2007), o estudo de caso possui uma natureza empírica, investigando um determinado fenômeno, geralmente contemporâneo, dentro de um contexto real de vida, com o objetivo de gerar uma análise que permita o aprofundamento do conhecimento acerca do problema de pesquisa. Yin (2010) concorda com esta afirmação.

O uso do estudo de caso vem ganhando força dentro da pesquisa científica, como identificado por Berto e Nakano (2000) e corroborado por Rocha (2011) em levantamento realizado nos artigos publicados na ENEGEP entre os anos de 1996 e 2009 – utilizando a porcentagem dos artigos contendo estudos de casos, pode-se traçar uma linha que demonstra o crescimento exponencial da relevância desse delineamento de pesquisa, conforme exposto na Figura 18.

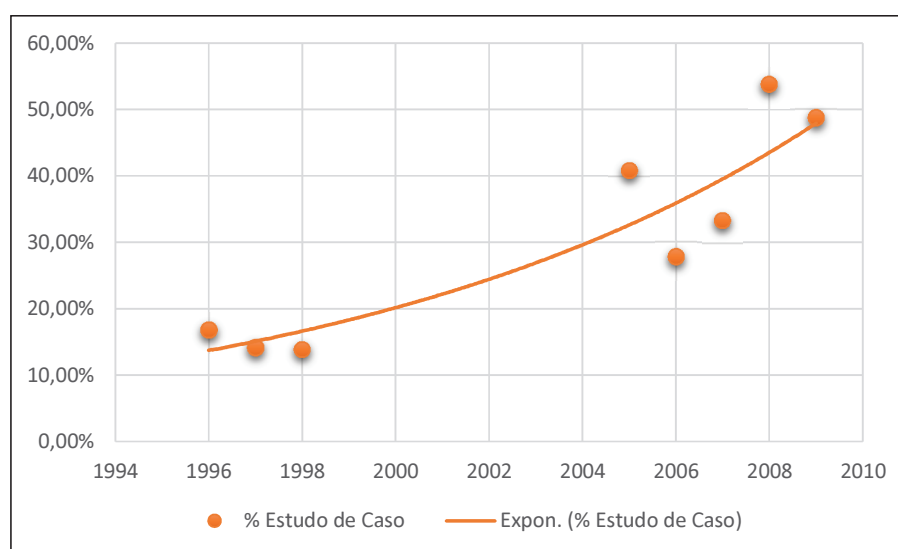


Figura 18: Estudo de caso em artigos do ENEGEP

Fonte: Adaptado de Berto e Nakano (2000) e Rocha (2011).

O modelo de estudo de caso aparece também como principal método científico em pesquisas indutivas, em que as generalizações derivam de observações de casos da realidade, estando presente em 64,8% dos artigos publicados em periódicos com classificação A no QUALIS, conforme estudo de Lacerda et al (2007).

3.2 Protocolo do estudo de caso

Para o estudo de caso, foram selecionadas empresas que utilizam ou estejam em níveis avançados de estudo para a implementação de soluções de Big Data. Adiciona-se ao estudo uma empresa conceituada no mercado que forneça soluções de Big Data.

As empresas pesquisadas estão caracterizadas no Quadro 2.

Quadro 2: Caracterização das empresas

Empresa	Caracterização da empresa
A	<p>Uma das 3 maiores empresas de soluções de armazenamento de dados, nos últimos anos trabalha para expandir seu portfólio de produtos para tornar-se uma fornecedora de soluções de Big Data. Possui presença global, com filiais em mais de 100 países e mais de 60.000 mil colaboradores.</p> <p>Possui como estratégia competitiva o alto investimento em pesquisa e desenvolvido, tendo destinado para este fim mais de 12 bilhões de dólares na última década. Em 2012, a empresa inaugurou no Brasil um centro de pesquisa e desenvolvimento de soluções com foco em Big Data.</p>
B	<p>Empresa do setor de telecomunicações. Fornece soluções de telefonia móvel e fixa, além de conexão de banda larga, tanto no mercado corporativo quanto no varejo. Com cobertura nacional, está entre as 5 maiores operadoras de telefonia móvel do Brasil.</p> <p>Empresa em constante busca por inovações tecnológicas, gosta de ser considerada pioneira na adoção de novas tecnologias e fornecimento de produtos inovadores aos seus clientes.</p>

C	<p>Empresa do setor de telecomunicações. Fornece soluções de telefonia móvel e fixa, além de serviços de conexão de banda larga e televisão por cabo. Com cobertura nacional, está entre as 5 maiores operadoras de telefonia móvel do Brasil.</p> <p>Expandiu recentemente seu portfólio de produtos, ingressando em novos mercados, como o de televisão por assinatura e internet de banda larga. Possui o desafio de ganhar clientes nesse novo mercado enquanto defende sua posição entre os líderes do setor de telefonia tradicional.</p>
D	<p>Empresa situada entre os líderes nas áreas de televisão por assinatura e Internet de banda larga no Brasil. Com presença em mais de 100 localidades, possui aproximadamente 5 milhões de clientes de TV por assinatura e internet de banda larga.</p> <p>Busca consolidar sua posição também no mercado de telecomunicações, com foco principalmente entre as classes média e alta.</p>
E	<p>Empresa de liderança do setor financeiro, oferece serviços bancários para milhões de brasileiros e está em intensiva competição com os demais líderes do segmento.</p> <p>Todas as empresas líderes possuem igualdade de recursos para investimento, forçando a empresa a buscar soluções que permitam obter melhoria nas dimensões competitivas por meio da inovação tecnológica.</p>

Fonte: Elaboração do autor.

Como ferramental para a pesquisa, foi utilizado questionário de características semiestruturado, que possibilitou aos participantes responderem de forma mais abrangente, permitindo capturar melhor a experiência de vida real destes e a complexidade dos cenários em que estão inseridos. O Quadro 3 apresenta as perguntas do referido questionário, que foi customizado para cada perfil de empresa: fornecedora ou consumidora de soluções de Big Data.

Quadro 3: Questões do estudo de caso

Perfil da empresa	Utiliza ou está em fase de implantação de soluções de Big Data	Oferece soluções de Big Data para outras empresas.
Empresas	B, C, D, E	A
Questão principal de pesquisa	Qual a contribuição da implantação de soluções de Big Data para melhoria das dimensões competitivas?	Qual a contribuição da implantação de soluções de Big Data para a melhoria das dimensões competitivas das empresas?
Questões do estudo de caso	<p>Qual a diferença percebida entre soluções de Big Data e soluções tradicionais de análise de dados?</p> <p>Qual a principal dificuldade encontrada na implementação de soluções de Big Data?</p> <p>Qual a solução de Big Data foi escolhida e qual o motivo que levou a esta escolha?</p> <p>Como foi determinado o uso da solução (qual problema de negócio objetiva-se resolver)?</p> <p>Por que as soluções tradicionais de análise de dados não conseguiam resolver esse problema?</p> <p>Como o Big Data mudou/irá mudar o relacionamento da empresa com os clientes?</p> <p>A empresa adotou protocolos específicos ou criou processos para controlar o acesso a dados privados dos usuários, como, por exemplo, redes sociais?</p> <p>A implantação do Big Data permitiu à empresa ficar mais competitiva? Se sim, como?</p>	<p>Qual a diferença percebida entre soluções de Big Data e soluções tradicionais de análise de dados?</p> <p>Existe por parte dos provedores de soluções de análise de dados uma tendência de classificar todos os produtos como soluções de Big Data? Ainda existe uma confusão quanto às diferenças entre análise de dados tradicionais e Big Data?</p> <p>Quais são exemplos de problema de negócio que soluções tradicionais de análise de dados não conseguiam resolver e por quê?</p> <p>Qual a principal dificuldade encontrada na implementação de soluções de Big Data?</p> <p>Como o Big Data mudou/irá mudar o relacionamento da empresa com os clientes?</p> <p>A implantação do Big Data permitiu à empresa ficar mais competitiva?</p>

Fonte: Elaboração do autor.

3.2.1 Coleta de dados

Yin (2010) coloca como um pré-requisito para o pesquisador que deseja realizar um estudo de caso ter uma mente questionadora durante a coleta de dados e possuir a capacidade de formular boas questões. O roteiro de pesquisa definido foi aplicado em contato direto com o entrevistado, preferencialmente em encontro pessoal e, quando este foi inviável, utilizou-se sistema de vídeo e/ou audioconferência.

Os entrevistados foram selecionados com base nos seguintes critérios:

- a) trabalhar em empresa de mercado que utilize ou esteja em níveis avançados de estudo da utilização de soluções de Big Data, tendo atuado diretamente com o projeto;
- b) trabalhar em empresa fornecedora de soluções de Big Data, atuando diretamente com tecnologia referente à solução, provendo consultoria para grandes empresas sobre o tema ou atuando como pesquisador de novas soluções, utilizando a tecnologia de Big Data.

Lacerda et al (2007) apontam que, em estudo realizado analisando 310 artigos publicados em periódicos nacionais e internacionais entre 2003 e 2006, as principais fontes de informação utilizadas para estudos de casos são: a documentação obtida, utilizada em 54% dos casos, e as entrevistas realizadas, fonte para 22% dos estudos de caso, como mostrado na Figura 19. Como mais de uma fonte pode ser utilizada, a soma das colunas da figura não totaliza 100%.

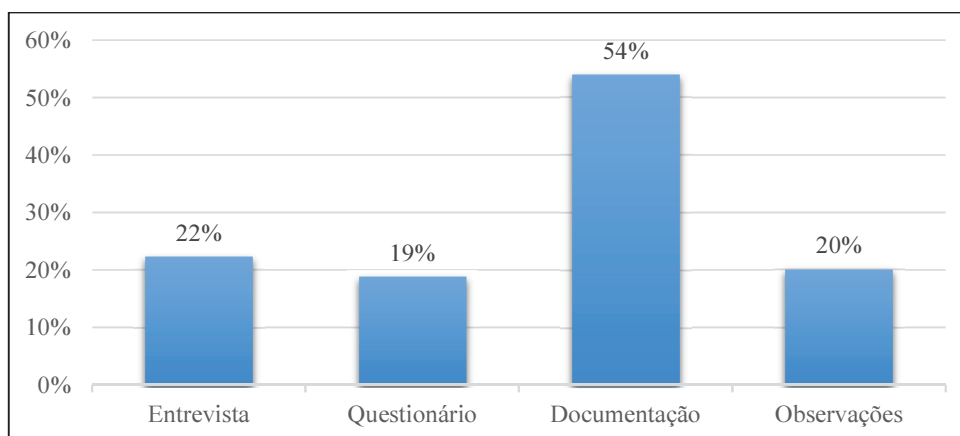


Figura 19: Fontes de dados em pesquisas de Estudo de Caso

Fonte: Adaptado de Lacerda et al (2007).

Lançou-se mão de visitas nas instalações da empresa, entrevistas por telefone e coleta de documentos, conforme descrito no Quadro 4.

Quadro 4: Dinâmica do processo de coleta de dados

Empresa	Entrevistas	Documentos
A	Diretor de soluções para clientes Engenheiro / Arquiteto de sistemas de Big Data Gerente do centro de pesquisa de soluções de Big Data Consultor de Soluções para Big Data	Materiais técnicos sobre os produtos Apresentações e arquiteturas de soluções desenhadas para clientes Web site da empresa e documentos públicos obtidos na internet.
B	Gerente de Projeto – TI Business Intelligence & Analytics Diretor TI Arquiteto – TI Business Intelligence & Analytics	Apresentações e arquitetura do projeto de Big Data Web site da empresa e documentos públicos obtidos na internet.
C	Gerente de Projeto responsável pela iniciativa de Big Data Analista Sênior de BI (1) Analista Sênior de BI (2)	Apresentações e arquitetura do projeto de Big Data Web site da empresa e documentos públicos obtidos na internet.
D	Gerente de Projeto responsável pela iniciativa de Big Data Gerente de Segmentação Analista de BI	Apresentações e arquitetura do projeto de Big Data Web site da empresa e documentos públicos obtidos na internet.
E	Gerente de Homologação de Novas Tecnologias Gerente de Desenvolvimento de Sistemas Diretor de Soluções	Apresentações e arquitetura do projeto de Big Data Web site da empresa e documentos públicos obtidos na internet.

Fonte: Elaboração do autor

3.2.2 Análise dos dados coletados

O estudo de caso requer um processo variado de análise dos dados coletados, justamente pela natureza variada da coleta de dados. Todavia, esta análise será predominantemente qualitativa, na qual o pesquisador, despido de quaisquer conclusões prévias, analisa sistematicamente os dados coletados (GIL, 2002).

Hoppen, Lapointe e Moreau (1996) já destacavam que o mais importante é que o pesquisador consiga entregar ao leitor um texto pertinente e verdadeiro sobre o tema que foi estudado, utilizando a informação gerada com base nos dados colhidos durante a pesquisa.

4 RESULTADOS E DISCUSSÕES

Aqui são analisados os resultados da pesquisa conduzida, objetivando responder à questão de pesquisa.

4.1 Estudo das empresas

Ao trazer os resultados dos múltiplos estudos de caso, é importante notar que a Empresa “A” possui perfil diferente das demais, por ter sido pesquisada como uma fornecedora de soluções de Big Data, enquanto as demais empresas pesquisadas são exclusivamente consumidores destas soluções. Não existe uma relação direta entre as empresas, ou seja, a Empresa “A” não é necessariamente a fornecedora das soluções usadas pelas demais empresas.

4.1.1 Empresa “A”

Como provedora de soluções de tecnologia, incluindo Big Data, a Empresa “A” foi estudada com o objetivo de identificar como a referida empresa está vendo a adoção pelas demais empresas usuárias/clientes.

A Empresa “A” possui a maior parte de suas receitas proveniente da venda de produtos e serviços para a infraestrutura de sistemas de tecnologia da informação. Entretanto, apesar do crescimento na demanda por esses produtos, tem havido uma diminuição significativa na margem de lucro das vendas, com a saturação do mercado devido à entrada de novos competidores.

No passado, os sistemas de computador eram baseados em sistemas mainframes e terminais “burros” com acesso restrito para apenas alguns milhões de usuários pelo mundo e contando com apenas alguns milhares de aplicativos diferentes. Surgiu, então, a tecnologia da segunda plataforma, baseada em computadores pessoais e processadores x86, com dezenas de milhares de aplicações e centenas de milhões de usuários. Neste momento, está ocorrendo uma

transformação semelhante, com a área de TI das empresas começando a adotar a chamada terceira plataforma, desenvolvida para suportar dispositivos móveis, internet de alta velocidade, compartilhamento de informações por redes sociais e com milhões de aplicações acessadas por bilhões de usuários ao redor do mundo. Dentro desse contexto, a Empresa “A” tem buscado novos segmentos do mercado de tecnologia, especificamente em área de tecnologias inovadoras. Essa mudança faz parte da estratégia competitiva da empresa para tornar-se líder de soluções para a chamada terceira plataforma, mostrada na Figura 20.

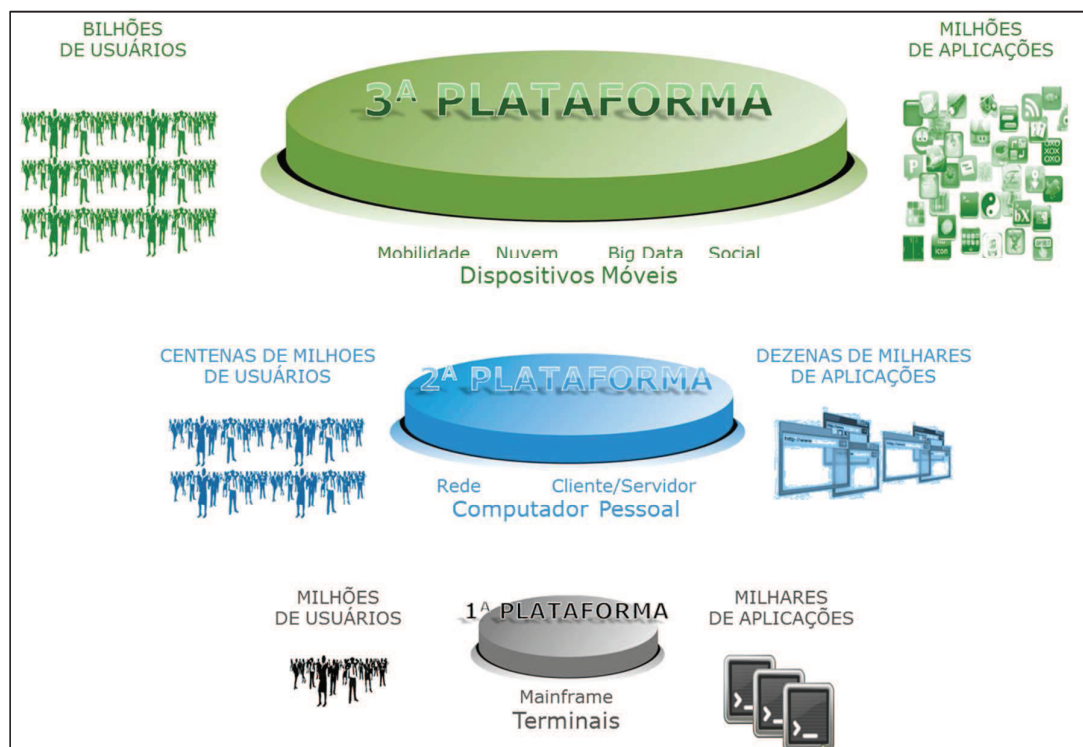


Figura 20: As três plataformas tecnológicas

Fonte: Empresa “A”.

Nessa nova terceira plataforma, soluções de Big Data terão uma participação fundamental, permitindo a manipulação da imensa massa de dados gerada por esses novos usuários, que acessam novas aplicações utilizando dispositivos móveis e redes sociais. Segundo estimativa dos executivos da Empresa “A”, até o final de 2016 a terceira plataforma irá criar um mercado de 44 bilhões de dólares para a venda de produtos (*hardware* e *software*) e serviços diretamente associados ao Big Data.

Dentro desse novo paradigma do mercado de tecnologia da informação e com o objetivo de possuir soluções inovadoras para ganhar espaço nesse novo mercado, a Empresa “A” investiu recentemente na construção de um centro de pesquisa e desenvolvimento exclusivo para a criação de soluções inovadoras com o uso do Big Data a fim de tratar problemas relevantes da indústria.

Um dos principais diferenciais de uma arquitetura tecnológica de Big Data é a capacidade de armazenar e processar grandes volumes de informações a um custo reduzido. Os três diferentes tipos de arquitetura para armazenamento e processamento de dados estão demonstrados na Figura 21 e detalhados na sequência:

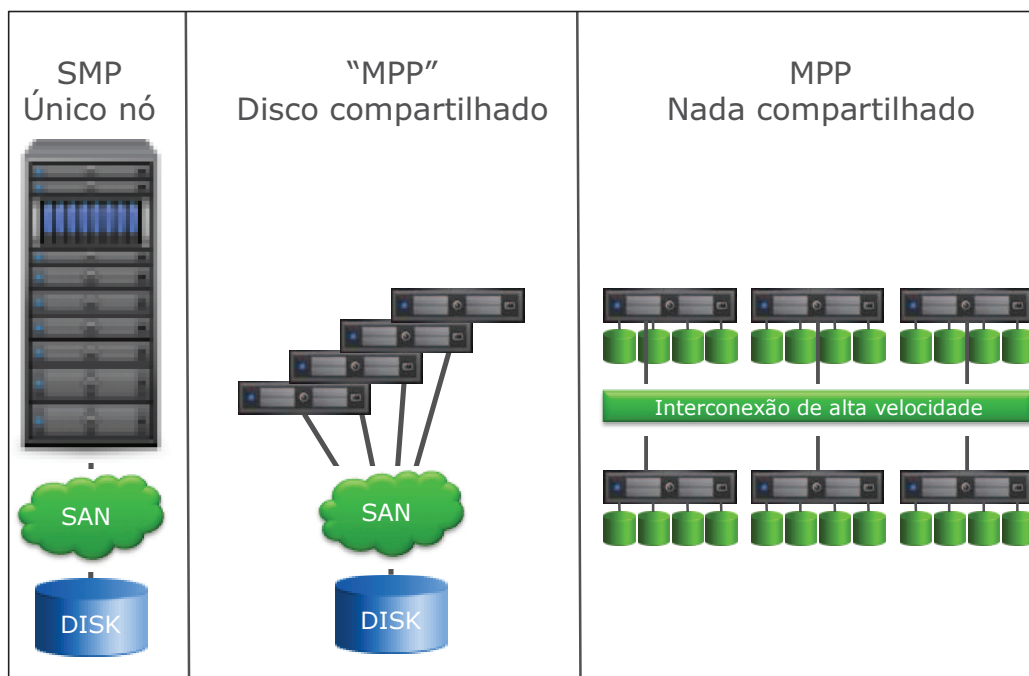


Figura 21: Arquitetura tradicional de processamento de dados *versus* Big Data

Fonte: Empresa “A”.

- a) SMP (*Symmetric MultiProcessing* ou Multi-Processamento Simétrico), arquitetura em geral empregada em sistemas tradicionais de análise de dados, quando um único equipamento com grande quantidade de processamento e memória analisa dados armazenados em um grande sistema de armazenamento. Em geral, esses sistemas possuem alto custo de processamento e de armazenamento, além de escalabilidade limitada;

- b) MPP (Massively Parallel Processing ou Processamento Paralelo Massivo), arquitetura de transição, encontrada em sistemas tradicionais de análise de dados e de Big Data. A carga de processamento é dividida entre vários computadores simples de menor capacidade, utilizando um repositório centralizado de armazenamento. Nesse caso, o custo de processamento é baixo, mas o custo de armazenamento em grande escala ainda é elevado;
- c) MPP com nada compartilhado: arquitetura de análise de dados utilizada por sistemas de Big Data, obtém grande capacidade de processamento por meio do uso de diversos servidores simples com armazenamento distribuído entre os servidores. Essa arquitetura é a que oferece o menor custo de processamento e o menor custo de armazenamento, além de possuir capacidade de crescimento superior as demais arquiteturas.

Para a Empresa “A”, além das diferenças técnicas mencionadas, o Big Data pode trazer mais benefícios para as empresas do que as soluções tradicionais de análise de dados. Esse ganho superior advém da possibilidade de obter informações como o que está acontecendo neste momento e prever a probabilidade de eventos futuros com elevado grau de assertividade, ao passo que soluções tradicionais estão limitadas a fornecer apenas detalhes sobre eventos passados. Exemplificando, soluções tradicionais de análise de dados podem responder a perguntas como “por que a empresa vendeu mal no último trimestre e o que pode ser feito para melhorar”, enquanto soluções preditivas como Big Data podem responder à pergunta de “qual o cenário de vendas mais provável para o próximo trimestre e quais ações tomar para garantir o resultado ótimo”. A Figura 22 mostra que, com o aumento do nível de inteligência das soluções analíticas, tem-se o aumento da vantagem competitiva da empresa frente às demais empresas.



Figura 22: Inteligência Potencial das soluções de análises de dados

Fonte: Empresa "A".

Contudo, dentro de tantas novas possibilidades, é comum que empresas se percam entre seus objetivos ou criem metas demasiadamente ambiciosas, levando à frustração e ao fracasso do projeto. É importante que, antes de começar um projeto de Big Data, as empresas tenham bem definidos os objetivos do projeto, como retorno financeiro, duração, e iniciativa estratégica da corporação alinhada ao projeto. O projeto deve ter objetivos de curto e longo prazo, alinhados com a maturidade analítica da empresa, permitindo, assim, a obtenção dos primeiros resultados logo nos primeiros meses ou semanas de projeto.

4.1.2 Empresa “B”

A Empresa “B” atua no mercado de telecomunicações, um mercado altamente competitivo, sofrendo pressão pela competição interna com outras grandes empresas do setor e também por novos entrantes e produtos substitutos, como, por exemplo, o WhatsApp e o Skype, ambas soluções que usam rede de dados para substituir, respectivamente, mensagens de textos por SMS e chamadas de voz.

A Empresa “B” cresceu ao longo dos últimos anos por meio da aquisição de novos clientes, aproveitando-se de um mercado em expansão no Brasil. Entretanto, esse modelo está saturado, havendo um baixo surgimento de novos clientes, por conseguinte forçando as empresas do setor a competir entre si pelos clientes já existentes. Adicionalmente, legislações da ANATEL, como a portabilidade numérica, facilitaram a troca dos clientes entre as empresas de telefonia, aumentando ainda mais a competitividade interna do mercado.

O mercado sofre também a pressão de novos entrantes, especialmente empresas oriundas do mercado de televisão e internet por assinatura, via cabo ou satélite, que expandiram suas ofertas para incluir também soluções de telecomunicações por voz e texto.

A Empresa “B” já investe há vários anos em soluções de análise de dados, como Teradata, sistemas de armazenamento de altíssima capacidade, como EMC VMAX, *software* de ETL²⁴ e *softwares* analíticos e de geração de relatórios como MicroStrategy e SAS.

Atualmente, a Empresa “B” conta com mais de 1 *petabyte*²⁵ de dados armazenados em sistemas de *data warehouse* e passíveis de serem analisados. Entretanto, devido às restrições dos sistemas tradicionais de análise de dados em contemplar o elevado volume de dados, a empresa possui limitações nessa área, sendo impossível consultar todos os dados disponíveis em uma única análise.

Pelo grande volume, a análise da massa total dos dados torna-se muito lenta; com isso, a empresa acabou restringindo a análise à apenas uma fração dos dados disponíveis. Adicionalmente, qualquer alteração na estrutura destes dados, por exemplo a inclusão de um

²⁴ ETL é o processo de extração dos dados de bases de dados transacionais, ou bases de dados de suporte aos sistemas de operação da empresa, transformação desses dados por meio de conversão ou filtragem de parte dos dados e carga dos dados transformados no *data warehouse* utilizado pelos sistemas analíticos. Esse processo é, em geral, demorado e complexo de ser alterado.

²⁵ 1 *petabyte* = 10^3 *terabytes* ou 10^{15} bytes (considerando base decimal).

novo campo, ou alteração de campos existentes, torna-se muito cara e demorada para ser realizada. Esses fatores obrigaram a Empresa “B” a criar ilhas analíticas, que possuem um subconjunto da massa total de dados. Cada ilha analítica pertence a um departamento distinto (marketing, engenharia, qualidade, financeiro), contendo os dados que esse departamento julga pertinente. Cada ilha foi construída sobre uma tecnologia distinta e não possui comunicação com as demais ilhas. Nesse modelo, a análise torna-se mais rápida, entretanto gera problemas, como:

- a) duplicidade do dado entre as ilhas;
- b) aumento do custo de infraestrutura;
- c) limitação das informações extraídas pela limitação na quantidade de dados analisados;
- d) diminuição da assertividade no resultado da análise;
- e) visões díspares entre áreas sobre um mesmo objeto ou contador de negócio, por exemplo, números de disponibilidade dos serviços divergentes entre engenharia e a área de qualidade.

Outro desafio de empresas desse segmento é a análise do CDR²⁶, por ser um arquivo baseado em texto e gerado em quantidades elevadas (milhões de registros por dia). O ETL do CDR para inseri-lo em um sistema analítico tradicional demanda grande poder computacional, ultrapassando os limites da solução de análise tradicional e obrigando a empresa a limitar os campos e quantidade de CDRs analisados.

Com o objetivo de resolver os problemas mencionados, a Empresa “B” adquiriu uma solução de Big Data que utiliza um repositório único de dados baseado na tecnologia EMC Isilon chamado *data lake*, ou, em tradução livre, lago de dados, que recebe esse nome por sua capacidade de armazenar dados provenientes de diversas fontes e em diversos formatos, como, dados estruturados oriundos de bases de dados corporativas e dados semiestruturados como CDRs.

Dentro dessa arquitetura, dados como o CDRs, mesmo sendo semiestruturados, podem ser armazenados diretamente nesse *data lake*, sem a necessidade de passar por um ETL. A análise dos dados é realizada utilizando um sistema de processamento de dados baseado em

²⁶ Call Detail Record ou Registro de Detalhes da Chamada é o registro contendo os detalhes da conexão realizada. É gerada e armazenada pela empresa prestadora do serviço de telefonia.

Hadoop. Tanto o EMC Isilon quando o Hadoop permite o aumento da capacidade de armazenamento e processamento utilizando nós paralelos de baixo custo (MPP), criando uma arquitetura com custo relativamente baixo, porém com elevada capacidade de armazenamento e processamento.

Dentro desse contexto, o projeto de Big Data da Empresa “B” busca, nessa primeira fase, a transformação do modelo analítico dos CDRs, com o objetivo de melhorar as seguintes dimensões competitivas:

- a) reduzir custos da área de TI, eliminando a duplicidade de informações existente entre as ilhas analíticas, resultando em redução na quantidade de *hardware*, *software*, energia, refrigeração e custos operacionais dos equipamentos;
- b) aumentar a agilidade na busca por informações, com o objetivo de cumprir determinações legais no rastreamento de chamadas, evitando, assim, multas da ANATEL;
- c) redução de custos operacionais por meio do aumento da flexibilidade na utilização e no formato da massa de dados, permitindo uma análise mais assertiva na identificação de falhas da rede, podendo, assim, apontar com maior precisão os usuários afetados por falhas, e, se necessário, realizar o ressarcimento pontual a esses usuários, evitando, dessa forma, pesadas multas do órgão regulador. Atualmente, a empresa, ao identificar uma falha, ressarcia todos os usuários de determinada área, pela impossibilidade de apontar exatamente os usuários e os serviços que foram impactados;
- d) medir com maior precisão o uso das antenas por usuários de outras operadoras, permitindo a cobrança pelos uso destas. Durante o piloto do projeto, foi identificado um valor a ser recebido equivalente a sessenta milhões de reais, que não havia sido identificado pelas ferramentas tradicionais de análise dos CDRs;
- e) aumentar a velocidade e capacidade de processamento, chegando próximo ao processamento em tempo real, permitindo, nessa primeira etapa do projeto, fornecer ao serviço de atendimento ao cliente informações precisas sobre o funcionamento dos serviços e os usuários impactados, permitindo, assim, um atendimento mais ágil ao usuário final, aumentando a satisfação dele, o que irá resultar em uma maior taxa de retenção.

A Figura 23 demonstra o entendimento da Empresa “B” sobre o ganho competitivo trazido pelo Big Data. O mercado de telecomunicações está em um momento no qual a estagnação das oportunidades está aumentando a competitividade. Por isso, ser capaz de analisar as informações o mais rapidamente possível para permitir uma melhor tomada de decisão é fundamental. A linha vermelha mostra que a demora entre um determinado evento e a reação da empresa diminuiu seu potencial de ganho naquela oportunidade. Uma solução, como o Big Data, que permita analisar e reagir mais rapidamente a um evento, pode trazer uma maior vantagem para a empresa na competição no mercado, como demonstrado pela linha azul.

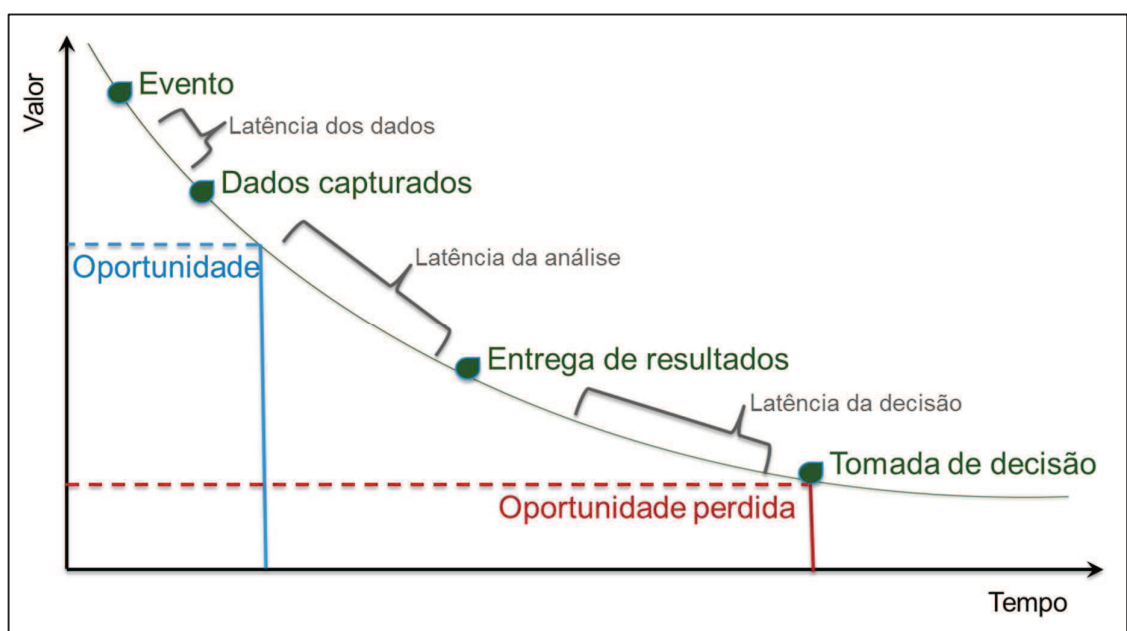


Figura 23: Curva de oportunidade

Fonte: Empresa “B”.

O principal desafio encontrado até o momento é a disponibilidade de mão de obra qualificada. Segundo a Empresa “B”, há uma grande escassez de profissionais que, além do entendimento da tecnologia, possuam também as demais dimensões que compõem o analista de Big Data, como o conhecimento em matemática, estatística e o conhecimento do negócio de empresas telecomunicações.

Um profissional com conhecimento em BI tradicional, ainda muito focado em análise do passado com dados estruturados, terá dificuldades em obter *insights* relevantes e extrair valor adicional da massa de dados, mesmo utilizando tecnologias analíticas de Big Data. Entretanto,

notou-se que a parte de operação e manutenção de sistemas de Big Data não ofereceram maiores desafios para a área de suporte da TI.

Nesse momento, não está no escopo da Empresa “B” o estudo de dados externos, como redes sociais, permanecendo o foco em analisar os dados já existentes e ainda subutilizados.

A Figura 24 resume as principais diferenças encontradas entre o uso da análise tradicional de dados e o Big Data, bem como as melhorias nas dimensões competitivas para a Empresa “B”.

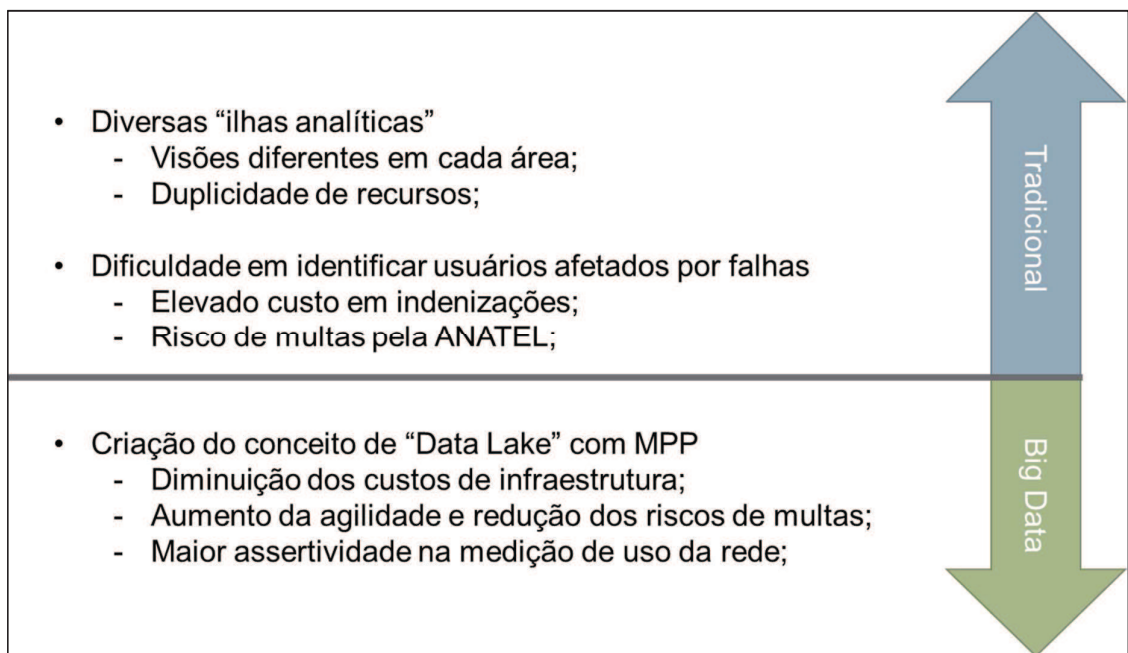


Figura 24: Resumo do estudo da Empresa “B”

Fonte: Elaboração do autor.

4.1.3 Empresa “C”

Atuando no mesmo mercado da Empresa “B”, concorre diretamente com essa, sofrendo também com a chegada de novos entrantes ao mercado e o crescimento de produtos substitutos como o WhatsApp e o Skype, previamente mencionados.

Possui uma área de BI forte e bem estruturada, e atualmente estuda como o Big Data pode complementar e adicionar capacidade de geração de *insights* ao BI tradicional. Está em

fase piloto a arquitetura mostrada na Figura 25, que junta em um repositório único, baseado em um sistema de arquivos distribuído de baixo custo (HDFS), dados oriundos de:

- a) registro de chamadas dos usuários (CDRs);
- b) dados de navegação e uso dos serviços pelos usuários;
- c) texto gerado pela interpretação dos chamados de voz do *Call Center*;
- d) dados do BI tradicional.

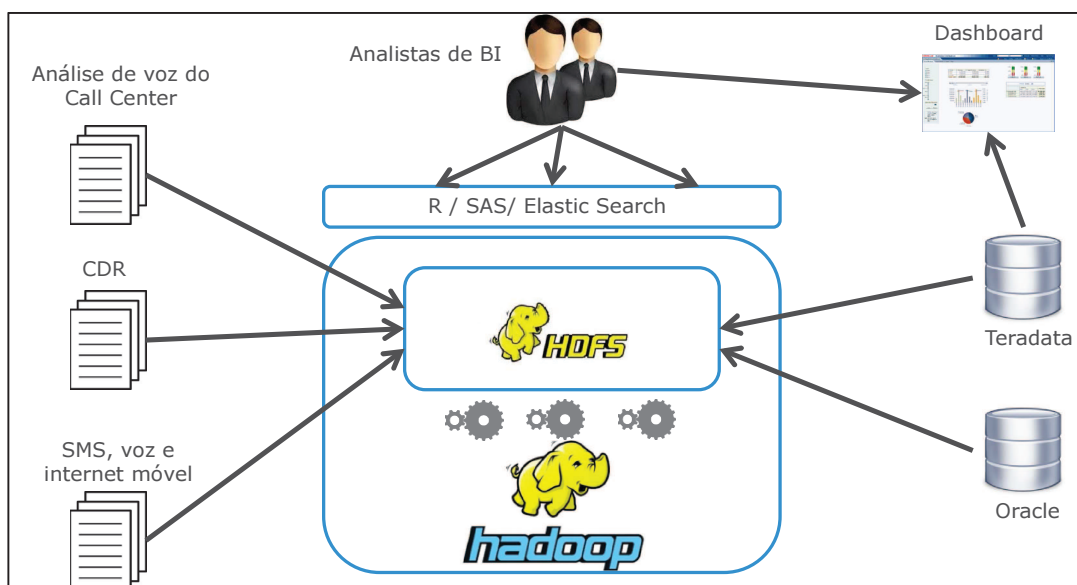


Figura 25: Estrutura do piloto para complemento do BI

Fonte: Empresa “C”.

Esses dados são analisados utilizando o paralelismo de performance (MPP) do Hadoop, em conjunto com ferramentas de análise estatística como R ou Elastic Search.

Esse modelo difere de sistemas tradicionais de análise de dados, que obriga que todos os dados sejam convertidos para um formato estruturado ou de tabela. Como mencionado, esse processo de ETL é demorado, possui alto custo e limita os dados que podem ser analisados.

O novo sistema existirá em complemento ao BI tradicional, sendo utilizado para complementar a solução atual (Teradata e Oracle) para análises que hoje não são possíveis de serem realizadas nos sistemas tradicionais:

- a) histórico maior dos CDRs, além da inclusão dos registros de chamadas não completadas, que, por falta de capacidade e pelo altíssimo custo de armazenagem na solução tradicional, eram descartadas. Atualmente, a empresa armazena apenas os registros das chamadas realizadas, o que impede uma visão mais completa sobre a intenção de uso dos serviços pelos clientes, bem como uma análise mais completa sobre a qualidade dos serviços prestados;
- b) dados semiestruturados, como DPI²⁷, que contém informações como o IP de destino do acesso à internet dos usuários, e se analisados podem ajudar a identificar padrões de comportamento dos usuários. Esses dados não eram analisados pelo seu grande volume e maior complexidade, visto que, eram dados semiestruturados;
- c) dados totalmente não estruturados ou textos livres, que são a transcrição das conversas entre os atendentes do *call center* e os clientes. A análise desses dados pode permitir um controle mais preciso da qualidade no atendimento ao cliente, levando a uma maior taxa de retenção desses clientes para a empresa. Esses são dados ainda mais complexos de serem analisados do que os CDRs e DPIs mencionados anteriormente, e, por isso, não faziam parte dos sistemas de BI da empresa.

Devido ao elevadíssimo volume de registros, chegando a 2 bilhões por dia, uma análise usando os sistemas atuais da empresa pode levar de 8 horas até a alguns dias, chegando, em alguns casos, a travar a base de dados, caso seja utilizada uma grande quantidade de filtros e variáveis. Com o novo sistema, esse tempo diminui para menos de 3 horas, mesmo com uma massa de dados maior e com um custo total inferior ao sistema original.

Para o Big Data, foram construídos 3 casos de negócio que estão sendo testados:

1. Análise de geolocalização dos usuários:

Por meio da análise dos registros dos celulares na rede da operadora e com a triangulação de antenas, é possível entender melhor o uso da rede e das antenas pelos usuários. Em um primeiro momento, essas informações serão usadas pela

²⁷ *Deep Packet Inspection* ou inspeção detalhada do pacote é o registro que contém informações básicas sobre o acesso que os usuários realizam aos sites de internet por meio da rede de dados móvel.

Empresa “C” para permitir um melhor posicionamento das antenas, melhorando a qualidade do serviço oferecida aos usuários, e criação de *hotspots*, para permitir o acesso à internet pelos clientes, sem a necessidade que esses usem a rede móvel, reduzindo, assim, os custos da empresa.

O objetivo secundário é conhecer mais sobre o cliente: em qual região ele mora, aonde trabalha, aonde ele vai no final de semana, quais os hábitos de viagem, lazer, entre outros. Com isso, a empresa pode reduzir seus custos de marketing e aumentar a relevância das campanhas para os clientes, oferecendo serviços mais aderentes às necessidades dos consumidores e, como consequência, aumentar os seus números de vendas.

2. Análise de sentimento dos usuários na interação com *Call Center*:

Análise do texto gerado pela transcrição automática da conversa de voz entre o atendente do *Call Center* e os clientes, com o objetivo de medir a satisfação do cliente e entender o que o leva a migrar para outras empresas. O objetivo final é diminuir a taxa de *churn*, ou seja, a quantidade de clientes que migram para outras operadoras.

Essa análise é realizada utilizando métodos estatísticos que atribuem valores para palavras, dependendo do posicionamento da palavra em uma frase e da proximidade com outras palavras. Por exemplo, a palavra “nunca” pode ter significados completamente diferentes: “nunca tive problema” ou “nunca mais quero ser clientes de vocês”.

Por meio do modelo criado, será possível antever clientes com maior probabilidade de deixar a Empresa “C”, que, associado a uma melhor categorização dos clientes, permitirá a realização de ações direcionadas de retenção dos clientes mais lucrativos.

3. Análise de navegação dos usuários:

Por meio da análise do DPI, é possível saber quais os sites de internet o cliente acessa. Porém, observa-se que não é possível acompanhar a navegação do usuário. Por exemplo, é possível saber que o cliente acessou o Facebook ou o Google, mas não é possível saber os relacionamentos do usuário ou o que ele pesquisou.

Com a informação obtida, a Empresa “C” está melhorando o seu sistema de clusterização ou agrupamento de clientes, criando segmentações mais precisas do que apenas faixa etária ou sexo, incluindo informações como hábitos de consumo, esportes preferidos, entre outros. É possível, inclusive, medir o uso de produtos concorrentes, como demonstrado na Figura 26 que aponta os principais clientes da empresa que usam também o WhatsApp para envio de mensagens de texto. Esses clientes, além de diminuir os gastos com os serviços de SMS da Empresa “C”, podem afetar outros clientes que passam a se comunicar da mesma forma. Ao juntar essa informação com dados de gasto médio e poder aquisitivo dos clientes, a Empresa “C” poderá criar campanhas e pacotes promocionais direcionados.

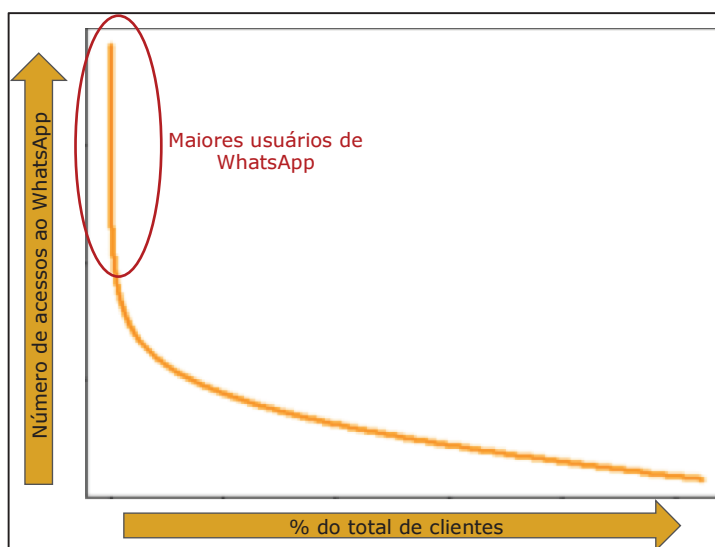


Figura 26: Identificação dos clientes que acessam o WhatsApp

Fonte: Empresa “C”.

A Empresa “C” já possui planos futuros de expansão do uso do Big Data, criando uma visão vinculada entre o perfil detalhado do cliente e a geolocalização dos usuários. Com isso, a empresa poderá gerar ofertas dentro do percurso do cliente ou quando esse está em uma determinada localidade. Essa capacidade pode gerar um novo produto de propaganda direcionada a ser vendido para outras empresas. Por exemplo, um cliente que acesse sites de venda de vinhos pode receber um SMS avisando de uma promoção em uma loja próxima a sua

localidade. Ressalta-se que não há a intenção de comercialização ou exposição dos dados dos usuários ou da individualização do cliente, entretanto há planos para enriquecer a massa de dados com fontes externas como redes sociais.

O envio de propaganda por meio de dispositivos móveis já é uma realidade, com legislações próprias que balizam a sua geração, entretanto ainda há um hiato na legislação sobre a classificação das informações privadas (como, por exemplo, registro de ligações e dados de navegação) e na permissão uso destes dados pelas empresas prestadoras dos serviços de telecomunicações. Sem entrar no mérito das possíveis implicações éticas e legais, o novo modelo buscado pela Empresa “C” permitiria a geração de campanhas de marketing ou propaganda muito mais direcionada e com maior relevância ao consumidor.

Os principais problemas enfrentados durante o projeto piloto foram:

- a) acesso aos dados: ao lidar com múltiplas fontes de dados, que estão armazenados em sistemas distintos, sob gestão de diferentes diretorias, é necessário transpor barreiras organizacionais;
- b) entendimento dos dados: como boa parte dos dados era armazenada sem um uso específico, não existe o metadado do dado, ou seja, não é claro o que cada campo e pedaço daquele dado representa;
- c) conhecimento do time: a Empresa “C”, apesar de suas buscas no mercado, está com grande dificuldade em encontrar profissionais que consigam congregiar os atributos de um cientista de dados: conhecimento nas novas tecnologias de Big Data, conhecimento matemático, estatístico e conhecimento do mercado de telecomunicações. Mesmo o treinamento de novos profissionais é algo que requer um tempo considerável, dado o extenso currículo necessário.

A Figura 27 resume as principais diferenças encontradas entre o uso da análise tradicional de dados e o Big Data, bem como as melhorias nas dimensões competitivas para a Empresa “C”.

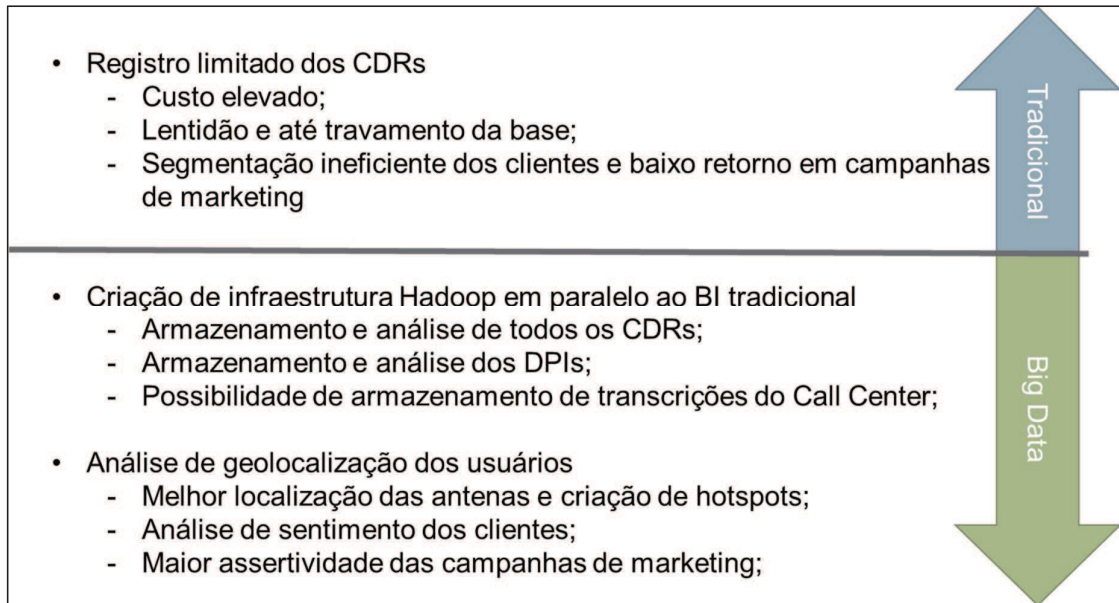


Figura 27: Resumo do estudo da Empresa “C”

Fonte: Elaboração do autor.

4.1.4 Empresa “D”

A Empresa “D” está em acirrada competição com novos entrantes no mercado, empresas de outros setores, como o de telecomunicações, que estão começando a oferecer serviços de internet de banda larga e TV por assinatura. Com menor relevância, produtos substitutos como provedores de filmes *on-line*, como a Netflix, e sistemas ilegais de *download* de filmes e séries também competem com os serviços oferecidos pela Empresa “D”.

A empresa possui dentro das diversas áreas, como engenharia, marketing e atendimento, sistemas e pessoal dedicado à análise de dados. Adicionalmente, foi criada uma nova área de inteligência analítica respondendo diretamente ao vice-presidente de estratégia, que de forma horizontal interage com os demais times analíticos. A análise de dados já faz parte da estratégia da empresa, porém trabalhando com silos de estruturas analíticas em cada uma das áreas.

Existia uma suspeita que a junção das informações contidas nas diferentes áreas poderia trazer novos *insights*, e, em um experimento desenvolvido internamente, com cruzamento manual dos dados, foi realizado um teste empírico duplo-cego correlacionando velocidade da internet dos clientes com a adimplência destes. O teste objetivava medir se, ao oferecer aos

clientes uma internet com velocidade superior àquela contratada, aumentaria a taxa de adimplentes destes clientes. Apesar de serem dois fatores que aparentemente não guardam nenhuma relação, o resultado do teste demonstrou que o grupo com a velocidade da internet aumentada apresentou uma maior taxa de adimplência em comparação ao grupo de controle. Esse experimento mostra o potencial de análises combinadas de bases de dados que hoje estão isoladas, como engenharia e financeiro. Esse primeiro teste instigou o vice-presidente a investir em novas tecnologias que permitissem o cruzamento dessas bases de dados e a criação de experimentações analíticas sobre essa grande massa de informações.

O principal diferencial no uso do Big Data para a Empresa “D” não está na análise de grandes volumes de dados, mas na capacidade de analisar uma grande variedade de dados em um mesmo modelo analítico.

O projeto piloto, descrito na Figura 28, junta as bases de dados usadas pelo departamento de atendimento ao consumidor com as bases de dados de engenharia. Essa massa de dados é, então, enriquecida com logs de sensores, geolocalização dos dispositivos e dados externos, como problemas na rede elétrica e condições meteorológicas.

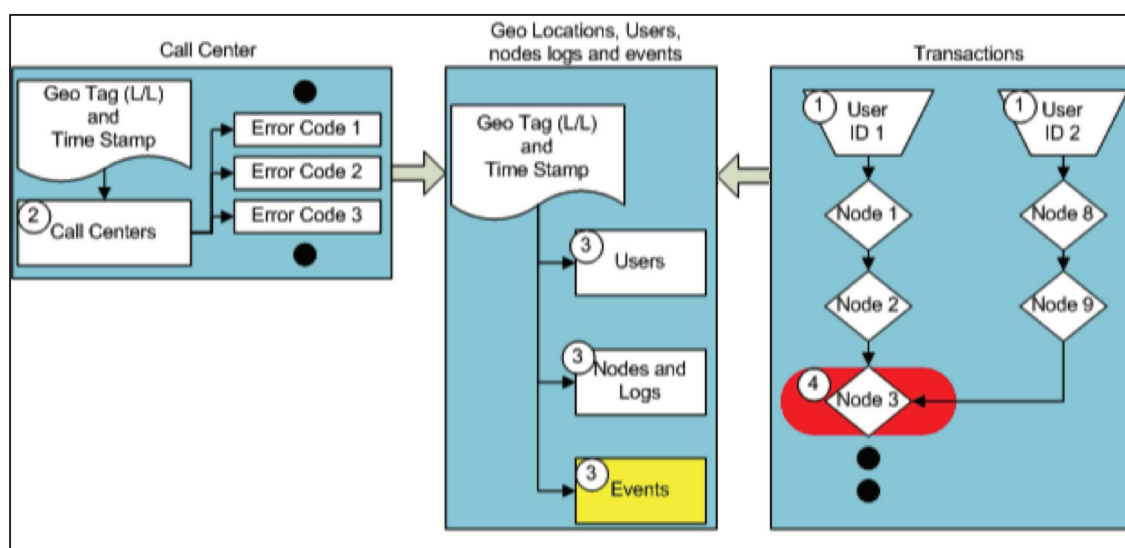


Figura 28: Interligação dos sistemas de atendimento e engenharia

Fonte: Empresa “D”.

O novo sistema, baseado em uma arquitetura MPP, permitirá cruzar os dados oriundos das diversas fontes de dados mencionadas e utilizar algoritmos específicos para determinar a relação entre eventos internos, como a quebra de equipamentos, ou reclamações recebidas no *call center* e eventos externos, como condições meteorológicas, com o objetivo de:

- a) evitar visitas desnecessárias do técnico, por um problema temporário, como falta de energia ou problemas meteorológicos;
- b) evitar visitas em duplicidade para diversos clientes que estejam sofrendo uma indisponibilidade por um problema em comum, como no exemplo da Figura 28, na qual a falha no “Node 3” afeta simultaneamente os usuários de ID 1 e 2;
- c) tornar visitas mais assertivas por meio de uma melhor identificação do componente defeituoso;
- d) manutenção preventiva de componentes e análise preditiva de indisponibilidade da rede, por exemplo por meio de sensores que identificam que determinado equipamento como um modem na casa do cliente está esquentando demais ou, então, a detecção de um sinal pirata na rede, gerando ruído e diminuindo a qualidade do serviço;
- e) munir o *Call Center* e os clientes de informações mais precisas sobre o status dos componentes e funcionamento dos serviços pertinentes a cada cliente.

Esse novo sistema trará como principais benefícios:

- a) redução nos custos provenientes de visitas técnicas, que hoje somam mais de cem milhões de reais ao ano;
- b) reduzir a taxa de *churn* por meio da melhoria no atendimento aos clientes em eventos de falhas e indisponibilidade dos serviços.

A maior dificuldade encontrada na criação do projeto foi a vinculação dele a uma área específica. A Empresa “D” trabalha com modelo de orçamento departamental, entretanto o sistema de Big Data implementado atravessa mais de uma área (atendimento e engenharia).

Pela mesma separação dos departamentos, o segundo ponto de maior dificuldade do projeto foi o acesso aos dados espalhados entre sistemas analíticos distintos.

O terceiro desafio foi a identificação e tradução dos dados de logs de dispositivos e atendimento de *call centers* que são armazenados, porém sem o metadado correspondente, por serem dados não utilizados por BI tradicional. A ausência do metadado dificulta a identificação da formatação e conteúdo do arquivo.

A Figura 29 resume as principais diferenças encontradas entre o uso da análise tradicional de dados e o Big Data, bem como as melhorias nas dimensões competitivas para a Empresa “D”.

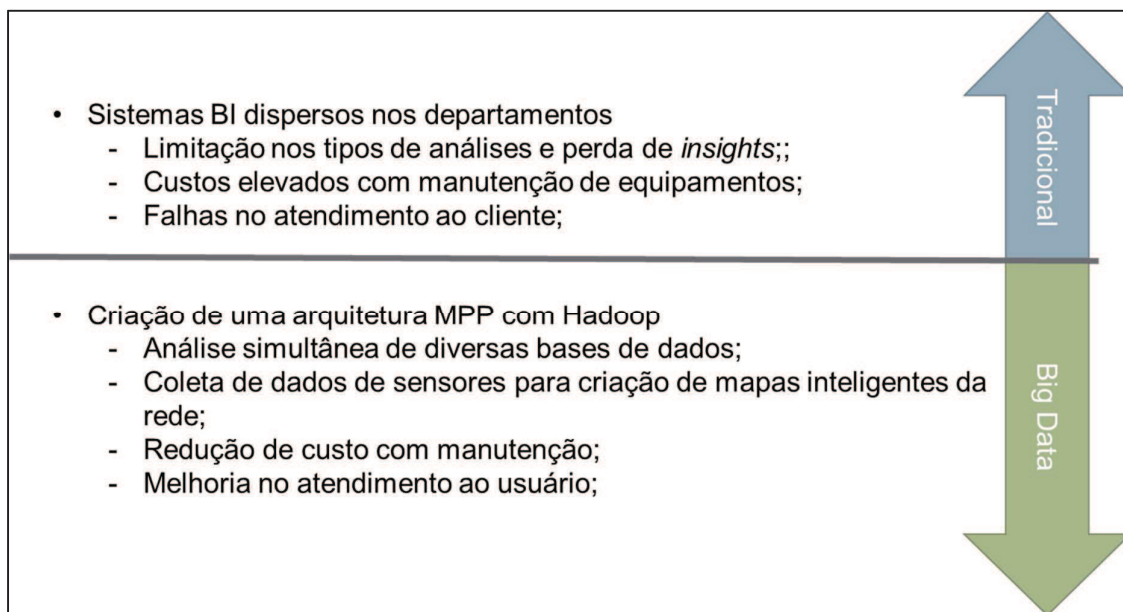


Figura 29: Resumo do estudo da Empresa “D”

Fonte: Elaboração do autor.

4.1.5 Empresa “E”

Nos últimos anos, a Empresa “E”, bem como as demais instituições financeiras de porte semelhante, gozou de excelente lucratividade, o que, entretanto, não diminui a competição interna no mercado.

Por atuar no mercado financeiro, a Empresa “E” herda desse a aversão a risco e o pragmatismo na adoção de novas tecnologias, sendo por definição própria uma empresa que

costuma adotar tecnologias quando essas já estão devidamente maduras e comprovadas por dezenas de casos de sucesso no mercado.

Entretanto, ao se deparar com um problema que não conseguiu resolver utilizando suas soluções tradicionais de análise de dados, a Empresa “E” buscou soluções inovadoras que conseguissem atender à demanda de processamento de grande volume de dados não estruturados de forma rápida e veloz. Nota-se que até este momento, a percepção da Empresa “E” sobre sistemas de Big Data era de algo utópico para a realidade deles, acreditando tratar-se apenas de uma solução voltada para análise de redes sociais e outras fontes de dados da internet.

Ao aprofundar os estudos sobre o Big Data, surgiu, então, a percepção de que esse pode ser usado para analisar o grande volume de dados já existente dentro das organizações, mas desprezado pelos sistemas de análise de dados tradicionais.

Ao iniciar o piloto de implementação do novo modelo de análise de dados, a Empresa “E” possuía dois fortes requisitos: um problema de negócio bem estabelecido e um prazo determinado para a obtenção de resultados. Esses requisitos permitiram que o projeto piloto fosse mais facilmente mensurável e tivesse sua efetividade e eficácia atestadas.

A Empresa “E” estava sofrendo uma série de fraudes nos caixas automáticos (ATMs) em que os ladrões utilizavam ataques sofisticados, com a quebra de alguns componentes dos ATMs e o uso de *softwares* maliciosos para sacar todo o dinheiro das máquinas, um processo denominado por eles como “sangrar o caixa eletrônico”. Os ataques eram realizados de forma coordenada e quase imperceptível, sem explosões ou algo semelhante, possibilitando aos ladrões o roubo de mais de mil caixas em um único final de semana, levando a um prejuízo aproximado de quase cem milhões de reais, que só era percebido pelo banco no próximo dia útil, na abertura da agência.

Os ATMs possuem diversos sensores, como temperatura, sensores sísmicos, controle de abertura do cofre, dispensadora de dinheiro, entre outros. Uma parte dessas informações é coletada e armazenada em uma base de dados relacional convencional. Porém, outra parte dos dados, por serem gerados em formato de log não estruturado, não eram armazenados na base dados, sendo simplesmente colocados em um repositório de rede. Adicionalmente, devido ao grande volume de informações geradas, de aproximadamente dois *terabytes* por dia, sem que houvesse um uso específico para esses dados, eles eram armazenados por apenas um dia, sendo descartados automaticamente depois disso.

O projeto piloto, implementado de acordo com a Figura 30, utiliza algumas características inerentes a sistemas de Big Data:

- a) sistema de processamento paralelo (MPP ou *Massively Parallel Processing*), de baixo custo e alta escalabilidade para poder processar as informações de forma rápida;
- b) sistema de armazenamento integrado que suporta dados estruturados e não estruturados simultaneamente. No total, foram 8 diferentes fontes de dados com 18 tipos diferentes de formatos de dados;
- c) Desenvolvimento de modelos analíticos de *machine learning* baseados em regressão logarítmica e árvores de decisão com uso de séries temporais e de *softwares* estatísticos para criar um modelo que permitisse caracterizar uma sequência de eventos atribuindo para determinar se um ATM estava sofrendo ataque.

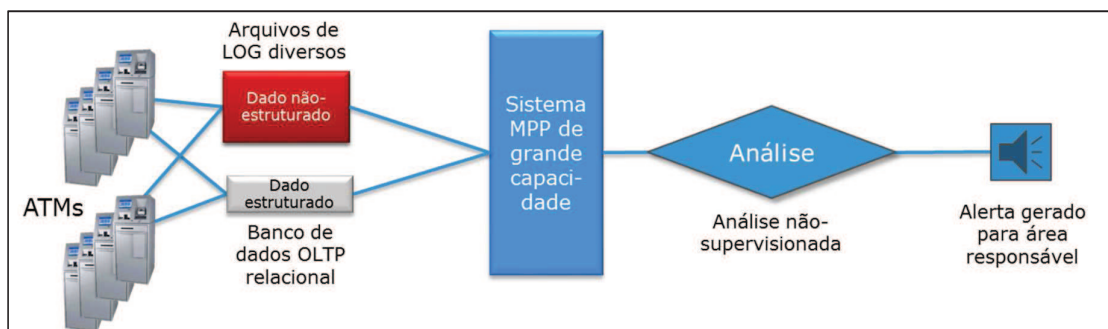


Figura 30: Estrutura do piloto para monitoração dos ATMs

Fonte: Empresa “E”.

O resultado foi um sistema que consegue identificar em questão de minutos e com quase 80% de acerto quando um ATM estava sendo atacado. Como o ataque demora algumas horas para ser realizado, torna-se possível a criação de alertas que possam parar o ataque e avisar as autoridades competentes. Foi possível também realizar análises preditivas para obter o padrão geográfico do ataque, identificando quais ATMs próximos poderiam também sofrer o ataque do grupo criminoso.

Além do foco específico em um problema de negócio conhecido, o fator de tempo do projeto era outro aspecto importante. Era essencial demonstrar que benefícios poderiam ser

obtidos em um prazo diminuto. O projeto piloto conseguiu trazer resultados em apenas oito semanas após seu início.

Durante esse projeto, a parte mais difícil foi no acesso aos dados, em especial os dados não estruturados, os quais representam 70% da massa total de dados e que até o momento não eram utilizados. Falta um processo de governança que cubra também esses dados provenientes de sensores e outros dispositivos, que até o momento eram desprezados ou simplesmente armazenados sem nenhum propósito específico. No caso dos dados provenientes dos ATMs, não havia o metadado desses dados, tornando o processo de análise desses dados mais complexo e trabalhoso.

Após a finalização do projeto piloto, a Empresa “E” está realizando a implementação do sistema em formato permanente. A principal preocupação é a natureza volátil de fraude. Quando o criminoso percebe que a fraude não está mais funcionando, ele muda e se adapta. Por isso, é importante um sistema que consiga aprender e dinamicamente apresentar novas defesas.

A Figura 31 resume as principais diferenças encontradas entre o uso da análise tradicional de dados e o Big Data, bem como as melhorias nas dimensões competitivas para a Empresa “E”.

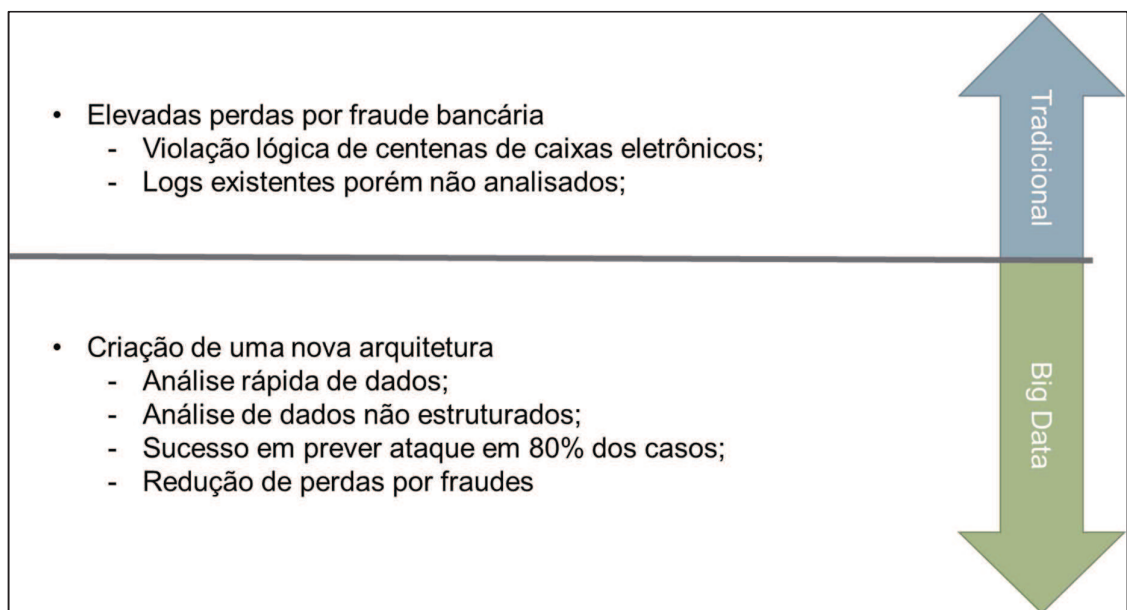


Figura 31: Resumo do estudo da Empresa “E”

Fonte: Elaboração do autor.

5 CONSIDERAÇÕES FINAIS

O termo Big Data define um novo paradigma no tratamento diferenciado e mais eficiente de grandes massas de dados, porém sua origem exata não pode ser precisada, apesar dos indícios sobre o período e grupo de pesquisadores que participaram do nascimento do termo. Esse é um termo que, até presente momento, possui real significado e abrangência ainda em formação, entretanto está claro o crescimento pelo interesse e estudo ao redor do tema.

Apesar de ser impossível prever se o termo estará presente no futuro, parece claro que os conceitos embarcados dentro do termo, que incluem o tratamento de grandes volumes de dados com grande velocidade e ampla variedade, estão criando uma mudança de paradigma na análise de dados.

Esta mudança ocorre porque as empresas estão nadando em um mar de dados, os quais são muito volumosos ou muito complexos para serem analisados da forma tradicional, e a análise dessa gigantesca massa de dados está se tornando a nova fronteira tecnológica para a criação e implementação de uma estratégia corporativa que permita a empresa triunfar em seu mercado.

Os fornecedores de soluções de análises de dados estão realizando grandes investimentos no desenvolvimento de soluções de Big Data e empresas de diversos segmentos estão obtendo resultados iniciais positivos, melhorando suas dimensões competitivas por meio do uso de Big Data.

Em todas as empresas pesquisadas, notaram-se semelhanças com relação ao entendimento da diferença entre soluções de Big Data e soluções tradicionais de análise de dados, sendo essa diferença caracterizada principalmente pela capacidade de soluções de Big Data:

- a) processarem dados de forma muito mais rápida do que soluções tradicionais, permitindo às empresas gerar respostas em tempo real para determinadas condições ou transações de negócio;
- b) analisarem grandes volumes de dados, que não poderiam ser analisados em sistemas tradicionais, seja pela limitação de capacidade a um custo acessível, seja pela agilidade no processamento de. Dessa forma, permite-se à empresa

análise extensiva dos dados, criando novas questões a serem investigadas utilizando pesquisa conduzida por método de experimentação analítica;

- c) incluir na análise dados não estruturados, e novas fontes de dados, como logs de sistemas, dados gerados por sensores e até informações de mídias sociais, permitindo obter *insights* mais precisos ou gerar novos estudos analíticos, como análise de sentimento dos consumidores sobre a marca da empresa ou sobre campanha de marketing.

Também em todos os casos, ficou claro que Big Data não se traduz apenas em aquisição de um novo produto ou na aquisição de mais infraestrutura para armazenar mais dados. É um conjunto de:

- a) novas tecnologias de armazenamento, como HDFS, bancos de dados não relacionais (noSQL), bancos de dados em memória (*in-memory database*);
- b) novas ferramentas de análise como R, ou Python;
- c) novos conhecimentos para o time de analistas da empresa, para uso das novas tecnologias e implementação de um novo modelo de experimentação de análise dos dados, incluindo conceitos matemáticos e estatísticos;
- d) ajustes na estrutura organizacional, com o surgimento do cientista de dados, o qual possui conhecimentos do ferramental tecnológico, conceitos estatísticos e conhecimento do negócio da empresa, permitindo a esse profissional criar as perguntas de pesquisa a serem investigadas pela análise na massa de dados.

Notou-se também que as maiores dificuldades enfrentadas pelas empresas originam-se justamente desse novo modelo organizacional e da necessidade de uma nova paleta de conhecimento dos profissionais.

O profissional de Big Data, chamado de cientista de dados, com conhecimentos em áreas distintas como tecnologia, matemática, estatística e negócios, é muito difícil de ser formado pelas empresas ou de ser encontrado no mercado profissional, causando uma lacuna na mão de obra especializada e limitando ou atrasando a adoção da solução.

A obsolescência da estrutura organizacional na relação entre as áreas das empresas evidencia-se nos relatos obtidos, especialmente no que tange à unificação das massas de dados departamentais ou ilhas analíticas.

A Tabela 1 traz um resumo dos principais ganhos nas dimensões competitivas das empresas pesquisadas ao realizarem projetos de adoção de Big Data. A Empresa “A” não faz parte do quadro por ter sido pesquisada como provedora de soluções de Big Data e não como empresa usuária de tais soluções.

Tabela 1: Principais ganhos com a adoção do Big Data

Empresa	Principais ganhos nas dimensões competitivas
B	<p>Redução de custos da infraestrutura de TI.</p> <p>Redução de custos com pagamentos de indenizações e multas.</p> <p>Aumento da receita por meio da medição mais precisa do uso dos serviços.</p> <p>Melhoria na qualidade do atendimento ao cliente.</p>
C	<p>Melhoria na qualidade do serviço para o cliente.</p> <p>Redução na perda de receita decorrente da diminuição de clientes.</p> <p>Aumento da eficiência de campanhas promocionais.</p>
D	<p>Reduzir o custo por visitas técnicas.</p> <p>Diminuir a taxa de perda de clientes.</p> <p>Melhorar a qualidade do atendimento ao cliente.</p>
E	<p>Diminuição de perdas por fraudes.</p>

Fonte: Elaboração do autor.

Como em todo modelo competitivo, ganha mais vantagem quem chega primeiro, daí a importância de as empresas encararem os dados como um ativo primordial e criarem projetos para investigação e subsequente implementação do Big Data como ferramenta de melhoria de suas dimensões competitivas, sempre alinhados aos objetivos corporativos e com metas bem estabelecidas.

O Big Data ainda é um conceito novo, estando as empresas pesquisadas ainda no estágio inicial de implantação do projeto. Por esse motivo, não se pode ainda chegar a uma conclusão sobre a real influência do Big Data na geração de melhoria para as dimensões competitivas de uma empresa. Existem, entretanto, fortes indícios constatados pelas empresas entrevistadas e alguns casos de resultados empíricos sobre as vantagens do uso do Big Data. Esses indícios estão levando essas empresas a realizarem grandes investimentos em projetos na área.

Como sugestão para trabalhos futuros podemos destacar:

- Ampliação do estudo sobre o Big Data em outras empresas, inclusive de outros segmentos industriais, bem como revisitar as empresas aqui estudadas para avaliar os progressos no uso do Big Data, ganhos obtidos e novas melhorias nas dimensões competitivas extraídas destes projetos.
- Expandir e detalhar o estudo sobre os aspectos de privacidade, incluindo as novas legislações que regulamentam o uso e acesso às informações disponíveis na Internet, adicionando a este estudo uma pesquisa quantitativa mais abrangente, obtendo uma parcela estatisticamente mais significativa dos usuários de redes sociais.

REFERÊNCIAS

- ALVES FILHO, A; PIRES, S. R. I; VANELLE, R. M. Sobre as prioridades competitivas da produção: compatibilidades e sequências de implementação. **GESTÃO & PRODUÇÃO**, v. 2, n. 2, p.173-180, ago. 1995.
- ANDRADE, M. **Introdução à metodologia do trabalho científico**. 10. ed. São Paulo: Atlas, 2010.
- AQUINO, J. Transforming social media data into predictive analytics. **Destination CRM**. 2012. Disponível em: <<http://www.destinationcrm.com/Articles/printArticle.aspx?ArticleID=85687>>. Acesso em: 01 jun. 2013.
- BERTO, R.M.v.S., NAKANO, D. N. A Produção Científica nos Anais do Encontro Nacional de Engenharia de Produção: Um Levantamento de Métodos e Tipos de Pesquisa. **Produção**, v. 9, n. 2, p. 65-76, 2000.
- BORBA, D. K; SANTOS, L. F. A; KAWAMOTO JR, L. T. Big Data: percepção dos usuários sobre vantagens e invasão de privacidade. In: Workshop de pós-graduação e pesquisa do Centro Paula Souza, 8., 2013, São Paulo. **Anais...** São Paulo: CEETEPS, 2013.
- BOYD, D; ELLISON, N. Social Network Sites: Definition, History, and Scholarship. **Journal of Computer-Mediated Communication**, n.13, p.210-230. 2008.
- BROWN, B; CHUI, M; MANYIKA, J. Are you ready for the era of ‘Big Data’?. **McKinsey Global Institute**, 2011. Disponível em: <http://www.mckinsey.com/insights/strategy/are_you_ready_for_the_era_of_big_data>. Acesso em: 10 jun. 2013.
- BRYNJOLFSSON, E; MCAFEE, A. Big Data – A Revolução da Gestão. **Harvard Business Review**, 2012. Disponível em: <<http://www.hbrbr.com.br/materia/big-data-revolucao-da-gestao>>. Acesso em: 15 mai. 2013.
- BUGHIN, J; LIVINGSTON, J; Marwaha, S. Seizing the potential of ‘Big Data’. **McKinsey Global Institute**, 2011. Disponível em: <http://www.mckinsey.com/insights/business_technology/seizing_the_potential_of_big_data>. Acesso em: 10 jun. 2013.
- CASONATO, R. et al. Top 10 Technology Trends Impacting Information Infrastructure, 2013. **Gartner**, 2013. Disponível em: <<https://www.gartner.com/doc/2340315>>. Acesso em: 10 jun. 2013.
- CEARLEY, D; CLAUNCH, C. The Top 10 Strategic Technology Trends for 2013. **Gartner**, 2013. Disponível em: <<https://www.gartner.com/doc/2335015/top--strategic-technology-trends>>. Acesso em: 10 jun. 2013.
- COMSCORE Announces Beta Release of Media Metrix® Multi-Platform, Providing a Unified View of Web, Smartphone and Tablet Audiences. Novembro de 2012. Disponível em: <http://www.comscore.com/Insights/Press_Releases/2012/11/comScore_Announces_Beta_Release_of_Media_Metrix_Multi-Platform>. Acesso em: 10 out. 2013.

DAVENPORT, T; BARTH, P; BEAN, R. How 'Big Data' is different. **MIT Sloan Management Review**, v. 54, n. 1, p.22-24, set. 2012.

DAVENPORT, T; HARRIS, J. **Competição Analítica: Vencendo Através da Nova Ciência**. Campus, 2007.

DEIBERT, R; CRETE-NISHIHATA, M. Blurred boundaries: probing the ethics of cyberspace research. **Review of Policy Research**, v. 28, n. 5, p.531-537, set. 2011.

DIEBOLD, F. "Big Data" dynamic factor models for macroeconomic measurement and forecasting. **Advances in Economics and Econometrics: Theory and Applications**, Eighth World Congress of the Econometric Society. p.115-122. 2003

DIEBOLD, F. **On the origin(s) and development of the term "Big Data"**. Penn Institute For Economic Research, 2012. Disponível em: <<http://economics.sas.upenn.edu/system/files/12-037.pdf>>. Acesso em: 12 out. 2013.

DIEHL, C. A; GONÇALO, C. R; MARTINS, G. A. Dimensões competitivas em organizações de serviços: Um modelo de pesquisa aplicada. In: ENEGEP, 21., 2001, Ouro Preto. **Anais...** Ouro Preto: ABEPRO, 2001.

FLICK, U. **Desenho da pesquisa qualitativa**. Porto Alegre: Artmed, 2009.

GALLANT, J. TIBCO CEO: How Real-Time Computing Will Change the Landscape. **ComputerWorld**, 2011. Disponível em: <<http://www.computerworld.com/article/2511353/business-intelligence/tibco-ceo--how-real-time-computing-will-change-the-landscape.html>>. Acesso em: 10 jun. 2013.

GANTZ, J; REINSEL, D. **The digital universe in 2020: Big Data, nigger digital shadows, and biggest growth in the far east**. IDC, 2012. Disponível em: <<http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>>. Acesso em: 2 jun. 2013.

GAYO-AVELLO, D. Don't turn social media into another 'literary digest' poll. **Communications of the ACM**, v. 54, n. 10, p.121-128. 2011.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GODOY, A. S. Introdução à pesquisa qualitativa e suas possibilidades, **Revista de Administração de Empresas**, São Paulo, v. 35, n. 2, p. 57-63, mar./abr. 1995.

HAYS, C. What Wal-Mart Knows About Customers' Habits. **The New York Times**, 2004. Disponível em: <<http://www.nytimes.com/2004/11/14/business/yourmoney/14wal.html>>. Acesso em: 7 ago. 2013

HEA, H; ZHAB, S; LI, L. Social media competitive analysis and text mining: A case study in the pizza industry. **International Journal of Information Management**, v. 33, n. 3, p.464-472. 2013.

HOPPEN, N.; LAPOINTE, L.; MOREAU, E. Um guia para a avaliação de artigos de pesquisa em sistemas de informação. **READ – Revista Eletrônica de Administração**, Porto Alegre, PPGA/UFRGS, 3 ed., v. 2, n. 2, nov. 1996.

KOSINSKI, M; STILLWELL, D; GRAEPEL, T. Private traits and attributes are predictable from digital records of human behavior. **Proceedings of the National Academy of Sciences**, v. 110, n. 15, p.5802-5805, apr, 2013.

KRISHNAMURTHY, B; WILLS, C. E. On the leakage of personally identifiable information via online social networks. **Proceedings of the 2nd ACM workshop on Online social networks**. New York: ACM, 2009.

_____. Privacy leakage in mobile online social networks. **Proceedings of the 3rd conference on Online social networks**. USENIX Association. 2010. Disponível em: <https://www.usenix.org/legacy/event/wosn10/tech/full_papers/Krishnamurthy.pdf>. Acesso em: 7 ago. 2013

LACERDA, D. P. et al. Algumas caracterizações dos métodos científicos em engenharia de produção: Uma análise de periódicos nacionais e Internacionais. In: ENEGEP, 27., 2007, Foz do Iguaçu. **Anais...** Foz do Iguaçu: ABEPRO, 2007.

LANEY, D. 3D data management: controlling data volume, Velocity, and Variety. Metagroup (Gartner), 2001. Disponível em: <<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>>. Acesso em: 2 jun. 2013

LESLIE, A. Social Analytics & Intelligence: Converting Contextual to Actionable Insights. **Hypatia Research & Advisory**, 2012. Disponível em: <<http://store.hypatiaresearch.com/socialanalyticsandintelligenceconvertingcontexttoactionableinsights.aspx>>. Acesso em: 2 jun. 2013

LOC. Library of Congress. Update on the Twitter Archive At the Library of Congress. Janeiro, 2013. Disponível em: <http://www.loc.govtoday/pr/2013/files/twitter_report_2013jan.pdf>. Acesso em: 10 out. 2013.

LOHR, S. The origins of Big Data an etymological detective story. **The New York Times**, 2013. Disponível em: <<http://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/>>. Acesso em: 2 jun. 2013

MALHOTRA, N. **Pesquisa de marketing**. 4. ed. Porto Alegre: Bookman, 2004.

MIGUEL, P. Estudo de caso na engenharia de produção: estruturação e recomendações para sua condução. **Produção**, 2007.

MURGEL, D; SILVA, J; NEVES, J. A ética nos negócios como diferencial competitivo. In: ENEGEP, 26., 2006, Fortaleza. **Anais...** Fortaleza: ABEPRO, 2006.

NARCIZO, R; SILVA, C. A utilização de questionários em surveys na engenharia de produção: algumas considerações. In: XVII SIMPEP, 17., 2010, Bauru. **Anais...** Bauru: UNESP, 2010.

NETO, J. P. B; FENSTERSEIFER, J. E; FORMOSO, C. T. Os critérios competitivos da produção: Um estudo exploratório na construção de edificações. **RAC**, v. 7, n. 1, p. 67-85, jan./mar. 2003.

NOVO, R. F; NEVES, J. M. S. Origem Do Big Data E Estudo Bibliométrico Sobre Sua Utilização Como Diferencial Competitivo Pelas Empresas. In: XX SIMPEP, 20., 2013, Bauru. **Anais...** Bauru: UNESP, 2013.

OLTSIK, J. The evolution of Big Data security analytics technology. **Enterprise Storage Group**. 2013. Disponível em: <http://www.esgnext.com/c/701C0000000eY2H/pdf/ESG_MLR_Big_Data_Security_Analytics_Mar_2013.pdf>. Acesso em: 2 jun. 2013.

OXFORD English Dictionary. Disponível em: <<http://www.oed.com/view/Entry/18833#eid301162177>>. Acesso em: 17 jun. 2013.

PETSCHULAT, S. Other people's data. **Communications of the ACM**, v. 53, n. 1, p. 53-57, jan. 2010.

PORTER, M. **Estratégia competitiva: Técnicas para análise de indústrias e da concorrência**. 7. ed. Rio de Janeiro: Campus, 1991.

PORTER, M. **Vantagem competitiva: Criando e sustentando um desempenho superior**. Rio de Janeiro: Campus, 1990.

ROCHA, M. K. et al. Métodos e tipos de pesquisa em artigos da engenharia de produção. In: COBENGE, 39., 2011, Blumenau. **Anais...** Blumenau: FSA, 2011.

ROESCH, S. **Projeto de estágio e de pesquisa em administração**. 3. ed. São Paulo: Atlas, 2007.

SANTOS, F. C. A; PIRES, S. R. I. Prioridades competitivas da estratégia de manufatura & dimensões competitivas da estratégia de recursos humanos: Estudo de casos. In: EnANPAD, 23., 1999, Foz do Iguaçu. **Anais...** Foz do Iguaçu: ANPAD, 1999.s

SCHMARZO, B. **Big Data: Understanding how data powers big business**. Wiley, 2013.

SMITH, A. 13% of online adults use Twitter. **Pew Research Center**. 2011. Disponível em: <<http://www.pewinternet.org/files/old-media/Files/Reports/2011/Twitter%20Update%202011.pdf>>. Acesso em: 10 out. 2013.

STEIN, J. Your Data, Yourself. **Time Magazine**, p. 40-46, mar. 2011.

TANKARD, C. Big Data security. **Network Security**, v. 2012, n. 7, p. 5-8, jul. 2012.

TEIXEIRA, R. et al. Fatores determinantes da competitividade na indústria de telecomunicações e repercussões para a estratégia. **BASE – Revista de Administração e Contabilidade da Unisinos**, v. 2, n. 1, p. 15-26, jan./abr. 2005.

TEIXEIRA, R. F; PACHECO, M. E. C. Pesquisa social e a valorização da abordagem qualitativa no curso de administração: a quebra dos paradigmas científicos, **Caderno de Pesquisas em Administração**, São Paulo, v. 12, n. 1, p. 55-68, jan./mar. 2005.

TENE, O; POLONETSKY, J. Privacy in the age of Big Data: a time for big decisions. **Stanford Law Review Online**. 2012. Disponível em: <http://www.stanfordlawreview.org/sites/default/files/online/topics/64-SLRO-63_1.pdf>. Acesso em: 10 out. 2013.

THE ECONOMIST. New rules for Big Data. **The Economist**, 2010. Disponível em: <<http://www.economist.com/node/15557487>>. Acesso em: 10 out. 2013.

TILLY, C. The Old New Social History And The New Old Social History. **CRSO Working Paper**, n. 218, 1980. Disponível em: <<http://deepblue.lib.umich.edu/bitstream/handle/2027.42/50992/218.pdf?sequence=1&isAllowed=y>>. Acesso em: 10 out. 2013.

TUCKER, W.; THALER, R. H. Informação mais inteligente, consumidor mais forte. **Harvard Business Review**. 2013. Disponível em: <<http://www.hbrbr.com.br/materia/informacao-mais-inteligente-consumidor-mais-forte>>. Acesso em: 10 out. 2013.

VASCONCELOS, F. C; CYRINO, A. B. VANTAGEM COMPETITIVA: Os modelos teóricos atuais e a convergência entre estratégia e teoria organizacional. **RAE - Revista de Administração de Empresas**. São Paulo, v. 40 n. 4, p. 20-37, out./dez. 2000.

VEJA. São Paulo, ed. 2321, abr. 2013.

VERGARA, S. **Projetos e relatórios de pesquisa em administração**. 13. ed. São Paulo: Atlas, 2011.

YIN, R. **Estudo de caso: Planejamento e métodos**. 4. ed. Porto Alegre: Bookman, 2010.