

CENTRO PAULA SOUZA

FACULDADE DE TECNOLOGIA DE AMERICANA
Curso de Segurança da Informação

Maiza Bueno da Silva

BIG DATA COMO FERRAMENTA DE APOIO
AO COMÉRCIO ELETRÔNICO

Americana, SP

2014

CENTRO PAULA SOUZA

FACULDADE DE TECNOLOGIA DE AMERICANA
Curso de Segurança da Informação

Maiza Bueno da Silva

BIG DATA COMO FERRAMENTA DE APOIO AO COMÉRCIO ELETRÔNICO

Trabalho de Conclusão de Curso desenvolvido em cumprimento à exigência curricular do Curso de Segurança da Informação, sob a orientação da Prof.^a Dr. Maria Cristina Aranda Batocchio

Área de concentração: Segurança da Informação

Americana, SP

2014

FICHA CATALOGRÁFICA – Biblioteca Fatec Americana - CEETEPS
Dados Internacionais de Catalogação-na-fonte

Silva, Maiza Bueno da

S581b

Big Data como ferramenta de apoio ao comércio eletrônico. / Maiza Bueno da Silva. – Americana: 2014.

63f.

Monografia (Graduação de Tecnologia em Segurança da Informação). - - Faculdade de Tecnologia de Americana – Centro Estadual de Educação Tecnológica Paula Souza.

Orientador: Prof. Dr. Maria Cristina Aranda Batocchio

1.Comercio eletrônico 2. Rede de computadores I.
Batocchio, Maria Cristina Aranda II. Centro Estadual de Educação Tecnológica Paula Souza – Faculdade de Tecnologia de Americana.

CDU: 658.845
681.519

Maiza Bueno da Silva

BIG DATA COMO FERRAMENTA DE APOIO AO COMÉRCIO ELETRÔNICO

Trabalho de graduação apresentado
como exigência parcial para obtenção
do título de Tecnólogo em Segurança da
Informação pelo CEETEPS/Faculdade
de Tecnologia - FATEC/Americana

Área de concentração: Segurança da
Informação

Americana, ___ de Junho de 2014.

Banca Examinadora:

Maria Cristina Aranda Batocchio
Doutora
FATEC Americana

Rogério Nunes de Freitas
Especialista
FATEC Americana

Maria Elizete Luz Saes
Mestre
FATEC Americana

AGRADECIMENTOS

Em primeiro lugar, agradeço esta oportunidade a FATEC Americana e seus professores por mostrarem um lado da tecnologia que eu não conhecia e, portanto, não teria meios de fazer este trabalho. Agradeço também a loja online, cujo nome fictício neste trabalho será Mundo Hipster por motivos de confidencialidade, pelas informações fornecidas a mim.

DEDICATÓRIA

A Deus, meus pais por todo o apoio dado a mim, meus amigos pela paciência em relação a minha ausência e minha orientadora Prof.^a Dr. Maria Cristina Aranda por todo o tempo dedicado a mim, ideias e ajuda.

RESUMO

Este texto conceitua sugestões de uso de um sistema Big Data em uma loja de *e-commerce* chamada ficticiamente para preservar a empresa em questão, de mundo Mundo Hipster. O Big Data é uma tecnologia que usa uma grande quantidade de dados para criar novas ideias e novos usos. É tratado sobre sua definição, como é usado, seus problemas e o controle dos mesmos e como o sistema funciona. É mostrado as relações dessa ferramenta em várias áreas do conhecimento, principalmente no Marketing e nas redes sociais. O estudo de caso explora o perfil de cliente e o nicho de mercado da loja Mundo Hipster, assim como os usos do Big Data em seu sistema, em técnicas de Marketing e nas redes sociais, que é seu principal foco. Este estudo também mostra que os usos do Big Data em uma loja de *e-commerce* são muito efetivos e versáteis, principalmente com o uso de redes sociais, aumentando a experiência de uma pessoa, cliente ou não, ao entrar no site da loja e ver seus produtos. Considerando os cuidados relacionados aos dados usados pela loja e a privacidade de seus clientes, conclui-se que o uso do Big Data em uma loja de *e-commerce* é viável e indispensável ao negócio.

Palavras chave: big data; e-commerce; segurança da informação.

ABSTRACT

This text conceptualizes suggestions for using a Big Data system in a e-commerce store called Mundo Hipster. It is treated about its definition, how it is used, its problems and their control and how the system works. Relations of this tool is shown in several knowledge areas, particularly in Marketing and social networks. The case study explores the customer profile and niche market of Mundo Hipster store, as well the uses of Big Data on its system, marketing techniques and social networks, which is its main focus. This study also shows that Big Data use in a e-commerce store are very effective and versatile, especially with use of social networks, enhancing the person experience, customer or not, on entering the store website and see its products. Considering the care related to the data used by the store and the privacy of its customers, it's concluded that the use of Big Data in a e-commerce store is viable and indispensable to the business.

Keywords: *big data; e-commerce; security information.*

SUMÁRIO

1. INTRODUÇÃO	10
2. BIG DATA.....	11
2.1. DEFINIÇÃO DO BIG DATA	11
2.2. OS CINCO Vs DO BIG DATA.....	15
2.3. SEGURANÇA DA INFORMAÇÃO DO BIG DATA	17
2.4. CONTROLE DO BIG DATA	25
2.5. FUNCIONAMENTO DO BIG DATA.....	31
2.6. BIG DATA NO COMÉRCIO.....	35
3. ESTUDO DE CASO.....	42
3.1. A EMPRESA.....	42
3.2. PESQUISA.....	42
3.3. MERCADO CONSUMIDOR.....	43
3.4. PERFIL DO CLIENTE	44
3.5. NICHOS DE MERCADO.....	45
3.6. SUGESTÕES DE USO DO BIG DATA NA EMPRESA.....	47
4. CONCLUSÃO	58

1. INTRODUÇÃO

O objetivo deste trabalho é criar maneiras e diferentes usos de um sistema de Big Data em uma loja de comércio, mais especificamente uma loja online real que será tratada aqui por um nome fictício. Esta loja, chamada Mundo Hipster, é do ramo de vestuário e em todos os seus produtos estão estampados motivos musicais.

Esta nova tecnologia, o Big Data, utiliza uma grande quantidade de dados e dá a eles uma nova finalidade. Finalidade esta que pode ser novas ideias obtidas a partir de tais dados, novos usos destes dados e novas ferramentas.

Este texto disserta sobre o que é o Big Data, como e onde ele pode ser usado e detalhar seu funcionamento. Assim como qualquer outra tecnologia, o Big Data também oferece riscos que serão exemplificados neste trabalho, e, a partir disto, será descrito como controlar e mitigar tais riscos.

Uma das maiores vantagens do Big Data é o seu amplo uso, onde todas as áreas do conhecimento podem tirar proveito de seus recursos. O foco neste texto é na área de Marketing e nas redes sociais, onde a tecnologia se mostra mais abrangente. Assim sendo, o estudo de caso avalia que tipo a loja tem ou poderá vir a ter, além de apontar estratégias de marketing implantadas no site da empresa, bem como nas redes sociais. Dessa forma a empresa conseguirá fidelizar seus clientes bem como angariar novos clientes.

O Big Data mostra-se muito versátil no uso com as diferentes redes sociais, podendo tirar proveito de várias de suas características individualmente ou combinadas com outras. Esta característica é tão efetiva que pode ser usada para melhorar a navegabilidade do cliente durante o processo de visita ao site bem como no momento da compra de produtos.

Com este trabalho é possível compreender esta nova tecnologia e entender o motivo de ser tão importante atualmente, não só no Marketing, mas em tantas outras áreas nas quais a ferramenta pode ser usada.

2. BIG DATA

O Big Data é uma grande tecnologia que está em expansão, e este capítulo trás suas principais características. Como toda tecnologia, o Big Data tem seus pós e contras, que serão definidos e exemplificados também neste capítulo, assim como os seus riscos e os relativos controles.

2.1. DEFINIÇÃO DO BIG DATA

Com o advento da Internet e seu grande crescimento nos últimos anos, uma grande quantidade de dados foram criados, sendo, em sua maioria, dados considerados “inúteis”. Inúteis não pelo fato de não terem um uso específico, uma finalidade, mas sim por não serem utilizados em nenhum momento.

Dados são gerados a todo momento. Dados são símbolos quantificáveis que têm uma sequência, que podem ser processados e armazenados por um computador, segundo Setzer (1999). Dentre esses dados, pode-se obter desde dados que são compartilhados no Facebook até cálculos matemáticos complexos processados em um super computador. Atualmente são gerados dados em larga escala. São mais de 2 exabytes de dados todos os dias no mundo todo, de acordo com Petry (2013) apud Matos; Braga; Kawamoto Jr. (2013), Sendo que um exabyte é o equivalente a 1 048 576 HDs externos de 1TB, por exemplo. Os números são tão grandes que é difícil imaginar essa quantidade de dados, e além disso se existe a capacidade de processamento para tamanha quantidade de dados.

O termo Big Data foi criado dentro das áreas de Astronomia e Genômica, após a grande quantidade de descobertas advindas dessas áreas nos anos 2000, gerando uma grande quantidade de dados e informação. Ou seja, no início, o termo Big Data foi cunhado em função da quantidade de dados a serem processados, sendo que os computadores não conseguiriam processá-los. A informação cresce quatro vezes mais rápido que a economia no mundo todo, apesar da capacidade de processamento crescer nove vezes mais rápido (MAYER-SCHÖNBERGER; CUCKIER, 2012).

Porém essa definição, em relação a quantidade de dados, mudou nos últimos anos. E hoje há várias definições para o termo Big Data. Isso aconteceu pelo fato desse termo ter diferentes utilidades dentro das diversas áreas do conhecimento.

Para a EMC Corporation [s.d.], uma empresa multinacional norte-americana que fornece infra-estrutura de informação, software e serviços:

O big data é a base para a criação de novos níveis de valor comercial. Com armazenamento, análise lógica e aplicativos integrados, o big data ajuda a impulsionar a eficiência e a qualidade, além de produtos e serviços personalizados, produzindo níveis superiores de satisfação e experiência do cliente.

Ou seja, é utilizar os dados disponíveis para alavancar o negócio, medindo esses dados para conhecê-los melhor, para então melhorá-los e agregar valor ao produto ou negócio. Por exemplo, em uma livraria física, onde os vendedores sabem quais livros vendem mais ou vendem menos; mas em uma livraria *online* é possível saber não apenas o que o cliente compra, mas sim o que ele procura, se é influenciado por promoções e resenhas e como navega no site. Essa livraria *online* desenvolve um algoritmo que prevê qual livro um cliente pode ler em seguida, baseado em suas buscas e sugestões. A livraria física, por sua vez, pode falir se não souber o que seus clientes querem. Mas para se usar o Big Data não é necessário um sistema automatizado para prever compras; a primeira livraria pode simplesmente anotar quais livros são mais vendidos, os vendedores perguntam para os clientes o que falta na livraria e o que gostariam de ver ali. Isso são dados, e é isso o que o proprietário desta livraria deve usar para melhorar seu negócio. Mas com a quantidade de informações e dados a serem manipulados, esse processo torna-se inviável.

Gartner Group [s.d.] diz que “Big Data - informações em tamanho extremo, diversificados e complicados - estão em todos os lugares. Esse fenômeno perturbador está destinado a ajudar as organizações a impulsionar a inovação, ganhando um olhar novo e mais rápido sobre seus clientes”, conduzindo à uma preocupação maior com os clientes.

Para Mayer-Schönberger e Cukier (2012):

[...] big data se refere a trabalhos em grande escala que não podem ser feitos em escala menor, para extrair novas ideias e criar novas formas de valor de maneiras que alterem os mercados, as organizações, a relação entre cidadãos e governos, etc. [...] O big data será uma nova fonte de valor econômico e inovação.

Pode-se perceber que o Big Data é uma tecnologia inovadora e de grande utilidade. É difícil imaginar para onde vão e o que é feito com os dados gerados todos os dias. Eles estão em toda parte, circulando em mais lugares do que se imagina. Os dados mais usados de uma pessoa são RG e CPF, assim como sua data de nascimento e sexo. Os planos de saúde têm esses dados e quantas vezes a pessoa vai ao médico durante o ano, fato este que pode determinar se esta pessoa lhes dá prejuízo ou não. Os cartões de crédito conseguem saber quanto o cliente ganha, se é consumista, suas preferências de compra, de onde compra, etc. Além disso, essa pessoa pode ser usuário de redes sociais e compartilhar fatos de sua vida nessas redes, curtir, compartilhar fatos, imagens e vídeos que acha interessante, manifestar seus sentimentos, o que viu, o que fez e até o que comeu. Os sites que essa pessoa visita também refletem o seu gosto pessoal. Jogos no celular, quantidade de chamadas durante o mês, conversas via SMS, Facebook e WhatsApp, fotos tiradas com o celular ou câmera pode definir o perfil de qualquer um. Multiplicando-se isso por mais de 7 bilhões de pessoas no mundo todo atualmente (ESTADÃO, 2013), conclui-se que as pessoas geram um quantidade muito grande de dados.

Porém não são apenas as pessoas que geram dados. Existe a chamada “Internet das Coisas”. Taurion [s.d.] alega que:

O mundo está cada vez mais interconectado e através da Internet não apenas as pessoas se comunicam, mas veremos mais e mais objetos “falando” uns com os outros, criando uma verdadeira Internet das coisas. A Internet das Coisas vai criar uma rede de centenas de bilhões de objetos identificáveis e que poderão interoperar uns com os outros e com os data centers e suas nuvens computacionais. A Internet das Coisas vai aglutinar o mundo digital com o mundo físico, permitindo que os objetos façam parte dos sistemas de informação. Com a Internet das Coisas podemos adicionar inteligência à infraestrutura física que molda nossa sociedade.

Qualquer tipo de aparelho que se conecta à Internet e gera algum tipo de dado faz parte da "Internet das Coisas". O acelerômetro do celular, o computador de bordo do carro onde é mostrado o consumo de combustível e a temperatura do motor, a geladeira com conexão *wi-fi*, a *smart* TV que acessa redes sociais, sensores de câmeras e de presença, o monitoramento de pressão de uma pessoa à distância, um tênis que transmite dados da corrida matinal para seu *smartphone*. Esses são alguns exemplos da "Internet das Coisas", que está em constante crescimento e tende a aumentar ainda mais.

A IBM [s.d.] alega que 90% dos dados que existe no mundo foram criados apenas nos últimos anos.

Mayer-Schönberger e Cukier (2012) dizem que:

Apesar de estarmos apenas nos primórdios do big data, nós o usamos diariamente. Filtros *anti-spam* são projetados para automaticamente se adaptarem às mudanças dos tipos de lixo eletrônico [...]. Sites de namoro formam pares com base em como suas várias características se correspondem às de relacionamentos anteriores. O corretor automático dos *smartphones* analisa nossas ações e acrescenta novas palavras em seus dicionários com base no que escrevemos

Empresas de grande porte usam o Big Data há algum tempo, como as empresas de fundo de risco, mais conhecidas como *hedge funds*, que usam redes sociais para prever o mercado de ações. A Amazon e a Netflix através do histórico de navegação de um usuário em seus sites, sugere a compra de produtos para seus clientes. O Flu Trends da Google prevê um surto de gripe a partir das buscas relacionadas ao assunto em seu site. A Farecast prevê o preço de passagens aéreas e a Targus descobriu que, quando um furacão se aproximava, um determinado doce é mais vendido que outros, fazendo com que o comércio os coloque perto de lanternas e ferramentas (MAYER-SCHÖNBERGER; CUCKIER, 2012). As empresas de pequeno e médio porte também podem, e devem, se beneficiar do Big Data. Porém são poucas as que se aventuram neste meio, e essas poucas inovam em áreas inéditas. Um bom exemplo é a BovControl na área de pecuária, que criou um aplicativo para dispositivos móveis onde o trabalhador registra a quantidade de animais da fazenda, seu peso, qual vaca produz mais leite, etc; e a partir desses

dados, o fazendeiro faz a gestão necessária, tendo uma visão mais ampla e clara de seu negócio (OLHAR DIGITAL, 2013).

Na Copa das Confederações de 2013 foi testado um novo uso do Big Data, com a criação de um algoritmo pela IBM que a partir de *tweets* e *posts* no Facebook, informavam o técnico da seleção, Felipe Scolari, sobre o que a torcida achava do jogo, e em tempo real. Com a associação da análise de dados e tecnologia de processamento de linguagem, a IBM desenvolveu o índice de sentimento social, capaz de criar *insights* instantâneos (em tempo real) sobre a opinião pública. Esta ferramenta ainda está disponível para qualquer empresa que queira saber sua reputação nas redes sociais (OLHAR DIGITAL, 2013).

A Amazon, varejista norte-americana, percebeu que poderia usar dados de seus mais de 150 milhões de usuários a seu favor. Há muitos anos eles usam os dados dos clientes para seu sistema de recomendações de produtos. No começo, os produtos recomendados eram apenas produtos parecidos com o que o cliente havia comprado: se ele comprou um livro de receitas de bolo, ele te recomendaria outros livros de receitas de bolo. Mas a Amazon procurou melhorar, e muito, esse sistema, baseando suas recomendações a partir da informação armazenada de todos os mais de 150 milhões de clientes: ao comprar o livro de receitas de bolo, a loja recomenda outros livros do mesmo tipo que muitas outras pessoas vêm comprando, ou recomenda uma forma de bolo e uma panela. É o famoso “Quem comprou isto, também comprou...” que muitas lojas de *e-commerce* passaram a usar também (BIG DATA STARTUP, [s.d.]).

E assim é possível perceber que o Big Data não é apenas uma tendência, mas sim o futuro. Com ele pode-se descobrir muitas coisas, criar relações e entender coisas que até então eram desconhecidas e criar oportunidades. Seu uso pode gerar ideias inovadoras ou simplesmente complementar as antigas, abrindo espaço para coisas que não haviam sido pensadas antes. Um dos pontos mais positivos do Big Data é o fato de ser universal, assim como a TI (Tecnologia da Informação), podendo ser usado em todas as áreas e trazendo muitos benefícios organizacionais.

2.2. OS CINCO Vs DO BIG DATA

O Big Data não é simplesmente uma quantidade imensa de dados à espera para serem usados, ele também está relacionado com Volume, Velocidade, Variedade, Veracidade e Valor desses dados, conhecidos como 5 Vs, aqui apresentados:

1. Volume: Como citado anteriormente, cria-se diariamente mais de 2 exabytes de dados (acredita-se que esse número duplica a cada 40 meses)(MATOS; BRAGA; KAWAMOTO JR.; 2013). Tudo o que é criado digitalmente entra neste montante, e quanto mais, melhor. Isso se deve ao fato de que, quanto mais dados, maior a possibilidade de encontrar informações relevantes para melhores tomadas de decisão.

2. Velocidade: Refere-se à velocidade em que os dados são criados, através de milhares de *tweets*, compartilhamentos em mídias, transações entre contas bancárias e cartões de crédito, cálculos matemáticos, etc. Em muitas situações os dados devem ser analisados em tempo real, como, por exemplo, na compra e venda de ações (CANALTECH, 2014). Outro fato que acelera a velocidade de dados é a concorrência organizacional (MATOS; BRAGA; KAWAMOTO JR.; 2013).

3. Variedade: É a composição dos dados estruturados e os não-estruturados que o Big Data usa. Dados estruturados são aqueles que podem ser colocados em tabelas e relações (CANALTECH, 2014) ou em banco de dados, e que atualmente são minoria nessa imensidão de dados. Já os dados não-estruturados são aqueles vindos de emails, fotos, vídeos, músicas, textos, sensores, *logs*, GPSs e mídias sociais. São em sua maioria gerados pelos seres humanos, e não por computadores ou máquinas, sendo assim difíceis de serem computados (SETTI, 2012). Este tipo de dado é o que existe em maior número, devido à suas características. Segundo Gartner (2014), dados não-estruturados somam mais de 85% nas coletas das empresas. Dados estruturados e não-estruturados relacionam-se com o SQL e o NoSQL, respectivamente. Segundo Moura (2013), ambos são bancos de dados, sendo que o SQL (*Structured Query Language*) lida com dados estruturados, que são dados organizados em relações e colocados em tabelas, e o NoSQL (*Not only SQL*) com dados relacionais não SQL, que, por sua vez, são organizados em

repositórios. O fato de o NoSQL não trabalhar com tabelas faz com que ele lide facilmente com vários tipos de bancos de dados, o que é importante para o Big Data e conseqüentemente para o uso do MapReduce.

4. Valor: Gera-se custos para implementar o Big Data e vê-lo funcionando, mas o quarto V não se refere a esse valor gasto. O valor do Big Data é quanto os dados valem para uma empresa. Um exemplo é o Facebook que, segundo Taurion (2013), "Seu valor não é sobre ativos fixos, mas pelo número de usuários que tem e principalmente pelo valor dos dados que ele mantém sobre estes usuários.". Ou seja, quanto mais sigilosos e importantes os dados são, maior valor eles terão.

5. Veracidade: Para que os dados tenham valor, é necessário que eles sejam verdadeiros, ainda mais em um ambiente em que a grande maioria de informações vêm de dados não-estruturados, entre eles *posts* do Facebook e blogs. Com isso é preciso ter certeza, ou uma margem aceitável de erro, que os dados façam sentido.

2.3. SEGURANÇA DA INFORMAÇÃO DO BIG DATA

É impossível não levar em consideração o tamanho, o poder e a versatilidade do Big Data. A implantação da metodologia pode ser a chave para muitas áreas do conhecimento, mas pode também levar essas áreas ao insucesso, pois, ao mesmo tempo que ele traz boas ideias, seu uso pode ser desastroso. O uso do Big Data dá uma grande sensação de poder, simplesmente pelo fato de armazenar uma grande quantidade de dados, principalmente se for dados pessoais.

Conclui-se que não há limites para o Big Data, pelo menos por enquanto, e ele pode ser de grande utilidade para quem o usar com inteligência. Mas o perigo está aí. O principal problema do Big Data não é seu custo de implementação ou a falta de profissionais na área, os chamados Cientistas de Dados, mas sim a liberdade e facilidade em minerar dados alheios. Com essa nova era, a principal questão é o quanto as empresas e governos vão tratar esses dados e até que ponto as leis de privacidade vão proteger os cidadãos.

É claro que nem todos os dados são pessoais. Os dados estruturados, que provêm de máquinas e computadores, também oferecem riscos. Aqui, os dados

gerados por máquinas podem ter grande valor para a empresa, fazendo com que certos produtos sejam mais eficientes que os do concorrente, por exemplo, e necessitam de cuidados. Mas como mencionado anteriormente, a grande maioria dos dados são não-estruturados, aqueles que provém de emails, redes sociais, fotos e vídeos em sua maioria. E são as organizações que usam esse tipo de dado que põem em risco a privacidade das pessoas, principalmente pelo fato de, acordo com Mayer-Schönberger e Cukier (2012), “[...] as empresas têm vários incentivos para captar ainda mais dados, mantê-los por mais tempo e reutilizá-los com mais frequência”. Então não é simplesmente coletar dados de pessoas que estão falando nas mídias sociais sobre uma marca, usar e pronto. Eles coletam esses dados, os guardam para possíveis futuras análises, e continuam coletando dados sobre essas mesmas pessoas, caso mencionem a marca novamente. Ou seja, há monitoramento o tempo todo e por vários motivos.

E é neste cenário que a Segurança da Informação torna-se extremamente importante. Este ramo da TI caracteriza-se por cuidar da informação. Ferreira [s.d.] a descreve da seguinte maneira:

Segurança da Informação está relacionada com proteção de um conjunto de dados, no sentido de preservar o valor que possuem para um indivíduo ou uma organização. São características básicas da segurança da informação os atributos de confidencialidade, integridade e disponibilidade, não estando esta segurança restrita somente a sistemas computacionais, informações eletrônicas ou sistemas de armazenamento. O conceito se aplica a todos os aspectos de proteção de informações e dados.

A Segurança da Informação tem três propriedades, citadas pelo autor: confidencialidade, integridade e disponibilidade. Confidencialidade não é sobre uma informação ser um segredo, mas sim sobre poder ser acessada apenas por quem tem direito a acessá-la, como o seu dono e as pessoas autorizadas por ele. A integridade refere-se ao objetivo de manter a informação íntegra, armazenando-a com as mesmas características originais. E por fim, a disponibilidade, que é manter a informação sempre disponível para ser acessada no momento que pessoas autorizadas necessitarem.

É interessante notar que a Segurança da Informação não é apenas sobre dados. Este ramo dedica-se tanto à segurança lógica de dados quanto a física, e também quanto a informações digitais e impressas. Ou seja, engloba-se todo e

qualquer tipo de informação. O Big Data faz o mesmo, englobando todo tipo de informação.

Mas o Big Data não é o vilão da história, assim como nenhuma tecnologia. O vilão é o poder, o poder que as organizações e governos querem ter por meio dessas tecnologias. Taurion (2013) afirma que:

[...] toda tecnologia é neutra por si, o uso que fazemos dela é que pode ser positiva ou negativa. Uma empresa pode usar a imensidão de dados que dispõe sobre seus clientes para uma promoção que melhora significativamente a experiência de seus clientes, como um governo com tendências ditatoriais pode usar os dados para controlar a vontade política de sua população.

As intenções de quem tem a tecnologia ao seu favor que são o problema.

A ética é uma grande questão relacionada à privacidade. A maioria das pessoas se sentem inseguras com as propagandas patrocinadas nas redes sociais, como no Facebook e nos serviços da Google. O jornal O Globo (2012) exemplifica muito bem esse tipo de situação:

Você vai ao mercado para comprar apenas o que falta para o jantar e, ao passar pelo corredor de produtos de higiene, seu celular o surpreende com uma mensagem. O remetente é o próprio varejista, que deseja atentá-lo para o desodorante em promoção na prateleira ali do lado. O SMS não diz, mas ele sabe que o seu estoque do produto está mesmo no fim e que, há duas semanas, você escreveu no Facebook o quanto gostava daquela marca.

Neste caso foram usados dados do GPS (*Global Positioning System*) do celular da pessoa e suas redes sociais. Mesmo se o perfil dessa pessoa no Facebook for privado, como o mercado pôde ver o que foi dito e usá-lo? Nesta nova era é preciso rever os princípios da ética e da privacidade, pois eles se tornam obsoletos com o passar do tempo. Taurion (2013) diz que o foco neste momento é saber se as leis e conceitos que temos hoje sobre a privacidade ainda valem, e se mudaram, as soluções que usamos talvez não façam mais sentido ao serem aplicadas. O conceito de privacidade tem que ser redefinido.

Para Mayer-Schönberger e Cukier (2012), a privacidade funciona a partir de três estratégias: consentimento individual, opção de exclusão e anonimização. E segundo os autores, essas estratégias perderam sua eficiência. Eles dizem que o consentimento individual é um problema, pois avisam quando seus dados estão

sendo coletados e o porquê, você aceita ou não, e então os dados começam a ser coletados; em tempos de Big Data, os autores questionam: “Como as empresas podem informar sobre um objetivo que não existe? Como as pessoas podem consentir em dar informações para o desconhecido?”. Esses modelos de “saber e consentir” precisam ser reavaliados, ou até mesmo substituídos. Taurion (2013) ainda completa dizendo que não dá para saber nem mesmo quais empresas coletariam os dados. Isso se deve ao fato de que uma empresa consegue armazenar os dados relativos ao tipo de compra que o cliente faz pela Internet, por exemplo, e uma segunda empresa, em vez de coletar dados novos, compra os dados já coletados da primeira empresa, mas para um outro objetivo. É mais fácil, barato e rápido. Mas ao aceitar os termos da primeira empresa, será que havia algo sobre permitir que uma outra empresa pudesse usar os mesmos dados? Se sim, a pessoa leu? Ambos os autores criticam este método de “saber e consentir” pois ninguém lê antes de aceitá-lo; Mayer-Schönberger e Cukier (2012) ainda complementam dizendo que:

Na prática, esse conceito gerou contratos gigantescos, raramente lidos e muito menos compreendidos [...] A alternativa, isto é, pedir aos usuários que concordem com qualquer uso futuro dos dados coletados, tampouco é útil. A permissão ampla castra a própria noção de consentimento informado. No contexto do big data, o testado e confiável conceito de conhecimento e consentimento é geralmente restritivo demais para se extrair o valor latente dos dados ou vazio demais para proteger a privacidade das pessoas.

Mas não é apenas o consentimento individual que está em perigo. Outro grande problema com o crescimento em grande escala do Big Data é a anonimização dos dados. Mayer-Schönberger e Cukier (2012) definem a anonimização como sendo o ato de apagar todos os tipos de dados pessoais em um banco de dados (nome, endereço, IP, etc). Porém essa estratégia simplesmente não é mais totalmente segura e efetiva em tempos de Big Data, ou seja, mesmo tendo os dados pessoais dos clientes apagados de um banco de dados, ainda é possível encontrar a pessoa. Isso pode ser completamente contraditório. Ohm (2009) apud Mali (2013) diz que “(...) temos tantos dados e os recombina tanto, que é relativamente fácil para especialistas recuperar informações pessoais a partir do cruzamento entre bases de dados.”. Isso significa que, cruzando um dado anônimo de uma base de dados com outra base de dados, é possível descobrir a identidade

real de uma pessoa. É como um quebra-cabeças, onde, juntando as peças certas, descobre-se o desenho.

Mali (2013) ainda descreve dois casos que a mídia acompanhou e que não poderiam exemplificar melhor o quanto a anonimização não é mais tão segura. Um deles é da AOL, onde a empresa divulgou uma lista anonimizada de centenas de milhares de usuários que usavam seus serviços para o uso de pesquisadores. Esta lista continha números no lugar dos nomes. Porém, em uma determinada reportagem, dois jornalistas pegaram um destes números da lista e descobriram a identidade da pessoa. Ao ver que o número realizou buscas sobre “homens solteiros com mais de 60”, “paisagistas em Lilburn” e indivíduos que tenham Arnold no sobrenome, ambos os jornalistas descobriram se tratar de uma senhora de 62 anos, viúva, chamada Thelma Arnold e que residia em Lilburn, (Geórgia, EUA). Com isso, a AOL demitiu três funcionários e fez o que pôde para retirar esta lista da internet.

Outro caso descrito por Mali (2013) é em relação à Netflix, que também liberou uma lista anonimizada de centenas de milhares de usuários e seus dados em relação à locação de filmes. A lista foi divulgada em um concurso onde programadores a usavam para melhorar o sistema de recomendações de filmes. Mas apesar dos dados estarem anonimizados, uma usuária mãe de família e secretamente lésbica foi encontrada. Ao ser identificada, ela processou a empresa temendo que sua privacidade fosse quebrada. Com isso, pesquisadores usaram essa mesma lista em um experimento e também conseguiram encontrar várias pessoas, comparando as notas anônimas nos mesmos filmes do serviço e do site *Internet Movie Database* (IMDb) com uma precisão de 84%.

Ou seja, encontrar uma pessoa em meio a tantos dados anônimos é completamente possível. Taurion (2014) diz que todas as pessoas deixam “pegadas digitais” o tempo todo. Ele diz que os *apps* usados (por exemplo, GPS do *smartphone*), os *posts* no Facebook, *tweets*, as buscas na Google, emails, compras na Internet e o uso dos cartões de crédito deixam pegadas. Muitas pessoas podem achar que isso é uma imensa falta de privacidade, mas essas mesmas pessoas devem lembrar que esse “rastreamento” não acontece apenas no mundo virtual, mas também no mundo real, como lembra Taurion (2014) ao dizer que todas as pessoas são filmadas várias vezes ao dia por câmeras de vídeo que estão em vários lugares,

“Como se manter anônimo com tanta pegada digital deixada a cada dia?”, ele questiona. Ele ainda diz:

[...] na prática nos acostumamos com a quebra de privacidade em troca de segurança (passamos pelos raios-x nos aeroportos sem reclamar e vemos as placas de 'sorria, você está sendo filmado' nas lojas e nem por isso deixamos de entrar nelas) e funcionalidade, ao usarmos apps, plataformas sociais e outros serviços na web. Talvez o próprio critério de privacidade deva ser repensado.

Com tudo isso em mente, é fato que é necessário repensar a privacidade. E mais além, precisa-se reeducar as pessoas para essa nova era que está chegando. O Big Data é como o petróleo, que oferece grandes avanços, mas também grandes problemas. É incontestável o fato de que ele faz parte da nova revolução tecnológica, porém é uma tecnologia que expõe os indivíduos de várias maneiras, obrigando-os a abrir os olhos para os perigos que a ganância traz para quem retém uma grande quantidade de dados, principalmente os pessoais. Que comerciante não ficaria extasiado ao saber que pode prever se seu novo produto será bem aceito entre os consumidores ou não? O Big Data beira a “clarividência” ao prever o comportamento de máquinas e pessoas, fato que até pouco tempo atrás era quase impossível. Essa “clarividência” conduz a outro ponto tão perigoso quanto a falta de privacidade citada anteriormente: definir ações através da previsão do comportamento humano.

Segundo Mayer-Schönberger e Cukier (2012), a tecnologia chegará a um ponto onde os dados tomarão conta da justiça, onde eles serão os juízes. Segundo eles, teremos uma realidade muito próxima à do filme *Minority Report* (2002), o filme de Steven Spielberg que mostra uma realidade onde a polícia retém criminosos antes mesmo de cometerem o crime, com a ajuda de paranormais que preveem quando o crime irá acontecer e quem está envolvido. É claro o paralelo que há entre o filme e o Big Data, onde a tecnologia entra no lugar dos paranormais, prevendo o comportamento de criminosos para a polícia. A questão é que, em uma sociedade como a do filme, as pessoas não são presas pelo crime que cometeram, mas sim pelo crime que iriam cometer. Os autores dizem que “[...] o filme retrata um perturbador futuro, no qual a descuidada análise do big data ameaça se tornar realidade, um futuro no qual os julgamentos de culpa se baseiam em previsões individualizadas de comportamento”.

Na verdade, os primeiros passos já estão sendo dados e a polícia está se beneficiando com esse tipo de análise preditiva. Mayer-Schönberger e Cukier (2012) dizem que mais da metade dos estados nos EUA usam a análise de dados em departamentos de liberdade condicional para decidir se um preso deve ser solto ou não. Outro exemplo que os autores dão é de uma cidade da Virginia onde a polícia cruzou vários bancos de dados (incluindo dados criminais, datas de pagamento de grandes empresas e eventos) que confirmaram várias tendências que os policiais suspeitavam em relação a criminalidade na cidade. A polícia suspeitava, por exemplo, que havia a tendência de crimes violentos aumentarem depois de feiras de armas e, após a análise dos dados, descobriu-se que a suspeita estava correta, e mais, os dados revelaram que o surto de crimes aumentava depois de duas semanas do evento, e não logo depois. Assim é possível que a polícia se organize e contenha o surto de violência na cidade.

Muitas pesquisas estão sendo realizadas em relação a análise preditiva. De acordo com Mayer-Schönberger e Cukier (2012) foi desenvolvido um projeto de pesquisa do *U.S. Department of Homeland Security* (Departamento de Segurança Interna dos Estados Unidos) que se aproxima ainda mais da realidade do filme *Minority Report*. O projeto chama-se FAST (*Future Attribute Screening Technology* - Tecnologia de investigação de Atributos Futuros) e seu objetivo é identificar supostos terroristas baseando-se no monitoramento de sinais vitais, linguagem corporal e padrões fisiológicos. Os cientistas esperam que, ao monitorar todas essas variáveis, seja possível prever a intenção da pessoa de fazer algo. Segundo o departamento, as pesquisas têm 70% de acerto nos testes realizados. Porém os autores questionam esses testes ao perguntar se essas pessoas testadas eram atores. E essa dúvida não deixa de estar errada pois não se sabe se o comportamento de um ator seria o mesmo de um terrorista real diante dos monitores de análise ou se as variáveis que denunciam o ato sejam as mesmas. Os autores complementam dizendo que “Apesar de os sistemas parecerem embrionários, a questão é que a polícia os leva bem a sério”.

Lidar com esse tipo de tecnologia é perigoso, pois o ser humano tende a querer cada vez mais o poder, e a influência que isso causa pode trazer severas consequências para a sociedade como um todo. Possuir uma tecnologia que pode prever não só os crimes, mas também se uma pessoa vai fazer algo errado é algo

extraordinário que tem tudo para dar certo. Porém, isso é um pensamento equivocado.

Com o Big Data na área criminal, pode-se prever crimes que acontecerão no futuro, prever quem serão os infratores, para então evitar que o ato aconteça. Mas, como bem exemplificado por Mayer-Schönberger e Cukier (2012), há a possibilidade de que não haja contentamento em simplesmente impedir que tal crime aconteça. É possível que, além de impedir que o ato aconteça, haverá o desejo de punir o provável infrator antes de fazer qualquer coisa. E isso é aparentemente óbvio de se pensar, pois, se apenas evitar que o crime aconteça, o infrator poderá tentar novamente, ser impedido, tentar outra vez e assim sucessivamente; sendo que, com o Big Data, pode-se detê-lo pelos seus futuros atos. Exatamente como sugerido no filme.

Apesar desse cenário parecer ficção, Mayer-Schönberger e Cukier (2012) fazem um paralelo com o que a sociedade tem atualmente. Eles dizem que essa futura realidade do Big Data é apenas a evolução do que já acontece:

Punições com base em previsão parecem um aprimoramento das práticas já aceitas. A prevenção de comportamentos não saudáveis, perigosos ou ilegais é uma das bases da sociedade moderna. Tentamos reduzir a liberdade de fumar para evitar o câncer; exigimos o uso de cinto de segurança para evitar fatalidades em acidentes de carros; não permitimos que as pessoas andem armadas no aviões para evitar sequestros. Medidas preventivas diminuem nossa liberdade, mas muitos as veem como um pequeno preço a se pagar para evitar um mal maior.

O problema será conter o avanço desenfreado e impedir que o Big Data tome decisões pelas pessoas. A exatidão em colher e analisar os dados poderá deixar as pessoas mais “preguiçosas”, deixando que o sistema tome a decisão em várias situações. Há situações em que isso não é um problema, como em relação à troca de uma peça em um motor de carro, por exemplo. O sistema analisa que uma peça essencial está com problema e decide pela troca, então o dono vai trocar a peça sem questionar. Mas como lidar em situações como a descrita anteriormente, em que o sistema usa a análise preditiva para decidir se um preso deve ser solto ou não? E se, nessa situação, vê-se que o preso tem um bom comportamento e que está pronto para voltar à sociedade, mas o sistema barra e diz que ele não deve ser solto? Dá-se uma segunda chance a ele ou coloca-o novamente na cela? Em

situações como essa, deve-se impor limites para que o Big Data não tome decisões sem consentimento. Em relação a isso, Mayer-Schönberger e Cukier (2012) concluem que “O problema é nos deixarmos guiar pelo resultado das análises mesmo quando tivermos motivos para suspeitar que haja algo de errado [...]. Ou ainda que atribuamos um grau de verdade que os dados não mereçam”.

Sendo assim, o uso do Big Data deve ser muito cuidadoso. O fato dele prever coisas e ser “inteligente” não o faz melhor que um ser humano com uma grande carga de experiência. Deixar que ele decida coisas pelas pessoas é extremamente errado. O que ele pode fazer é mostrar estatísticas que podem fazê-los mudar de ideia, mas jamais decidir algo. O Big Data é uma tecnologia que, por ter um grande potencial, pode ser usada para coisas erradas. Cabe-se evitar isso, regulando-o e controlando-o para que seja usado da maneira correta.

2.4 CONTROLE DO BIG DATA

Com o uso do Big Data pode-se gerar novas ideias, mostrar coisas que antes não haviam sido vistas ou pensadas, diminuir custos, aumentar lucros, facilitar algumas coisas e melhorar outras, e muitas outras situações. Mas assim como todas as tecnologias, o Big Data tem seu lado negativo e, neste caso, um lado preocupante. Diz-se isso pois o Big Data trabalha com todos os tipos de dados, e entre eles os dados pessoais e confidenciais. O uso de dados de alto risco traz perigos, como apontado anteriormente. O Big Data pode acabar com a privacidade e até mesmo prever comportamentos. A questão é como lidar com a intrusão da tecnologia na vida das pessoas e se é possível evitar isso.

Conclui-se que isso é possível, mas não será fácil. Não será fácil pois haverá a necessidade de mudar a maneira de pensar, mudar “tradições”. Em seu livro, Mayer-Schönberger e Cukier (2012) têm algumas ideias interessantes em relação ao controle do Big Data, sendo em sua maioria em relação à segurança da informação.

Primeiramente, a privacidade deve ser totalmente remodelada. A privacidade como é conhecida hoje não protege mais os dados, é extremamente falha. Não há como ter quando há observação constante por parte de empresas e governos, que terão lucro e informações a partir da análise do que se fala e do que se faz. O princípio de privacidade mudou ao longo das décadas, e deve mudar novamente na

era do Big Data. Durante muito tempo, colocar as pessoas no controle da situação era o princípio essencial das leis em relação à privacidade. Com as pessoas no controle, elas podiam decidir se seus dados poderiam ser processados, por quem seriam coletados e como seriam usados. Essa solução era válida apenas em um mundo de poucos dados, onde era possível saber quem poderia usá-los e para quê, o que é impossível nos dias de hoje. Mas o mundo mudou, e quando a Internet chegou às massas o princípio de deixar a pessoa no controle já era impossível de ser aplicado. Não é possível cada empresa sentar e conversar com todos os usuários para falar como os dados serão usados. Criou-se, então, o sistema de “leitura e consentimento”, onde a empresa mostra todos os seus termos e condições antes da pessoa instalar seu programa ou cadastrar em seu site. Os autores chamam esse sistema de “imediatista”, e não estão errados, pois ao se deparar com os termos, a grande maioria das pessoas imediatamente os aceita, sem nem mesmo lê-lo. E como mencionado anteriormente, isso pode ser um problema. Esses documentos são muito longos e técnicos, então é normal que um usuário final não se interesse em ler. E ao não ler e consentir com os termos da empresa, a pessoa pode estar aceitando algo que não queira. E agora o mundo está novamente mudando. De acordo com Mayer-Schönberger e Cukier (2012), com o Big Data nada disso é válido, nenhum desses princípios de leis de privacidade funcionam mais. Não funcionam pois o principal uso é o uso secundário dos dados, e não a coleta dos mesmos. O valor dos dados não está mais no momento em que se pede autorização para coletá-los (uso primário), mas sim o que a empresa vai fazer com eles depois disso (uso secundário). Dar o poder de escolha para o usuário, se pode coletar, quem coleta e para quê, não é mais válido pois não há como saber quem irá usar seus dados e como eles serão usados. O usuário pode consentir para a empresa X usar seus dados para uma finalidade, mas uma outra empresa, a Y, pode comprar esses dados da empresa X e usá-los para alguma outra finalidade. O mesmo vale para o “saber e consentir”, onde não é possível aceitar a coleta de dados para empresas e finalidades que não se sabe quais serão no momento da coleta.

Em relação a este problema, os autores dizem que “Imaginamos uma mentalidade de privacidade bem diferente para a era do big data, menos focada no consentimento individual na hora da coleta e mais na responsabilização dos

usuários por seus atos”. Em um mundo como imaginado, administrar corretamente o Big Data e usar as ideias descobertas de maneira certa, trás vantagens palpáveis para as empresas que usam dados (os usuários de dados), onde não será preciso buscar o consentimento dos usuários para os usos secundários dos dados pessoais. Mas a falta de proteção dos dados em más implementações trazem riscos para os usuários, em que eles podem ser expostos a leis, multas e processos. “A responsabilidade dos usuários de dados só funciona se tiver força” (MAYER-SCHÖNBERGER; CUCKIER, 2012). Transferir a responsabilidade de privacidade para os usuários de dados faz sentido, pois eles sabem exatamente o que fazer com essas informações e dados coletados. Ao fazer uma avaliação desses mesmos dados, também evita-se que estratégias confidenciais da empresa sejam divulgadas. E principalmente o fato de que, aproveitando o uso secundário desses dados, é simplesmente justo ser responsabilizados por eles, já que aproveitam o seu máximo.

Para que isso tudo funcione, deve-se criar uma classe reguladora de dados, onde esses reguladores monitoram as empresas e o que elas fazem com os dados. Dentre suas obrigações, está a de classificar os dados de acordo com sua importância, onde eles decidem o quanto tais dados podem ser reutilizados e avaliam o seu risco. Essa limitação de tempo que as empresas terão para usar os dados, as incentivam a usá-los antes de sua perda. Mas não é só isso, esses usuários de dados deverão ter suas obrigações, como a de se responsabilizar pelo uso dos dados e apagar os dados pessoais após um determinado tempo (que o regulador definirá).

Mas essas estratégias não protegem a privacidade em todas as situações, e em relação a isso, os autores definem o que chamam de “privacidade diferencial”. Esse tipo de privacidade é basicamente usar um grande banco de dados de dados pessoais e as pesquisas retornarem um resultado aproximado, onde não é possível associar a pesquisa a uma única pessoa. Um exemplo prático que eles dão é o Facebook, que divulga dados pessoais para anunciantes em potencial, onde os números são apenas aproximados: “Procurar por mulheres asiáticas em Atlanta interessadas em ioga Ashtanga gerará um resultado de ‘aproximadamente 400”, não um número exato, tornando impossível que se use essa informação para encontrar um alvo específico”. Ou seja, identidades pessoais não são reveladas nesse tipo de busca, que é o objetivo da privacidade diferencial.

Lidar com a privacidade, ou a falta dela, nessa nova era não é o único problema. Outra grande questão é como vamos lidar com a justiça nesses tempos de Big Data, que também deverá mudar e ser atualizada. O sistema jurídico funciona de uma maneira relativamente simples, responsabilizando as pessoas pelo o que elas fizeram após avaliar seus atos. Mas como a justiça irá funcionar em um mundo onde sistemas preveem o que as pessoas podem fazer? A justiça deixará de julgar os atos que as pessoas fizeram para julgar o que elas poderão fazer? Não. Segundo os autores, o governo deverá responsabilizar as pessoas pelo o que elas fizeram no passado, e não por previsões do que elas podem vir a fazer. E ao julgar por ações passadas, o governo não poderá se basear apenas em sistemas de Big Data, em que seus resultados podem interferir no julgamento. É fácil entender o motivo do Big Data estar fora de decisões judiciais. Com ele e suas previsões, a sociedade vai querer basear suas decisões apenas em estatísticas, deixando de responsabilizar as pessoas. Os autores insistem que a responsabilidade individual deve ser protegida, pois é somente com ela que os atos devem ser julgados. Um exemplo claro dessa situação seria usar o Big Data para encontrar possíveis criminosos, usando algum algoritmo que mostra se a pessoa tem probabilidade de cometer crimes ou não, e caso sim, colocá-la em tratamentos especiais para conter essa “criminosidade”. Tudo isso em nome da contenção de riscos. A sociedade pode não ver muitos problemas nesse caso, pois não se importa com pequenos sacrifícios para evitar riscos maiores (como quando são proibidos de fumar para evitar o câncer de pulmão). Mas o problema não é a sociedade, mas sim as pessoas que estão em situações como a mencionada. Ao ser qualificada na categoria “possível criminoso” e induzida a tratamentos, a pessoa se sentirá punida por algo que não fez (ainda) e verá a intervenção como um tipo de castigo. E pior, seus amigos e familiares podem fazer julgamentos a partir disso e mudar seu comportamento em relação a essa pessoa, até mesmo distanciando-se, piorando a situação como um todo. “Quanto mais deixamos de responsabilizar as pessoas por seus atos em nome de intervenções com base em dados para reduzir os riscos na sociedade, mais desvalorizamos o ideal de responsabilidade individual” (MAYER-SCHÖNBERGER; CUCKIER, 2012). O Big Data não pode ser colocado em um pedestal, onde a verdade está contida apenas em seus resultados. Caso contrário, a responsabilidade pelos atos que cada indivíduo tem e o livre-arbítrio serão

esquecidos e viveremos em uma ditadura governada pelo sistema de previsões. Mayer-Schönberger e Cukier (2012) dizem que:

Um pilar fundamental da governança com big data deve ser uma garantia de que continuaremos a julgar as pessoas pela responsabilidade pessoal e pelo comportamento de fato, não pela análise 'objetiva' dos dados para determinar se são criminosas em potencial. Somente assim as trataremos como seres humanos: como pessoas que têm liberdade de escolha sobre suas ações e o direito de serem julgadas por elas.

Porém, dizer que o Big Data será controlado apenas com essas novas soluções é errado. De acordo com os autores, deveremos evitar o que chamam de "barões dos dados". Esses barões nada mais são que as empresas que retêm uma grande quantidade de dados, como o Facebook e a Google, por exemplo, assim como os barões da borracha nos EUA no século XIX. A solução aqui é adaptar as leis para que não haja o monopólio por parte de empresas que trabalham com grandes quantidades de dados, e as usam em suas tecnologias. A ausência de leis e debates com os usuários aumenta o risco de que empresas e governos ditem o que pode ou não ser feito em relação aos dados. Evitar o monopólio de tecnologias baseadas em Big Data e concorrência desleal seriam os principais objetivos dessas novas leis.

A privacidade, no entanto, é algo que já começou a mudar. Não evoluiu o suficiente, mas as empresas dedicam um pouco mais de seu tempo para melhorar esse quesito para seus clientes. Principalmente pelo fato de que, segundo Rodotà (1995) apud Doneda [s.d.], a privacidade não é mais relacionada ao eixo "pessoa-informação-segredo", mas sim pelo eixo "pessoa-informação-circulação-controle". Ou seja, para a privacidade evoluir de fato, deve-se focar no seu controle, pois não se trata mais de um segredo, e sim de informações. E as empresas e governos que entenderem esse fato e buscarem a privacidade para as pessoas, serão bem-sucedidas. Um exemplo é de empresas como Apple, Facebook, Google e Microsoft que passarão a avisar seus clientes quando o governo dos EUA solicitar os dados que elas recolheram (LANDIM, 2014). Quanto mais a sociedade solicitar direitos de privacidade, menor será o comércio de dados e maior será o controle da sociedade sobre o poder abusivo relacionado aos dados (SILVEIRA, 2014).

Os próximos anos nessa era do Big Data deverão ser de muitas mudanças. Mudanças em conceitos, regras, leis e tradições que a sociedade construiu com o passar do tempo. Assim como o controle do fogo, a roda, o combustível, a imprensa e tantas outras tecnologias importantes que o ser humano descobriu e que mudou o mundo em que vivemos, não devemos deixar que elas fiquem maiores que o próprio ser humano, guiando e moldando o futuro (MAYER-SCHÖNBERGER; CUCKIER, 2012).

2.5. FUNCIONAMENTO DO BIG DATA

Acredita-se que o Big Data é uma tecnologia que terá vida longa, pois mostrou-se uma valiosa ferramenta para os negócios e para a TI, fazendo com que muitas organizações se interessassem em sua implantação.

O Big Data, além de uma ótima metodologia para os negócios, é uma grande oportunidade para a TI e seus serviços. É difícil falar sobre ele e não mencionar Hadoop e MapReduce, que são as principais e mais importantes ferramentas que fazem o Big Data funcionar.

Hadoop é um *framework open-source* criado pela Yahoo em 2005 (TAURION, 2012), que processa e armazena a grande quantidade de dados que recebe por meio de um *cluster* de computadores, sendo licenciado pela Apache (REZENDE, 2014). Um *framework* é, basicamente, uma coleção de partes de códigos que podem ser utilizados em vários outros códigos. Ao desenvolver aplicações e perceber que elas têm algo em comum, é possível fazer o uso de um *framework*, criando um código genérico e comum a todas elas, e depois adicionando as particularidades de cada uma. Müller (2008) exemplifica bem um uso de *framework* dizendo que "Um ótimo exemplo é um formulário de login, muitos sites utilizam, por que você não cria um formulário genérico e reutiliza da mesma forma para os sites que for desenvolver, afinal login aqui e em qualquer lugar do mundo será o mesmo". *Open-source* por sua vez, é, segundo Koch [s.d.]: "Projetos de Open Source são aqueles com a finalidade de criar e manter softwares que incluem o código fonte para seus usuários. Dependendo da licença usada, ao baixar e instalar um desses softwares, o usuário ganha o direito de usá-lo para qualquer fim, modificá-lo e redistribuí-lo.". Então, pelo fato do Hadoop ser um *framework open-souce*, ele é totalmente customizável para atender necessidades específicas das organizações que o usarem. O MapReduce, por sua vez, é um *software* criado pela Google que acelera as pesquisas em seu navegador (TAURION, 2012).

Apesar de ambas as ferramentas estarem sempre relacionadas ao grande volume de dados do Big Data, elas não existem apenas para servir ou que o Big Data depende exclusivamente dessas ferramentas para existir. O que acontece é o fato de que o Hadoop foi o responsável por trazer o MapReduce à superfície do

conhecimento (GASPAROTTO, 2014) e o Big Data demanda serviços que as duas ferramentas oferecem, e casaram-se muito bem. Os três, o Big Data, o Hadoop e o MapReduce, coexistem em perfeita harmonia, mas não dependem um do outro.

O Hadoop é composto por alguns módulos, segundo Rezende (2014):

- Hadoop *Common*: é onde estão as bibliotecas e utilitários usados em módulos do Hadoop;
- Hadoop *Distributed File System* (HDFS): é um sistema de arquivos que armazena os dados e usa *hardware* barato em um *cluster* de computadores;
- Hadoop YARD: administra os recursos que são responsáveis pela gestão de recursos dos *clusters* e o uso destes mesmos recursos no agendamento de aplicações para os usuários;
- Hadoop *MapReduce*: modelo de programação que processa uma grande quantidade de dados.

Dentre esses quatro módulos do Hadoop, o Big Data funciona com dois principais, o HDFS (*Hadoop Distributed FileSystem*) e o HMR (*Hadoop MapReduce*).

O HDFS é um sistema de arquivos que armazena arquivos muito grandes e de forma distribuída. Ele foi criado a partir do conceito “*write-once, read-many-times*” (“escreva uma vez, leia várias vezes”). É exatamente o que o Hadoop precisa, pois os dados são processados muitas vezes, mas, geralmente, são escritos uma vez apenas. O que diferencia o HDFS dos sistemas de arquivos comuns é o fato de que ele armazena arquivos grandes, pois ele em si, é maior que os outros sistemas de arquivos, sendo que um bloco de dado do HDFS tem 64MB e enquanto um comum tem apenas 512 bytes. Ele tem um *log* que edita dados, permitindo um controle maior dos dados do sistema. Ele é um sistema de arquivos como os outros, que tem permissões de escrita, leitura e execução para os arquivos, que podem ser modificados pelos seus respectivos donos e pelo administrador do sistema. Em relação ao seu funcionamento, ele precisa de um *namenode* e de um ou mais *datanodes*. O *namenode* é o mestre que comanda o sistema de arquivos e que sabe quais blocos de dados cada *datanode* tem. Os *datanodes* são os trabalhadores que guardam e recuperam blocos de dados apenas quando o mestre pede e o enviam relatórios periódicos sobre quais blocos estão armazenando (GASPAROTTO, 2014).

O HMR, por sua vez, é indispensável para o *framework*, pois é ele o responsável pelo processamento dos dados. O que o torna tão importante é o fato de que seu processamento é em paralelo, fazendo com que o programador não precise de esforço extra para programar e garantir o processamento desses dois. Isso faz do MapReduce a melhor ferramenta para processamento de dados em grande escala. O próprio nome revela as duas fases de seu processamento: o Map seguido do Reduce. O Map é a primeira fase, onde os dados que entram são mapeados e processados. Depois os dados vão para a fase Reduce, onde são reduzidos como entradas para, então, serem enviados para os arquivos que armazenarão os resultados obtidos nessas duas fases. Um fato interessante é que após essas fases o Hadoop comanda a situação internamente ao escalonar os processos de forma completamente independente, onde nem mesmo o programador sabe como é feito o processo. O que torna o MapReduce tão importante para o Big Data é que ele tem a capacidade de processar qualquer tipo de arquivo, até mesmo arquivos de texto, desde que tenha-se uma *stream* de dados de entrada que os transforme em pares chave/valor. No MapReduce a entrada da aplicação é a lista de pares chave/valor, onde cada um dos pares são processados e transformados em pares de chave/listas de valor. Essa nova lista é a lista de entrada para a segunda parte do processamento do HMR, onde são reduzidos e agregados para gerarem arquivos de saída (GASPAROTTO, 2014).

Com a definição dos dois mais importantes componentes do Big Data, o Hadoop e o MapReduce, é fácil entender o motivo de estarem sempre juntos: ambos facilitam, e muito, o trabalho no Big Data. Não apenas nos dados em si, mas também na parte física, já que o Big Data não é apenas um servidor de banco de dados, mas sim um *cluster* de computadores. Um *cluster*, segundo Pitanga (2003), é:

[...] um sistema que compreende dois ou mais computadores ou sistemas (denominados nodos) na qual trabalham em conjunto para executar aplicações ou realizar outras tarefas, de tal forma que os usuários que o utilizam tenham a impressão que somente um único sistema responde para eles, criando assim a ilusão de um recurso único (computador virtual).

O Hadoop, apesar de ser complexo para os desenvolvedores o aplicarem (TAURION, 2012), ele facilita a administração do *cluster* onde todos os dados são

processados. Outra vantagem é o fato dele ser *open-source* e, assim como as distribuições Linux, pode ser configurado para se adequar melhor a determinados tipos de uso (GASPAROTTO, 2014). Bednarz (2012) afirma que o Hadoop é atrativo pois "As organizações podem coletar mais dados e realizar análises que antes não eram práticas por causa do custo, complexidade e falta de ferramentas."

O MapReduce, por sua vez, é o que faz o Big Data ser tão versátil. Ele pode fazer qualquer coisa que o desenvolvedor queira (GASPAROTTO, 2014), adaptando-se tanto aos dados estruturados como os não-estruturados, atingindo perfeitamente ao seu objetivo junto ao Big Data.

Visto como é o funcionamento do Big Data, é importante salientar as diferenças em relação ao Data Warehouse, pois ambos são frequentemente confundidos. Os dois sistemas são muito diferentes entre si. O primeiro, como definido há pouco, é um sistema extremamente versátil e veloz, que é o contrário do segundo. Data Warehouse é definido como um sistema de organização de dados em modelo estruturado, que é uma primeira diferença. O Big Data trabalha com dados estruturados e, principalmente, os não-estruturados. O Data Warehouse é usado quando é necessária a precisão e qualidade dos dados (PELOSINI, 2012). Ou seja, é necessário um projeto longo e detalhado para a construção de um Data Warehouse, pois ele é lento e não volátil (por usar apenas dados estruturados) (TAMAKI, 2012). Em português seu nome é traduzido como "armazém de dados" ou "depósito de dados", o que demonstra o que ele faz. O Data Warehouse é usado para o armazenamento de dados relacionados ao que a organização faz para, então, ajudar em decisões dos executivos. Outra característica é que por ser não volátil, dados são sempre adicionados ao sistema e nunca apagados (RAMOS, 2009). Portanto, com seus usos diferenciados, é possível entender o porquê o Big Data não tem relações com o Data Warehouse, mesmo sendo repetidamente confundidos.

2.6. BIG DATA NO COMÉRCIO

Sabe-se que o Big Data é uma ferramenta que pode ser usada em todas as áreas do conhecimento, sendo assim, uma ferramenta de infinitas possibilidades e com grandes resultados. Desde simplesmente fazer medições nas máquinas de uma fábrica têxtil para evitar problemas mecânicos, até mesmo para ganhar mais clientes e, conseqüentemente, mais dinheiro.

O Big Data, sendo usado corretamente, leva ao crescimento das empresas em vários sentidos. No exemplo da fábrica têxtil, citado anteriormente, o uso dessa metodologia aumenta a vida útil das máquinas, a economia com o concerto (sendo ao contratar mecânicos ou terceirizadas) e a não necessidade de troca de peças em pouco tempo. Pode parecer pouco, mas no fim do mês é possível ver que todo esse dinheiro e trabalho economizado pode ser investido em outras áreas em favor da fábrica, dando a ela uma vantagem competitiva em relação às concorrentes.

Todos os benefícios que uma boa implementação de Big Data traz, faz com que muitas empresas se interessem por ele. Ao investir tempo e dinheiro, elas desejam um bom retorno, sendo ele novas ideias ou simplesmente estatísticas desconhecidas pelos executivos que poderão mudar o rumo da empresa.

Dois áreas que estão se beneficiando muito desta metodologia são as de Marketing e Publicidade e Propaganda. Em seus respectivos ramos, ambas as áreas estão conseguindo ter grandes *insights* e estão se beneficiando muito com isso. O Big Data permite ter uma visão muito mais ampla dos clientes, e, ao mesmo tempo, extremamente focada. Isso faz com que as empresas que o usam tenham uma grande vantagem, pois sabem com mais precisão do que antes o que seu cliente quer e o que ele espera da empresa e dos produtos. Sendo assim, essas áreas incluem o Big Data em várias de suas técnicas, aumentando o desempenho das mesmas e tendo resultados mais satisfatórios e personalizados.

O comércio em geral terá benefícios com o Big Data, tanto o físico quanto o *e-commerce*, ao usar a metodologia para que dados relevantes sejam descobertos e fazer com que a experiência do cliente melhore. Há várias técnicas que podem ser usadas com o Big Data nessas áreas, e que se tornam muito importantes para o crescimento do negócio.

As varejistas retêm muitos dados sobre seus clientes, desde nome e endereço até seu histórico de compras. O problema é que a grande maioria não usa esses dados ao seu favor. Usar essa tecnologia requer a coleta de dados, que já é feita pelas lojas há muito tempo, e o retorno ao cliente, que poucas fazem.

O objetivo do Big Data no Marketing é melhorar a experiência do cliente com a loja, sendo ao prever quais produtos são mais interessantes para ele ou para simplesmente aumentar as vendas.

Segundo Garcia (2014), duas coisas que as varejistas procuram hoje é ter um relacionamento íntimo com o cliente, mas que seja ao mesmo tempo um atendimento customizado.

Empresas que têm um relacionamento próximo com seus clientes são aquelas que mais atendem suas necessidades. Isso se deve ao fato de que essas empresas entendem o que o cliente quer, e também o que ele espera da empresa. Essas lojas podem usar o Big Data para criar programas de fidelidade personalizados para seus clientes (em parceria com uma determinada academia, por exemplo, a loja pode oferecer descontos em roupas de ginástica para o clientes daquela academia que escolherem essa loja) e melhorar o ambiente físico da loja (se a maioria dos clientes são mulheres, colocar produtos femininos logo na entrada, por exemplo), segundo Garcia (2014). O mesmo vale para o *e-commerce*, onde ele diz que “Nos EUA, as grandes redes investem mais em conhecer esta pessoa e enviar mensagens menos intrusivas e mais comunicativas”, ou seja, não lotar a caixa de email do cliente com ofertas a todo o momento.

Ter um relacionamento customizado é mais complicado, pois, apesar das empresas terem muitos dados relacionados aos seus clientes, a maioria não consegue tirar algo produtivo desses dados. Para Garcia (2014), é necessário ter uma certa maturidade do negócio, e que seus executivos consigam interpretar as análises corretamente. Garcia (2014) exemplifica muito bem ao dizer que um evento comercial nos EUA que reuniu uma grande quantidade de lojistas do país todo, usaram os dados das inscrições para criar mapas que priorizavam a área de atuação de cada lojista, ou seja, cada um tinha um mapa personalizado de acordo com as informações coletadas na inscrição para o evento.

As possibilidades do uso do Big Data são muitas, principalmente no *e-commerce*, onde, até então, era mais difícil selecionar produtos específicos para cada cliente. Em um site o cliente não tem um vendedor para lhe mostrar as peças de roupa que estão na moda e ajudar a encontrá-las, por exemplo. Em uma loja física, o cliente pede por uma camiseta ao vendedor, que trás vários modelos e em várias cores. Mas se o cliente não falar do que gosta e do que não gosta, o vendedor não vai saber e pode acabar mostrando apenas o que o cliente não quer, e assim é fácil perder este cliente e conseqüentemente a venda. Em um comércio eletrônico todos os itens estão lá e cabe ao cliente procurar o que quer, e isso pode ser um problema se o cliente não encontrar o que procura. Com o Big Data a navegação no site pode ser personalizada de acordo com cada pessoa que o acessa.

Tecnicamente, um rastreo no histórico do cliente é feito assim que ele entrar no site, e a partir disso o acesso é personalizado. Segundo Mamede (2013):

[...] Embora os consumidores sejam diferentes entre si e respondam a estímulos distintos, a maioria dos sites de venda insiste em exibir a todos a mesma vitrine. Mas [...] já é possível seduzi-lo com aquilo que ele mais gosta. Para moldar a oferta, pode-se analisar de forma instantânea, por exemplo, seu histórico de compra, seus interesses nas redes sociais e a lista de itens em que ele costuma clicar. Tudo isso também pode ser comparado com o perfil de outros usuários.

É fácil, então, encontrar a solução para o exemplo dado anteriormente. Ao acessar uma loja *online* que usa o Big Data para melhorar a experiência de consumo, o cliente tem seu histórico analisado. Procura-se pelo histórico de compras, gostos pessoais falados nas redes sociais, marcas curtidas no Facebook, produtos que visualizou em outras lojas, etc. Todas essas informações criarão um perfil bem exato do tipo de consumo desse cliente. Com o perfil criado, o sistema relaciona as preferências do cliente com os itens que há na loja, exibindo no navegador apenas aquilo que o cliente pode gostar. Isso faz com que a compra seja muito mais simples, direta e rápida para o cliente. A Amazon é a maior loja que usa esse tipo de sistema para as recomendações em seu site (MAMEDE, 2013).

A possibilidade de personalização na navegação do cliente é tão grande que pode ser um problema. De acordo com Mamede (2013), há uma prática do uso do

Big Data que pode trazer confusão e muita polêmica. Ele se refere à variação dos preços, que variam de cliente para cliente:

Os dados do Big Data, como é de imaginar, permitem inferir o perfil socioeconômico do consumidor virtual. Internautas que acessam determinada loja por meio de iPads e Macbooks e cujo histórico de compras inclua viagens internacionais tendem a aceitar pagar mais por um produto do que quem acessa de PC e tenha chegado à loja por meio de site de comparação de preço.

Essa prática discriminatória julga o cliente a partir do seu histórico de compras. E não é o Big Data que faz esse julgamento, mas sim os executivos da empresa, que podem escolher focar-se em clientes com maior poder aquisitivo para lucrar mais. Essa é uma das razões porquê o Big Data deve ser usado com parcimônia. E a Amazon teve duras consequências ao usar esse tipo de ideia, como exemplificado por Mamede (2013) onde, em 2000, a varejista decidiu variar os preços de DVDs de acordo com o navegador que o cliente usava no momento. Com isso, ela precisou compensar os clientes que pagaram mais caro pelos DVDs.

Mas o uso do Big Data no comércio não é apenas para aumentar os lucros do negócio, mas também para cortar custos desnecessários. Uma boa prática é o controle de estoque. Segundo o Canaltech (2013), a predição do Big Data pode ajudar ao aumentar a precisão na encomenda de produtos, mesmo que sejam sazonais, ao analisar que tipo de produto corre o risco de ficar parado na prateleira ou se esgotar rapidamente. Assim, há mais organização no setor e os custos ficam menores, pois a empresa não irá gastar com produtos que podem não ser vendidos.

O Big Data pode ter uma relação muito íntima no quesito de vendas, pois ele pode decidir diretamente o que mostrar ou não a um cliente. Com um sistema relativamente avançado, o Big Data simplesmente poderá mostrar ao cliente o que ele quer ver e ignorar produtos que o cliente não se interessa.

Com isso em mente, dois outros grandes usos do Big Data na área de Marketing são nos chamados *cross-selling* e no *up-selling*. Ambos são técnicas que visam convencer o cliente a comprar algo a mais do que havia planejado em um primeiro momento.

Cross-selling é vender um produto que tem algum tipo de relação com aquele que o cliente está comprando. Um exemplo básico é um cliente que entrou em uma loja para comprar uma camisa e uma calça, e saiu com uma gravata e um cinto

também. Ou seja, no *cross-selling* as vendas são “casadas” com outros produtos relacionados. O mesmo pode acontecer ao comprar um sapato e o vendedor sugerir uma meia também, ou um travesseiro e uma fronha. Segundo a Marketdata [s.d.], existem algumas técnicas usadas para esse tipo de venda, entre elas:

- Técnicas de Associações: Identificam o quanto frequentemente certos atributos ocorrem juntos. É, assim, uma ferramenta efetiva para identificar que produtos um consumidor tipicamente combina.

- Análises de Clusters: Agrupam clientes similares, podem mostrar que produtos têm maior atratividade para cada segmento específico de clientes. Bancos podem, assim, construir portfólios de produtos e serviços específicos para cada *cluster*.

- Modelagem Preditiva: Pretende estimar a probabilidade de um dado cliente comprar um determinado produto.

A primeira técnica, sobre associações, é sobre identificar quais produtos têm mais chances de serem comprados juntos. É mais provável um cliente comprar um sofá e decidir levar algumas almofadas também do que levar um lustre para enfeitar a sala. A segunda técnica, a análise de *clusters*, não tem relação com *cluster* de computadores, mas sim com um grupo de pessoas. Aqui são identificados grupos que tendem a comprar o mesmo tipo de produto. Um livro, por exemplo, exerce maior atratividade para pessoas tímidas e introspectivas do que para pessoas que gostam de festas e têm muitos amigos. Na modelagem preditiva, a terceira técnica, calcula-se a probabilidade de um cliente levar óculos de natação ao comprar uma sunga, por exemplo. A probabilidade é maior ao comprar uma bolsa e sapato do que um urso de pelúcia e um café. A frase “Quem comprou esse produto também comprou...” é um uso resumido de todas essas técnicas do *cross-selling*.

Já o *up-selling* é vender algo relacionado ao produto, mas que seja mais caro. Para Frachetta (2008), esse tipo de venda é “[...] uma estratégia [...] na qual um vendedor ou um site sugere um adicional para o produto ou serviço que está sendo comprado”. Como, por exemplo, ao comprar um carro e pagar a mais para ter *air-bags* e freios ABS. Ou seja, o vendedor convence o cliente a pagar uma quantia a mais para ter um produto ou serviço melhor. Outro exemplo seria a venda de um

computador, onde o vendedor oferece uma garantia estendida de um ano para a máquina, fazendo com que o produto seja tecnicamente melhor e a compra mais cara, e ainda assim mantendo a relação entre os dois produtos.

Resumindo, ambas as técnicas agregam valor ao produto original, melhorando a experiência do cliente ao mesmo tempo que aumenta o lucro da loja. Frachetta (2008) exemplifica muito bem o uso das duas técnicas com um pequeno diálogo típico de *fast-foods* como o McDonald's:

-Atendente: O senhor não gostaria de batatas grandes, por apenas R\$ 0,50 a mais?

-Atendente: Pedido feito, senhor. Acompanha um Sunday de morango?

Este é um exemplo muito bom pois mostra as duas técnicas sendo usadas ao mesmo tempo e em uma única compra. Na primeira pergunta o vendedor usa o *up-selling*, pois oferece um complemento à compra por R\$ 0,50 a mais. Já na segunda pergunta, o vendedor usa o *cross-selling*, onde é provável que o cliente queira uma sobremesa para acompanhar o lanche.

É interessante o uso de ambas as técnicas pois, de acordo com Frachetta (2008), elas adicionam conveniência ao cliente, fazendo com que o cliente se fidelize à loja. Mas isso não funciona se os vendedores não entenderem o conceito das técnicas, onde ele também diz que “A equipe de vendas deve ter sempre em mente que estes conceitos só funcionam se eles puderem economizar o precioso tempo de seus clientes.”. E complementando tudo isso, o site Administradores (2009) diz que:

O executivo da Brasanitas diz que tanto a venda cruzada como a venda com valor adicional reforçam a percepção de cumplicidade com as necessidades do cliente, que geralmente prefere contar com serviços bem realizados por um só fornecedor a contratar experts em vários segmentos e multiplicar o número de interlocutores.

Ou seja, quanto mais a loja facilitar para os clientes, melhor se torna o negócio (incluindo aumento dos próprios clientes e do faturamento).

Assim como o Big Data, o *cross-selling* e o *up-selling* devem ser usados com cuidado. O principal problema é irritar o cliente. Isso pode acontecer de duas maneiras segundo Frachetta (2008): oferecendo as coisas erradas ou oferecendo

coisas demais ao cliente. Devido a essa dificuldade, deve-se, então, colocar em prática ideias como: manter as sugestões de ambas as técnicas simples, não oferecer muitos produtos ao cliente para não confundí-lo, colocar produtos relacionados próximos uns dos outros, oferecer produtos com preços diferentes, criar coleções de produtos (como *boxes* de livros e DVDs) e, principalmente, preparar os vendedores para o uso dessas técnicas.

É possível ver que o uso do Big Data no comércio é grande e há inúmeras maneiras de implantá-lo. Empresas que buscam não só conhecer seu cliente, mas também entendê-lo e descobrir o que ele quer, são as que conseguem usufruir melhor da metodologia nesse ramo de atuação. A maior vantagem do Big Data no comércio é o fato de que seu uso traz benefícios para ambos os lados, para a empresa e para o cliente, melhorando o faturamento e a experiência de compra.

3. ESTUDO DE CASO

Este estudo de caso irá focar em melhorias que o uso do Big Data pode trazer para uma empresa. Melhorias estas que podem ser aplicadas em empresas de grande, médio e pequeno porte, nas mais diversas áreas e ramos de atuação.

O Big Data em si não traz melhorias. É o uso que os administradores da empresa dão para a ferramenta que o torna indispensável para o negócio. Seu uso pode ser aplicado nas finanças, estatísticas comerciais, Marketing e Publicidade, entre outros.

3.1. A EMPRESA

A empresa estudada neste estudo de caso é uma loja *online*, no ramo de vestuário, chamada Mundo Hipster. Criada e idealizada por Tony Strauss em 2004, tem como slogan a frase “Música que Veste”. A loja é focada em música, mais especificamente no estilo indie e alternativo, que estampa a maioria de suas camisetas. As estampas variam de um estilo mais rock e pesado como Nine Inch Nails, Queens Of The Stone Age e AC/DC até os mais leves e alternativos como Arctic Monkeys, Florence + The Machine e Arcade Fire. Apesar do foco ser na música, a loja também oferece estampas de seriados como Friends, Lost e How I Met Your Mother, filmes como Perks Of Being a Wallflower, V de Vingança e A Clockwork Orange, entre outros.

3.2. PESQUISA

Pesquisa é o uso de meios conhecidos (textos, perguntas, testes, livros, Internet) para criar ou analisar algo existente. Segundo Demo (2000) apud Motter [s.d.], “Na condição de princípio científico, pesquisa apresenta-se como a instrumentação teórico-metodológica para construir conhecimento”. Sendo assim, um meio de construção de novas ideias.

Sendo uma técnica versátil, a pesquisa pode ser feita a partir de indivíduos ou grupos de pessoas sobre os mais variados assuntos, como o Censo que o IBGE (Instituto Brasileiro de Geografia e Estatística) faz sobre a população do Brasil.

A pesquisa para este trabalho foi feita a partir de um questionário, onde, a partir das perguntas, foi possível ver o mercado de consumidores que a empresa abrange. O questionário foi respondido por trinta e três pessoas e continha dezessete perguntas, que abrangeram desde o sexo e idade do indivíduo até opiniões sobre a loja.

3.3. MERCADO CONSUMIDOR

A palavra "mercado" tem dois significados. O primeiro é o sentido literal da palavra, sendo assim, um local onde produtos ou serviços são vendidos para as pessoas (consumidores), um exemplo é de uma loja onde se vende doces. O segundo significado refere-se ao grupo de pessoas que compram produtos ou serviços (como os produtos da loja de doces, por exemplo), sejam elas vendedores ou consumidores, em lojas físicas ou virtuais (MIRANDA, 2012). No ponto de vista empresarial, mercado consumidor é o grupo de pessoas que procuram ou que podem comprar determinado tipo de produto ou serviço que a empresa oferece (SEBRAE, [s.d.]).

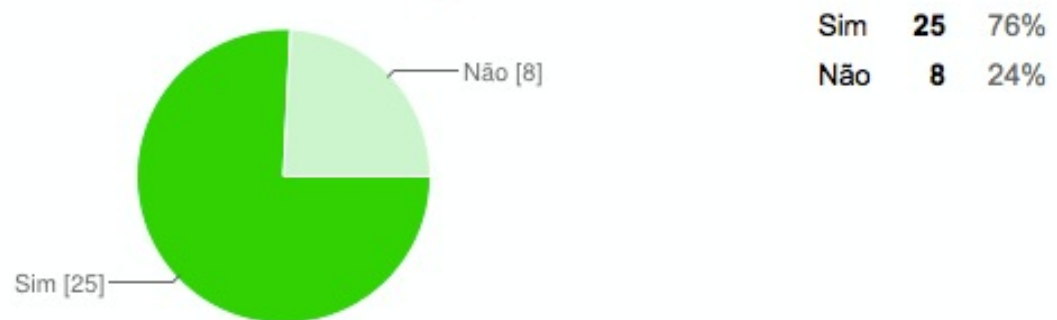
Ou seja, produtos e serviços são vendidos por pessoas. Pessoas estas que também são consumidores de outros produtos. É um ciclo, onde todo indivíduo em uma sociedade é de alguma maneira consumidor.

Mercado consumidor é, portanto, o conjunto de pessoas com poder aquisitivo suficiente para comprar um produto ou serviço oferecidos por uma empresa.

Dois questões da pesquisa definem bem o mercado consumidor da loja, seus gráficos podem ser vistos a seguir, na Figura 1 e Figura 2:

Figura 1: Compras online

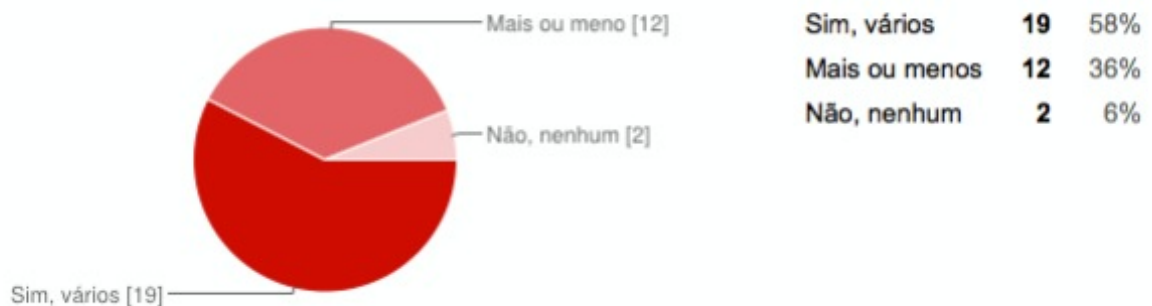
Você tem o costume de fazer compras online?



Fonte: Próprio autor

Figura 2: Estilo de música das camisetas

Você curte algum estilo de música representado nas camisetas?



Fonte: Próprio autor

De acordo com as respostas obtidas nessas perguntas do questionário, o mercado consumidor da Mundo Hipster são pessoas que têm o costume de fazer compras *online* e que gostam de um estilo de música mais alternativo.

3.4. PERFIL DO CLIENTE

O perfil do cliente define-se pelas características das pessoas que compram produtos ou serviços de determinada empresa, o tipo de pessoa que compra em uma loja (LI, [s.d.]). Essas características envolvem muitas variáveis, entre elas o sexo, idade, faixa salarial, local de residência, afiliações, interesses, gostos, entre outros.

Para definir o perfil do cliente, é necessário conhecê-lo. O principal meio de se fazer isso é através de pesquisas com questionários, onde é possível ver que a maioria dos clientes da empresa são homens solteiros com idade entre 20 e 30 anos e que gostam de esportes, por exemplo. O uso de redes sociais também ajuda a criar um perfil, principalmente pela interatividade direta da empresa com os clientes e os possíveis novos clientes.

Definindo o perfil, a empresa consegue visualizar o grupo de consumidores que deve atender. Se a maioria dos clientes são homens jovens e solteiros que gostam de esportes, como citado anteriormente no exemplo, a empresa saberá que o foco de seus produtos ou serviços não serão produtos femininos (como maquiagem e bolsas, por exemplo), mas sim masculinos e esportivos.

Saber que tipo de cliente uma empresa tem não implica apenas em definir o foco dos produtos ou serviços, mas também definir novidades, promoções e linguagem usada nas redes sociais que a empresa vai usar. É esse conjunto de coisas que define a fidelização ou não de consumidores.

Ao acompanhar o site e as redes sociais da Mundo Hipster, é fácil notar que o perfil de cliente da loja é um público mais jovem, descolado e com um gosto musical específico. Isso é visto principalmente nas redes sociais, onde seu linguajar é jovem, extrovertido e bem humorado.

3.5. NICHOS DE MERCADO

O nicho de mercado pode ser definido como a escolha de um mercado específico para um público específico, ou seja, tornar-se especialista em um ramo do mercado que não atende um público geral, mas apenas determinados consumidores (FELIPINI, [s.d.]). Por exemplo, uma loja de acessórios que é especializada em chapéus (sendo uma loja de acessórios, poderia oferecer bolsas, jóias, lenços e outros, mas seu nicho de mercado são os chapéus, tendo como principal clientela as pessoas que usam chapéus).

Felipini (2008) afirma que a escolha do nicho de mercado é basicamente “o que' vai ser vendido e 'para quem' vai ser vendido”. Isso significa que a escolha do nicho de mercado está intrinsecamente ligada à definição do perfil do cliente. A partir do momento em que a empresa identifica o tipo de consumidor que tem, é fácil focar

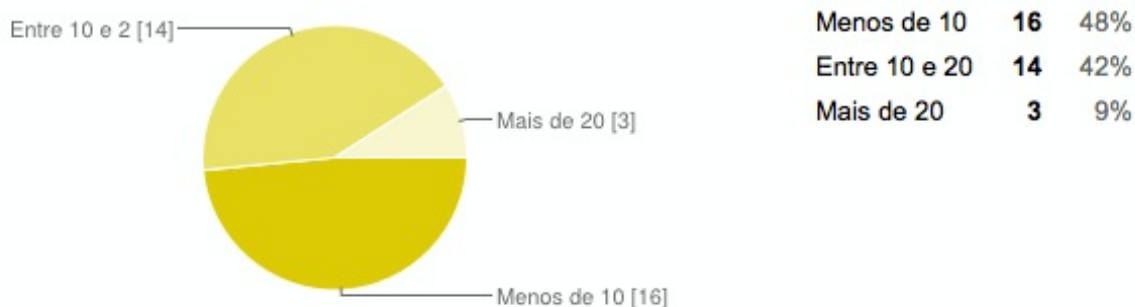
os produtos ou serviços oferecidos para esse tipo específico de público. A loja de acessórios pode ter escolhido focar seus produtos em modelos de chapéus por ter notado que sua clientela são mulheres e homens que moram em capitais focadas em moda e que tenham uma renda mensal maior que a média da população, por exemplo.

Segundo Mattar e Auad (1997), encontrar um nicho de mercado é se adaptar a um segmento e oferecer seus produtos ou serviços com um bom atendimento e muito bem ajustados às necessidades e desejos do cliente. Assim sendo, a empresa que deseja focar-se em um determinado nicho, deve fazê-lo com grande determinação para atender o cliente da melhor maneira possível. Isso implica em, no caso da loja de acessórios focada em chapéus, ter uma grande variedade de modelos, cores, matérias-primas, marcas e até mesmo oferecer um serviço de alfaiataria para futuros concertos ou ajustes, por exemplo. “A prática do nicho de mercado garantirá a sua sobrevivência na comunidade empresarial” (MATTAR e AUAD, 1997), ou seja, focar-se em um nicho e ser o melhor neste mercado específico é garantia de diferencial no mercado.

Ao navegar pelo site da Mundo Hipster, nota-se que o nicho de mercado é bem específico, sendo ele jovem e alternativo. Isso acontece pois pode-se ver que muitas bandas estampadas nas camisetas da loja não são *mainstream*, ou seja, são pouco conhecidas do grande público. Exemplos de bandas pouco conhecidas e que estão estampadas nas camisetas da loja são Haim, The xx, Björk, The Kooks, Cage The Elephant e Jake Bugg, para citar algumas. A pesquisa feita para este estudo de caso pediu para as pessoas acessarem o site da loja e responderem quantas bandas, seriados e filmes conheciam, e o resultado pode ser visto na Figura 3:

Figura 3: Bandas, seriados e filmes da loja

Quantas bandas, seriados e filmes representados na loja você conhece?



Fonte: Próprio autor

O fato de que as estampas representadas na loja não são muito conhecidas pelo público geral, principalmente as bandas, é comprovado neste gráfico, onde quase metade das pessoas (48%) responderam que conhecem pouco das estampas da loja. Mas isso não significa que as bandas não sejam famosas, mas sim que seu público é um público mais alternativo e que gosta de coisas diferentes. E esse público alternativo são geralmente jovens que buscam conhecer novas e boas bandas que ou estão no começo ou tem um estilo mais diferente.

3.6. SUGESTÕES DE USO DO BIG DATA NA EMPRESA

O Big Data sendo uma ferramenta tão versátil, faz com que seja facilmente empregado em vários ramos. O comércio, como dito anteriormente, é uma área que está colhendo bons frutos ao adotar essa metodologia. O objetivo deste trabalho é sugerir benefícios de uso do Big Data para uma empresa, no caso, a loja de vestuário *online* Mundo Hipster.

Através de contato com a loja, soube-se que, até o presente momento, a loja não usa nenhum tipo de ferramenta focada em Big Data. Portanto, as sugestões a seguir foram feitas a partir da pesquisa realizada com potenciais clientes, da observação do site e do comportamento da loja nas redes sociais.

Usar um sistema de Big Data não é simplesmente implantá-lo e obter logo em seguida respostas sobre como melhorar a empresa e ganhar mais dinheiro. Primeiramente, precisa-se de foco. É o foco sobre o que fazer, que decisões tomar. Esse rumo poderá guiá-los a um patamar mais alto, ou simplesmente levá-los ao fracasso. E isso depende dos executivos, e não do Big Data. Depende de como eles lidarão com as estatísticas, como os dados e informações serão explorados e como serão usados.

SISTEMA

Toda loja, física ou *online*, usa algum tipo de sistema. Independentemente do tipo de sistema, todos eles devem acumular dados. Esses dados são relativos aos preços, descontos, quantos produtos foram vendidos na última semana ou no último

mês, trocas, devoluções, clientes, etc. É um sistema com dados e informações básicas. Ao implementar um sistema de Big Data, essas informações serão apenas uma parte dentro de um todo. Isso acontece pois o Big Data coleta uma quantidade muito maior de informações, pois tudo deverá ser transformado em dado. Desde, por exemplo, quantas vezes uma máquina quebra, a quantidade de produto que é usado em determinado tempo, em que período gasta-se mais energia elétrica e quais aparelhos gastam mais energia, em que pontos um vidro tende a ser mais facilmente quebrado, etc; até relacionados às pessoas, como rotatividade de funcionários, assuntos que as pessoas comentam mais, experiência de funcionários mais velhos, etc. O Google, por exemplo, transforma em dados suas buscas e informações, o Facebook faz o mesmo com a rede de amizades das pessoas, o Twitter também com que as pessoas falam no momento, o LinkedIn transforma também em dados os currículos e dados profissionais (LÁZARO, 2013). Tudo pode ser transformado em dado, principalmente em tempos de Big Data, pois quanto mais dados o sistema puder analisar, melhores serão os resultados.

Lázaro (2013) diz que:

Cada processo de negócio agora é 'datafocado' – podemos monitorá-lo, armazenar a sua história e começar a identificar padrões de uso e, portanto, ter recomendações de como podemos melhorar ou monetizar isso, que não é, realmente, um novo fenômeno.

Assim sendo, a primeira coisa que a Mundo Hipster deve fazer é o negócio ser transformado em dado. Fazer isso com tudo o que entra na loja e o que sai, a produção (quantidade de defeitos em camisetas e estampas, quantidade e qualidade de tecido, quantos metros de linha são usados, defeitos em máquinas de costura e estamperia, etc.), acessos ao site, o quanto clientes e possíveis clientes socializam com a marca nas redes sociais, curtidas no Facebook e seguidores no Twitter e Instagram, novos clientes, vendas etc. Esse trabalho todo será a base onde o Big Data irá trabalhar. O sistema pode identificar, por exemplo, que o óleo usado nas máquinas de costura dura mais tempo do que se imaginava, trazendo uma economia para a empresa. Coisas simples podem ser convertidas em dados e serem o motivo de boas descobertas.

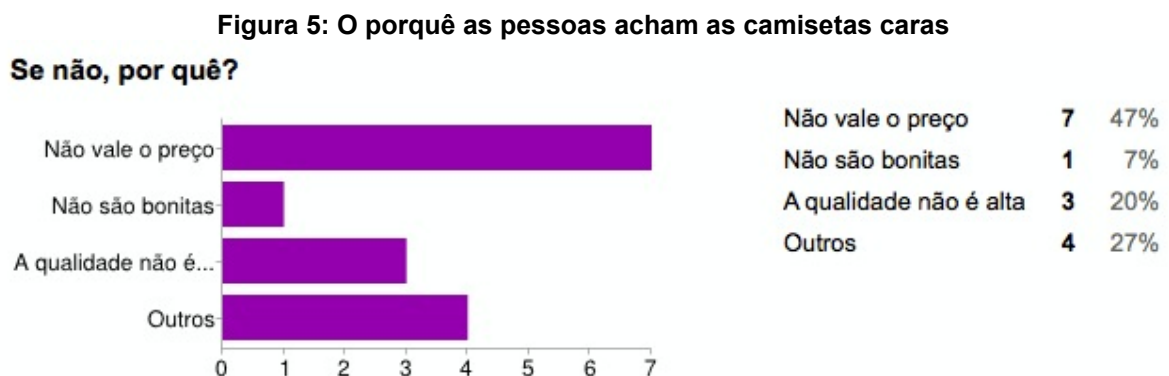
Com tudo devidamente convertido em dado, o próximo passo é usar o Big Data para analisar as informações que estão armazenadas. A Mundo Hipster é uma

loja, e o objetivo de uma loja é vender. Com isso em mente, a loja deve reunir informações acerca do que é relevante para que um cliente realize uma compra, que tipo de informações fazem com que o cliente compre o produto ou voltem a visitar o site da loja. Os financeiro pode dizer que o preço é relevante para a compra, e após uma análise, o Big Data pode dizer que além do preço, informações importantes para os clientes realizarem a compra são, principalmente, detalhes do produto e métodos de pagamento. Ou seja, a loja deverá encontrar meios para melhorar esses quesitos. Segundo a pesquisa realizada, a loja deve ter um cuidado especial em relação ao preço dos produtos, conforme pode-se observar na Figura 4:



Fonte: Próprio autor

De acordo com o gráfico, mais da metade das pessoas considera o preço alto ou relativamente alto (63%), então a loja deve encontrar meios de contornar isso. A pesquisa também mostrou que as pessoas acham que as camisetas não valem o preço atual, conforme é visto na Figura 5:

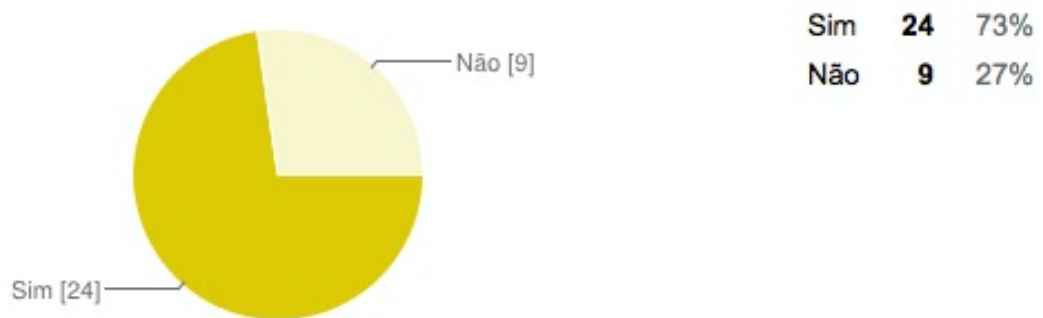


Porém, a Mundo Hipster tem o hábito de criar promoções no site. Sempre há promoções que têm a duração de um fim de semana em datas especiais, como o Natal e o Dia das Mães (mesmo não sendo o perfil de cliente e seu público-alvo). Há também o que a loja chama de Camiseta do Dia, onde todo dia uma camiseta é escolhida para ter desconto. Isso acaba amenizando o problema.

Um dado interessante que a pesquisa mostrou é que, mesmo a loja tendo foco em música alternativa, as pessoas gostariam de mais produtos. É possível visualizar isso nas Figuras 6 e 7:

Figura 6: Outros tipos de produtos

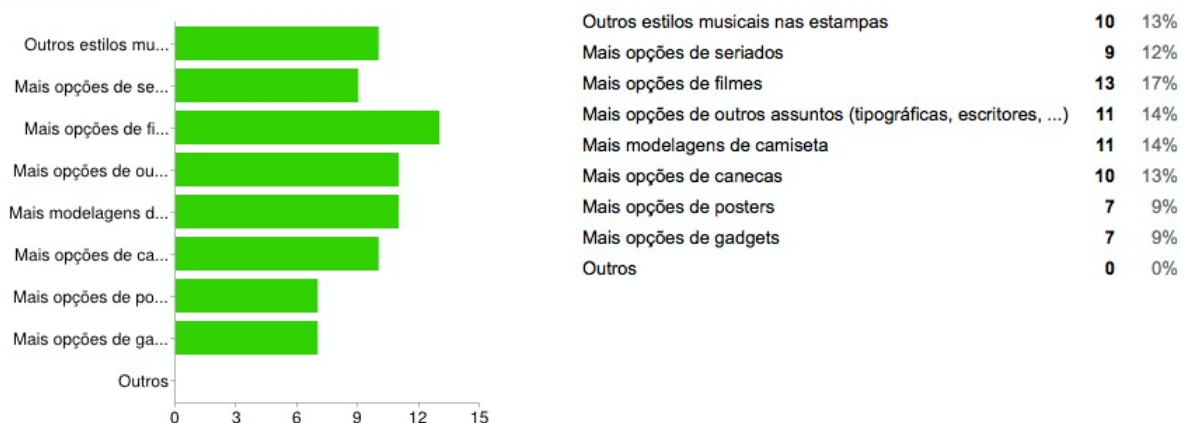
Você acha que a loja poderia vender outros tipos de produtos?



Fonte: Próprio autor

Figura 7: Quais produtos

Se sim, quais?



Fonte: Próprio autor

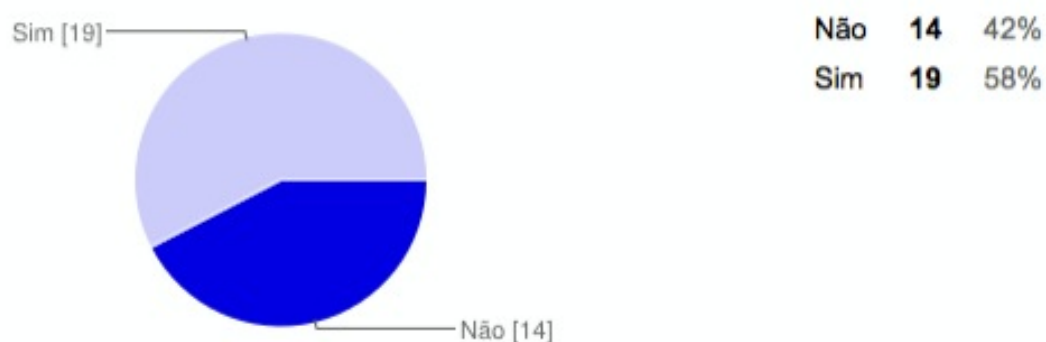
Observa-se no gráfico, que as pessoas que acessaram a loja acham que devem ser disponibilizados uma variedade maior de produtos. Porém, os indivíduos

que responderam o questionário referem-se a outras opções de camisetas: mais opções de filmes (17%), outros assuntos e outras modelagens de camiseta (14%) e outros estilos musicais nas estampas (13%). Ou seja, muitas vezes uma pessoa pode não comprar algum produto na loja por não ter o que ela goste. O foco deve ser no perfil do cliente, e esse perfil são de pessoas jovens com gosto musical alternativo. Porém, a loja pode adicionar mais modelos de camisetas, que são bem limitados, criando modelos regata e com manga longa, por exemplo. A customização pode ajudar nas vendas. Se a loja adicionar novos modelos de camiseta e novas cores, é possível criar um método onde o cliente escolhe o modelo da camiseta, tecido e cor e a estampa que preferir. Neste caso, o Big Data avaliará a camiseta customizada pelo cliente, permitindo a compra ou não. Por exemplo, se a pessoa escolher uma camiseta azul escuro e uma estampa predominantemente preta, o sistema não permitirá a compra, pois não será possível ver a estampa na camiseta. É possível assim, multiplicar a variedade de modelos, as chances de agradar um cliente e, conseqüentemente, as vendas. Se a customização de um produto for recorrente, pode-se produzir o produto em maior escala.

Outro dado constatado com a pesquisa é em relação aos métodos de pagamento. Foi perguntado se a loja deveria ter outros métodos, e a maioria das pessoas respondeu que sim, conforme a Figura 8:

Figura 8: Outros métodos de pagamento

Você acha que deveria ter outros métodos de pagamento?



Fonte: Próprio autor

Em relação aos métodos de pagamento, na Figura 9, observa-se que:

Figura 9: Quais métodos de pagamento

Se sim, quais outros métodos você gostaria?



Fonte: Próprio autor

PayPal e PagSeguro foram os mais votados. Ao adicionar o PayPal como método de pagamento no site, é possível que a clientela fora do Brasil aumente, por exemplo.

Mas o Big Data pode ir além. Com a análise de dados, o sistema pode identificar muitas coisas que podem ser úteis no negócio. Identificar segmentos de mercado, tendências e perfis de cliente são apenas alguns exemplos. Com o Big Data identificando essas características do negócio, é possível criar estatísticas sobre a indução de compra. Essas estatísticas dizem o que mais induz uma pessoa a fazer a compra, seja isso o seu perfil de cliente, preço, poder aquisitivo, entre outros.

O sistema pode ajudar a manter o preço acessível, de acordo com o perfil do cliente da loja e em comparação aos concorrentes. O mesmo vale para as promoções, onde o sistema pode mostrar as melhores épocas do ano para lançar grandes promoções, se as promoções aumentam a receita da loja e que tipo de promoção os clientes gostam mais. Mas não adianta ter um bom preço e boas promoções regularmente se a qualidade dos produtos não for boa. Isso vale principalmente para produtos vendidos que, se a qualidade for alta, pode-se justificar o preço mais alto. As camisetas da Mundo Hipster têm uma qualidade alta, são muito bem feitas, com diferentes tipos de malhas, bem trabalhadas, de ótima

qualidade, não desbotam e as estampas são bem produzidas, coloridas e duráveis. E esse é um diferencial que a loja pode focar. Clientes satisfeitos divulgam a loja de boca em boca e até mesmo pela Internet, seja simplesmente por um comentário sobre o tecido que é diferente, pela estampa bonita ou pelo fato de ter encontrado uma camiseta de uma banda pouco conhecida em um site com uma qualidade tão boa. É difícil encontrar camisetas no mercado com temas diferentes, principalmente bandas, e isso faz com que o cliente da loja mostre sua aquisição para amigos que têm um potencial maior para se tornar um cliente.

O uso de um sistema Big Data favorece a loja, sendo possível encontrar processos a serem mudados ou simplesmente melhorados. Cabe aos executivos definirem um alvo e focar nesse objetivo, traduzindo as estatísticas para melhor atendê-los.

MARKETING E REDES SOCIAIS

Uma boa estratégia de Marketing é imprescindível para uma loja, em qualquer ramo de atuação. Sendo o produto de boa qualidade ou não, se o Marketing é efetivo, as pessoas irão prestar mais atenção na marca e, conseqüentemente, as vendas poderão ser melhores.

É sabido que o Marketing teve a ascensão com as redes sociais, principalmente o Facebook, onde é possível promover o negócio, compartilhar novidades e ter contato direto e rápido com clientes e possíveis clientes. E um ótimo uso do Big Data é relacionado a ambos: usar as redes sociais para divulgar o negócio e conquistar clientes.

Em contato com a Mundo Hipster, soube-se que o foco deles é o relacionamento com os clientes através das redes sociais. E é fácil notar isso pois, ao observar o comportamento da loja na internet, é possível ver que são muito ativos. Além de promover as promoções e itens da loja, eles são muito sociáveis. Ao entrar no perfil da loja no Twitter, por exemplo, é possível ver muitas respostas e conversas com seus seguidores (que podem ou não ser clientes da loja), sempre solícitos, simpáticos e bem humorados. Tanto no Twitter, quanto no Facebook, o linguajar é jovem e moderno, com uso de brincadeiras com artistas e memes do momento, atingindo seu público alvo. Pode-se definir um meme como um fenômeno

na Internet, que ao ser compartilhado por muitas pessoas acaba tendo grandes proporções (RIBEIRO, 2012). Um meme pode ser um vídeo, uma nova gíria, uma frase, ou simplesmente uma imagem. É uma piada ou brincadeira que se propaga na rede de maneira muito rápida (RIBEIRO, 2012), virando uma moda por um determinado tempo (até as pessoas se cansarem da piada). Um exemplo de meme utilizado pela loja é o 4:20, onde todos os dias, às 16h20, é compartilhado uma foto ou um *gif* "louco", tosco, sem muito sentido, no Twitter da loja. E com isso eles, de uma maneira inusitada, acabam fidelizando e conquistando novos clientes.

Com o foco em redes sociais, a Mundo Hipster está no caminho certo. Mas o Big Data pode ajudar ainda mais nesse quesito, criando ligações e tendências que a loja não havia notado antes.

A loja tem sua própria rede social, chamado ReverbME. Toda pessoa que se cadastrar no site já tem um perfil na rede social, que pode ser usada para adicionar amigos, conversar com a equipe da loja, receber informações sobre promoções, lançamentos e pedidos. No perfil de cada pessoa é solicitado algumas informações, como o nome, idade, cidade onde mora e bandas que gosta, mas existe a opção de colocar o *link* do perfil do Facebook e o usuário do Twitter também. Assim, alguns usuários do ReverbME podem ser amigos em outras redes sociais. Mas a loja pode ir além. Com as informações de idade e bandas que gosta, a loja consegue criar um perfil dos usuários. Porém, se adicionar o perfil de outras redes sociais como o Last.fm (rede social de música) e Filmow (rede social de cinema e seriados), é possível detalhar ainda mais esse perfil, podendo saber quais bandas que seus usuários mais ouvem, quais tipos de filme gostam e quais seriados assistiram. Para aprofundar ainda mais, pode-se usar estatísticas do site Banco de Séries para saber quais os seriados mais assistidos no momento. Tudo isso pode ser usado para pesquisa de novas estampas de camisetas, e é o Big Data que se encarrega disso. O sistema poderá identificar comentários sobre a loja, bandas, seriados e filmes através do Twitter. Páginas de bandas curtidas e vídeos compartilhados no Facebook também serão identificados. No Last.fm é possível ver os artistas e bandas que a pessoa mais ouviu na última semana, no último mês, ou até mesmo no último ano. O Filmow mostra os filmes mais populares do momento e o Banco de Séries mostra os seriados mais assistidos. O Big Data usará todas essas informações para criar estatísticas que podem mostrar o que é popular e o que os

clientes gostam, sendo o meio que a loja usará para criar novas estampas e produtos.

Com todos esses dados é possível melhorar os emails que a loja manda para seus clientes. A empresa cria um email e manda para todos as pessoas cadastradas no site, contendo novidades sobre novos produtos, serviços ou promoções. É o mesmo email para todos, não importando seus gostos pessoais. Mas com o Big Data isso pode ser totalmente personalizado, um email diferente para cada pessoa. O sistema identifica os gostos da pessoa, baseado nos dados retirados das redes sociais, colocando apenas informações relevantes para esta pessoa no email. A Mundo Hipster pode lançar uma promoção válida para dez camisetas, por exemplo, e há usuários que gostam apenas de uma dessas bandas, e outros que gostam de cinco delas; o Big Data irá identificar essas diferenças e personalizará o email para cada um desses usuários.

É comum lojas *online* mostrarem sugestões de produtos, e a Revebcity usa esse sistema também. Porém, a implantação do Big Data pode melhorar muito esse sistema de recomendações, tanto para usuários, quanto para novos clientes. Para os novos clientes, que ainda não têm cadastro no site, o Big Data pode simplesmente mostrar camisetas de artistas, filmes e seriados que estão em alta naquele momento. Para os clientes da loja, que têm cadastro, esse sistema será personalizado, mostrando estampas de bandas, filmes e seriados relacionados ao gosto da pessoa (através do Last.fm, Facebook, etc).

Outra sugestão para a loja é permitir que os usuários “curtam” as estampas do site. Com este recurso, é possível ver quais estampas agradam mais os usuários e quais não agradam. O Big Data identificaria quais modelos, cores e estilos de estampas são mais populares, onde os designers usariam esses dados para criar estampas novas com menos chances de serem rejeitadas. Apenas usuários cadastrados no site poderiam avaliar as estampas dessa maneira, sendo que os resultados seriam visíveis para todos. Com os resultados tornando-se públicos, novos clientes poderiam pesquisar apenas as camisetas mais bem votadas, ou seja, as mais bonitas, influenciando a decisão de compra.

Como já citado, a Mundo Hipster é uma loja focada em música indie e alternativa, portanto, artistas pouco conhecidos do grande público. Apesar do perfil de cliente e o nicho de mercado serem específicos desse estilo, o site não é apenas

acessado por indivíduos com esse perfil alternativo. Pessoas com um gosto totalmente diferente daqueles representados na loja, podem se deparar com o site, achá-lo interessante e buscar por produtos. Pessoas que não conhecem as estampas representadas no site, podem achá-las bonitas e querer conhecer sua origem. A sugestão aqui seria incorporar um videoclipe da banda ou artista na página da camiseta, assim como o trailer do seriado ou filme, complementando as informações que o site já dá sobre o assunto da estampa. Com isso é possível mostrar novos artistas para antigos e novos clientes.

A Mundo Hipster usa intensamente as redes sociais, compartilhando vídeos de bandas relacionadas à loja e letras de músicas nesse mesmo estilo. É possível perceber o gosto musical diferenciado da loja. Com isso em mente, a loja poderia criar um perfil no site Spotify. O Spotify é um serviço de *streaming* de música que tem milhões de usuários. *Streaming* é uma tecnologia que transmite dados pela internet de modo instantâneo, sejam dados de áudio ou vídeo (COUTINHO, 2013). Ou seja, não há a necessidade de download, e muitas pessoas utilizam esse meio para assistir seriados e shows *online*, por exemplo. O Spotify contém milhões de músicas em seu acervo, e é possível a criação de *playlists* (listas de reprodução) que podem ser compartilhadas nas redes sociais. A Mundo Hipster poderia criar *playlists* com músicas que seguem o estilo da loja, ou com músicas apenas das bandas representadas nas camisetas, *playlists* temáticas (músicas alegres, românticas, natalinas, etc) seguindo o estilo indie e alternativo, entre outros. Seria mais uma maneira de divulgar a loja nas redes sociais e interagir com os clientes.

Apesar de todas as lojas, sejam elas *online* ou físicas, terem suas particularidades, todas elas têm um mesmo objetivo além de ganhar dinheiro. Todas as lojas procuram agradar o cliente. Esse agrado pode ser um vendedor simpático, um ambiente agradável e bonito, bolachas e café, poltronas confortáveis, entre outros. Percebe-se que é fácil agradar um cliente em uma loja física, mas o mesmo não acontece em lojas *online*. Isso não acontece pois não há vendedores, nem espaço físico, nem a possibilidade de oferecer um café para o cliente. Mas em ambos os casos, agradar o cliente da maneira correta é sinônimo de diferencial. E as lojas *online* têm meios interessantes de se diferenciar. A maioria delas oferece frete grátis, por exemplo, e isso não é o bastante. A Mundo Hipster oferece frete grátis em compras a partir de duzentos e sessenta reais, um valor alto e que poderia

ser diminuído. Outra sugestão seria incluir brindes nas compras. Em toda compra, a loja envia junto à camiseta um *paper toy* (modelos de papel que são cortados e encaixados ou colados para formarem uma personagem, por exemplo) de um dos dois mascotes da loja, mas são apenas duas variações de modelo. Outros brindes poderiam ser adesivos da loja ou relacionados à estampa comprada, marca páginas também relacionados à estampa, um ocasional “vale-camiseta” em uma compra aleatória, etc. Esses pequenos agrados “mimam” os clientes, e os fazem voltar a comprar na loja, tornando-se clientes fiéis.

Sendo o foco da Mundo Hipster o contato com os clientes através das redes sociais, as possibilidades de ideias são maiores. O Big Data pode identificar bandas novas e bem comentadas para fazerem parte da coleção de camisetas, por exemplo. Pode-se também identificar quais bandas e artistas estão perto de lançar um novo álbum ou uma nova turnê, onde a loja pode colocar em promoção as estampas relacionadas. O mesmo vale para seriados que estão retornando com novas temporadas e bandas e artistas aniversariantes. Sorteios de produtos na página da loja no Facebook chamam a atenção e futuros clientes. A escolha da camiseta para sorteios e promoções pode ser baseada em Big Data.

As possibilidades são muitas se tratando da parceria entre Big Data, Marketing e redes sociais. O Marketing está sempre mudando suas técnicas, e as redes sociais têm maneiras distintas de mostrar o que as pessoas gostam no momento. Acompanhar isso tudo é difícil, mas não para o Big Data, que além de monitorar essas constantes mutações, consegue criar associações que antes eram impensadas.

As sugestões dadas neste estudo de caso não são absolutas. São ideias para introduzir o sistema em uma empresa de comércio *online*, o que não impede das mesmas evoluírem e se tornarem melhores e ainda mais precisas. E isso é um dos fatos mais interessantes do Big Data. Ele não é implantado e simplesmente usado. Ele é um sistema que pode estar em constante evolução para melhor atender as necessidades do negócio de acordo com o foco de seus executivos.

4. CONCLUSÃO

Este trabalho teve como objetivo introduzir e implantar o conceito de Big Data em uma loja de *e-commerce*. A partir da definição sobre o que é, para que serve, como e onde usar o sistema, foi possível identificar vários usos para a loja citada.

A loja poderá usá-lo de várias maneiras. No sistema, o Big Data pode descobrir e sugerir produtos, preços e promoções. Mas sendo o objetivo da loja o foco nos clientes em redes sociais, pôde-se notar que os usos do sistema são maiores e mais versáteis nesse sentido.

Ao implantar o Big Data, a loja deve atentar-se aos possíveis problemas relatados no começo deste trabalho. Problemas relacionados aos dados e à Segurança da Informação. Ao transformar em dados os produtos e métodos, deve-se usar táticas (físicas e lógicas) da Segurança da Informação para protegê-los, evitando o roubo, perda e acesso indevido desses dados. Evitando, então, que lojas concorrentes tenham acesso às particularidades da loja estudada para promover-se. Além do cuidado com os dados, existe a questão da privacidade, que, como visto neste trabalho, é um grande problema. Foi sugerido diversas vezes a análise de dados dos clientes da loja, dados em relação ao sexo, idade e profissão e dados retirados das redes sociais que usam. Ou seja, dados pessoais. A loja não deve cometer o erro de usar indevidamente esses dados, invadindo a privacidade dos clientes. O uso desses dados pessoais é apenas para a personalização da experiência do cliente no site, email e redes sociais.

O Big Data é uma ferramenta que pode trazer grandes benefícios para a sociedade como um todo, desde que usada com grande cuidado e não deixando que faça decisões no lugar das pessoas. Com atenção a estes detalhes, o uso do Big Data será próspero e de vida longa.

REFERÊNCIAS BIBLIOGRÁFICAS

ADICIONE valor às vendas cruzadas. Administradores, João Pessoa, 24 de nov de 2009. Disponível em: <<http://www.administradores.com.br/noticias/negocios/adicione-valor-as-vendas-cruzadas/28006/>>. Acesso em: 10 de mai de 2014.

AGÊNCIA Estado. **ONU: População mundial é de 7,2 bilhões de pessoas.** Estadão, São Paulo, 13 de jun de 2013, Internacional. Disponível em: <<http://internacional.estadao.com.br/noticias/geral,onu-populacao-mundial-e-de-7-2-bilhoes-de-pessoas,1042156>>. Acesso em: 20 de mar de 2014.

BEDNARZ, A. **Hadoop conquista TI das companhias.** Computerworld, [S.l.], 22 de feve de 2012. Disponível em: <<http://computerworld.com.br/tecnologia/2012/02/17/hadoop-conquista-ti-das-companhias/>>. Acesso em: 25 de abr de 2014.

BIG Data Brasil. **O big data.** Disponível em: <<http://bigdatabrasil.wordpress.com/2012/12/20/o-big-data/>>. Acesso em: 15 de mar de 2014.

BIG Data: os cinco Vs que todo mundo deveria saber. Canaltech, São Bernardo do Campo, 28 de mar de 2014. Disponível em: <<http://corporate.canaltech.com.br/dica/big-data/Big-Data-os-cinco-Vs-que-todo-mundo-deveria-saber/>>. Acesso em: 15 de abr de 2014.

COMO o big data pode aumentar as vendas no e-commerce. Canaltech, São Bernardo do Campo, 19 de dez de 2013. Disponível em: <<http://corporate.canaltech.com.br/noticia/e-commerce/Como-o-Big-Data-pode-aumentar-as-vendas-no-e-commerce/>>. Acesso em: 10 de mai de 2014.

CONHECENDO o mercado. SEBRAE, Santa Catarina, [200-?]. Disponível em: <<http://www.sebrae-sc.com.br/ideais/default.asp?vcdtexto=4782&%5E%5E>>. Acesso em: 25 de mai de 2014.

CONHEÇA as aplicações reais do big data. Olhar Digital, [S.l.], 14 de dez de 2013. Disponível em: <<http://img.olhardigital.uol.com.br/pro/video/39376/39376?nohits>>. Acesso em: 2 de abr de 2014.

COUTINHO, M. **Saiba mais sobre streaming, a tecnologia que se popularizou na web 2.0.** TechTudo, [S.l.], 27 de mai de 2013. Disponível em: <<http://www.techtudo.com.br/artigos/noticia/2013/05/conheca-o-streaming-tecnologia-que-se-popularizou-na-web.html>>. Acesso em: 30 de mai de 2014.

DONEDA, D. **Da privacidade à proteção de dados pessoais.** E-Gov, Santa Catarina, [201-?]. Disponível em: <<http://egov.ufsc.br/portal/sites/default/files/anexos/29536-29552-1-PB.pdf>>. Acesso em: 17 de mai de 2014.

EMC. EMC Corporation. Disponível em: <<http://brazil.emc.com/campaign/bigdata/index.htm>>. Acesso em 15 de mar de 2014.

FELIPINI, D. **Qual é o seu nicho de mercado?** eCommerceOrg, [S.l.], [200-?]. Disponível em: <http://www.e-commerce.org.br/artigos/nicho_de_mercado.php>. Acesso em: 25 de mai de 2014.

FERREIRA, M. **O que vem ser segurança da informação?** APInfo, [S.l.], [201-]. Disponível em: <<http://www.apinfo.com/artigo81.htm>>. Acesso em: 3 de mai de 2014.

FRACHETTA, A. **Up-selling e cross-selling.** Inteligência Mercadológica, [S.l.], 24 de set de 2008. Disponível em: <<http://inteligenciamercadologica.info/2008/09/24/up-selling-e-cross-selling/>>. Acesso em: 10 de mai de 2014.

IBM. IBM Corporation. **Aproveitando o valor de seus dados em um planeta mais inteligente.** 1 fotografia, colorido. In: IBM. A sua infraestrutura está pronta para o Big Data? [S.l.], [201-?]. Disponível em: <http://www.ibm.com/midmarket/br/pt/infografico_familia_ibm_storwize.shtml>. Acesso em: 20 de abr de 2014.

GARBIA, B. **Big Data já é realidade. Conheça modelos de implementação.** Mundo do Marketing, Rio de Janeiro, 5 de fev de 2014. Disponível em: <<http://www.mundodomarketing.com.br/reportagens/pdv/29858/big-data-ja-e-realidade--conheca-modelos-de-implementacao.html>>. Acesso em 10 de mai de 2014.

GASPAROTTO, H. M. **Hadoop mapreduce:** introdução ao big data. DevMedia, [S.l.], 30 de jan de 2014. Disponível em: <<http://www.devmedia.com.br/hadoop-mapreduce-introducao-a-big-data/30034>>. Acesso em: 25 de abr de 2014.

GARTNER. Gartner Group. Disponível em: <<http://www.gartner.com/technology/topics/big-data.jsp>>. Acesso em: 15 de mar de 2014.

GARTNER. Gartner Group. In: Conferência Gartner Business Intelligence & Information Management 2014. São Paulo: Gartner Summits, 13 e 14 de mai de 2014.

HOW Amazon is leveraging big data. Big Data-Startups, Holanda, [2012?]. Disponível em: <<http://www.bigdata-startups.com/BigData-startup/amazon-leveraging-big-data/>>. Acesso em: 4 de abr de 2014.

KOCH, D. **O que é um projeto open source.** HowStuffWorks, [S.l.], [200-?]. Disponível em: <<http://tecnologia.hsw.uol.com.br/projetos-open-source1.htm>>. Acesso em: 25 de abr de 2014.

LANDIM, W. **Apple, facebook e google vão avisar quando o governo solicitar seus dados.** Tecmundo, Curitiba, 3 de mai de 2014. Disponível em: <<http://www.tecmundo.com.br/privacidade/54213-apple-facebook-e-google-va-avisar-quando-o-governo-solicitar-seus-dados.htm>>. Acesso em: 19 de mai de 2014.

LÁZARO, D. **Você sabe o que é dataficação?** Computerworld, [S.l.], 26 de ago de 2013. Disponível em: <<http://computerworld.com.br/blog/analytics/2013/08/26/voce-sabe-o-que-e-dataficacao/>>. Acesso em: 28 de mai de 2014.

LI, A. **Definição de perfil de clientes**. eHow, [S.l.], [200-?]. Disponível em: <http://www.ehow.com.br/definicao-perfil-clientes-fatos_68157/>. Acesso em: 25 de mai de 2014.

MAIA, M. A. **O que é segurança da informação**. Blog de Segurança da Informação, [S.l.], 19 de ago de 2013. Disponível em: <<http://segurancadainformacao.modulo.com.br/seguranca-da-informacao>>. Acesso em: 3 de mai de 2014.

MALI, T. **Por que os seus dados te denunciam**. Revista Galileu, São Paulo, 19 de abr de 2013. Disponível em: <<http://colunas.revistagalileu.globo.com/colunistas/2013/04/19/anonimizacao-de-dados-uma-mentiraa/>>. Acesso em: 17 de mai de 2014.

MAMEDE, G. **Big Data invade o e-commerce do estoque ao marketing**. E-Commerce News, [S.l.], 12 de jun de 2013. Disponível em: <<http://ecommercenews.com.br/artigos/cases/big-data-invade-o-e-commerce-do-estoque-ao-marketing>>. Acesso em: 10 de mai de 2014.

MARKETDATA. Disponível em: <<http://www.marketdata.com.br/Artigo.aspx?idConteudo=32>>. Acesso em: 9 de mai de 2014.

MATOS, T. M.; BRAGA, H.; KAWAMOTO JR.; L. T. **Big Data: quais os impactos e o que impulsiona**. In: VIII Workshop de Pós-Graduação e Pesquisa do Centro Paula Souza, São Paulo: Centro Paula Souza, out de 2013. pp 606-614.

MATTAR, F. N.; AUAD, M. **Nicho de mercado - um conceito ainda indefinido**. Fauze Najib Mattar, São Paulo, 21 e 22 de out de 1997. Disponível em: <<http://www.fauze.com.br/DOCUMENTOS/Nicho%20de%20mercado.pdf>>. Acesso em: 25 de mai de 2014.

MAYER-SCHÖNBERGER, V; CUCKIER, K. **Big Data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana**. 1 ed. Rio de Janeiro, Elsevier, 2013. 163 p.

MIRANDA, M. B. **Mercado consumidor**. Direito Brasil, [S.l.], [200-?]. Disponível em: <<http://www.direitobrasil.adv.br/arquivospdf/revista/revistav61/ensaios/mc.pdf>>. Acesso em: 25 de mai de 2014.

MOTTER, J.E. **O que é pesquisa? Como se faz pesquisa em educação?** Instituto de Matemática UFRGS, Rio Grande do Sul, [200-?]. Disponível em: <<http://www.mat.ufrgs.br/~vclotilde/disciplinas/pesquisa/pesquisa.pdf>>. Acesso em: 24 de mai de 2014.

MÜLLER, N. **Framework, o que é e para que serve?** Ofivina da Net, [S.l.], 20 de nov de 2008. Disponível em: <http://www.oficinadanet.com.br/artigo/1294/framework_o_que_e_e_para_que_serve>. Acesso em: 23 de abr de 2014.

O QUE É um framework? [S.l.], [200-?]. Disponível em: <<http://www.leffa.pro.br/textos/abnt.htm#4.7>>. Acesso em: 23 de abr de 2014.

PELOSINI, R. **Big Data, datawarehouse e BI.** Pelosini, [S.l.], 27 de set de 2012. Disponível em: <<http://pelosini.com.br/big-data-datawarehouse-e-bi>>. Acesso em: 30 ma de 2014.

PITANGA, M. **Computação em cluster.** Clube do Hardware, [S.l.], 30 de mai de 2003. Disponível em: <<http://www.clubedohardware.com.br/artigos/153>>. Acesso em: 26 de abr de 2014.

RAMOS, L. D. **Data warehouse.** DevMedia, [S.l.], 18 de out de 2009. Disponível em: <<http://www.devmedia.com.br/data-warehouse/12609>>. Acesso em: 30 de mai de 2014.

REZENDE, R. **Big Data e hadoop:** descubra o que é. SQL Magazine, n. 120, p. 26-30, [2014].

RIBEIRO, C. **O que é meme?** TechTudo, [S.l.], 7 de mai de 2012. Disponível em: <<http://www.techtudo.com.br/artigos/noticia/2012/04/o-que-e-meme.html>>. Acesso em: 30 de mai de 2014.

RICARTE, I. V. L. **Banco de dados relacionais.** Campinas, 15 de out de 2002. Disponível em: <<http://www.dca.fee.unicamp.br/cursos/PooJava/javadb/bdrel.html>>. Acesso em: 19 de abr de 2014.

SETTI, R. **De “moneyball” a davos:** o big data se abra para o mundo. O Globo, Rio de Janeiro, 4 de abr de 2012. Disponível em: <<http://oglobo.globo.com/sociedade/tecnologia/de-moneyball-davos-big-data-se-abre-para-mundo-4460918>>. Acesso em: 11 de abr de 2014.

SETZER, V. W. **Data, informação, conhecimento e competência.** *DataGamaZero*, Rio de Janeiro, n. 0, dez de 1999. Disponível em: <http://www.dgz.org.br/dez99/Art_01.htm>. Acesso em: 14 de mar de 2014.

SILVEIRA, S. A. **Marco civil e a proteção da privacidade.** Com Ciência, [S.l.], 10 de mai de 2014. Disponível em: <http://www.comciencia.br/comciencia/?section=8&edicao=99&id=1208&utm_source=hootsuite&utm_campaign=hootsuite>. Acesso em: 19 de mai de 2014.

TAMAKI, G. **Você realmente sabe o que é big data?** Computerworld, [S.l.], 24 de set de 2012. Disponível em: <<http://computerworld.com.br/blog/opiniao/2012/09/24/voce-realmente-sabe-o-que-e-big-data/>>. Acesso em: 30 de mai de 2014.

TAURION, C. **A internet das coisas.** [S.l.], [2012?]. Disponível em: <http://www.ibm.com/midmarket/br/pt/pm/internet_coisas.html>. Acesso em: 27 de mar de 2014.

TAURION, C. **Como repensar a privacidade com as novas tecnologias?** iMasters, [S.l.], 10 de fev de 2014. Disponível em: <<http://imasters.com.br/gerencia->

de-ti/tendencias/como-repensar-a-privacidade-com-a-novas-tecnologias/>. Acesso em: 17 de mai de 2014.

TAURION, C. **Conhecendo o hadoop**. iMasters, [S.I.], 27 de mar de 2012. Disponível em: <<http://imasters.com.br/artigo/21026/banco-de-dados/conhecendo-o-hadoop/>>. Acesso em: 25 de abr de 2014.

TAURION, C. **Privacidade em tempos de big data**. IBM, [S.I.], 25 de nov de 2013. Disponível em: <https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/privacidade_em_tempos_de_big_data?lang=en>. Acesso em: 17 de mai de 2014.

TAURION, C. **V de valor em big data**. IBM, [S.I.], 28 de out de 2013. Disponível em: <https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/v_de_valor_em_big_data?lang=en>. Acesso em: 25 de abr de 2014.

TLCBrazil. **Banco de dados noSQL**. IBM, [S.I.], 24 de mai de 2013. Disponível em: <https://www.ibm.com/developerworks/community/blogs/tlcbr/entry/bancos_de_dados_nosql?lang=en>. Acesso em: 24 de abr de 2014.