

ANÁLISE DE NEGÓCIO COM BIG DATA¹

NATÁLIA MENDES PETRUCELLI²

Resumo: Este trabalho tem por objetivo explicar o conceito de Big Data, mostrar suas características, utilidades e exemplificar os segmentos utilizados. Além disso, procura mostrar a importância do sistema para as empresas que ainda não conhecem Big Data e desejam ampliar seus horizontes buscando formas rápidas e inteligentes que auxiliam na tomada de decisão. Com o mercado cada vez mais competitivo, as empresas que possuem esses diferenciais poderão ter maiores oportunidades de sucesso.

Palavras-chave: Big Data, Análise de dados

Abstract: This study aims to explain the concept of Big Data, it shows its features, utilities and exemplify its used segments. Moreover, attempts to show the importance of the system for companies that do not yet know Big Data and wish to enlarge their horizons seeking quick and clever ways that help to make decisions. With the increasingly competitive market, companies that have such differentials may have greater opportunities for success.

Keywords: Big Data, Data analysis

1 INTRODUÇÃO

Nossa geração está vivendo e presenciando a era da informação. A todo o momento estamos gerando dados sobre que fazemos, seja algo relevante ou não. A quantidade de dados gerados está atingindo escalas imensuráveis dentro de um cenário que, quanto mais rápido processarmos essas informações e obtermos um resultado relevante, melhor será a tomada de decisão. É exatamente essa a ideia de Big Data dentro de uma organização.

Pode-se citar como exemplos de grandes quantidades de dados: transações bancárias, tráfego de dados de celulares, gerenciamento de redes sociais, entre outros.

Vale ainda mencionar que o acúmulo de dados não é algo atual, mas com as mudanças nas organizações e com softwares mais precisos, a análise de dados se tornou mais rápida.

¹ Trabalho orientado pelo Prof. Ms. Felipe Diniz Dallilo

² napetrucelli@gmail.com

Universidade de Araraquara (UNIARA)

Muitas informações são geradas diariamente num ambiente corporativo. Além do dinamismo que já conhecemos, existe uma preocupação crescente em acompanhar o mercado, tendências, finanças, previsões de vendas e, acima de tudo, as decisões precisam ser as mais corretas e ágeis possíveis. Segundo Taurion (TAURION, 2013), no levantamento feito pela IDC em 2013 o total de dados criados e replicados foi de quatro ZB (zetabytes), quase 50% a mais que 2012 e quatro vezes maior que em 2010. A previsão é que em 2020 esse número chegue em até 40 ZB (LOBO, 2012).

2 CARACTERÍSTICAS

Como principais características de Big Data tem-se o conceito dos três “V’s”. Esse conceito ganhou prospecção em 2001, quando o analista Doug Laney introduziu em uma publicação no MetaGroup (KNILANS, 2014).

- **Volume:** é a quantidade de dados em grande escala e que cresce exponencialmente;
- **Velocidade:** o tratamento de dados deve ser feita em tempo real para garantir a veracidade das informações e ter um melhor resultado nas soluções dos problemas das organizações;
- **Variedade:** com o grande volume de dados existentes atualmente, sabe-se que é possível encontrar dados em diversos formatos, como estruturados, ou seja, armazenados em bancos de dados, além de outros localizados em outras fontes, como imagens, vídeos e documentos.

Com a quantidade de dados que cresceram nos anos posteriores e com o que se aprendeu sobre Big Data, a IBM incluiu um quarto “V”, conforme descrito abaixo:

- **Veracidade:** não vale somente a quantidade de dados. É preciso que exista veracidade nas informações. Isso acrescenta um novo desafio para o sistema, que é organizar as grandes quantidades de dados.

O volume, velocidade, variedade e veracidade que estão gerando atualmente, vão além do que os sistemas tradicionais podem lidar de forma eficiente. A consequência disso é o quinto “V” que as organizações buscam em seus dados:

- **Valor:** resultado de todos os aspectos citados acima procurando benefícios que compense todo o investimento, através da mineração e análise complexa de dados. Esse conceito traz às empresas a capacidade de melhor análise de dados com o objetivo de acelerar o crescimento rentável.

2.1 Importância

Uma das maiores importâncias do Big Data não é a quantidade de dados que são aglomerados, e sim o que se faz com eles. Através da análise correta dos dados, podem-se encontrar respostas como: redução de custo, redução de tempo, desenvolvimento de novos produtos, decisões mais inteligentes. Apesar disso, segundo Hekima (HEKIMA, 2015), ainda são poucas as organizações que estão se beneficiando com tal tecnologia. Esse paradigma tende a mudar com a grande concorrência entre as empresas. Hekima (HEKIMA, 2015) afirma também que até 2024, os servidores empresariais do mundo vão processar anualmente o equivalente digital a uma pilha de livros que se estenderia para mais de 437 anos luz. Essa quantidade de dados é imensa e quase impensável de reunir ou analisar.

Para tal análise das informações que existem atualmente, necessita-se de um Cientista de Dados, profissional capaz de entender e transformar os dados disponíveis. As etapas que devem ser concluídas para auxiliar na tomada de decisão incluem negócios, compreensão e modelagem de dados.

2.2 Utilidades

Através das análises dos dados, pode-se criar diversos cenários empresariais para aplicar os resultados obtidos, dependendo do foco. O Big Data auxilia os gestores a entenderem e preverem o comportamento e percepções dos consumidores, além de elaborar estratégias para fidelizá-los. É capaz de analisar o histórico de vendas e suas sazonalidades, podendo apontar a quantidade ideal de mercadorias que deverá manter em estoque, além dos produtos mais vendidos. Com as grandes informações capturadas de blogs e redes sociais, Big Data também é capaz de avaliar e mostrar o que realmente os clientes esperam dos produtos. Além disso, com a coleta de dados relacionados à idade, renda, localização dos consumidores, a empresa economiza recursos com campanhas mais direcionadas.

Vale ressaltar que Big Data não é só útil em organizações, mas também em setores públicos e privados. Nas eleições auxiliam nas análises dos eleitores e na área da saúde, pode se tornar uma importante ferramenta para diagnosticar pessoas com determinadas doenças, quantidades de remédios por cidades, além de regiões por onde uma epidemia se espalhou.

A informação sozinha não vale nada, mas através de análises inteligentes, boa interpretação, modelagem de dados e uso das ferramentas corretas, amplia as possibilidades e recursos para que a tecnologia seja de grande utilidade para diversos segmentos.

2.3 Vantagens

Os consumidores são tratados com exclusividade e ganham importância no que diz respeito a anseios, desejos e comportamento. Já as empresas se beneficiarão dessas informações e entenderão qual cliente precisa de qual produto ou serviço. Basicamente depende da criatividade

de quem cria e absorve corretamente as informações, atrelando as necessidades de quem usa. Segundo Stela (STELA, 2015) uma das maiores vantagens da tecnologia é a capacidade de transformar o negócio mudando a maneira como as informações são analisadas e utilizadas na tomada de decisão.

2.4 Desvantagens

Como desvantagem dessa tecnologia, vale citar que dados imprecisos, incompletos ou manipulados de forma fraudulenta, poderão representar um risco para tomadores de decisão. Se não houver consistência nos dados, por consequência, as análises serão imprecisas e não terão credibilidade. Por isso é tão importante que o cientista de dados seja capacitado para esta atividade.

3 BIG DATA

O termo Big Data tem sido usado para definir grande quantidade de dados armazenados, mas, além disso, é importante mencionar que são principalmente dados não estruturados, como por exemplo: logs de acesso, frases escritas por clientes ou funcionários, dados de redes sociais, posições geográficas, e-mails, arquivos no servidor, entre outros (SANTOS, 2015). Os profissionais de T.I e todos que acompanham a tecnologia sabem que a evolução tecnológica tem seguido a Lei de Moore, que prevê a capacidade de processamento e armazenamento dobrando a cada 18 meses. Em contrapartida, a Lei de Parkinson, quando aplicada à computação, decide que toda a capacidade disponível tende a ser ocupada, gerando mais dados armazenados para analisar, mas sendo inviável em ferramentas convencionais (SANTOS, 2015).

Poderoso (PODEROSO, 2016) cita que para atender a demanda de dados necessita-se de uma arquitetura compatível. O atual momento tecnológico permitiu estabelecer formas de armazenar dados não estruturados. Para tal, foi criado um novo padrão de banco de dados chamado de NOSQL (Not Only SQL). Com isso, arquivos de textos, imagens, voz, que são armazenados, voltam a fazer parte do contexto de análise de dados. Sabe-se que dados estruturados são facilmente tratados em banco de dados padrão SQL, que são formatados com base em linhas e colunas e uma sintaxe robusta, além de modelagem consistente. Já para dados não estruturados, utiliza-se o próprio sistema de arquivos, seja Linux ou Windows.

3.1 Arquitetura

Em termos de arquitetura, Big Data utiliza um conjunto de técnicas e ferramentas com capacidade de processamento e armazenamento em paralelo, sem necessidade do uso de um único computador com capacidade maior.

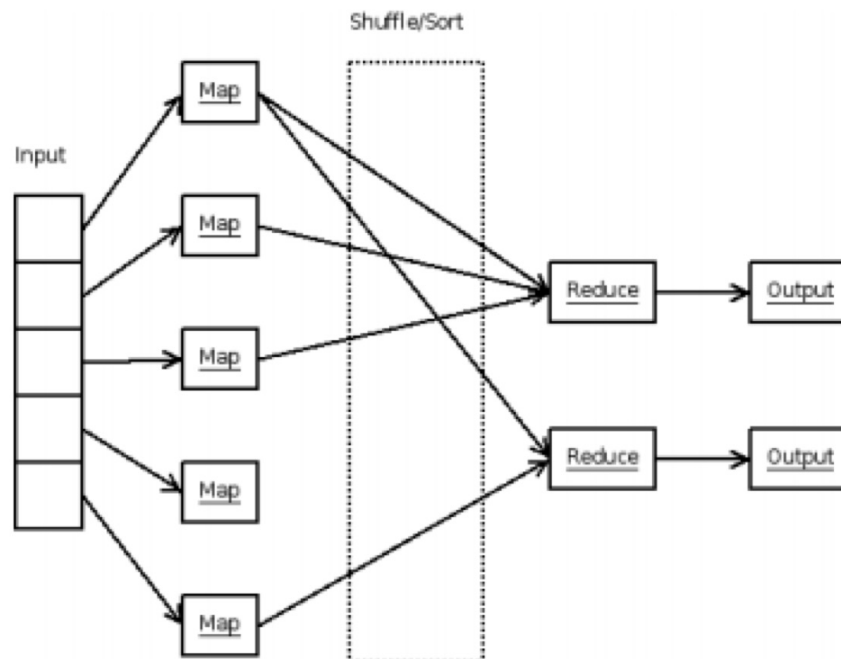
A base arquitetônica deverá ser feita com base na natureza de análises que as empresas desejam. Mesmo que as análises sejam em tempo real, também terão necessidade de armazenar

uma quantidade de dados. Também se deve pensar na quantidade certa de redundância para que esteja protegido de eventuais latências e inatividades.

Segundo Poderoso (PODEROSO, 2016), os principais componentes da arquitetura são:

- **Hadoop**: plataforma open source de armazenamento e processamento de grandes volumes de dados utilizando hardware simples e que utilizam clusters para agilizar o acesso e manipulação de dados;
- **MapReduce**: modelo de programação paralela, escalável e que permite a utilização de um hardware simples na realização de trabalhos complexos conforme mostra a figura 1. A Função do operador Map é dividir o problema em subconjuntos pequenos e que possam ser distribuídos por outros nós de clusters. Já a operação Reduce tem o objetivo de agregar e tratar as informações dos subconjuntos e responder ao problema original. Ao contrário do SQL, o MapReduce armazena os dados de forma persistente para consultas posteriores.

Figura 1 – Fluxo do processo MapReduce - Fonte - <http://www.lecturemaker.com>



- **NoSQL**: banco de dados que permite armazenar e recuperar dados com restrições inferiores aos bancos de dados relacionais. Permite aumentar a escalabilidade e disponibilidade do ambiente. A necessidade de uma nova tecnologia e banco de dados surgiu como consequência da ineficiência dos bancos de dados relacionais em lidar com essa tarefa.
- **SQL**: bancos de dados tradicionais que armazenam a maior parte dos dados estruturados das organizações. Os dados normalmente têm origem em sistemas ERP, SCM (Supply Chain), CRM, etc.;

- **DW:** o Data Warehouse é um banco de dados dos sistemas tradicionais que são modelados para auxiliar na tomada de decisão.

Para trabalhar com Big Data, necessita-se trabalhar com algum dos componentes citados e em conjunto, pois isoladamente não conseguem o resultado esperado.

As tecnologias Hadoop e MapReduce são projetadas os três Vs citados anteriormente (volume, velocidade e variedade) de dados (INTEL, 2012). Essas tecnologias também exigem mais esforços de infra-estrutura para atender o processamento distribuído de análise de dados não estruturados.

3.2 Mapa Mental

Com o mapa mental (figura 2) foi possível evidenciar os principais elementos, preocupações e necessidades de um ambiente Big Data (PODEROSO, 2016). Um aspecto interessante de Big Data é a comparação com o modelo convencional de armazenamento de dados para tomada de decisão.

Figura 2 – Mapa Mental Big Data - Fonte: Visão geral da arquitetura de Big Data



A visão de negócio é essencial para um projeto desse porte, afim de identificar dados que permitam uma visão de competitividade de mercado envolvendo conhecer o cliente, o produto, a cadeia de distribuição, fornecedores e concorrentes. O armazenamento de dados engloba o

recolhimento dos dados e o armazenamento em ambientes, em bancos de dados SQL e Data Warehouses.

A escalabilidade tem relação com a demanda que o volume representa. Quanto maior o volume, maior capacidade deverá ser oferecida. Além disso, envolve a variedade das origens encontradas no projeto. Coleta e integração de dados são tarefas realizadas pela ETL (Extract Transform Load), ou seja, Extração Transformação Carga. Segundo (Ribeiro) ,o ETL visa trabalhar com toda a parte de extração de dados de fontes externas, transformação para atender às necessidades de negócios e carga dos dados dentro do Data Warehouse. Desta forma é possível mover dados de um sistema para outro com rapidez e eficiência, organizando de maneira estruturada os dados “sujos”, mantendo a consistência das análises.

Em relação a segurança, não está relacionada apenas à questão de Governança e seus acordos de acesso, mas também está ligada à questão ética. Neste caso, os dados que são públicos são avaliados e destinados de maneira correta.

A visualização de dados determina o uso e a prática de técnicas estatísticas adequadas para responder às questões de negócio que justificarão o desenvolvimento do projeto. Já as análises avançadas tem relação com o aprendizado de máquina, utilizando algoritmos testados e desenvolvidos para que o modelo de previsão permita que a velocidade atenda às necessidades do negócio.

Todos os elementos citados são de tamanha importância e devem ser considerados em projeto de Big Data.

3.3 Algoritmos

Apesar do foco em Big Data estar voltado para volume e variedade de dados, sem a analítica ou algoritmos, de nada serviria o grande volume de informações e as análises não seriam possíveis. Um dos principais algoritmos utilizados em Big Data é o chamado algoritmo preditivo, muito presente em setores como meteorologia.

O modelo preditivo é uma função matemática que aplicada em massa de dados, tem capacidade de identificar padrões ocultos e prevê o que poderá ocorrer no futuro com base em probabilidades (TAURION, 2015). Existem dois tipos de modelos preditivos, os supervisionados e os não supervisionados. No primeiro modelo, os dados de entrada e saída desejada são apresentados juntos. Neste momento ocorre uma fase chamada de treinamento, cujo tempo de duração é até que o modelo aprenda a mapear os dados e identificar os padrões. Já os modelos não supervisionados recebem apenas os dados de entrada e sua tarefa é descobrir os relacionamentos entre os dados apresentados.

Um exemplo prático do uso desse tipo de algoritmo foi um trabalho realizado pela empresa Hekima para o Governo Federal durante a Copa do Mundo no Brasil. Utilizando

algoritmos preditivos, a empresa analisou a movimentação das ruas e redes sociais para prever quais manifestações ficariam mais cheias, quais seriam mais violentas e quais seriam pacíficas. Com essas análises, as ações dos órgãos federais deixaram de ser intuitivas e ganharam maiores chances de sucesso.

Quando englobam análise preditiva, formulas matemáticas e estatísticas possibilitam a descoberta de padrões ocultos nos dados que os humanos não enxergariam (HEKIMA, 2016).

4 ANÁLISE DE NEGÓCIO

São poucas empresas que contam com um projeto de Big Data e possuem todo o benefício prometido pela indústria de tecnologia, principalmente no que diz respeito à tomada de decisões e previsões nos mais diversos cenários de negócios (CAVALCANTI, 2016). Na visão de Bill Schmarzo, CTO de Global Services em Big Data da EMC, a resposta desse problema está na forma com que as organizações levam o tema. A discussão precisa deixar de ser apenas da área de TI e deverá ser um debate de negócios.

Além de trabalhar na visão da organização, outro desafio das empresas é encontrar o profissional adequado. Como já citado anteriormente, o cientista de dados é o responsável por pela iniciativa em Big Data na empresa. Bill Shmarzo avalia que o profissional deverá pensar como uma pessoa de negócios para gerar análise e fazer a leitura de dados.

Um fato que encontramos no mundo é que as empresas criadas após a internet possuem Big Data em seu “core bussiness”. Pode-se citar como exemplo a Amazon, Facebook, Google, e-bay, LinkedIn, AirBnB, TripAdvisor, Groupon, entre outras. Para essas empresas, os modelos e processos foram criados sob o conceito de imensas quantidades de dados para explorar e, naturalmente, ser base dos negócios (TAURION, 2015). Entre as empresas que surgiram antes da internet, apenas uma pequena parcela aderiu ao conceito de Big Data e aplica de forma sistemática.

Mendes (MENDES, 2012) afirma que para obter as vantagens trazidas pelo Big Data é ter o conhecimento profundo do negócio para chegar a combinação ideal de dados e informações sobre o cliente e o mercado e, dessa forma, possa favorecer a estratégia da empresa.

4.1 Desafios

De acordo com uma pesquisa realizada em 2013 pela EMC Brasil, em 73% das empresas a maior barreira é a própria cultura. Essa pesquisa destaca também que em 88% das companhias entendem que o maior desafio é capacitar seus trabalhadores para essa tecnologia (LOBO, 2013). O grande entrave é que o Brasil não possui cultura de compartilhar informações. O Marketing não quer compartilhar seus dados com profissionais de TI ou com a equipe de Negócios.

Se olhar o mesmo cenário em 2016, pode-se ver que ainda o país está aquém do que

poderia se tornar em relação à tecnologia. Segundo a ASM (ASM, 2016), assim como existe em outros países, o Brasil sofre uma escassez de recursos para projetos de Big Data. Isso se deve ao pouco investimento das universidades nessa área por conta de custos de infraestrutura, falta de financiamento, no caso em universidades públicas e falta de informações sobre o tema. Apenas em 2014 iniciaram treinamentos ou cursos voltados ao tema Big Data (ASM, 2016). Além desses fatores, outro que afeta diretamente é a crise econômica atual que desacelerou o crescimento da demanda, mesmo com investimentos na área.

Os profissionais que tiverem capacidade de modelagem, análise e interpretação terão grandes oportunidades em Big Data. Nos próximos anos, a tecnologia Big Data será uma das principais fontes competitivas para as empresas em praticamente todos os setores, com destaque para marketing, vendas, RH e finanças (ASM, 2016).

4.2 Oportunidades

A grande oportunidade está em saber gerenciar grandes volumes de dados em grandes volumes em questão de armazenamento e acesso não é problema; a grande sacada está na capacidade de tratar os dados armazenados e transformá-los em informações, na velocidade, veracidade e qualidade que as empresas precisam (BARBOSA, 2016). Para atingir esse sucesso é necessário que hajam profissionais capacitados em realizar processo de análise de dados. Quanto mais especializações em conhecimento, elaboração e implementação de algoritmos, mais preparado o profissional estará nessa área que se mostra promissora no Brasil e no mundo.

5 CONCLUSÃO

Ao final do trabalho pode-se concluir que Big Data é uma tecnologia em grande crescimento no mercado, mas que ainda muitas empresas são carentes de infraestrutura e de profissionais capacitados. Como forma de mudar as culturas das empresas, existem iniciativas, tanto do governo quanto de fundações, como por exemplo a DataViva, que oferece ferramentas de visualização, cruzamento e estudo de informação no Ministério do Trabalho junto com dados de exportação de bens e serviços. Também vale citar o DataRio, que disponibiliza dados públicos à qualquer pessoa que deseja criar seu aplicativo com base nesses dados (MENDES, 2012).

Sabe-se que as ferramentas não farão todo o trabalho por si só. A empresa necessitará de um profissional capacitado para analisar, modelar e mostrar o que é relevante. O profissional usará as ferramentas para ter um resultado que auxilie na tomada de decisões. Com esse cenário de possibilidades e crescimento, empresas que adotam o Big Data tendem a evoluir da mesma maneira. As empresas que souberem aproveitar da melhor forma a demanda de dados que vem aumentando ao longo dos anos poderão levar vantagem no mercado.

REFERÊNCIAS BIBLIOGRÁFICAS

ASM. **Big Data: Profissionais envolvidos e desafios impostos pela escassez.** [S.l.], 2016. Disponível em: <<http://www.asmlatin.com/2016/06/big-data-desafios-impostos-pela-escassez-de-profissionais/>>.

BARBOSA, M. **Especialistas explicam como aplicar Big Data na prática.** [S.l.], 2016. Disponível em: <<http://startupi.com.br/2016/04/especialistas-explicam-como-aplicar-big-data-na-pratica/>>. Acesso em: 18 de Novembro de 2016.

CAVALCANTI, V. **Big Data precisa ser abraçado pelas áreas de negócios.** [S.l.], 2016. Disponível em: <<http://itforum365.com.br/noticias/detalhe/119683/big-data-precisa-ser-abracado-pelas-areas-de-negocios/>>. Acesso em: 25 de outubro de 2016.

HEKIMA. **Qual a Importância de ter uma cultura de Big Data em todas as áreas da empresa?** [S.l.], 2015. Disponível em: <<http://www.bigdatabusiness.com.br/qual-a-importancia-de-ter-uma-cultura-de-big-data-em-todas-as-areas-empresa/>>. Acesso em: 22 de setembro de 2016.

HEKIMA. **O que é análise preditiva?** [S.l.], 2016. Disponível em: <<http://www.bigdatabusiness.com.br/o-que-e-analise-preditiva/>>. Acesso em: 25 de outubro de 2016.

INTEL, I. E. C. **Big Data 101: Unstructured Data Analytics A Crash Course on the IT Landscape for Big Data and Emerging Technologies.** [S.l.], 2012. Disponível em: <<http://www.intel.com/content/www/us/en/big-data/unstructured-data-analytics-paper.html>>. Acesso em: 25 de outubro de 2016.

KNILANS, E. **The 5V's of Big Data.** [S.l.], 2014. Disponível em: <<http://ats.avnet.com/na/en-us/news/Pages/The-5-Vs-of-Big-Data.aspx>>. Acesso em: 24 de outubro de 2016.

LOBO, A. P. **Big Data: Em 2012, 2,8ZB de dados criados e replicados.** [S.l.], 2012. Disponível em: <<http://convergenciadigital.uol.com.br/cgi/cgilua.exe/sys/start.htm?UserActiveTemplate=site&inford=32617&sid=97>>. Acesso em: 18 de junho de 2016.

LOBO, A. P. **Falta de profissionais é o desafio do big data no Brasil.** [S.l.], 2013. Disponível em: <<http://convergenciadigital.uol.com.br/cgi/cgilua.exe/sys/start.htm?UserActiveTemplate=site&inford=34475&sid=97>>.

MENDES, D. **Big Data trazendo novas oportunidades e vantagens competitivas.** [S.l.], 2012. Disponível em: <<http://imasters.com.br/artigo/24944/banco-de-dados/big-data-trazendo-novas-oportunidades-e-vantagens-competitivas/?trace=1519021197&source=single>>. Acesso em: 24 de Outubro de 2016.

PODEROSO, C. **Visão geral da arquitetura de Big Data.** [S.l.], 2016. Disponível em: <<https://corporate.canaltech.com.br/coluna/big-data/Visao-geral-da-Arquitetura-de-Big-Data/>>. Acesso em: 03 de outubro de 2016.

SANTOS, F. U. d. **O que é Big Data e como funciona?** [S.l.], 2015. Disponível em: <<https://www.profissionaisti.com.br/2015/03/o-que-e-big-data-e-como-funciona/>>. Acesso em: 24 de outubro de 2016.

STELA, J. **Tudo sobre Big Data: a revolução da análise de dados.** [S.l.], 2015. Disponível em: <<http://www.t-systems.com.br/imprensa-analistas/tudo-sobre-big-data-a-revolu-o-da-an-lise-de-dados/1303392>>. Acesso em: 24 de Outubro de 2016.

TAURION, C. **Big Data.** Rio de Janeiro: Editora Brasport, 2013. 781 p.

TAURION, C. **Algoritmos preditivos no desenvolvimento de projetos Big Data.** [S.l.], 2015. Disponível em: <<http://litterisconsulting.com.br/algoritmos-preditivos-no-desenvolvimento-de-projetos-big-data/>>. Acesso em: 25 de outubro de 2016.