

Fundamentos e aplicação do Big Data: como tratar informações em uma sociedade de *yottabytes*¹

Max Silva Caldas

Graduado em Sistema de Informação pela Universidade CEUMA. Analista de Negócios/Implantação na TOTVS/MA.

Emanoel Costa Claudino Silva

Mestre em Ciência da Computação pela Universidade Federal do Maranhão. Atualmente é Professor de nível superior e Analista/DBA.

Resumo: Há algum tempo, uma questão que os cientistas da computação tentam resolver é o que fazer para administrar a incalculável quantidade de dados gerados a todo momento. Isso remete à análise e busca de respostas em relação às áreas onde o Big Data é utilizado. Pois cada vez mais se tem a necessidade de processar dados de forma eficaz e mais rápida. Esporte, saúde e varejo são alguns dentre os incontáveis campos onde o Big Data já faz parte do cotidiano. Com isso, o objetivo desta revisão é conceituar e analisar o Big Data e apresentar os recursos tecnológicos envolvidos, já que diante do crescimento irrefreável de produção de dados, verifica-se que a administração do Big Data pode-se constituir na menina dos olhos da informatização nos próximos anos.

Palavras-chave: Big Data. Compartilhamento em nuvem. Geração de Conhecimento. Data Warehouse. Competitividade.

1 Introdução

Todas as informações que são geradas ininterruptamente pelos mais diversos meios – em imensa quantidade, velocidade, variedade, veracidade, valor –, como arquivos de textos,

¹ Um yottabyte é uma unidade de medida equivale a 10^{24} bytes. Ou a 1.000 Zettabytes, 1 milhão de Exabytes ou 1 bilhão de Petabytes, onde 1 Petabyte equivale a 1.000 Terabytes, ou 1 milhão de Gigabytes. O nome "Yotta" é originado da nona letra do Alfabeto grego(I ou ι), iota.

códigos, CEFV, sistemas corporativos, redes sociais, entre outros, se tornam de certa forma um grande leque de dados desordenados, mas que podem ser analisados, transformados e utilizados na solução de diversos problemas. A essa avalanche de dados, desconexos entre si, deu-se o nome de Big Data.

Nos anos de 1970, no Brasil era comum e mesmo prático guardar e recuperar informações em diários, livretos e cadernetas, por períodos curtos de tempo. Como exemplo, pode-se citar o uso de pequenos pedaços de papéis utilizados por comerciantes para registrar as vendas realizadas no dia. Porém, já não é possível gerenciar uma empresa que não esteja informatizada em razão da grande quantidade de dados e da velocidade com que as informações se processam. E com essa informatização crescente, as informações são geradas, cada vez mais, em maior quantidade e variedade. É nessa realidade que houve a necessidade de tratar essas informações de forma a agregar valor às organizações.

Preocupações que vão desde conhecer o comportamento do consumidor a, por exemplo, melhorar o desempenho de um atleta, analisando padrões encontrados em cada partida para que no decorrer da carreira o nível de excelência aumente cada vez mais. Na saúde, tornou-se muito mais rápido o acesso a dados comparativos entre as doenças: uma questão essencial, nos dias de hoje, é poder analisar e comparar resultados de exames de pessoas com sintomas semelhantes, como aquelas diagnosticadas com câncer, com exames de outros pacientes com a mesma doença, para chegar a um diagnóstico mais preciso e rápido, tendo como consequência melhor precisão no tratamento. São questões como essas que há alguns anos poderiam ficar sem resposta, ou, cujas informações levariam muito tempo para poder ser reunidas, tempo que um paciente não tem. Hoje o cenário foi modificado graças ao Big Data, podendo, assim, não só obter respostas rápidas e precisas, mas mostrar fatos e resultados.

O objetivo geral deste estudo é conceituar e analisar o Big Data, buscando desmistificar o que hoje é visto como tema complexo para o público, em geral, e apresentar os recursos tecnológicos envolvidos, mostrando como o Big Data se faz presente e imprescindível, e como é necessário saber lidar com isso na atualidade.

Tendo em vista o interesse e as dificuldades para compreensão do Big Data, a existência de um trabalho, que aborde e analise o tema, torna-se relevante pelo fato de auxiliar na

introdução de profissionais ou leigos ao tema, ampliando sua grade de conhecimento voltada para o armazenamento de dados.

O intuito desta pesquisa é sintetizar os principais conceitos relacionados ao Big Data e tecnologias que o abrangem, como: Data Mining, Data Warehouse, Data Mart e BI; analisar as aplicações do Big Data.

2 Fundamentos Tecnológicos do Big Data

Para aprofundar-se no estudo de Big Data é necessário abordar quais são seus predecessores e as tecnologias que ainda são utilizadas, de forma paralela, a complementar e a suportar o uso desse novo conceito.

Com o crescente volume de dados e a demanda para extrair informações sobre esses dados, que na maioria das vezes são armazenados sem destino e ou objetivo certo, surge à necessidade de evoluir para uma tecnologia mais ampla.

Nas seções a seguir, serão abordadas tecnologias-base, como Data Warehouse, Data Mart, Data Mining, BI e Computação em Nuvens, além de uma das maiores fontes de dados para o Big Data, as Redes Sociais.

2.1 Data Warehouse (DW)

Um Data Warehouse possui apenas duas operações básicas, que são a inserção de dados inicial e incremental e a consulta a esse repositório de dados, por meio de leitura. Com isso, pode-se dizer que um DW é diferente de sistemas tradicionais, portanto, os requisitos de SGBD desses dois sistemas são diferentes.

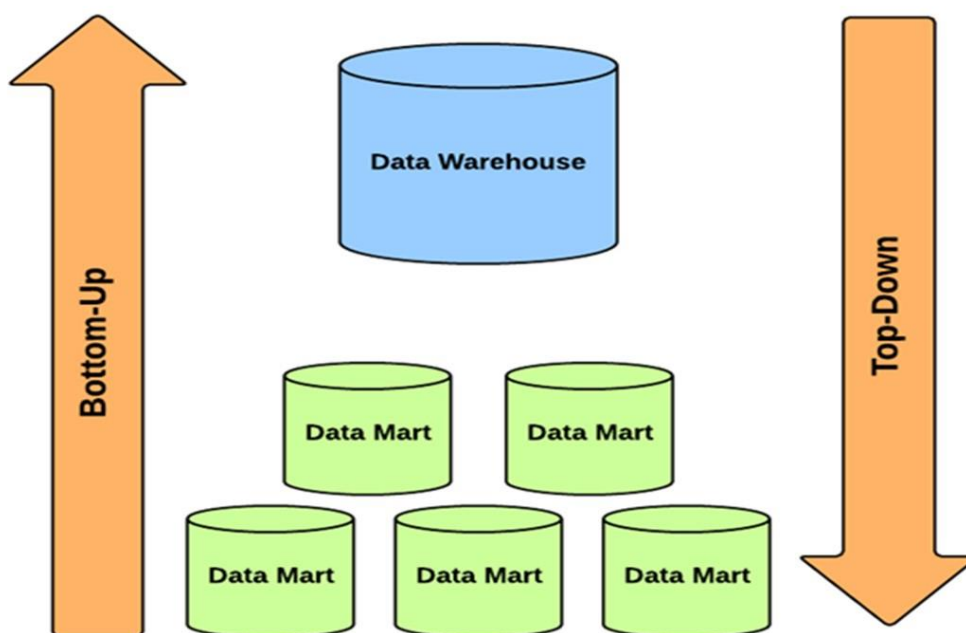
Em um DW, não é preciso preocupar-se com impasses de *deadlock* ou até mesmo de atualizações de registro a registro. Os dados que são trazidos e/ou alimentados de ambientes operacionais ficam no DW apenas enquanto são necessários, assim que perdem a função para análises e tomadas de decisões são descartados.

Como no DW não existe bloqueio de *deadlock* e as operações se restringem a consultas, não há necessidade de bloqueio por concorrência de usuários no acesso às informações.

2.2 Data Mart (DM)

Pode-se dizer que os Data Marts são Data Warehouse orientados a departamentos, como departamentos de uma organização (comercial, administrativo, financeiro, etc.), em que cada departamento poderá ser considerado um Data Mart individual. Depois de um período estipulado um DM, poderá constituir ou formar um DW – neste caso, pode-se dizer que um DM é um subconjunto de DW.

Um DW após ser implantado por vários DM passa a realizar o caminho inverso de um DM, alimentado, assim, os DM separadamente. Com isso, pode-se afirmar que os DM surgem de duas formas, como na Figura 1: *Top-down* e *Bottom-up*.



Fonte: Fonte: ELIAS, (2014).

Como exemplo, pode-se dizer que um DM surge de forma *Top-down* quando uma organização cria um DW e depois o segmenta para os departamentos, ou seja, divide o DW

em áreas menores, formando, assim, pequenos bancos de dados orientados por departamento e/ou assunto.

O DM surge de forma *Bottom-up* quando a situação é inversa à *Top-down*, quando por estratégia de negócio se criam pequenos bancos de dados e depois os unem para formar uma área inteira (DW). Assim que os resultados são alcançados, esses pequenos bancos (DM) são integrados e formam um DW.

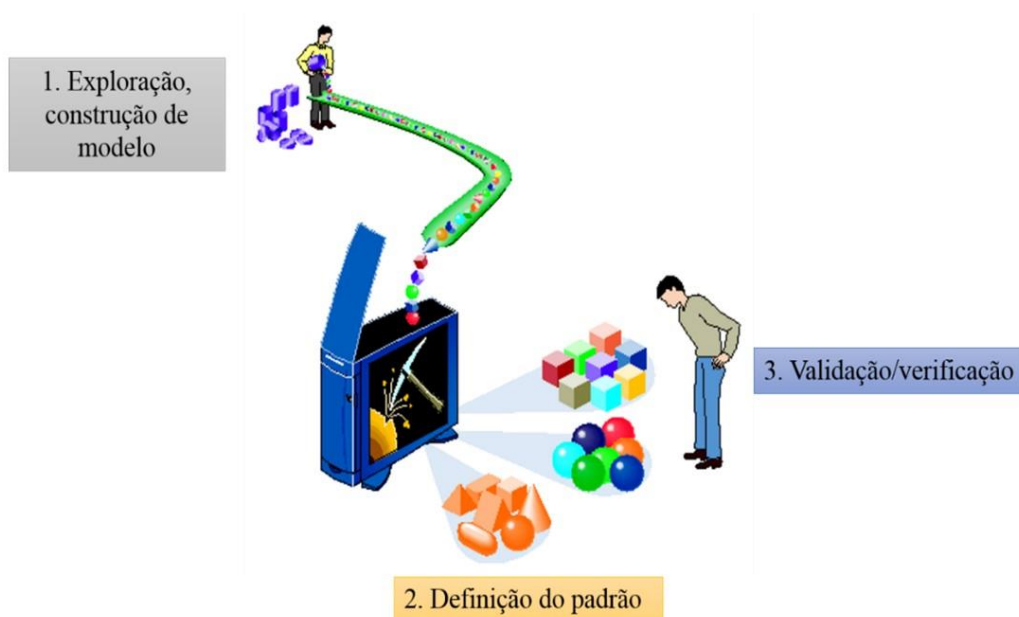
Podem-se encontrar inúmeras semelhanças entre DM e DW, entretanto, a infraestrutura de software e hardware é bem diferente. Em um DM o modelo de dados é dimensional, em que as informações se relacionam como um cubo, podendo-se fatiar esse cubo e aprofundar-se em cada dimensão ou eixo para extrair mais detalhes sobre os processos internos que ocorrem numa organização. Já no DW pode haver modelos híbridos ou completamente relacionais.

Segundo Machado (2010), a principal vantagem do uso de um DM é o retorno rápido que ele possibilita, além de garantir maior participação do usuário final, que é capaz de avaliar seus benefícios.

2.3 Data Mining (DN)

O termo DN, utilizado nessa pesquisa para designar Data Mining, é conhecido também como mineração de dados, consiste em um processo analítico projetado para explorar grandes quantidades de dados (tipicamente relacionados a negócios, mercado ou pesquisas científicas), na busca de padrões consistentes e/ou relacionamentos sistemático entre variáveis para, então, validá-los aplicando os padrões detectados a novos subconjuntos de dados.

Visando transformar esses dados em conhecimento, criou-se o processo conhecido por Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases – KDD*). (Figura 2)



Fonte: do autor.

Devido à disponibilidade de enormes quantidades de dados em formato digital e a necessidade iminente de extrair desses dados informações e conhecimentos úteis a diversas aplicações, o Data Mining foi popularmente tratado como sinônimo de descoberta de conhecimento em bases de dados.

A premissa de um Data Mining é uma argumentação ativa, em que, uma vez definido o problema, os dados e a ferramenta de análise, o Data Mining pesquisa, automaticamente, nesse montante de dados, anomalias e prováveis relacionamentos, encontrando possíveis problemas que não foram identificados anteriormente pelos usuários.

Usando técnicas de estatística e inteligência artificial, além de reconhecimento de padrões e recuperação de informações, como também o Data Mining é possível fazer uso de algoritmos de aprendizagem ou classificação baseada em redes neurais, para explorar um conjunto de dados, retirando e auxiliando na identificando de padrões, favorecendo, assim, a descoberta por meio do conhecimento.

Um DN usa repositórios, como DW e DM, que permitam selecionar colunas e dados a ser cruzados no processo de mineração.

2.4 Business Intelligence (BI)

Business Intelligence é uma tecnologia que tem como conceito básico a entrega da informação coletada a partir de dados do Data Warehouse, de forma exata e útil, para a tomada de decisões.

Uma solução de BI permite monitorar o desempenho dos processos operacionais, táticos ou estratégicos por meio de indicadores de desempenho e apresentá-los em painéis de controle ou *dashboard*, com recursos analíticos e interativos que permitem cruzar e analisar informações, no tempo em que se precisa, transformando o processo de decisão em algo simples, rápido e eficiente.

As soluções de BI também englobam diversas técnicas e ferramentas para coletar dados de diversas fontes de dados. Entretanto, o BI coleta dados relacionais; já o Big Data usa dados não relacionais e fontes inusitadas.

2.5 Redes Sociais

Nas últimas décadas, grande parte das mudanças na sociedade se deu ao avanço tecnológico e à forma de comunicação mais acessível e rápida. Isso impactou diretamente na forma de pensar e agir das pessoas, principalmente, em seu comportamento social. Esse avanço tecnológico também influenciou muito em todas as áreas sociais, tais como a política e a economia. Um exemplo disso é o uso das redes sociais como plataforma de campanhas políticas por partidos eleitorais.

Pode-se dizer que as redes sociais são, de certa forma, uma estrutura social composta por pessoas e/ou organizações, conectadas por um ou vários tipos de relação, que partilham valores e objetivos comuns.

Não se pode contestar que as redes sociais já são uma influência importante na vida das pessoas. Antes as redes sociais tinham como objetivo a comunicação entre os membros (sociedade), além de partilhar valores e sentimentos em comum, entretanto, hoje há redes

sociais para quase tudo, e, a cada momento, mais recursos são adicionados às inúmeras redes sociais existentes.

Além dos portais das principais redes de rádio e TV, jornais e revistas tradicionais, as redes sociais como Facebook, Twitter entre outras, também permitem a rápida disseminação de informações: “Por estarem conectadas, todas as informações que os usuários visualizam (notícias, novidades, vídeos, leituras, links, entrevistas, entre outros) podem ser rapidamente compartilhadas, o que vem a caracterizar a participação do sujeito na rede, criando sua identidade no ciberespaço” (MARCON; MACHADO; CARVALHO, 2013, p. 15).

Com a utilização das redes sociais em tempo real, muitas barreiras são estreitadas e distâncias diminuídas, desde relacionamentos pessoais, namoros e amizades ao uso pelas corporações, as quais aproveitam cada vez mais dos recursos para se beneficiarem e ampliarem os lucros.

2.6 Computação em Nuvem

A denominação *cloudcomputing* chegou ao conhecimento público em 2008. Pode-se dizer que computação nas nuvens é todo conceito que se tem hoje do que é computação (processamento, armazenamento e softwares), só que armazenado na rede, podendo ser acessadas, remotamente, de qualquer lugar do mundo, e independente de plataforma, nas mais variadas aplicações por meio da internet com a mesma facilidade de tê-las instaladas no próprio computador.

O armazenamento dos dados é realizado por serviços que poderão ser acessados de qualquer lugar do mundo e a qualquer hora, não importando se o usuário possui ou não o software ou até mesmo espaço em disco para isso. O acesso é feito, principalmente, por meio de navegadores, que acessam remotamente servidores físicos localizados em qualquer lugar do mundo.

3. Um conceito chamado Big Data

A expressão Big Data está cada vez mais popular, porém, ainda não se têm bem claro seu significado, aplicabilidade e finalidade.

O Big Data vem chamando atenção pela acelerada escala em que os volumes cada vez maiores de dados são criados pela sociedade. No entanto, existem muitas dúvidas de como

sair do conceitual e criar soluções de negócio, que mineralizem essa massa de dados, já que, a cada dia, são geradas dezenas de *petabytes* de dados, em escala real e não mais imaginária e futurista.

O Big Data permite a utilização de diversas tecnologias de gerenciamento dessa massiva geração de dados, como: Data Warehouse, Data Mart, Data Mining, BI (*Business Intelligence*) e *CloudComputing*.

Com o decorrer do tempo, o volume de informações produzidas, diariamente, mundo afora cresceu com velocidade tão alta que se chegou a um ponto em que as técnicas utilizadas para análise e administração dos dados ficaram ultrapassadas. Esse montante de dados, variedade e, principalmente, velocidade com que crescem e trafegam estão revolucionando cada vez mais a sociedade no trabalho, lazer, saúde, educação, etc.

A consequência da produção exponencial e crescente de dados chega a afetar todos os setores econômicos, desde um simples consultório médico até grandes indústrias varejistas. Profissionais da área de informática (engenheiros e cientistas da computação) designaram esse fenômeno de Big Data, como se verá a seguir.

3.1 Apresentando o Big Data

De acordo com Lima Junior (2011), a expressão Big Data se refere ao conjunto de dados produzidos atualmente no mundo, cujo volume está além dos padrões e da capacidade das ferramentas utilizadas por bancos de dados para capturá-los, analisá-los e gerenciá-los. Mas não se pode dizer que Big Data seja apenas um imenso volume de dados.

A princípio, pode-se definir o conceito de Big Data como sendo um conjunto de dados extremamente amplos e que, por este motivo, necessitam de ferramentas especialmente preparadas para lidar com grandes volumes, de forma que toda e qualquer informação processada por esses meios possa ser encontrada, analisada e aproveitada em tempo hábil: “O valor real do Big Data está no insight que ele produz quando analisado – buscando padrões, derivando significado, tomando decisões e, por fim, respondendo ao mundo com inteligência” (INTEL, 2013, p. 3).

Gartner² *apud* Silva (2013) define como Big Data o termo adotado pelo mercado para descrever problemas no gerenciamento e processamento de informações extremas, as quais excedem a capacidade das tecnologias de informação tradicionais ao longo de uma ou várias dimensões.

O Big Data está focado, principalmente, em conjunto de dados extremamente grandes, gerados a partir de processos e práticas tecnológicas, tais como mídia social, tecnologias operacionais, acessos à Internet e fontes de informações distribuídas, telefonia, etc.

Geram-se muitos dados a todo instante no mundo atual. No princípio, produzia-se o *bit* (equivalente a 1 impulso elétrico) e o *byte* (conjunto de 8 *bits* que representa um comando entendido pelo computador). Hoje já há *exabytes* de dados gerados diariamente, chegando a acumular em um ano cerca de 1,8 *zetabytes* de dados, e a cada dia esse número cresce à medida que se produzem ainda mais dados.

De maneira mais simplista, a ideia de Big Data também pode ser compreendida como a análise de grandes quantidades de dados para a geração de resultados importantes que, em volumes menores dificilmente seriam alcançados.

Em 1980, surgiram os conceitos de Data Warehouse, até chegar à ideia de Big Data, o que ocorreria em 1990. A diferença de Data Warehouse e Big Data é a velocidade com que os dados precisam ser disponibilizados, uma vez que, em um projeto de Data Warehouse, o processo de ETL torna-se mais lento até que as informações estejam disponíveis. Segundo Taurion (2013), o Big Data está muito além de um grande Data Warehouse. Entretanto, pode-se dizer que o Big Data se relaciona com o Data Warehouse.

Já na década de 1980, o volume de dados crescia exponencialmente. Foi o momento em que surgiram dúvidas de o que fazer com os dados armazenados. Na época, as técnicas utilizadas já não supriam as necessidades de tratar a maioria dos repositórios de dados. Com essa demanda crescente, no fim da década de 1980, surge o conceito e a utilização de Data Mining ou Mineração de Dados.

² Disponível em <<http://www.gartner.com/>>.

É certo que se trata de dados muito significativos, volumes enormes e informações diversas. Esses dados são extraídos em grande parte das redes sociais, onde a cada minuto os servidores dessas plataformas são bombardeados com milhões de dados, além do *Enterprise Resource Planning* (Planejamento de Recursos Empresariais - ERP), que concentra dados de toda uma organização.

3.2 As bases do Big Data

Pode-se dizer que de certa forma há como analisar todos os tipos de dado, tanto dados estruturados (de um sistema ERP, por exemplo) como não estruturados, hoje representados pela grande massa de informações que se podem extrair, principalmente, das redes sociais.

Esse modo de pensar muda a forma como se analisam os dados, passando-se a dar atenção a dados que antes eram considerados sem valor. Pode-se dizer que, em vez de analisar apenas um percentual de dados, como uma amostragem, por exemplo, seriam analisados além da amostragem, dados de diversas fontes nunca utilizadas antes, como um comentário numa rede social.

Um exemplo prático disso, são as pesquisas de boca de urna, em que os dados são analisados baseados em levantamentos de amostragem – se a apuração dos eleitores fosse completa, teoricamente, seria feita a eleição de fato.

Quando se fala em Big Data, logo se imaginam “grandes dados” relacionados ao imenso volume de dados a ser analisados. Entretanto, a expressão é bem mais abrangente, tendo como base os 5 Vês: volume, velocidade, variedade, veracidade, valor.

3.3 Big Data Além do Conceito

Como já abordado, existem muitos dados que irão gerar informação para todos os setores econômicos possíveis.

De acordo com Davenport (2014), da mesma maneira que acontece com a maioria das tendências tecnológicas estratégicas, o Big Data adentra inúmeras funcionalidades específicas que distinguem os tipos de sistema, que representa os componentes típicos do *stack* de Big Data.

O contexto atual do Big Data possui no centro de atenção os dados e o potencial analítico, e ao redor orbitam as aplicações, utiliza-se um novo contexto de análises. Passa a ser importante desenvolver a capacidade de produzir novos questionamentos. Uma aplicação prática é a fidelização de clientes com base em interesses e hábitos de compra e, simultaneamente, considerar as previsões econômicas e monitorar os passos da concorrência (TAURION, 2013).

3.3.1 Plataformas do Big Data

As plataformas Big Data são constituídas, geralmente, de um conjunto de funções que levam à alta performance do processamento de dados. Tais plataformas devem permitir que se possam interagir com dados armazenados, além de administrá-los e aplicar técnicas computacionais avançadas, a fim de extrair o máximo possível de informação de qualidade.

O crescimento do processamento do Big Data vem se tornando amplo e complexo. Hoje existem infinitas opções e maneiras para uso de Big Data. Dados que podem afetar de forma positiva e/ou negativa todos os setores da economia.

Os dados gerados são muitos e diversos, coletados de todo lugar, de todos os setores, econômico, educacional, lazer, saúde, etc. De um simples sensor num viaduto, uma compra no supermercado, uma ida ao médico para uma consulta rotineira, tudo gera dados.

Nessa pilha de “setores” que formam o Big Data, os dados se categorizam como o principal setor, pois constituem a matéria-prima de todas as análises e interpretações que podem ser geradas.

As informações extraídas de um Big Data, geralmente, são apresentadas em aplicações destinadas aos usuários, com interface amigável e de fácil entendimento. Hoje se fala muito em aplicações que utilizam a técnica de *narrativescience* e *automated insights*, em que as informações extraídas dos dados analisados, a partir da ideia de inteligência artificial (IA), são apresentadas como narrativas automatizadas em formato de texto, em que se cria uma história de dados brutos (NARRATIVESCIENCE, 2015).

Não há diferenças entre o armazenamento de quaisquer dados para o armazenamento de dados para fins de Big Data, a diferença que se pode citar é o custo ainda mais baixo para armazenar dados não estruturados, que em grande parte utiliza aplicações para armazenamento *open source*.

3.4 Big Data na nuvem

Modelos de armazenamento em nuvem oferecem flexibilidade excepcional, permitindo avaliar a melhor abordagem para cada necessidade do usuário de negócios. Por exemplo, as organizações que já mantêm um ambiente interno de nuvem podem criar sistemas de análise de Big Data para suas ofertas, usar um provedor de serviços em nuvem ou construir um sistema híbrido, que protege determinados dados sensíveis em uma nuvem particular, usufruindo a vantagem de contratar fontes externas de dados valiosos e aplicações fornecidas em nuvens públicas (INTEL, 2015).

Utilizar a infraestrutura de nuvem para analisar Big Data é importante, como se verá nos itens a seguir.

Os investimentos em análise de Big Data podem ser significativos e demandam infraestrutura eficiente e econômica – os recursos para apoiar modelos de computação distribuída internamente, normalmente, estão alocados em centros de grande e médio porte. As nuvens privadas podem oferecer um modelo mais eficiente e econômico para implantar a análise de Big Data, enquanto permitem ampliar os recursos internos com serviços de nuvem pública. Esse sistema híbrido de nuvem permite às empresas usar o armazenamento de espaço *on-demand* e capacidade de informatização por meio de serviços de nuvem pública para algumas opções de análise (por exemplo, projetos de curto prazo) e fornecer capacidade adicional e escala conforme necessário (INTEL, 2015).

O Big Data pode mixar fontes internas e externas – enquanto as empresas costumam manter dados mais sigilosos internamente (*in-house*). Grande parte do Big Data (de propriedade da empresa ou gerados por terceiros e prestadores públicos) pode ser alocada, externamente, já em ambiente de nuvem. Processar fontes de dados relevantes por trás do *firewall* pode ser um investimento significativo de recursos. Analisar os dados onde eles residem – quer em centros

de dados de nuvem internos ou públicos, ou em sistemas de ponta e dispositivos do cliente – muitas vezes faz mais sentido (INTEL, 2015).

Os serviços de dados são necessários para extrair valor de Big Data – dependendo dos requisitos e do ambiente de utilização, o melhor investimento do orçamento em TI pode ser concentrar-se em Serviço de Análise (AaaS), apoiado por nuvem privada interna, nuvem pública ou num modelo híbrido (INTEL, 2015).

Modelos de computação em nuvem podem ajudar a acelerar o potencial para soluções de análise em escala. Nuvens oferecem flexibilidade e eficiência de acesso aos dados, fornecendo insights e agregando valor. No entanto, análises de Big Data baseadas em nuvem não são solução para tudo.

As organizações que utilizam infraestrutura de nuvem para fornecer AaaS dispõem de várias opções. Por questões como: carga de trabalho, pesagem, custo, segurança e interoperabilidade de dados; a TI pode optar por utilizar a nuvem privada para mitigar o risco e manter o controle; usar a infraestrutura de nuvem pública, serviços de plataforma ou de análise para aumentar ainda mais a escalabilidade; ou implantar um modelo híbrido que combine recursos e serviços de nuvem privada e pública.

3.5 Aplicações do Big Data

À medida que os dados são gerados e permitem produzir informações úteis, pode-se dizer que se está no mundo do Big Data. Hoje ele já faz parte de todos os setores econômicos e, a cada dia que passa, afeta ainda mais cada envolvido. Exemplos desse envolvimento é a participação mais ativa de tecnologias de Big Data no setor de saúde, educação e, principalmente, nas redes de varejistas e no mundo dos negócios.

A cada dia, empresas inovadoras como a IBM (2012), LinkedIn, Amazon, entre outros, vêm desenvolvendo tecnologias e formas de utilização do Big Data, para as mais diversas áreas.

As organizações cujo o produto fim não é a tecnologia, a usam cada vez mais, como também os serviços de Big Data para ficar na frente das concorrentes.

3.6 Bibliotecas Universitárias

O Big Data se expande para todos os campos do conhecimento humano. Apesar de seu conceito não ser tão claro, como já abordado, seu crescimento é volumoso, decorrente do enorme número de informações não estruturadas, além de ser necessário processamento em grande escala para que seja possível a extração de informações que possam estabelecer valores que impactam na economia, nas organizações, no governo e nas relações interpessoais. Tudo isso tem grande influência no universo bibliotecário. Desde os anos 1940, já existia a preocupação em quantificar o grande volume de dados bibliotecários, que já cresciam a cada ano mais e mais.

De acordo com Rider (1944), naquela época já se estimava de que as bibliotecas universitárias norte-americanas iriam dobrar de tamanho (em volume de informações) a cada 16 anos. Rider se baseou nessa estimativa para projetar o crescimento da Biblioteca da Universidade Yale, onde o mesmo estimava que no ano de 2040, a biblioteca teria aproximadamente 200 milhões de volumes, distribuídos em cerca de 6.000 km de prateleiras, e um corpo laborar composta de mais de 6.000 catalogadores.

Nas bibliotecas universitárias brasileiras o cenário de Rider, não é muito diferente, estima-se um crescimento médio em cerca de 3.000 volumes por ano por biblioteca, dependendo do porte da universidade.

Com esse crescimento, as tecnologias do Big Data, podem auxiliar as bibliotecas universitárias na criação de repositórios digitais, para que seja possível a o gerenciamento de forma mais rápida e ágil.

4 Considerações Finais

O trabalho realizado permitiu adquirir conhecimento sobre onde e como o Big Data trabalha.

Como visto, o fenômeno do Big Data vem se desenvolvendo exponencialmente. Tecnologias novas são apresentadas com potencial para administrar grandes volumes de dados. Entretanto, ainda é incipiente o consenso sobre o real potencial do Big Data e quais

tecnologias melhor o norteiam. Além de que, ainda existem várias dúvidas do muito que se pode fazer com o Big Data para agregar valor a uma empresa, por exemplo.

Ao lembrar que o Big Data é formado por volume – basta pensar que o que era futuro há uma década, *terabytes*, hoje temos dentro de casa –, velocidade, variedade, valor e veracidade, pode-se observar que já se utiliza esse conceito desde o princípio da informatização, com a diferença, que só agora se passou a tentar organizar tudo que se gera e produz.

Juntos, esses fundamentos oferecem a oportunidade de analisar quase tudo que é veiculado nos meios eletrônicos, dados estruturados e não estruturados, como: imagens, vídeo, áudio e documentos.

A capacidade de cruzar imensos volumes de dados traz poder às organizações, para obter conhecimento sobre regiões mais afetadas por determinadas doenças, monitoração e prevenção de terremotos, utilizando tecnologias que auxiliam a tomada de decisões.

Durante a elaboração desta pesquisa, foram encontradas algumas dificuldades, dentre elas a restrição do tema, por abranger muitas áreas acabou atrapalhando o foco do trabalho.

Com essa ampla gama de assuntos que poderiam ser abordados, pode-se deixar como sugestão de trabalhos futuros o uso de Big Data para entender e viabilizar soluções para macroproblemas urbanos, como grandes congestionamentos que ocorrem no Brasil, como tratar esses engarrafamentos usando dados estatísticos em tempo real; ou como adequar os horários de trabalho nas empresas por meio de uma agenda diferenciada de modo a evitar o mesmo horário de entrada e saída dos trabalhadores, uma vez que essa rotina contribui para criar o caos no transporte público.

Além da utilização do Big Data na área da saúde, onde se pode usá-lo para detecção e antecipação de infartos, desenvolvimento de vacinas e até mesmo identificação e prevenção de surto de doenças tropicais, por exemplo: dengue, malária, tuberculose e ebola. Hospitais já trabalham com o Big Data para tratar de doenças como o câncer, processando em segundos uma enorme gama de informações que permite chegar ao diagnóstico mais preciso, no momento, sobre determinado estágio da enfermidade – o que seria impossível para qualquer médico fazê-lo apenas com planilhas e textos.

É amplo, como já se sabe, o uso do Big Data nas mais diversas áreas, setores e situações, de forma a evitar e prever acontecimentos que interferem diariamente com toda a vida existente na terra.

Esse futuro vai depender quase exclusivamente da tecnologia para administrar o Big Data e pode estar próximo em termos de tempo e espaço.

O elenco de soluções oferecidas pelo Big Data é praticamente infindável e pode ser benéfico tanto para o mundo empresarial como para o âmbito individual, deverá revolucionar a computação, e a vida nunca mais será a mesma como se conhece.

Fundamentals and Application of Big Data: Information on how to treat a society yottabytes

Abstract: For some time, a matter which computer scientists try to solve is what to do to manage the incalculable amount of data generated every time. This refers to the analysis and search for answers in relation to areas where big data is being used. For increasingly have the need to process data effectively and more quickly. Sports, health, retail, are a few among countless fields where Big Data is already part of everyday life. Thus, the aim of this review is to conceptualize and analyze Big Data and present the technological resources involved since before the unstoppable growth data production, it appears that the management of Big Data can be in the eyes of the girl computerization In the next years.

Keywords: Big Data. Share Cloud. Knowledge generation. Data storage. Competitiveness.

Referências

DAVENPORT, Thomas H. *Big Data no trabalho: derrubando mitos e descobrindo oportunidades*. Rio de Janeiro: Elsevier, 2014.

ELIAS, Diego. *A abordagem Top-Down e Bottom-Up no Data Warehouse*. Canaltech, 12 maio 2014. Disponível em: <<http://corporate.canaltech.com.br/materia/banco-de-dados/a-abordagem-top-down-e-bottom-up-no-data-warehouse-21108/>>. Acesso em: 10 maio. 2015.

IBM. *Saiba o que é o Big Data e os desafios que as empresas enfrentam*. IBM, 2012. Disponível em: <http://www.ibm.com/midmarket/br/pt/infografico_bigdata.html>. Acesso em: 28 ago. 2015.

INTEL. *Guia de Planejamento: saiba mais sobre Big Data*. 2013. Disponível em <https://dialogoti.intel.com/sites/default/files/documents/90318386_1.pdf>. Acesso em: 04 jun. 2015.

LEITÃO, Bárbara Júlia Menezello. A biblioteca universitária. In: _____. *Avaliação qualitativa e quantitativa numa Biblioteca Universitária. Interciência*. Rio de Janeiro, 2005.

LIMA JUNIOR, W.T. *Jornalismo Computacional em função da Era do Big Data*. SBPJor – Associação Brasileira de Pesquisadores em Jornalismo. In: 9º. Encontro Nacional de Pesquisadores em Jornalismo. Rio de Janeiro, Universidade Federal do Rio de Janeiro, 3 a 5 nov. 2011.

MACHADO, Felipe Nery Rodrigues. *Tecnologia e projeto de Data Warehouse: uma visão multidimensional*. 5. ed. rev. e atual. São Paulo, 2010.

MARCON, Karina; MACHADO, Juliana Brandão; CARVALHO, Marie Jane Soares. Arquiteturas pedagógicas e redes sociais: uma experiência no Facebook. *Revista de Informática Aplicada*, v. 9, n. 2, 2013. Disponível em: <<http://ria.net.br/index.php/ria/article/download/104/99>>. Acesso em: 15 maio 2015.

NARRATIVE SCIENCE. Disponível em: <<https://www.narrativescience.com/>>. Acesso em: 18 maio 2015.

PRASS, Fernando Sarturi. *Big Data: do Data Warehouse ao Data Mining em um único lugar*. Sopa de Letrinhas – O Blog do Business Intelligence. Disponível em: <<http://fp2.com.br/blog/index.php/2012/um-visao-geral-sobre-fases-kdd/>>. Acesso em: 13 maio 2015.

ROGENSKI, Renato. Uma entrevista didática sobre Big Data. *Tecnologia, Exame.com*, 25 jul. 2013. Disponível em: <<http://exame.abril.com.br/tecnologia/noticias/uma-entrevista-didatica-sobre-big-data>>. Acesso em: 23 abr. 2015.

RIDER, Arthur Fremont. *The Scholar and the Future of the Research Library. A Problem and Its Solution*. Hardcover, 1944.

SILVA, Alex. *Big Data e seus impactos*. Trabalho de conclusão de curso (TCC em Engenharia Elétrica) – Instituto Nacional de Telecomunicações (Inatel). 2012. Disponível em: <<http://pt.slideshare.net/AlexSilva8/big-data-apresentao-final-inatel>>. Acesso em: 07 mar. 2015.

TAURION, Cezar. *Big Data*. Rio de Janeiro: Brasport Livros e Multimídia Ltda, 2013.

WEITZEL, Simone da Rocha. *Elaboração de uma política de desenvolvimento de coleções em bibliotecas universitárias*. 2. ed. Rio de Janeiro: Interciência, 2006.

Informação bibliográfica deste texto, conforme a NBR 6023:2002 da Associação Brasileira de Normas Técnicas (ABNT):

CALDAS, Max Silva; SILVA, Emanoel Costa Claudino. Fundamentos e aplicação do Big Data: como tratar informações em uma sociedade de yottabytes. *Bibliotecas Universitárias: pesquisas, experiências e perspectivas*, Belo Horizonte, v. 3, n. 1, p. 65-83, jan./jun. 2016.

Recebido em: 22.11.2015.

Aceito em: 17.02.2016.