

BIG DATA

Bruno Dutra Cardozo Murata¹, Luis Gustavo B. P. Gomes²

Curso Análise e Desenvolvimento de Sistema – Faculdade de Tecnologia de Gravataí
(FAQI) – 94.030-001 – Gravataí – RS – Brasil.

brunodcm73@hotmail.com¹, Luis.gomes@tbsa.com.br²

***Resumo.** Esse artigo apresentará uma visão sobre as aplicações do conceito de Big Data, que consiste em práticas para lidarmos com grandes massas de dados, que atualmente tem um crescimento absurdo, pois, estamos passando pela era da informação tecnologia automatizada.*

1. Introdução

O termo Big Data foi introduzido em 2005 por Roger Magoulas, após o primeiro ano da criação do termo Web 2.0, ele referenciava a um conjunto de dados que era quase impossível de processar com as ferramentas de inteligência negócios existentes daquela época. Porém, no mesmo ano de 2005, foi lançado pela Yahoo, o Hadoop que é uma plataforma open-souce, desenvolvida para o processamento e análise de grandes volumes de dados, podendo ser dados estruturados ou não estruturados.

Apesar do termo “Big Data” ser relativamente novo, ou ao menos, começou a aparecer na mídia recentemente, os questionamentos sobre o volume de dados tem alguns marcos na história o primeiro deles é datado dos anos 40 com a publicação do livro “The Scholar and the Future of the Research Library” do Autor Fremont Rider, que relata o crescimento das bibliotecas nos EUA, outro marco importantíssimo, é o artigo “How Much Information Is There In the World?”, em 1997 pelo programador Michael Lesk, que concluiu uma estrutura de alguns milhares de petabytes, até os anos 2000.

Atualmente, estamos passando por uma fase de crescimento exorbitante de volume de dados, pois, a mobilidade e a internet of things estão ao nosso redor e de maneira acessiva, e financeiramente fácil a todas as faixas econômicas da sociedade atual. Não é difícil a compreensão desse volume de dados e sua variedade. Presentemente, temos operações bancárias, atividades das companhias aéreas, redes sociais, e-commerce, app's e muitos outros exemplos que diariamente geram grandes volumes de dados. Mas apenas ter os dados não é o suficiente, se faz necessário o poder de analisá-los, pois, com a informação correta e em tempo hábil é possível criar estratégias de marketing com mais eficiência, indicar cortes de gastos, aumentar a produtividade, superar concorrentes, logo, são fatores que podem relativamente ser decisivos para o futuro de empresas e também para usos de controle de doenças. Para todos esses fatores foi criado essa metodologia, que está iniciando suas atividades, pois, a taxa de crescimento de volume de dados será cada vez maior.

Iremos estruturar os principais conceitos do Big Data da seguinte forma: 2. Motivo da criação, 3. Os V's da Big Data, 4. Sistemas e técnicas, 5. Considerações finais.

2. Motivo da criação

Lidamos com dados desde os primórdios da história humana, porém, desde a nossa última década estamos passando por uma era de muitos volumes dados e temos que utilizarmos de forma correta e de fácil acesso, por esse motivo foram criados alguns conceitos para lidarmos com esse problema. Para termos noção da quantidade de volumes que estamos falando, em 2008 a IBM já estava preocupada com a possível falta de espaço para o armazenamento, pois, já tínhamos produzidos cerca de 2,5 quintilhões de bytes todos os dias, e surpreendentemente mais de 90% foram produzidos em 2006 e 2007, já o estudo da EMC Empresa líder no segmento internacional de armazenamento, segundo o site o globo.com, prevê um crescimento exorbitante até o ano de 2020, conforme figura 1.

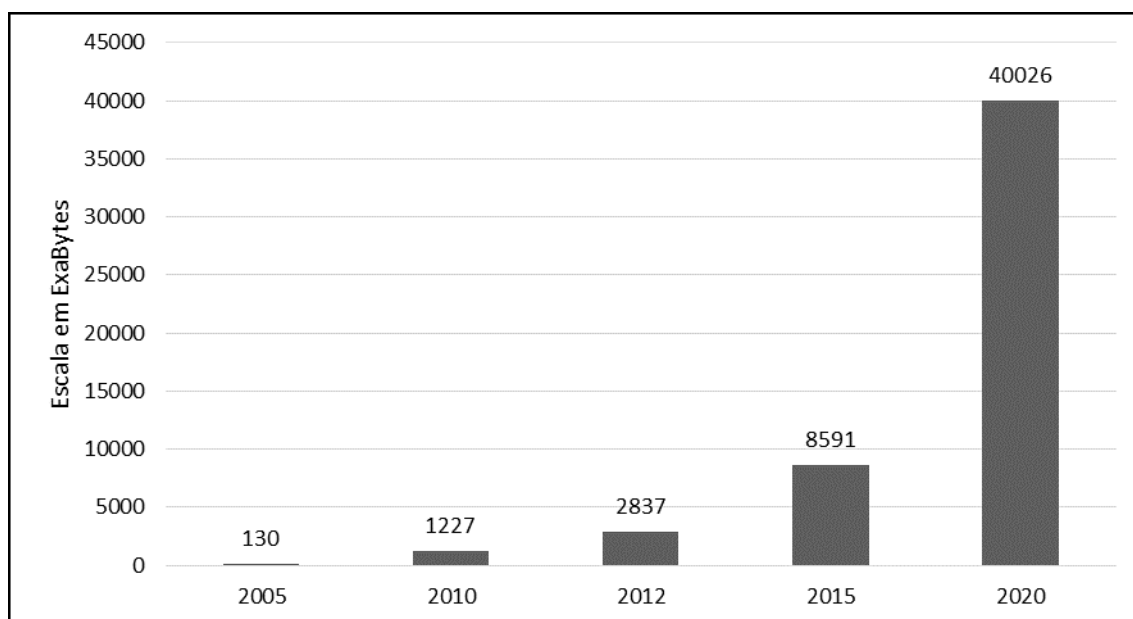


Figura 1. Prospecção de crescimento de dados

Fonte: Elaborado pelo os autores (com base no estudo da EMC)

Na Figura 1, podemos visualizar uma linha de tendência, sobre o crescimento do volume de dados que estamos passando, logo, o mesmo estudo da EMC traz um aspecto interessante sobre dados armazenados na nuvem, destes 40.026 EB serão 45 %, ou seja, 14.996 EB, estariam locados na nuvem.

Esse volume massivo de dados decorre a adesão de empresas a globalização e cada ano que passa, essa informação tende a aumentar cada vez mais. Por esses motivos o termo está sendo abordado com uma maior frequência.

3. Os V's da Big Data

Afim de deixar o conceito de Big Data mais claro, muitos autores assim como Cezar Taurion, tendem a tratar o assunto subdividindo-o em 5 V's:

1. Volume: Trata-se sobre a quantidade de dados gerados diariamente por pessoas ou empresas. Para termos uma ideia dessa grandeza, a Cisco (uma das empresas líder no mercado), realizou uma pesquisa e concluiu que o volume de trafegados no ano de 2016, superou um zettabytes correspondente à $1.250000e+20$ Bytes, essa quantidade é maior que a soma dos dados produzidos entre os anos de 1984 a 2012.

2. Velocidade: Trata-se da velocidade que os dados são gerados e processados, os dados devem ser processados em tempo hábil para a operação, imaginamos uma operação bancária com cartão de crédito demorasse, as consequências são muitas, começando na diminuição de vendas à possíveis ações ilícitas. Sabe-se também que o tempo é crucial para a tomada de decisões dentro de uma organização.

3. Variedade: Refere-se as diferentes formas de processamento e análise dos dados encontrados, há duas maneiras de dados os estruturados e os não estruturados. Os estruturados são dados presentes na maioria dos bancos de dados de empresas e sites, por exemplo, e eles possui o conceito chave criado pelo programador Edgar Frank Codd que é a relação entre os dados, onde temos uma maior facilidade de análise, já os não estruturados referem-se a maior quantia de volume de dados encontrados no mundo, são referentes as imagens, mensagens, vídeos, e-mail, áudios e muitas outras fontes.

4. Veracidade: Refere-se à validação dos dados coletados, pois, como já dito anteriormente, as informações extraídas dos dados são utilizadas para tomadas de decisões, e muitos dados são irrelevantes, sendo necessário a utilização de ferramentas para filtrar dados verídicos que agreguem valor.

5. Valor: Está relacionado ao valor da informação obtida pelas ferramentas, pois, não adianta termos uma imensidade de dados (várias espécies) processamos imediatamente se o resultado não servir para as organizações ou para os usuários finais.

4. Sistemas e técnicas

Para que seja possível aplicar os conceitos de Big Data, temos que relatar algumas soluções que se encontram no mercado atual.

4.1 Apache

Para começarmos temos o Hadoop, e como dito anteriormente, foi criado pela Yahoo, posteriormente, transferida para a Apache. Ele é uma plataforma open-source

desenvolvida em Java para processamentos de grandes escalas, foi projetado para ser altamente confiável e utiliza programação de auto nível. Basicamente, sua funcionalidade é dividida em módulos sendo eles: Hadoop Common (Biblioteca e repositório de arquivos), Hadoop Distributed File System (Sistema de fontes distribuídas e gerenciador de dados), Hadoop FIO (Framework e gestor de recursos) e o Hadoop MapReduce (Responsável pela análise e resultado).

4.2 Microsoft

Nesse âmbito a Microsoft tentou primeiramente criar uma plataforma independente do Hadoop, o SQL Server 2012, porém, não obteve o resultado esperado e posteriormente, lançava a sua nova solução, o Windows Azure que é baseada nas técnicas que são utilizadas no Hadoop, além de herdar alguns componentes do SQL Server 2012. O Azure traz uma solução cloud computing e tem alguns outros benefícios bem interessantes, fora o poder de utilizar todos os recursos do Hadoop, ele flexibiliza para o usuário a possibilidade de alteração de códigos de execução para análise, outra questão, é a possibilidade que o programador tem como escolher a linguagem na qual deseja trabalhar.

4.3 SAP

Outra ferramenta disponibilizada é SAP HANA, foi desenvolvida com base no Hadoop, que permite que o programador tenha o poder de escolha da linguagem de programação, além desses requisitos o HANA é uma plataforma completamente imaginada para negócios em tempo real. Outro recurso importante do HANA é que ele converge banco de dados e plataformas de aplicações em memória para poder transformar transações, lógica analítica e outras para proporcionar resultados em um menor tempo.

5. Considerações Finais

Desde os tempos primórdios o ser humano deixa para seu legado a informação, sendo sua disponibilidade de diversas maneiras, tais como o desenho e a escrita, e com isso foi possível a evolução do homem devido ao fato de obter a informação e tratá-la de forma correta, nesse pensamento, nos remetendo a nossa época onde a globalização e a informação são momentaneamente atualizadas e disponibilizadas, neste âmbito, podemos dizer que estamos em plena evolução para o que...esse é o “x” da questão.

Também podemos citar que o estudo de conceito estão recentemente começando, pois, a grandeza da proporção dos dados estão tomando um rumo para novos estudos, sem contar as diversas possibilidades de emprego que estarão disponíveis no mercado, não somente nessa área de volume de dados, pois, quando maiores os dados a se-

rem armazenado são maiores as estruturas físicas a serem construídas, sem contar área de segurança de informação que deverá agir de forma incansável. Por fim, podemos visualizar um grande ramo dentro da Tecnologia da Informação a ser explorado.

Referências

IBM, “DeveloperWorks”

https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/big_data_vol_ume_variedade_velocidade_de_dados?lang=en [acessado em Abr 17]

DataFloq, “A Short History Of Big Data”

<https://datafloq.com/read/big-data-history/239> [acessado em Abr 17]

SAS, “O que é Big Data? “

https://www.sas.com/pt_br/insights/big-data/what-is-big-data.html [acessado em Abr 17]

INFOWESTER, “Big Data“

<https://www.infowester.com/big-data.php> [acessado em Abr 17]

APACHE, “Hadoop“

<http://hadoop.apache.org/> [acessado em Abr 17]

MICROSOFT, “Azure“

https://azure.microsoft.com/pt-br/?&wt.mc_id=AID_sem_gp6EE5UH [acessado em Abr 17]

EMC, “O Universo Digital“

<https://brazil.emc.com/infographics/digital-universe-business-infographic.htm> [acessado em Abr 17]

SAP, “Bigdata Solucion“

<https://www.sap.com/solution/big-data.html> [acessado em Abr 17]