



Data Science Academy

www.datascienceacademy.com.br

Microsoft Power BI Para Data Science

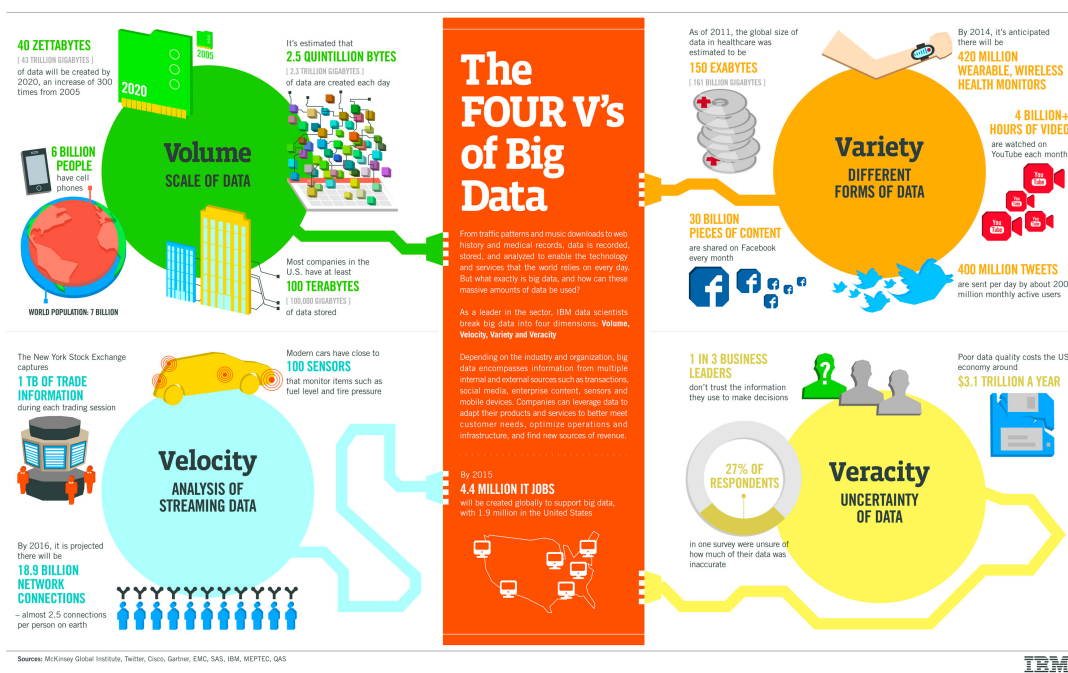
Big Data Analytics

Vemos com frequência pela internet, definições incorretas de Big Data, Data Science e Analytics e uma mistura dos conceitos. É importante que o profissional compreenda estes conceitos, para uma escolha correta das ferramentas e técnicas que se devam utilizar.

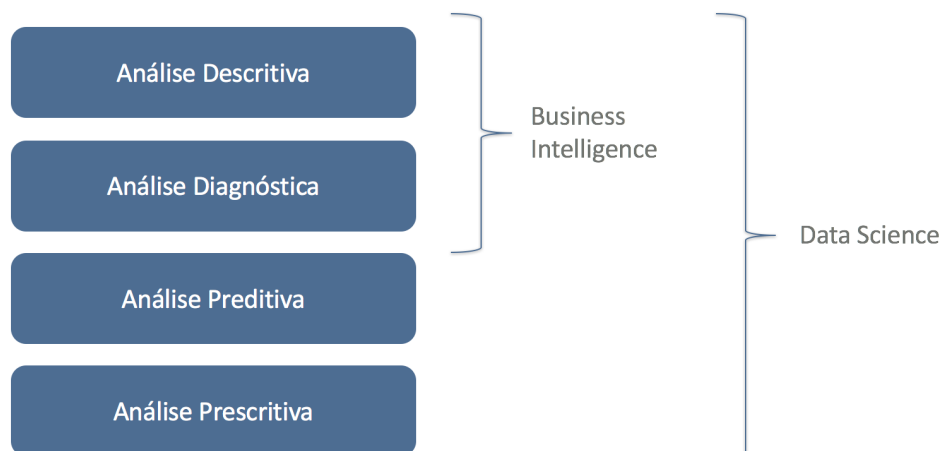
É importante deixar claro o conceito de Big Data, diferenciando-o do Data Warehouse tradicional. O risco de referir-se a Big Data como um grande banco de dados é confundi-lo com outros bancos de dados de grande porte como Data Warehouses por exemplo. As duas plataformas apresentam características distintas e fundamentais que definem a importância de cada uma na geração de conhecimento. Toda cadeia produtiva de conhecimento a partir de dados digitais se diferencia de uma plataforma e outra. Desde as estruturas de geração, armazenagem, segurança e compartilhamento, até os processos de indexação e análise.

Muitas vezes ouvimos falar de Big Data como uma metodologia, tecnologia ou ferramenta capaz de extrair valor de uma grande quantidade de dados digitais. **Este conceito é incorreto.** Big Data se refere ao conjunto de dados definido pelos 4 V's: volume, variedade, velocidade e veracidade. A extração do conhecimento se dá através de técnicas como Data Science por exemplo. A melhor maneira para falar de metodologias e tecnologias em Big Data é utilizar a expressão Analytics, que pode ser definida como o uso da tecnologia para a compreensão, comunicação e utilização inteligente dos dados digitais.

Alguns autores consideram o valor como mais uma dimensão crítica do Big Data, o quinto V. Nós consideramos esta abordagem incorreta pois essa dimensão não se trata de uma característica intrínseca dos dados digitais e sim uma dimensão que se alcança após o esforço de inteligência exercido sobre eles. A definição do Big Data com 4 V's é a definição cunhada pela IBM, conforme pode ser visto no gráfico abaixo.



E quando abordamos o termo Analytics o mesmo pode ser dividido em 4 áreas fundamentais:



- Análise Descritiva: se encarrega de analisar o que aconteceu;
- Análise Diagnóstica: analisa porque determinado evento aconteceu;
- Análise Preditiva: aponta o que irá (ou poderá) acontecer (predição);
- Análise Prescritiva: identifica o que se deseja que aconteça e o que pode ser feito para alcançar esse objetivo (reforçar, modificar ou evitar uma predição).

De modo igual, nas organizações, mesmo antes da existência dos computadores os gestores já praticavam os quatro tipos de analytics citados, utilizando tabelas, análise de cenários, simulações, brainstorms e também muita experiência e intuição. O processo é o mesmo em todos os contextos, só mudam as ferramentas e a metodologia.

Em Big Data os quatro tipos de Analytics são possíveis embora os maiores valores agregados aos dados estejam nas duas últimas (preditiva e prescritiva) já que as duas primeiras (descritiva e diagnóstica) já eram amplamente utilizadas e com sucesso antes do fenômeno Big Data. Nesse ponto, aparece a primeira das importantes diferenças básicas entre os dados em Big Data e bancos de dados estruturados de grande porte. Quando falamos em análise preditiva, estamos falando de uma metodologia completamente diferente das previsões que estamos acostumados a ver a partir de Data Warehouses ou de pesquisas de opinião. O que determina a mudança na metodologia é a natureza dos dados utilizados. A análise preditiva é feita a partir da utilização de grandes quantidades de dados de mesma natureza. A metodologia consiste em uma projeção, em que os dados viajam para dimensões maiores de tempo ou espaço. O resultado de uma previsão é uma estimativa e a ciência que fundamenta a metodologia é a Estatística.

Big Data Analytics, portanto, é a aplicação de análise preditiva sobre grandes conjuntos de dados!

