

INTRODUÇÃO À CIÊNCIA DE DADOS

Ciência de Dados e a Indústria 4.0

DADO - INFORMAÇÃO - CONHECIMENTO

Dado é o menor e mais simples elemento de um sistema.

É uma unidade indivisível, extremamente objetiva, geralmente abundante e que tem o papel de registrar um fato (evento).

Considerado um elemento de fácil manipulação e transporte

Basicamente **informação** é um conjunto de dados dentro de um contexto

Conhecimento: Uma informação que, devidamente tratada, muda o comportamento do sistema.

CENÁRIO

Evidência (e por que não efervescência) de temas como:

- Indústria 4.0 (Quarta revolução industrial);
- Big Data;
- IOT (Internet das Coisas);
- Certeza do uso de dados por grande partes das grandes empresas;
- LGPD (Lei Geral de Proteção de Dados);

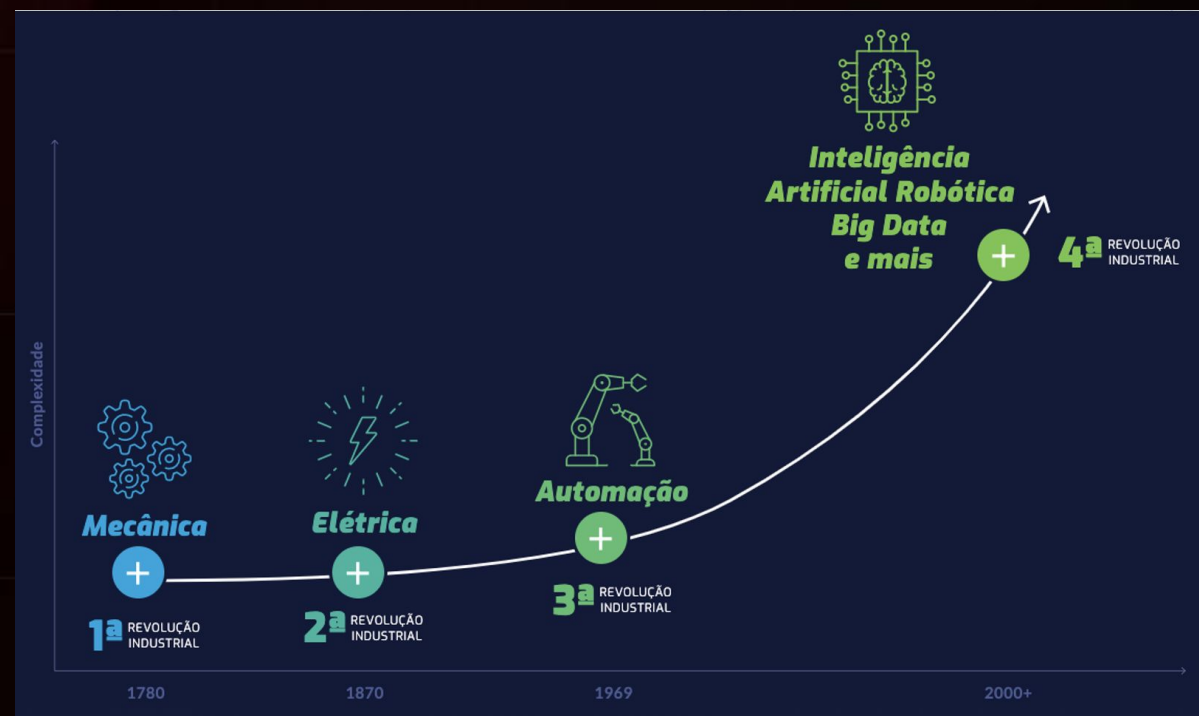
INDÚSTRIA 4.0

1ª - Mecanização da Produção

2ª - Tecnologias com eletricidade e produção em massa.

3ª - Automação industrial, nanotecnologia, biotecnologia, telecomunicações, mecatrônica.

4ª - Inteligência Artificial, Big Data, Computação em Nuvem, Realidade Aumentada, Internet das Coisas, Robótica Autônoma...



CENÁRIO

The five largest companies

in 2010

Exxon Mobil \$343

PetroChina \$316

Apple Inc. \$269

ICBC \$248

Microsoft \$237

vs 2020

Apple Inc. \$1,576

Microsoft \$1,551

Amazon.com \$1,433

Alphabet Inc. \$980

Facebook Inc. \$676



Largest companies in terms of market capitalization.
All figures in 2020 billion US dollars; 2020 companies as of July 1st.
Source: ycharts.com

The \$1 trillion club

Apple, Microsoft, Amazon, and Alphabet are worth a combined \$4.7T



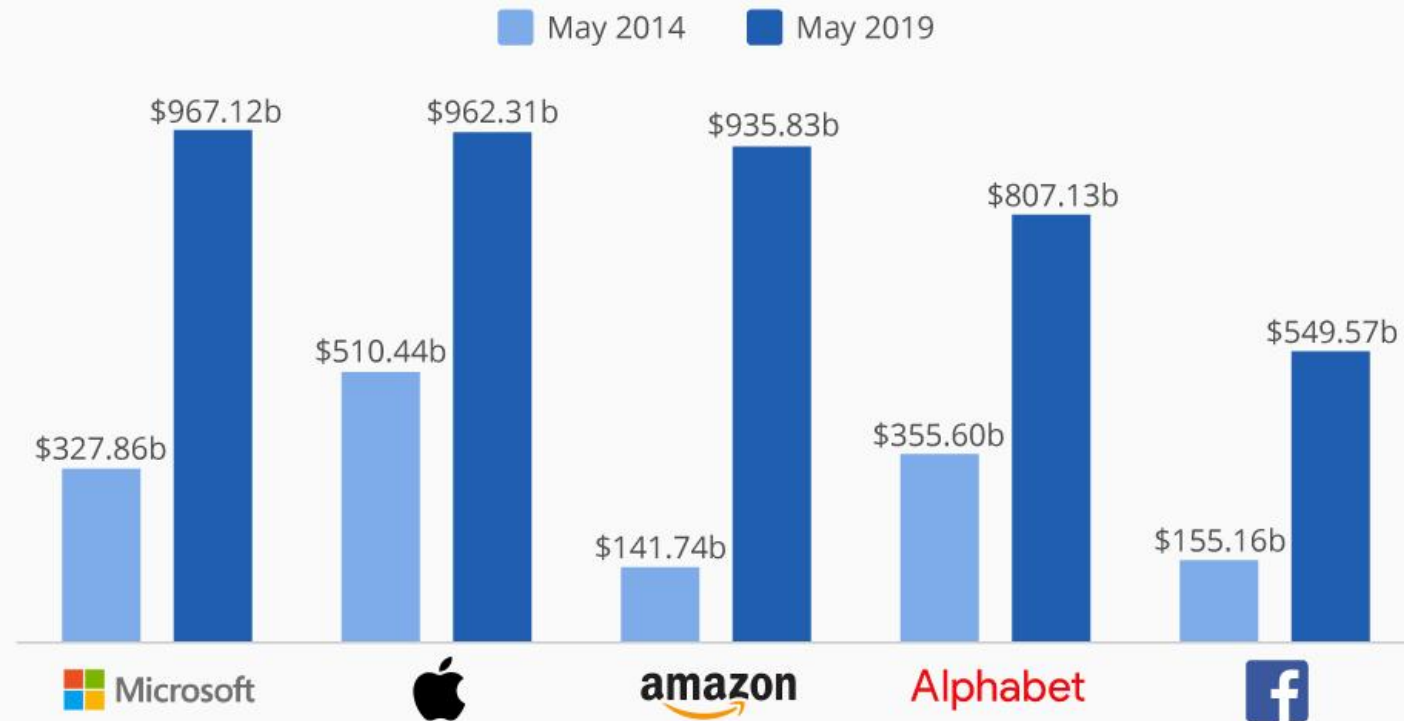
SOURCE: FactSet. Data as of 1/31/2020.



CENÁRIO

GAFAM Market Cap Jumps \$2.7 Trillion Over Past Five Years

Market capitalization of selected tech companies in May 2019 and May 2014*



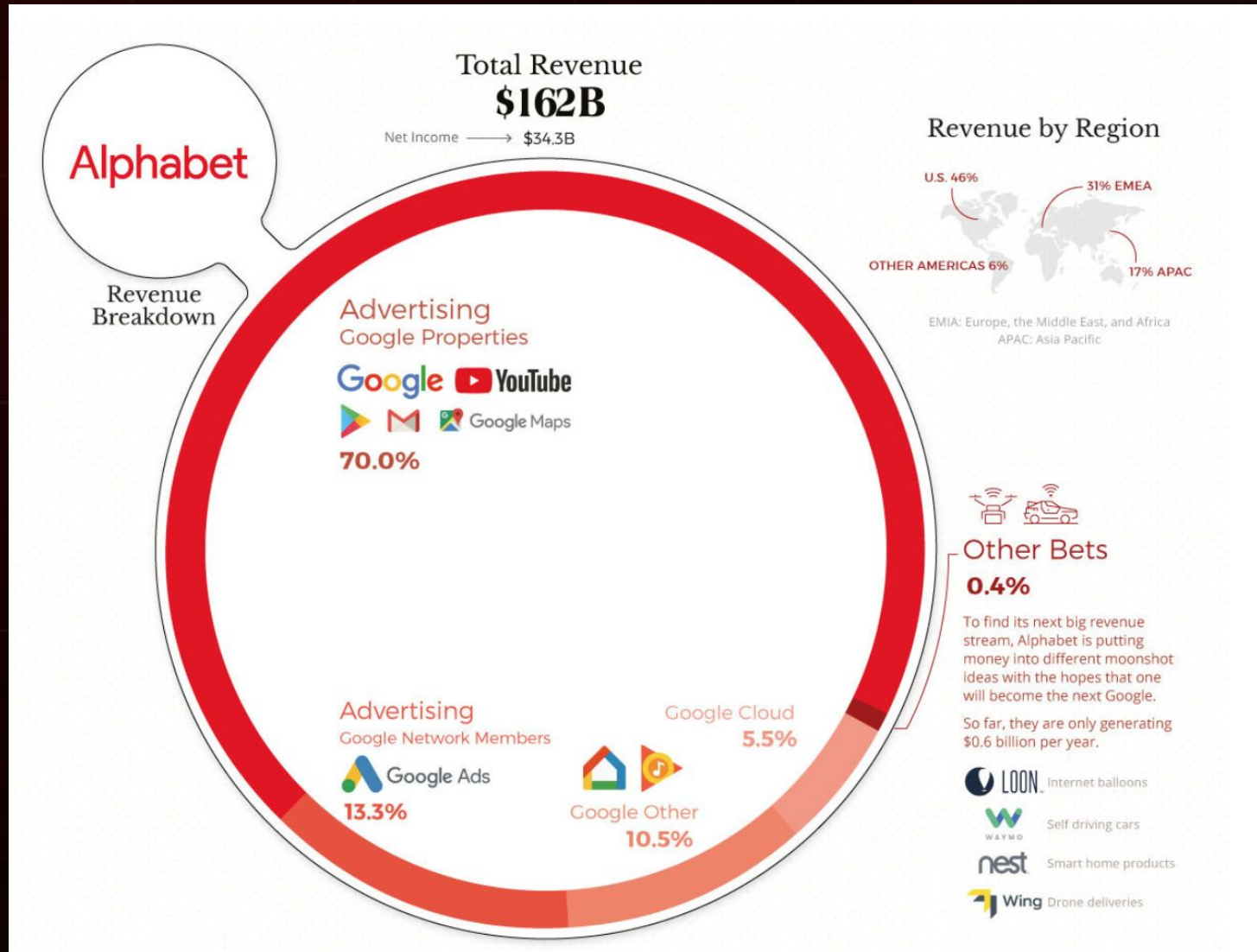
@StatistaCharts

* data as of May 2 of the respective year

Source: YCharts

statista

CENÁRIO



Fonte:

<https://www.visualcapitalist.com/how-big-tech-makes-their-billions-2020/>

TRANSFORMAR DADOS EM ATIVOS

- Todo esse contexto apresentado tem como premissa algo em comum, grande parte das empresas transformou os **dados em seus principais ativos**.
- Não falamos mais com tanto entusiasmo sobre petróleo, bancos, bebida, comida entre outros tipos de consumo, as principais empresas do planeta só pensam em DADOS!

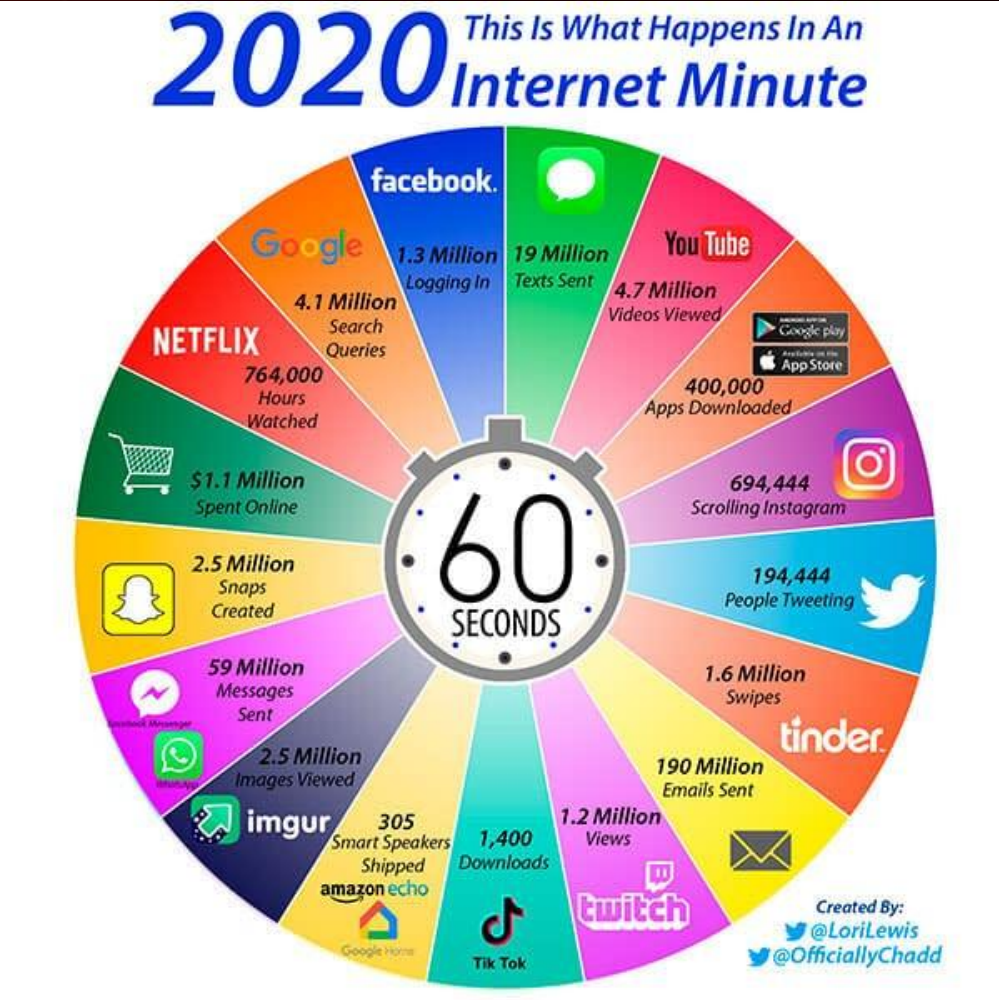
BIG DATA

- Muitas definições, mas podemos pensar em:
 - termo utilizado para nomear um paradigma caracterizado por **grandes conjuntos de dados**, que **aplicativos tradicionais** não são capazes de lidar.
- Identificado geralmente por 3V's (Volume, Variedade, Velocidade)
- Alguns consideram até 10 Vs atualmente
- Todo processo de manipulação e análise de dados fica mais complicado e em grande parte das vezes muito mais caro quando se trabalha com Big Data.
- Dados podem ser estruturados, semi-estruturados ou não estruturados.

BIG DATA x CIÊNCIA DE DADOS

- Ainda há muita confusão e dificuldade de compreensão sobre a relação existente entre DS e Big Data.
- DS pode estudar qualquer conjunto de dados (mesmo que não for Big Data)
- Entretanto a evolução do DS aconteceu a partir do momento que foi possível realizar análises sobre grandes conjuntos de dados, principalmente porque quanto mais dados melhores são os resultados das análises.
- Os termos estão diretamente conectados e relacionados.

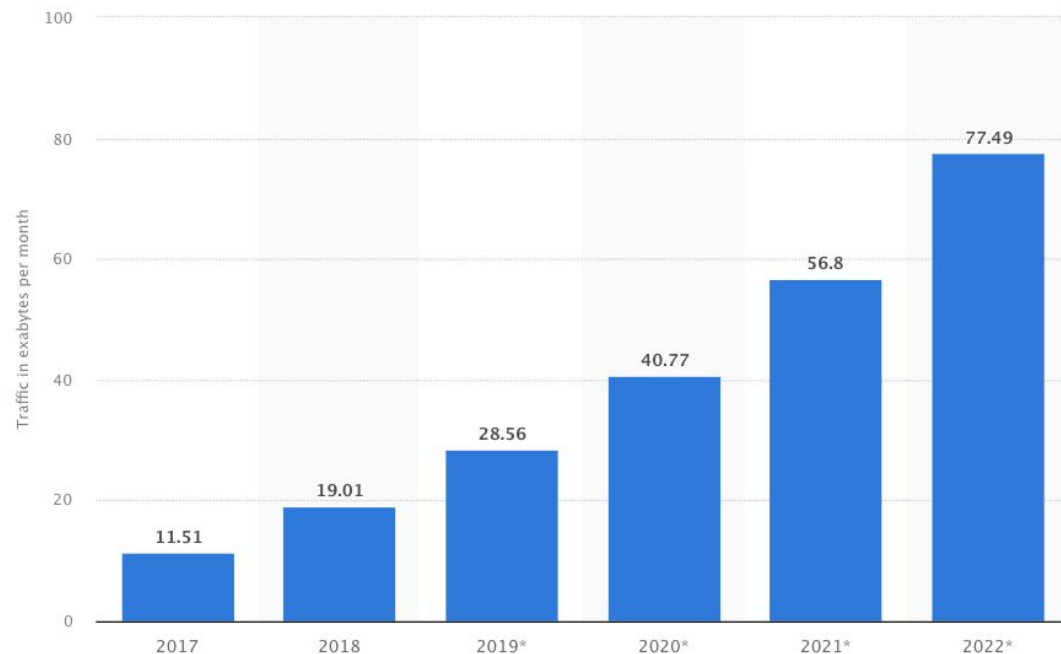
BIG DATA (DADOS NÃO PARAM...)



DADOS NÃO PARAM (5G chegando)

Global mobile data traffic from 2017 to 2022

(in exabytes per month)



© Statista 2021

[Additional Information](#)

[Show source](#)

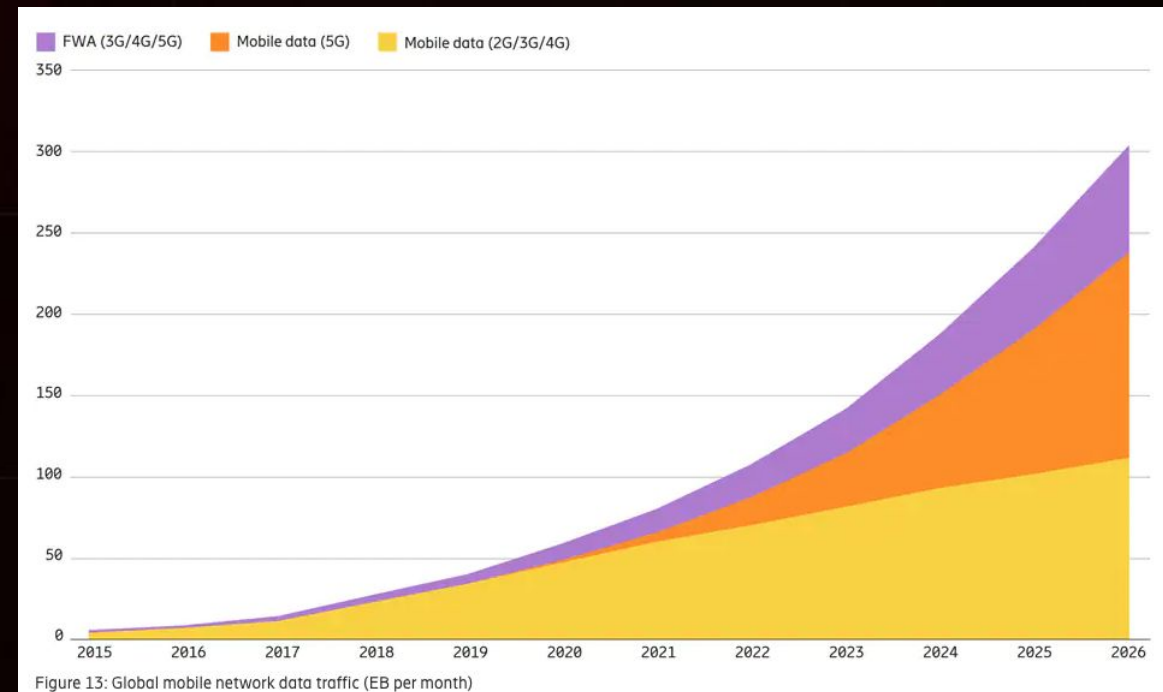


Figure 13: Global mobile network data traffic (EB per month)

DADOS NÃO PARAM

- GPS (pessoas e carros).
- Sensores dos mais variados.
 - Jogador de futebol usa um top;
 - Gado usa sensores (colar, brinco);
 - Toda navegação em e-commerce (e em muitos outros sites é monitorada);
 - Cada compra gera uma infinidade de dados;
 - Onde está, valor que pagou, quantos ingressos comprou, como pagou, onde estava, que bandeira de cartão usou....
 - Já observaram dados de voos?
 - <http://www.flightradar24.com>
 - Cada curtida, cada comentário, cada foto!

DADOS ESTRUTURADOS

- Aqueles que possuem formato e comprimento definido, como por exemplo, números, datas e grupos de palavras.
- Consistem de um conjunto de dados definidos a partir de um esquema formalmente definido.
- Exemplo: dados armazenados em bancos de dados relacionais; dados numa estrutura XML regida por um documento XSD; dados de planilhas com clareza estrutural; dados oriundos de sensores e equipamentos, desde que com uma estrutura de metadados bem definida.

DADOS NÃO ESTRUTURADOS

- Consistem em conjuntos de dados que não têm uma estrutura definida.
- Em razão de não haver uma estrutura formal, a extração de informações nesses conjuntos de dados torna-se complexa do ponto de vista computacional.
- Exemplos de dados são: áudios, vídeos, documentos em formato texto, imagens (e fotos), dados de mídias sociais, entre outros.

DADOS SEMI-ESTRUTURADOS

- Consistem em dados com uma estrutura implícita e flexível, geralmente um meio-termo entre a estruturação e a falta total de estruturação.
- Mesmo a estrutura não sendo rígida, a existência de uma mínima estrutura implícita facilita a gestão dos dados.
- Exemplos são arquivos tabulares em planilhas (formatos TSV e CSV), arquivos XML, conteúdos Web acompanhados de tags;

IOT (INTERNET OF THINGS)

- Internet das coisas.
- Termo utilizado para caracterizar a ideia de que a internet pode estar presente em todas as coisas.
- Tem como conceito fundamental a fusão do mundo real com o mundo digital, fazendo com que os indivíduos estejam em constante comunicação e interação com pessoas e objetos.

IOT (INTERNET OF THINGS)

- IOT possui funções de reconhecimento inteligente, localização, rastreamento e gerenciamento dos diversos dispositivos, trocando informações a todo momento.
- Soluções de IOT são implementadas com redes de identificação por radiofrequência, sensores, tecnologias inteligentes e nanotecnologias, possibilitando uma grande variedade de serviços.

BUSINESS INTELLIGENCE (BI)

- Muita gente não consegue distinguir a diferença entre BI e DS, e é fato que no ambiente corporativo há uma grande semelhança, principalmente porque os dois se utilizam da **conversão de dados brutos em conhecimento**, para tomada de decisão.
- Basicamente o princípio do BI preza pela possibilidade de visualização de dados de forma interativa, de forma que seja possível manipular esses dados com intuito de analisar e comparar situações e desempenhos históricos e atuais, objetivando gerar insights que podem servir como base para tomada de decisão.

BI e DS

- DS por sua vez também transforma dados brutos em cenários (que também podem ser interativos como o BI), entretanto no caso do DS o objetivo principal é **projetar situações**, o que podemos entender como um **olhar para o futuro**.
- De certa forma há muito em comum entre BI e DS, e tecnologias que nasceram exclusivas para cada um deles tornam-se ainda mais poderosas quando passam a fazer parte de um grande arsenal para ser utilizado com os dois propósitos.
- Exemplo: uma empresa que investe em marketing digital, o BI informa se tá dando retorno, o DS indica como investir nas próximas oportunidades!

BI e DS

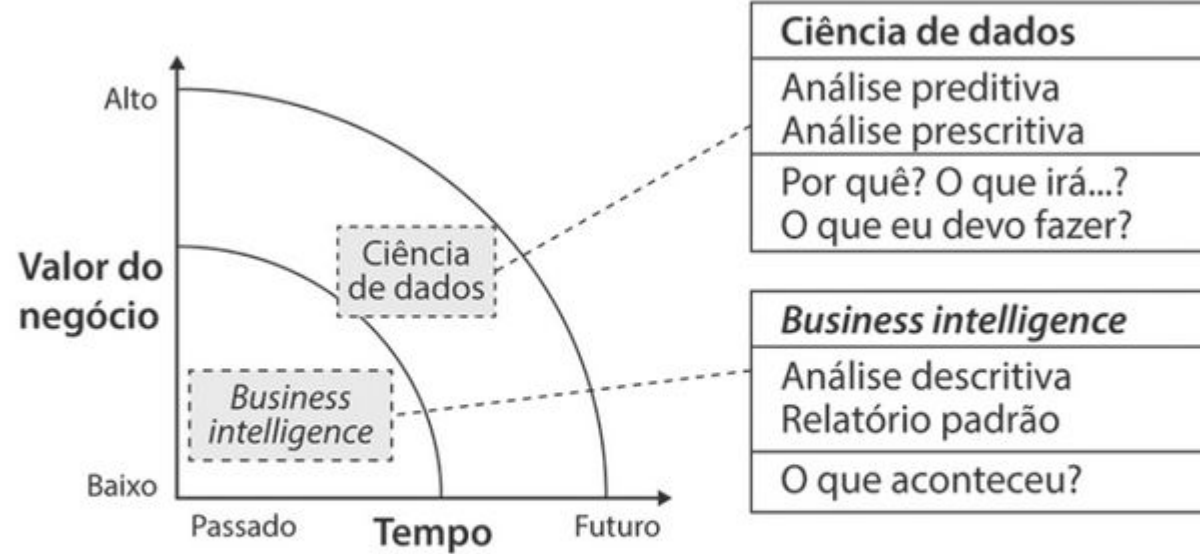


Figura 2. *Business intelligence* versus ciência de dados.

Fonte: Adaptada de Schmarzo (2016).

FINALIZANDO

- Dado, informação e conhecimento!
- Indústria 4.0 e os dados passando a ser ativos valiosos.
- Big Data.
- Dados que não param.
- IOT
- Data Science x Business Intelligence, combinar os dois é sensacional!

INTRODUÇÃO À CIÊNCIA DE DADOS

Ciência de Dados e a Indústria 4.0