



# Agenda



- Entendendo o Big Data
- Carreiras em Ciência de Dados e Big Data
- Áreas de Atuação



# Entendendo o Big Data...

“**DADOS SÃO O  
NOVO PETRÓLEO.**”  
Clive Humby



**The  
Economist**

MAY 6TH–12TH 2017

Theresa May v Brussels

Ten years on: banking after the crisis

South Korea's unfinished revolution

Biology, but without the cells

## The world's most valuable resource



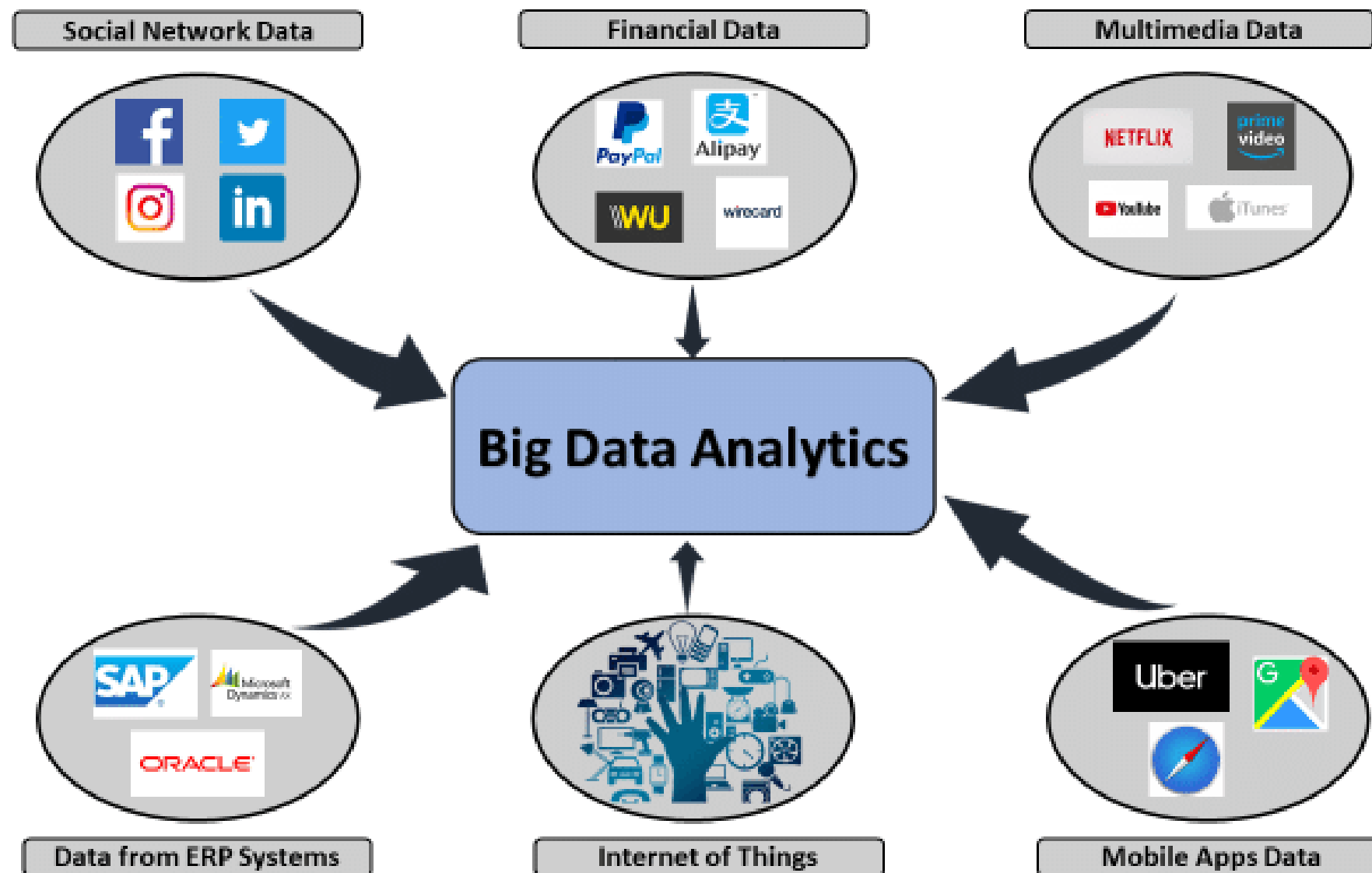
**Data and the new rules  
of competition**

Maio/2017

BIG DATA são dados com maior variedade (*diferentes formatos*) que chegam em volumes crescentes (*tamanho*) e com velocidade cada vez maior (*gerados em tempo real*).

De forma simplificada, BIG DATA é um conjunto de dados maior e mais complexo, especialmente de novas fontes de dados. Esses conjuntos de dados são tão volumosos que o software tradicional de processamento de dados simplesmente não consegue gerenciá-los. No entanto, esses grandes volumes de dados podem ser usados para resolver problemas de negócios que você não conseguiria resolver antes.

# Entendendo o Big Data : fontes de dados





## HOW BIG IS A PETABYTE?

**11,000 4k movies**



It would take you over 2.5 years of nonstop binge watching to get through a petabyte's worth of 4k movies



**20+ PB of data**  
in the Library of Congress



If you took a petabyte's worth of 1GB flash drives and lined them up end to end, they would stretch over

**92 football fields**



**4,000 digital photos**  
every day for the rest of your life

### A ESCALA DOS BYTES

1 BYTE = 1 CARACTERE (A LETRA "A", POR EXEMPLO)

KILOBYTE = 1 000 BYTES ( $10^3$ )

MEGABYTE = 1 000 000 BYTES ( $10^6$ )

GIGABYTE = 1 000 000 000 BYTES ( $10^9$ )

TERABYTE = 1 000 000 000 000 BYTES ( $10^{12}$ )

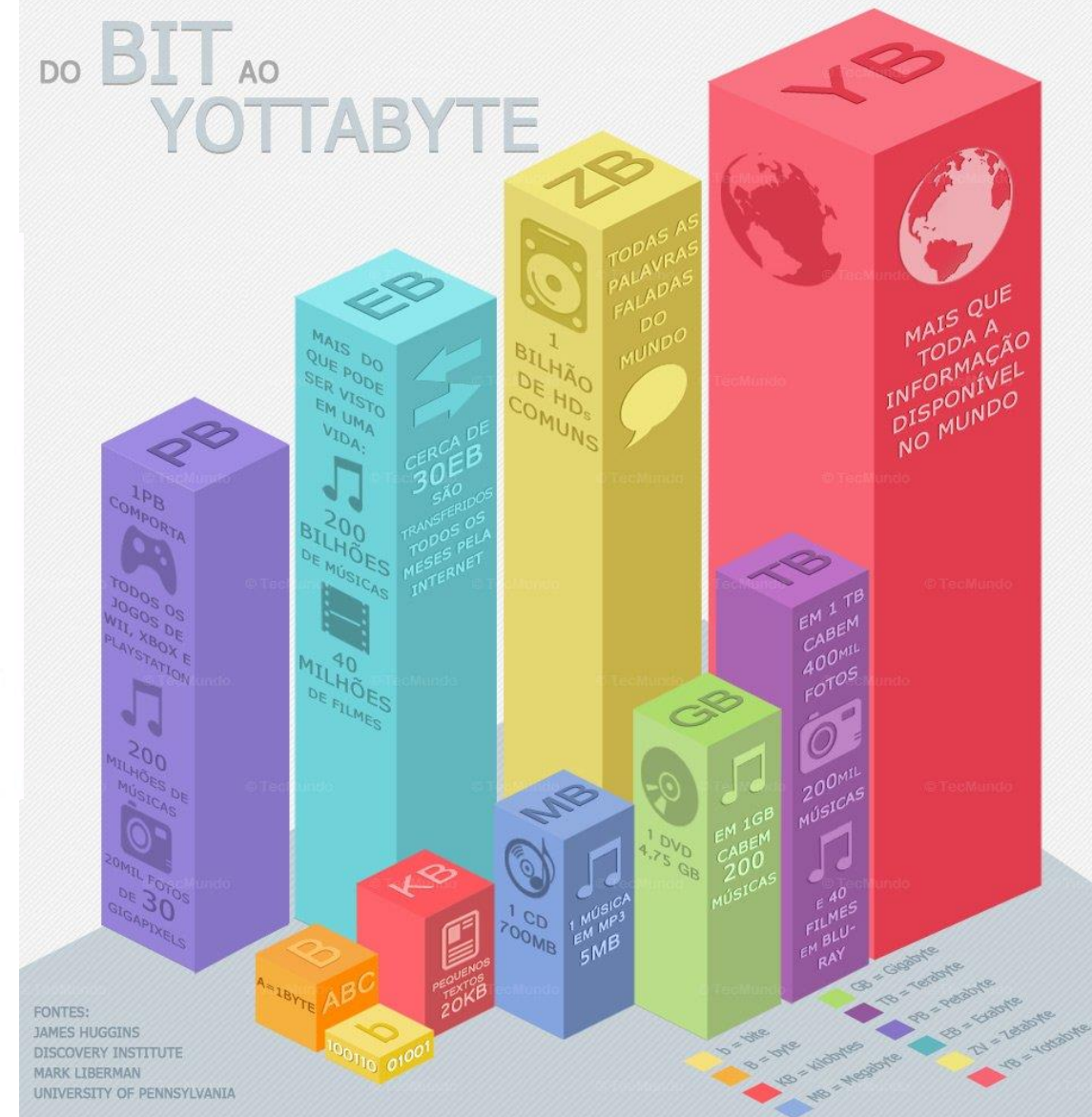
PETABYTE = 1 000 000 000 000 000 BYTES ( $10^{15}$ )

EXABYTE = 1 000 000 000 000 000 000 BYTES ( $10^{18}$ )

ZETTABYTE = 1 000 000 000 000 000 000 000 BYTES ( $10^{21}$ )

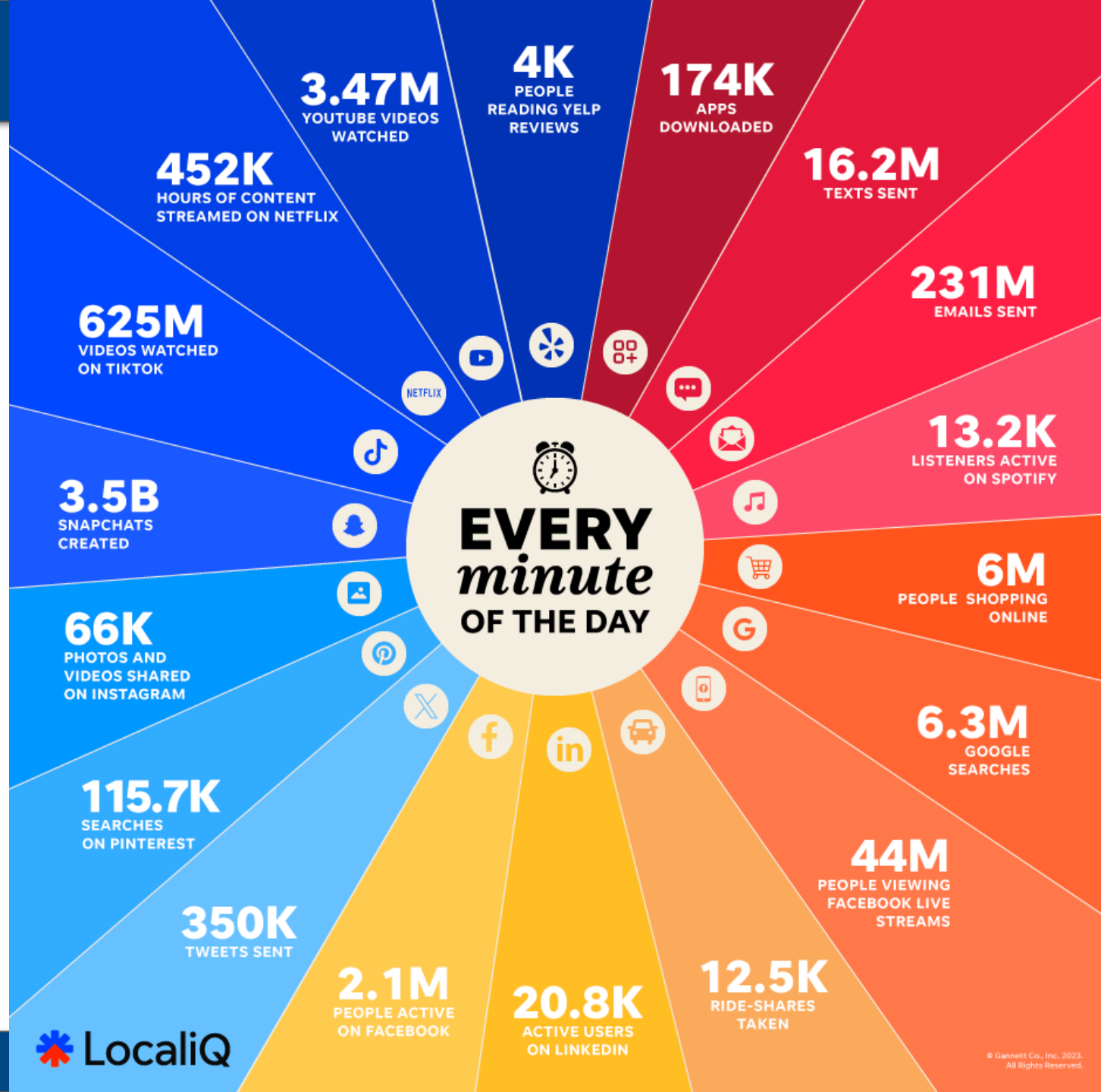
YOTTABYTE = 1 000 000 000 000 000 000 000 000 BYTES ( $10^{24}$ )

## DO BIT AO YOTTABYTE



COPYRIGHT

# Entendendo o Big Data : muitos dados a cada minuto





# Entendendo o Big Data : variedade dos dados

## Dados estruturados

**Ex.:** Banco de dados

**Estrutura rígida**  
Projetada previamente  
Representação homogêna

Cada campo de dados tem um formato bem definido.

Formato é um padrão aceito pelo campo.

Dados de um mesmo registro possuem relação entre eles.

Registros possuem valores diferentes, mas mesmos atributos.

Atributos ou campos são definidos por um esquema.

## Dados semi estruturados

**Ex.:** XML, JSON, RDF, OWL.

**Estrutura flexível**  
Representação heterogêna

Cada campo de dados tem uma estrutura, mas não existe uma imposição de formato

O esquema é criado com a definição de elementos internos dos arquivos (nós), legíveis para seres humanos









## Dados não estruturados

**Ex.:** Textos, arquivos, documentos, imagens, vídeos, áudios, redes sociais etc.

**Sem estrutura**  
(ou com estrutura mínima de arquivo)

Mais de 80% dos dados gerados no mundo é deste tipo

### Unstructured data types

 Text files and documents	 Server, website and application logs	 Sensor data	 Images
 Video files	 Audio files	 Emails	 Social media data



JAN  
2024

# ESSENTIAL DIGITAL HEADLINES

OVERVIEW OF THE ADOPTION AND USE OF CONNECTED DEVICES AND SERVICES



TOTAL  
POPULATION



**8.08**  
**BILLION**

URBANISATION

**57.7%**

we  
are  
social

UNIQUE MOBILE  
PHONE SUBSCRIBERS



**5.61**  
**BILLION**

vs. POPULATION

**69.4%**

Meltwater

INDIVIDUALS USING  
THE INTERNET



**5.35**  
**BILLION**

vs. POPULATION

**66.2%**

KEPIOS

SOCIAL MEDIA  
USER IDENTITIES



**5.04**  
**BILLION**

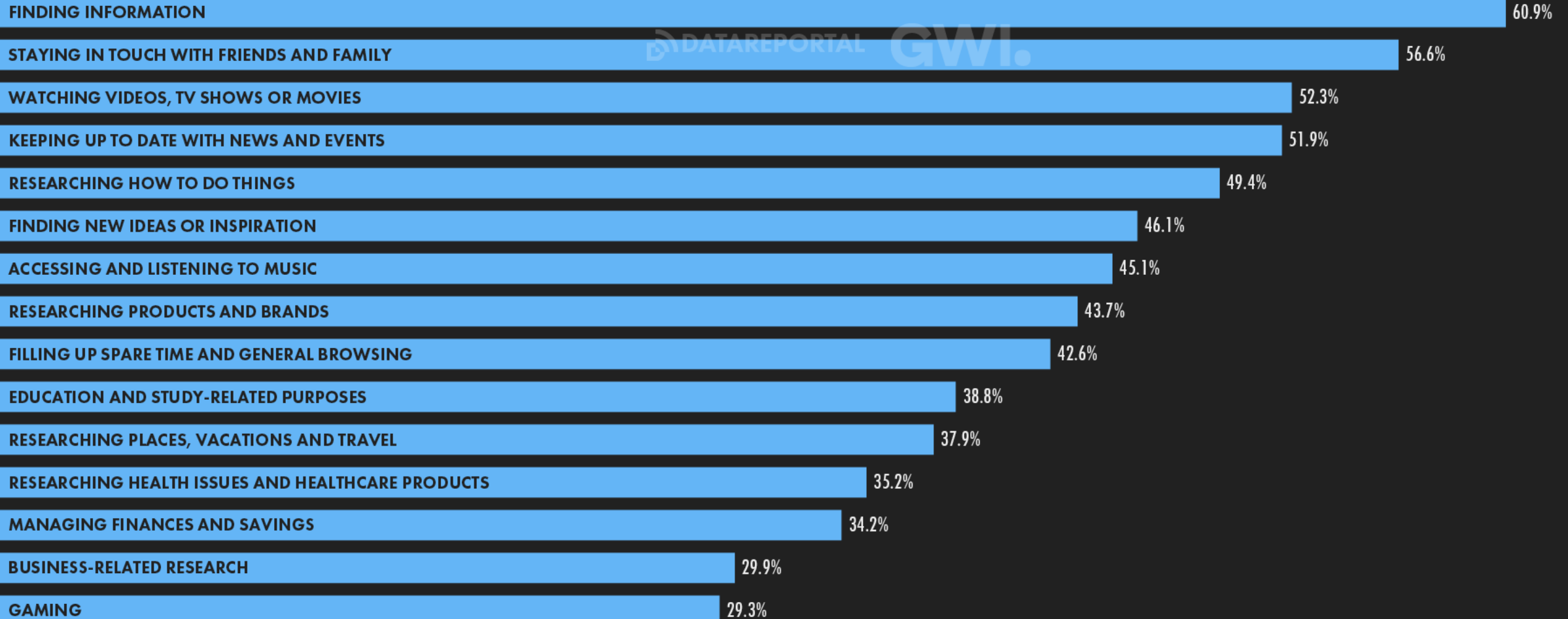
vs. POPULATION

**62.3%**

JAN  
2024

# MAIN REASONS FOR USING THE INTERNET

PRIMARY REASONS WHY INTERNET USERS AGED 16 TO 64 USE THE INTERNET



JAN  
2024

# ECOMMERCE: CONSUMER GOODS CATEGORIES

ESTIMATED ANNUAL SPEND IN EACH CONSUMER GOODS ECOMMERCE CATEGORY (B2C ONLY, U.S. DOLLARS, FULL-YEAR 2023)



GLOBAL OVERVIEW

ELECTRONICS



**\$781.3**  
BILLION

YEAR-ON-YEAR CHANGE  
**+11.8% (+\$82 BILLION)**

FASHION



**\$673.6**  
BILLION

YEAR-ON-YEAR CHANGE  
**+8.4% (+\$52 BILLION)**

FOOD



**\$370.7**  
BILLION

YEAR-ON-YEAR CHANGE  
**+16.0% (+\$51 BILLION)**

BEVERAGES



**\$209.3**  
BILLION

YEAR-ON-YEAR CHANGE  
**+14.2% (+\$26 BILLION)**

DIY & HARDWARE



**\$201.8**  
BILLION

YEAR-ON-YEAR CHANGE  
**+8.5% (+\$16 BILLION)**

FURNITURE



**\$188.0**  
BILLION

YEAR-ON-YEAR CHANGE  
**+11.6% (+\$20 BILLION)**

PHYSICAL MEDIA



**\$182.7**  
BILLION

YEAR-ON-YEAR CHANGE  
**+0.8% (+\$1.5 BILLION)**

BEAUTY &  
PERSONAL CARE



**\$151.2**  
BILLION

YEAR-ON-YEAR CHANGE  
**+5.1% (+\$7.4 BILLION)**

TOBACCO  
PRODUCTS



**\$104.4**  
BILLION

YEAR-ON-YEAR CHANGE  
**+1.9% (+\$1.9 BILLION)**

TOYS &  
HOBBY



**\$80.53**  
BILLION

YEAR-ON-YEAR CHANGE  
**+5.2% (+\$4.0 BILLION)**

HOUSEHOLD  
ESSENTIALS



**\$72.98**  
BILLION

YEAR-ON-YEAR CHANGE  
**+11.9% (+\$7.8 BILLION)**

OVER-THE-COUNTER  
PHARMACEUTICALS



**\$59.65**  
BILLION

YEAR-ON-YEAR CHANGE  
**+7.6% (+\$4.2 BILLION)**

LUXURY  
GOODS



**\$54.42**  
BILLION

YEAR-ON-YEAR CHANGE  
**+9.7% (+\$4.8 BILLION)**

EYE-  
WEAR



**\$23.48**  
BILLION

YEAR-ON-YEAR CHANGE  
**+5.9% (+\$1.3 BILLION)**



statista

we  
are  
social

statista

Meltwater

statista

statista

Meltwater

statista

KEPIOS

statista

we  
are  
social





Big Data designa dados e processos cuja escala, distribuição, diversidade e velocidade de criação requer o uso de novas tecnologias de armazenamento, processamento e análise dos dados. Isto envolve competências em computação, matemática e negócios → Faltam profissionais com a formação completa envolvendo estas competências



# Ciência de Dados x Big Data

**Ciência de dados** combina estatística, matemática, programação, inteligência artificial e resolução de problemas para capturar dados de maneiras engenhosas, com uma visão diferente para encontrar padrões e realizar análises, juntamente com as atividades de limpeza, preparação e organização dos dados.



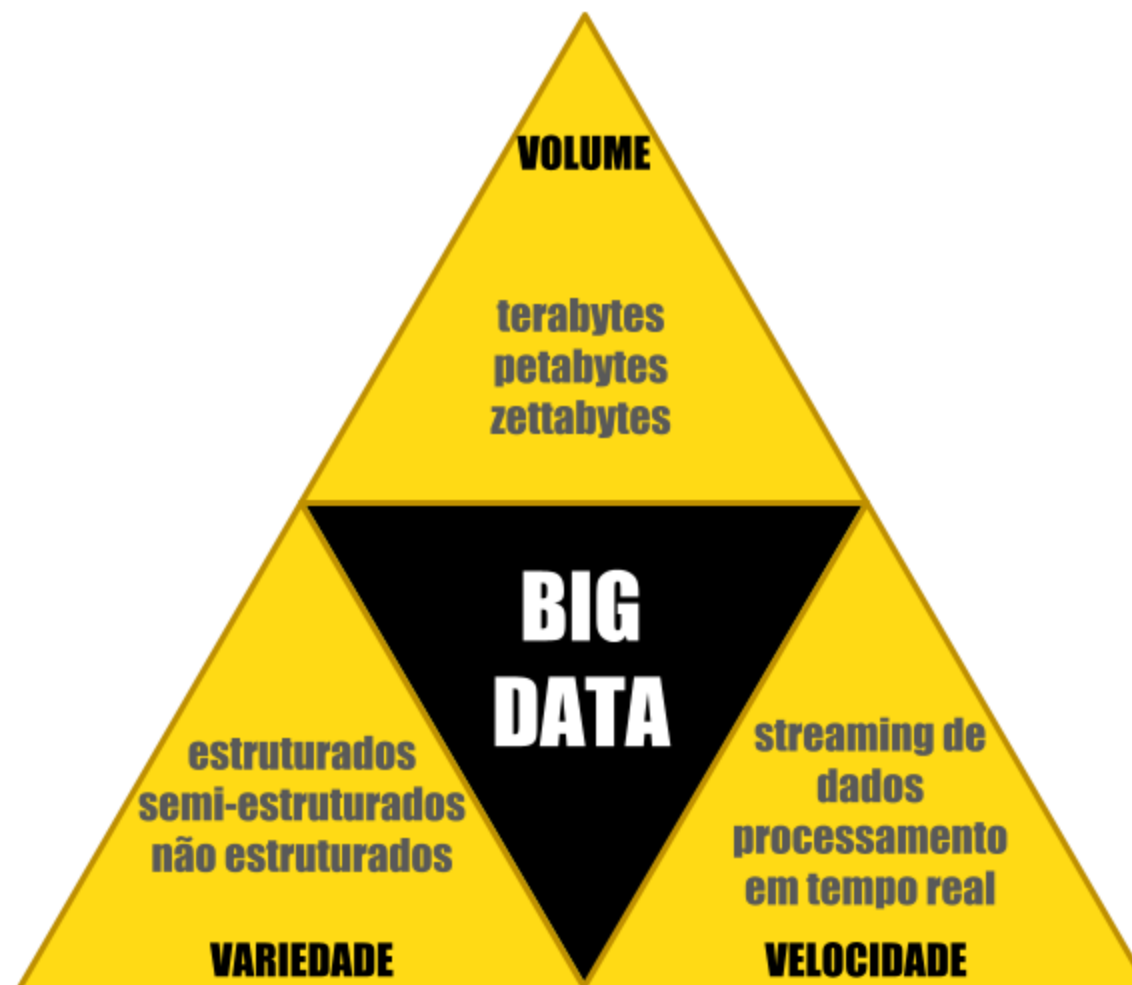
**Métodos Científicos** para transformar, extrair e analisar informações do grande volume de dados

Conjunto de técnicas e ferramentas capazes de armazenar, extrair, organizar e transformar **grandes volumes de dados**, gerados em **diferentes formatos** e **grande velocidade** em informações que possibilitam uma análise ampla e em tempo hábil.



**Infra-estrutura de Tecnologia**  
(ferramentas para armazenamento, processamento, análise e visualização)

# Inicialmente 3Vs



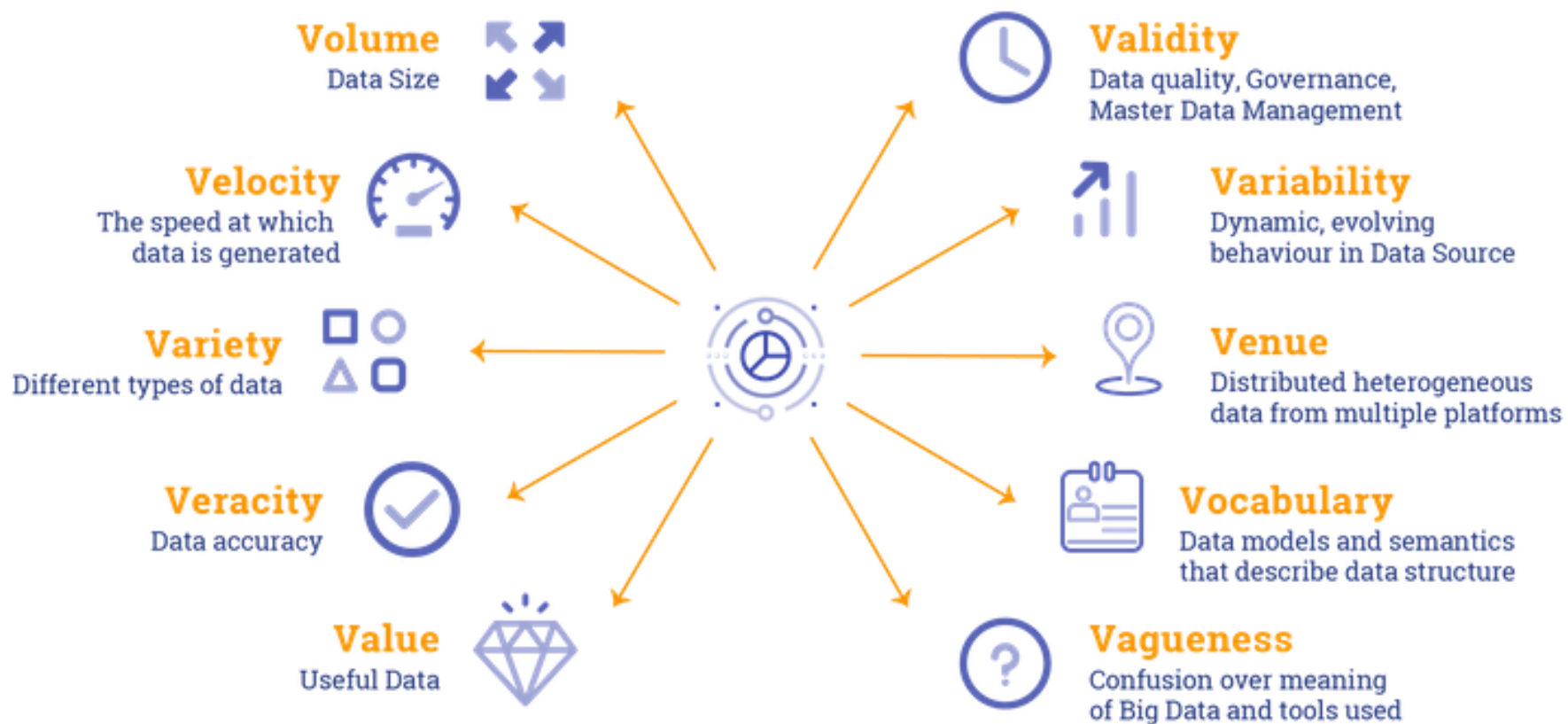


## QUAIS SÃO OS 5 V DO BIG DATA?



SPEEDIO

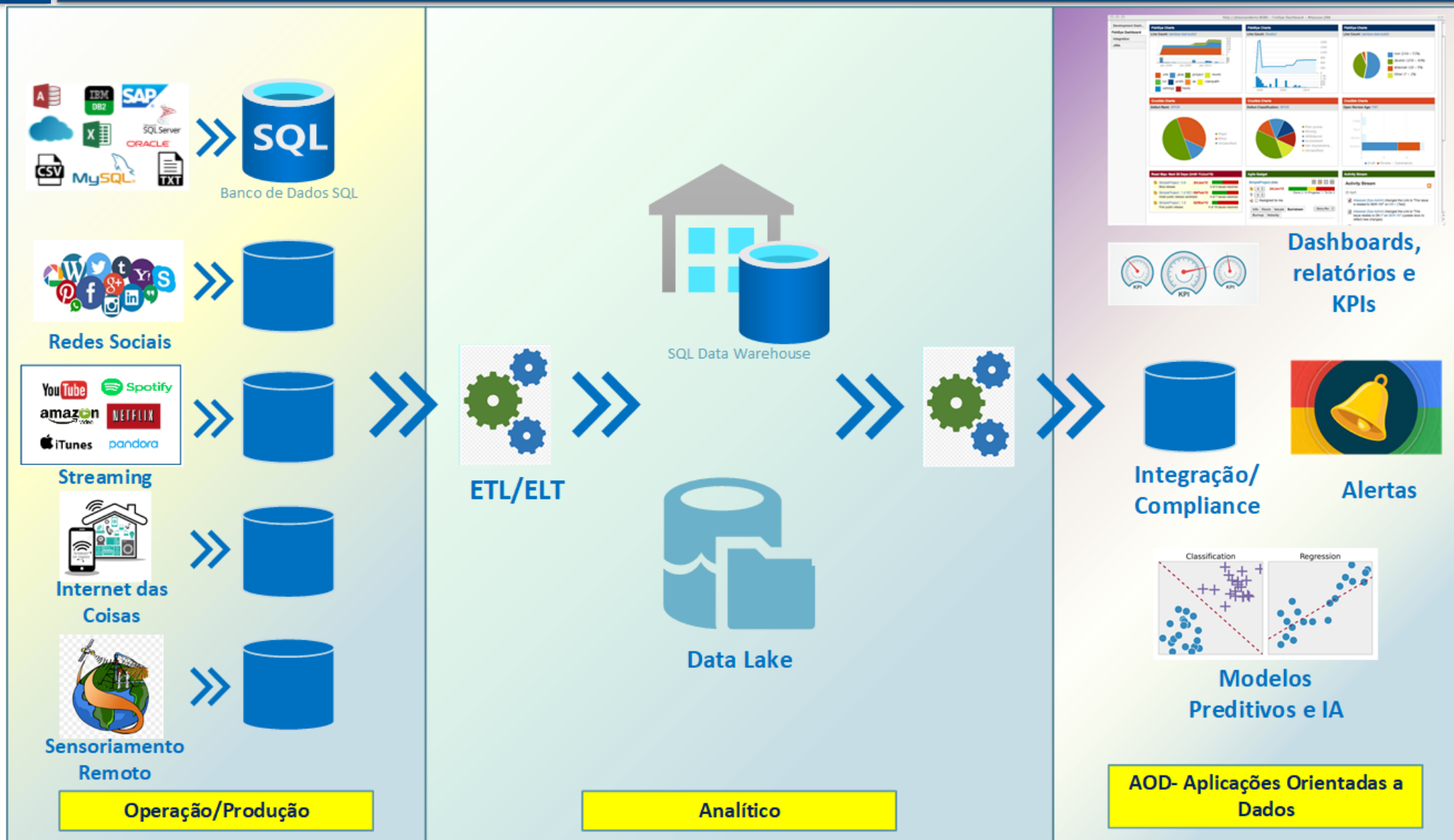
## THE 10 Vs OF BIG DATA



Source: xenonstack.com



# Visão Geral do Processo





# Carreiras em Ciência de Dados/Big Data

- Cientista de Dados 
- Analista de Negócio
- Analista de Dados
- Analista de Business Intelligence 
- Analista de Big Data
- Arquiteto de Dados
- Engenheiro de Dados 
- Engenheiro de Machine Learning 
- Analistas de Marketing
- Estatístico
- Desenvolvedor de Visualização de Dados ( Data Storytelling)

Em alta nos  
últimos dois  
anos (LinkedIn)



<https://business.linkedin.com/talent-solutions/resources/talent-acquisition/jobs-on-the-rise-us>



# Áreas de Atuação e Aplicações

## Saúde

- Prevenção de Epidemias
- Telemedicina
- Tecnologia “vestível”



## Educação

- Diminuir evasão
- Indicadores de desempenho docente e discente
- Ensino Adaptativo



## Negócios

- Fidelização dos clientes
- Maximizar ROI
- Marketing



## Setor Público

Combater corrupção e desvio de receitas

Cidades Inteligentes



## Agronegócio

- Histórico de consumo de água e rotação de cultura
- Geolocalização
- Análise do solo para plantio inteligente



## Mercado Financeiro

- Reduzir as taxas de churn
- Personalizar serviços
- Detecção de Fraudes

