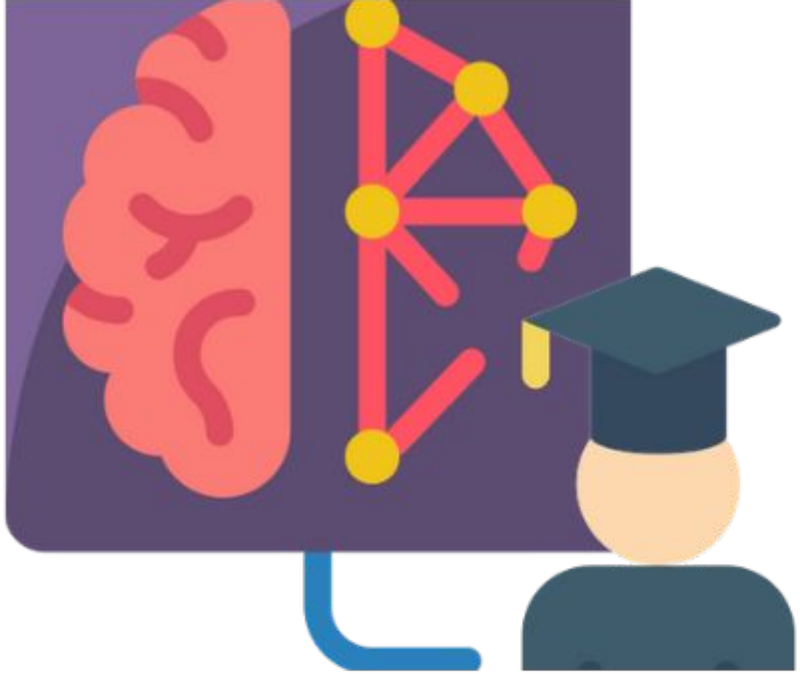


AI-school

Data Science - Big Data Analytics

marcusRB

Presentación Programa

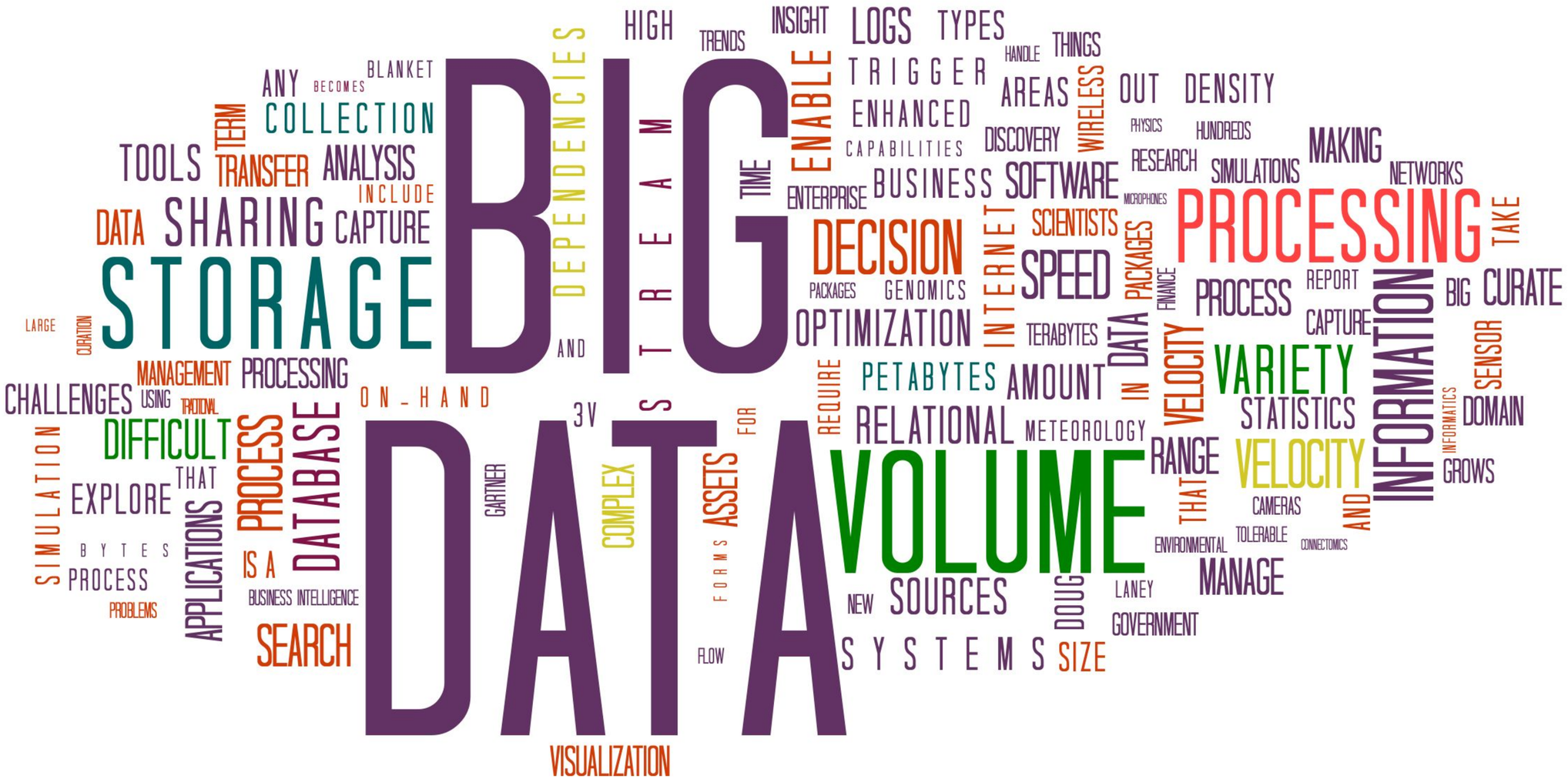
AI-school
Data Science - Big Data Analytics

marcusRB

Big Data

AI-school
Data Science - Big Data Analytics

¿Qué es Big Data?



AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Qué es Big Data?

¿Qué inició la revolución de los datos masivos?

Principios s. XXI...

- El proyecto Sloan Digital Sky Survey
- El proyecto del genoma humano

Después...

- Internet, redes sociales, etc
- IoT (Internet Of Things)

AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Qué es Big Data?

¿Qué inició la revolución de los datos masivos?

Problemas en las empresas de Internet

- **Gran cantidad de datos** acumulados hacía inviable su procesamiento en un único ordenador
- **Heterogeneidad de datos.** Necesidad de insertar, consultar y procesar datos en diferentes estructuras
- Necesidad de **procesar rápidamente los datos**
  - Ejemplo: queries del buscador web

AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Qué es Big Data?

Hasta el 2005 se han generado...

130 exabytes de datos

AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Qué es Big Data?

A

1 Byte

8 bits

10010010

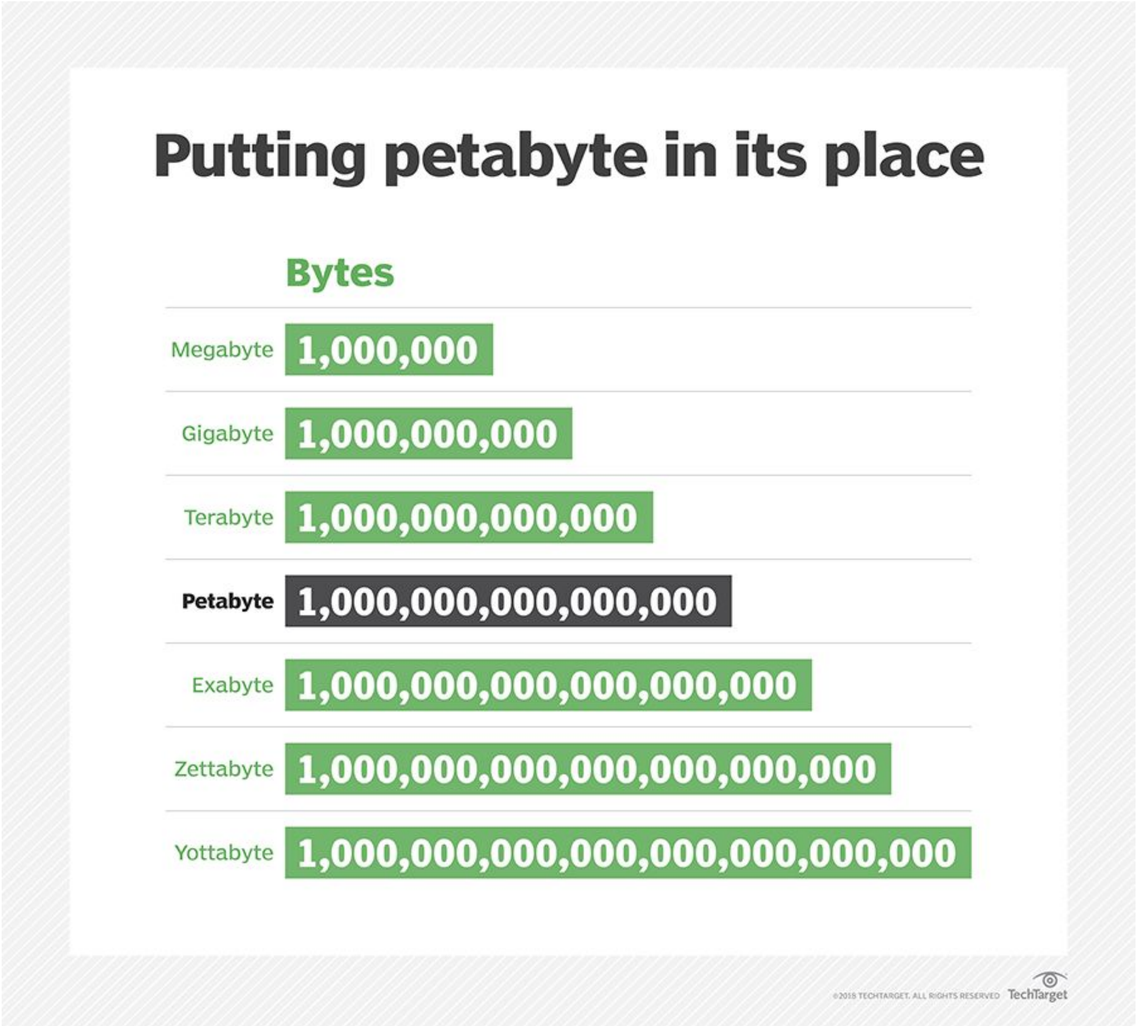
1 bit

0 = false/cerrado 1

= true/abierto

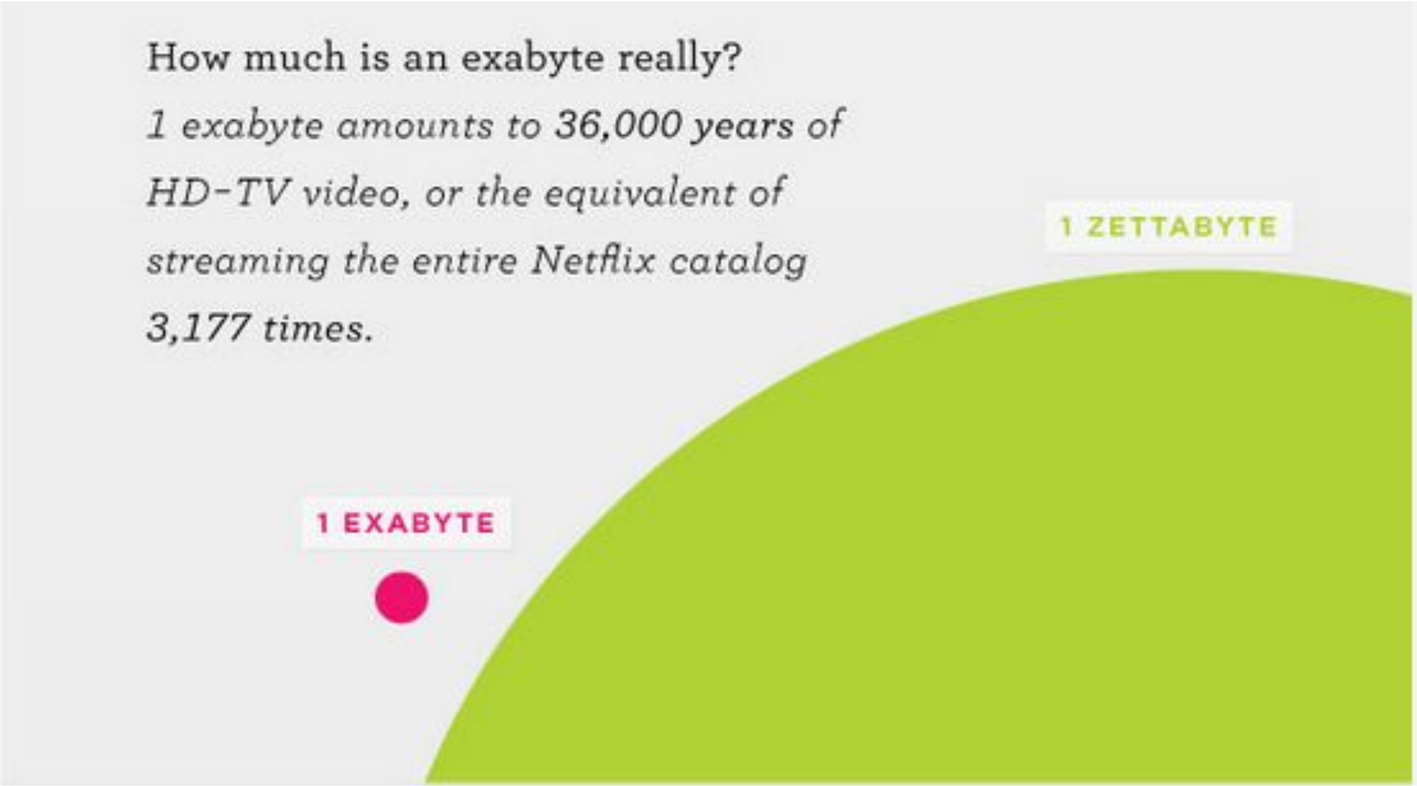
AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Qué es Big Data?



AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Qué es Big Data?



AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Qué es Big Data?

marcusRB

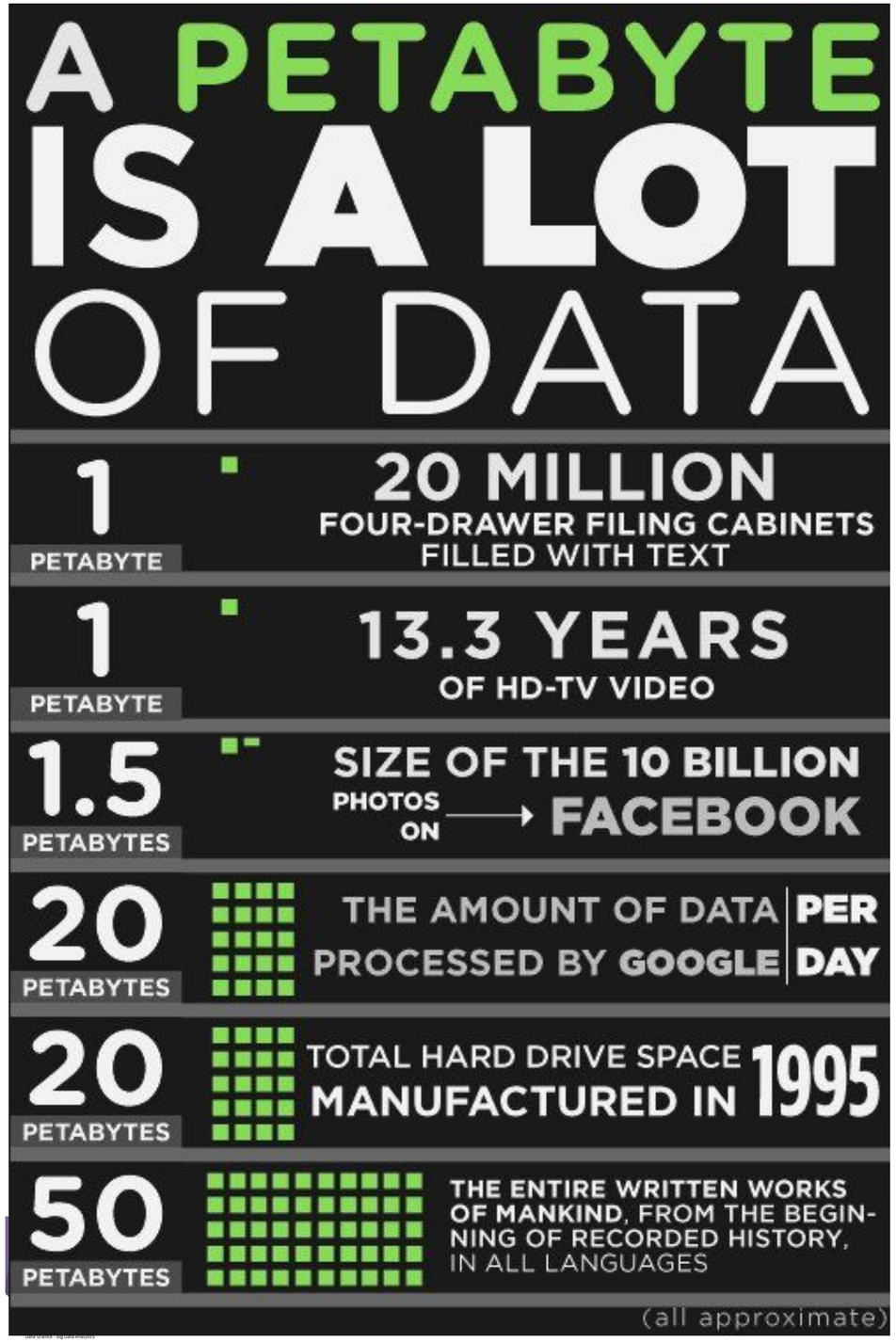
Real-World Data Storage Capacity

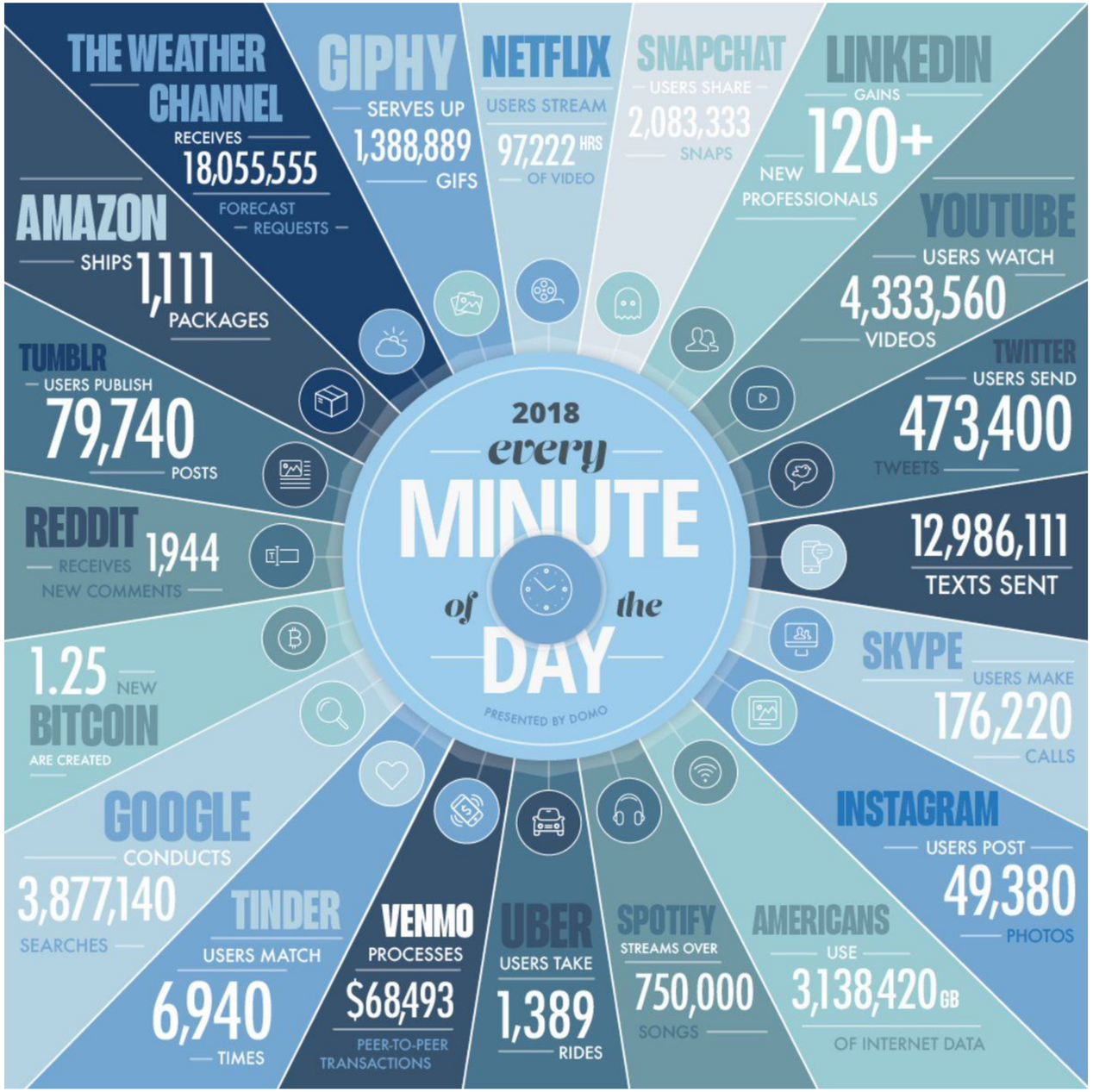
ITEM	CAPACITY REQUIRED
Amount of text on an average print page converted to HTML	1 kilobyte
Typewritten page	2 kilobytes
Short novel	1 megabyte
Downloadable MP3 music file	2 megabytes to 5 megabytes
Pickup truck filled with books	1 gigabyte
50,000 trees made into paper and printed	1 terabyte
Information contained in all U.S. academic research libraries	2 petabytes
All words ever spoken by human beings	5 exabytes

©2017 TECHTARGET. ALL RIGHTS RESERVED TechTarget

AI-school

¿Qué es Big Data?





A DAY IN DATA

The exponential growth of data is undisputed, but the numbers behind this explosion – fuelled by internet of things and the use of connected devcies – are hard to comprehend, particularly when looked at in the context of one day

500m

tweets are sent every day

Twitter

4PB

of data created by Facebook, including

350m photos

100m hours of video watch time

Facebook Research

320bn

emails to be sent each day by 2021

306bn

emails to be sent each day by 2020

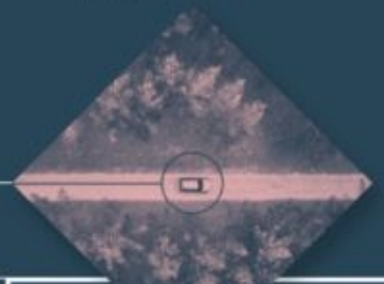
294bn

billion emails are sent

Radicati Group

3.9bn

people use emails



4TB

of data produced by a connected car

Intel

65bn

messages sent over WhatsApp and two billion minutes of voice and video calls made

Facebook

DEMYSTIFIYING DATA UNITS

From the more familiar 'bit' or 'megabyte', larger units of measurement are more frequently being used to explain the masses of data

Unit		Value	Size
b	bit	0 or 1	1/8 of a byte
B	byte	8 bits	1 byte
KB	kilobyte	1,000 bytes	1,000 bytes
MB	megabyte	$1{,}000^2$ bytes	1,000,000 bytes
GB	gigabyte	$1{,}000^3$ bytes	1,000,000,000 bytes
TB	terabyte	$1{,}000^4$ bytes	1,000,000,000,000 bytes
PB	petabyte	$1{,}000^5$ bytes	1,000,000,000,000,000 bytes
EB	exabyte	$1{,}000^6$ bytes	1,000,000,000,000,000,000 bytes
ZB	zettabyte	$1{,}000^7$ bytes	1,000,000,000,000,000,000,000 bytes
YB	yottabyte	$1{,}000^8$ bytes	1,000,000,000,000,000,000,000,000 bytes

*A lowercase "b" is used as an abbreviation for bits, while an uppercase "B" represents bytes.

463EB

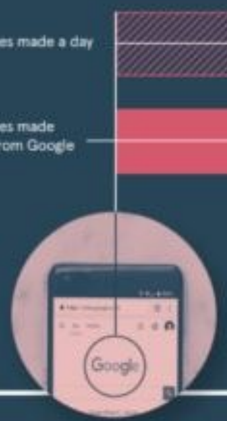
of data will be created every day by 2025

IDC

95m

photos and videos are shared on Instagram

Instagram Business



Searches made a day → 5bn

Searches made a day from Google → 3.5bn

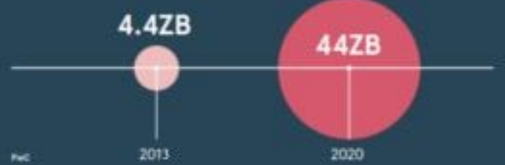
Smart Insights

28PB

to be generated from wearable devices by 2020

Statista

ACCUMULATED DIGITAL UNIVERSE OF DATA



PwC

RACONTEUR

¿Qué es Big Data?

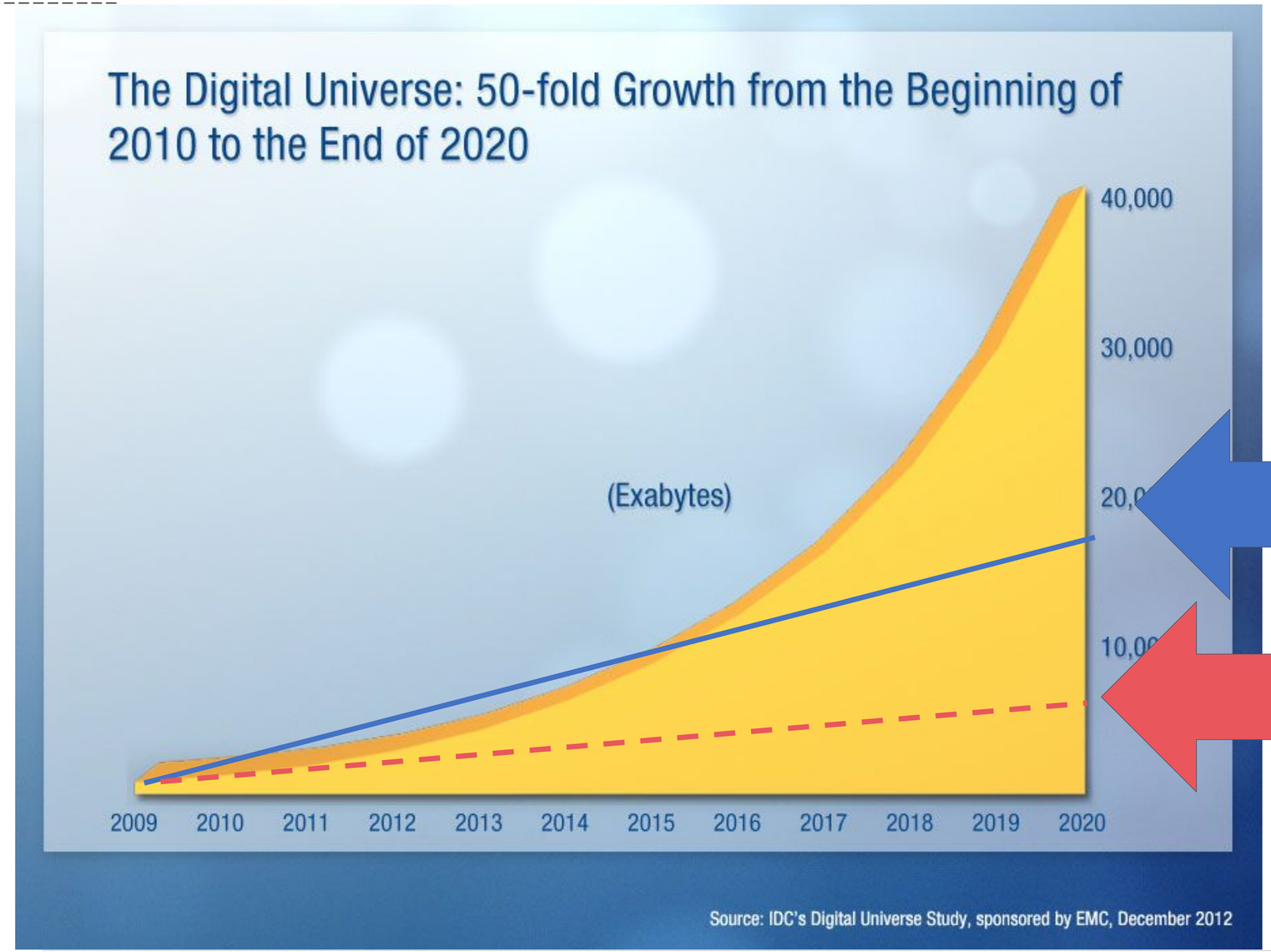
2005 - 130 exabytes

2010 - 1.200 exabytes

2015 - 7.900 exabytes

2020 - 40.900 exabytes

¿Qué es Big Data?



Machines

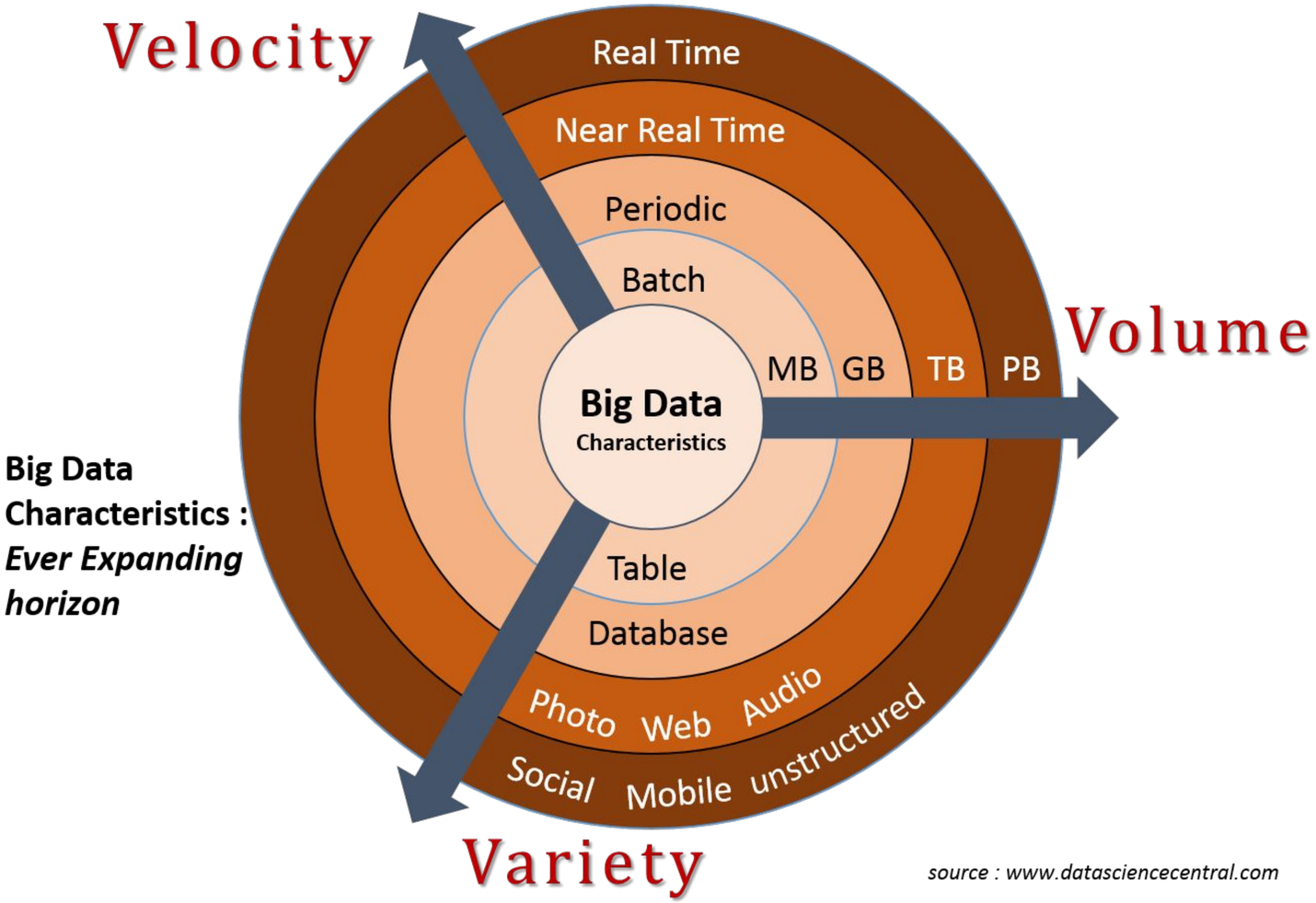
Data Roles

AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Qué es Big Data?

2001

Definición de las 3 v's



Big Data Characteristics : ***Ever Expanding horizon***

source : www.datasciencecentral.com

AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

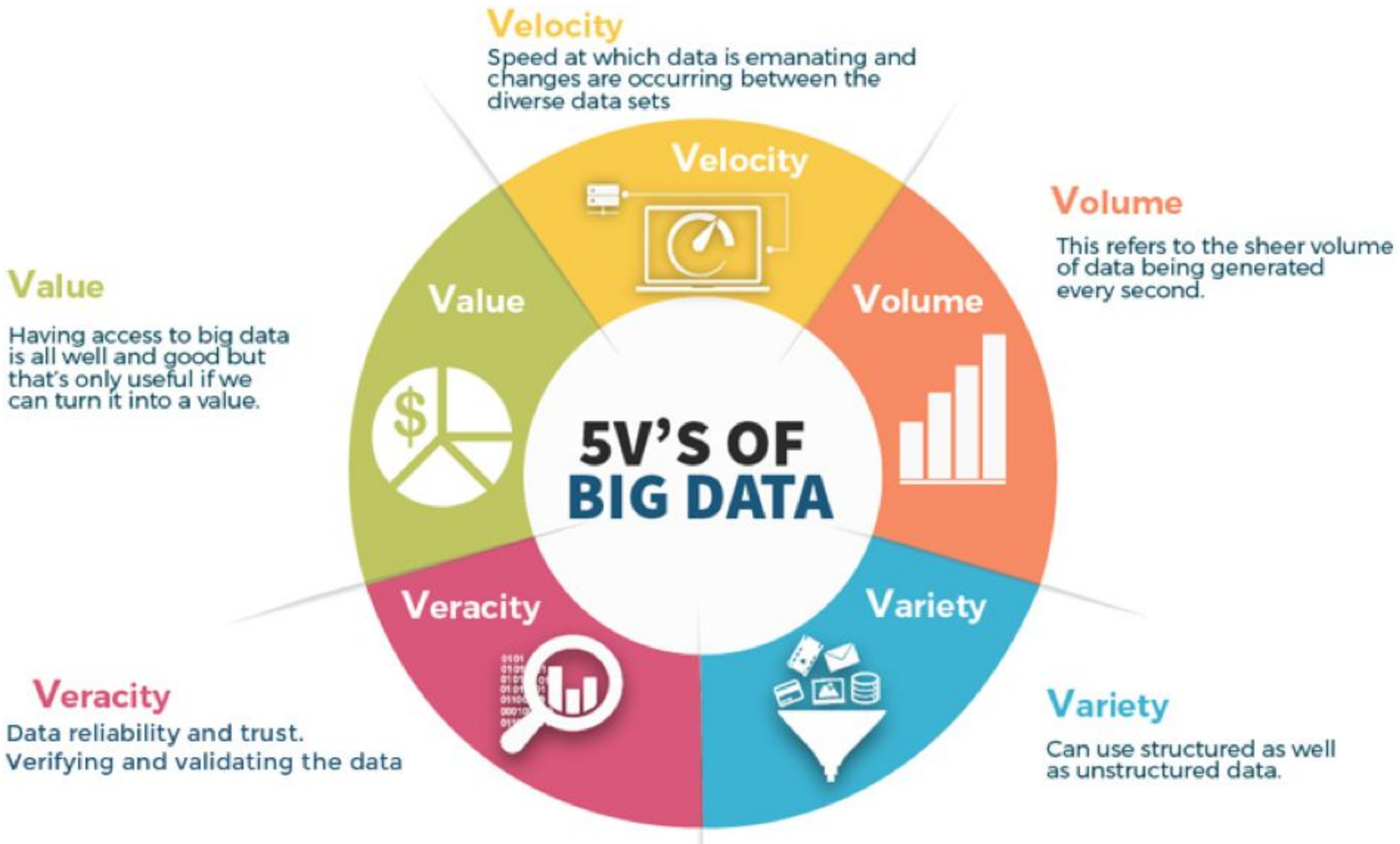
¿Qué es Big Data?

Definición
de las
4 v's

AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Qué es Big Data?

Definición de las 5 v's



AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Qué es Big Data?

...y también...

AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

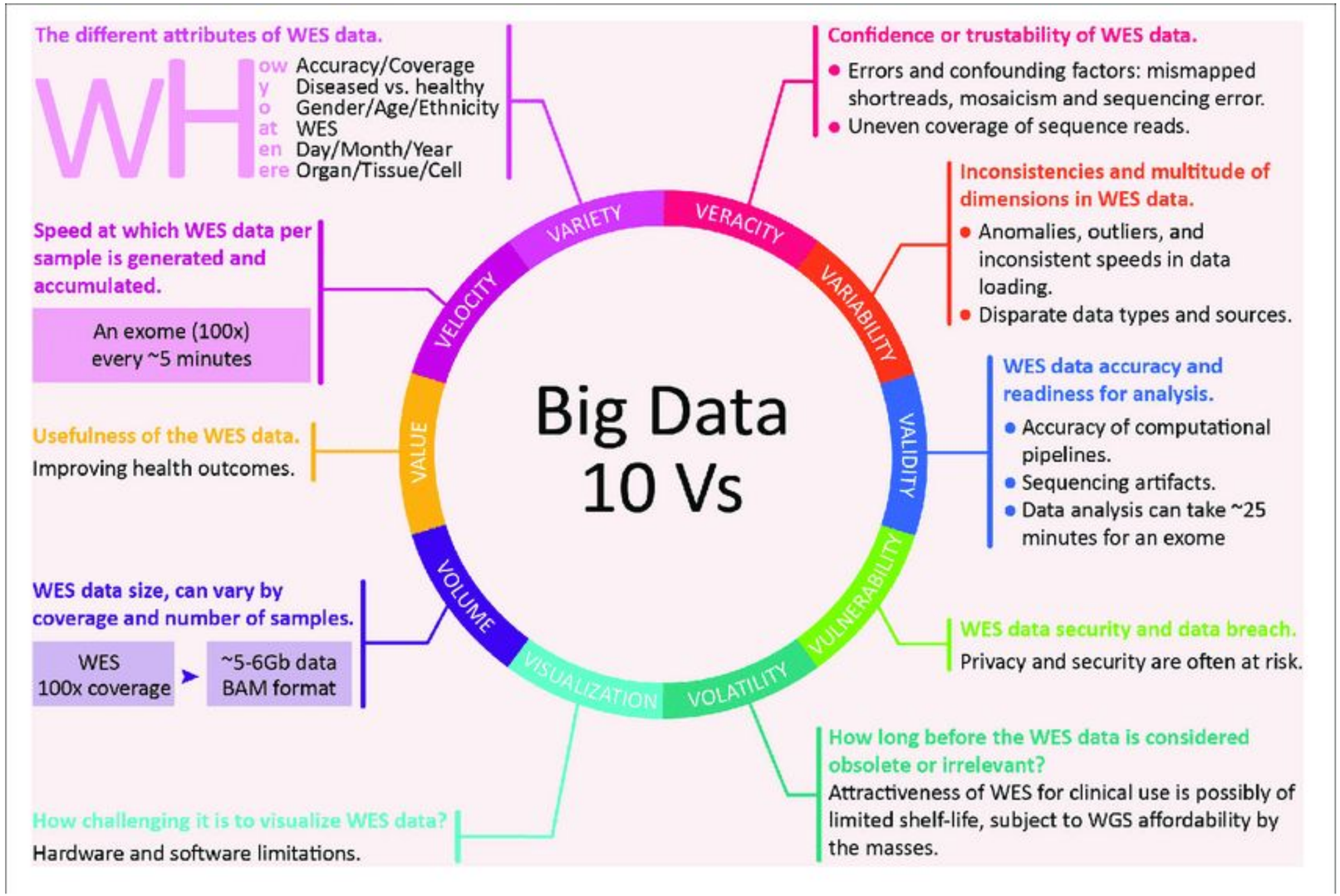
¿Qué es Big Data?

Definición de las 8 v's

AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Qué es Big Data?

Definición de las 10 v's



AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

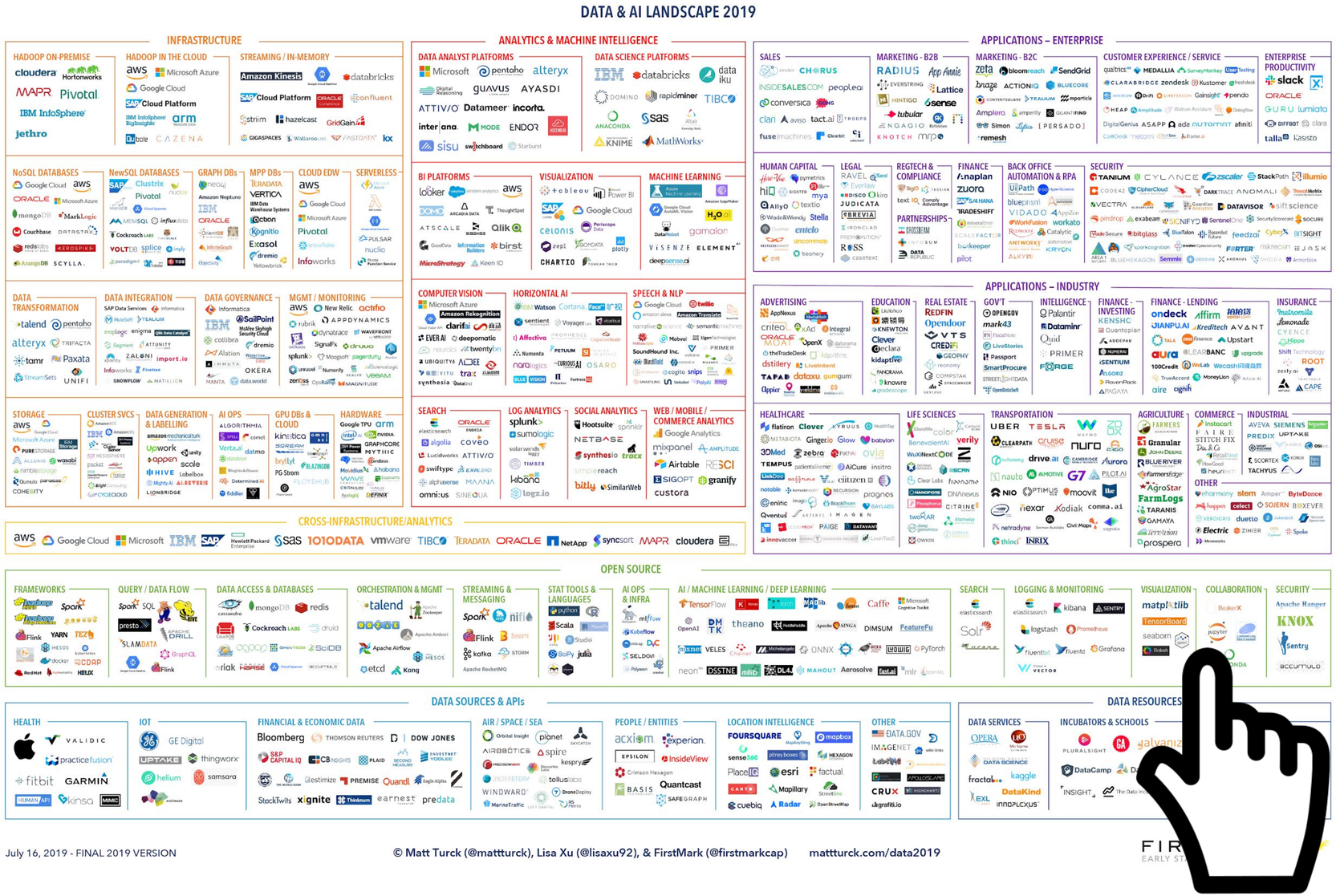
¿Qué es Big Data?

En el 2001, el analista Doug Laney de META Group (ahora Gartner) utilizaba y definía el término Big Data como:

*"el conjunto de **técnicas** y **tecnologías** para el tratamiento de **datos**, en entornos de **gran volumen**, **variedad** de orígenes y en los que la **velocidad** de respuesta es crítica".*

AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

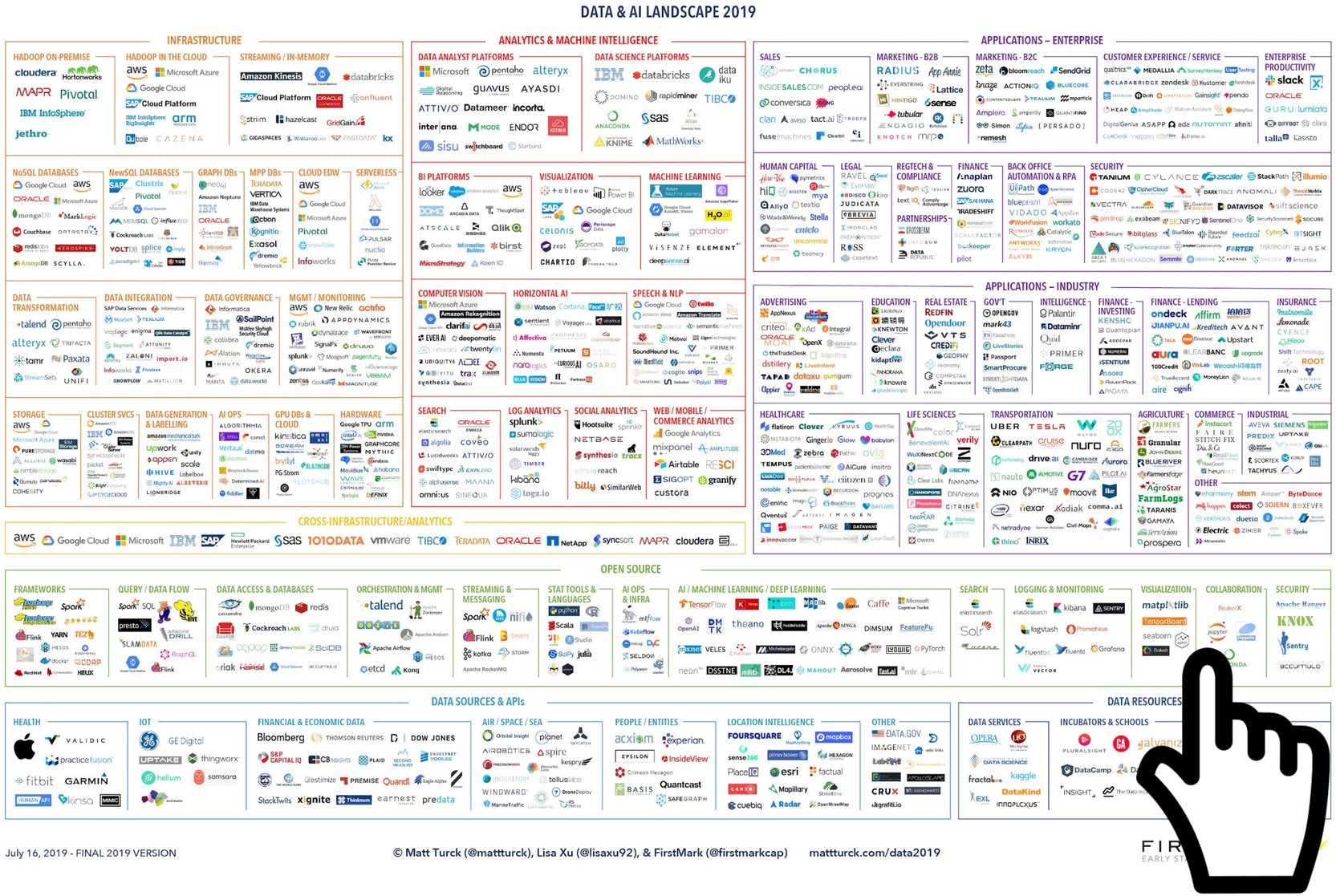
Big Data ecosystem 2016



AI-school
Data Science - Big Data Analytics

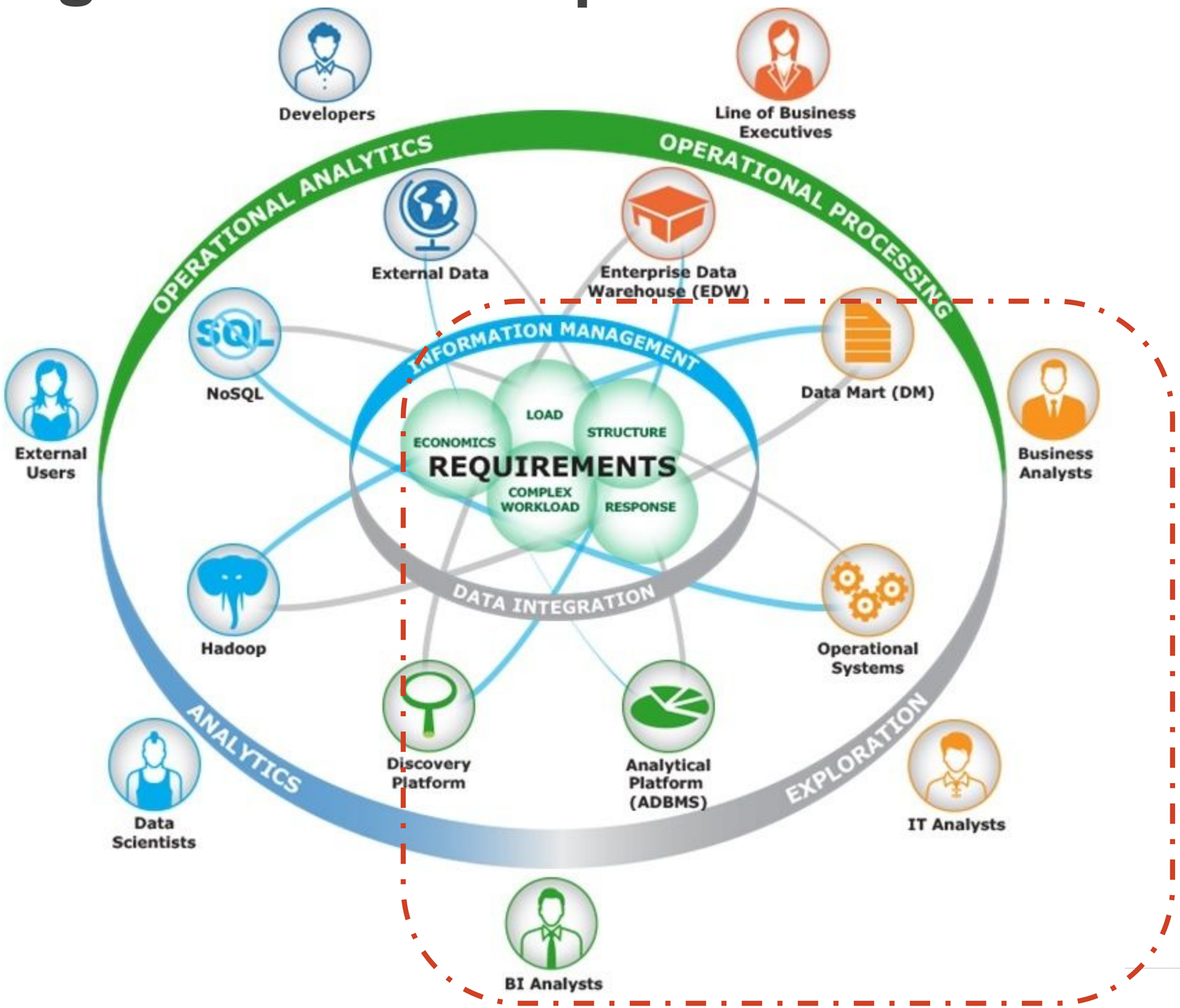
AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

Big Data ecosystem ...2020

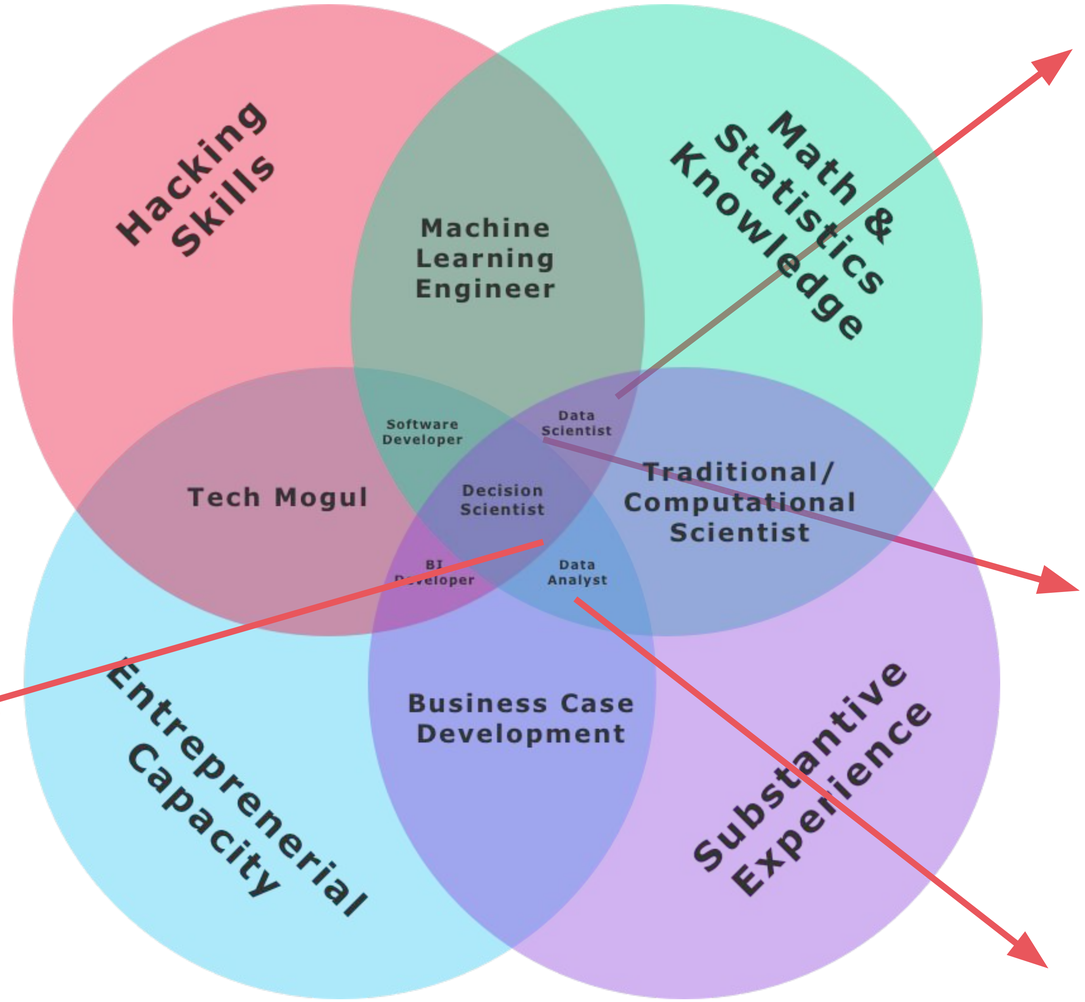


AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Quién son las figuras en una empresa?



¿Qué rol tiene un Data Analyst? ... 4 áreas



Pensamiento analítico / Algoritmos

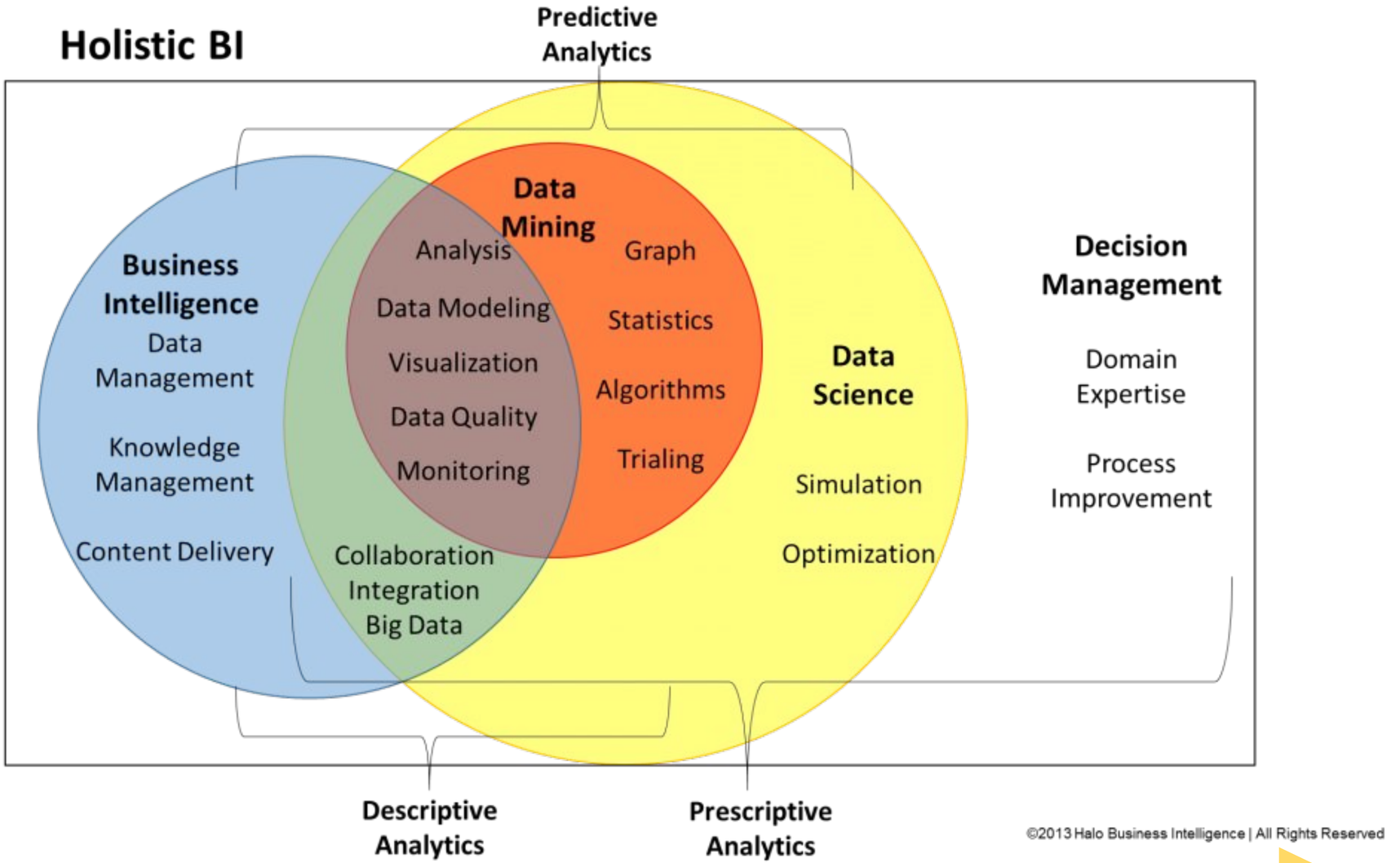
Pensamiento Computacional

Capacidad de análisis y visualización

Business / Comunicación

AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

...y, un Business Analyst?



AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

¿Qué es capaz de realizar?

TRANSFORMAR INFORMACIÓN DE LOS DATOS EN **CONOCIMIENTO**

NUEVOS **INSIGHTS** DE ANÁLISIS DE LOS DATOS

CREAR, TRANSFORMAR Y SER CAPAZ DE **MODELAR LOS DATOS**

CONTEXTUALIZACIÓN & ÁREA DE NEGOCIO - CASO DE USO

AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

marcusRB

AI-school
Data Science - Big Data Analytics

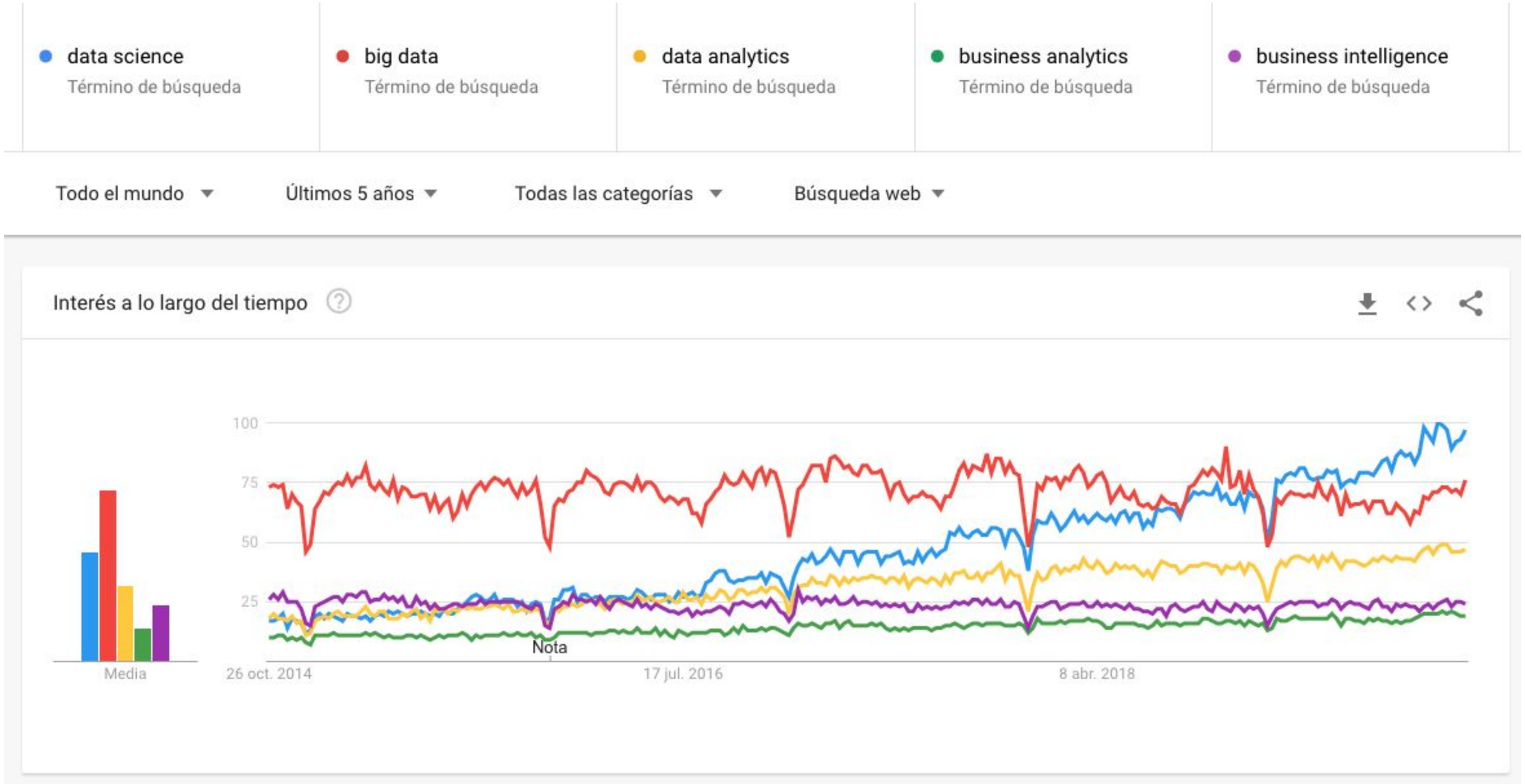
AI-school
Data Science - Big Data Analytics

marcusRB

TENDENCIAS

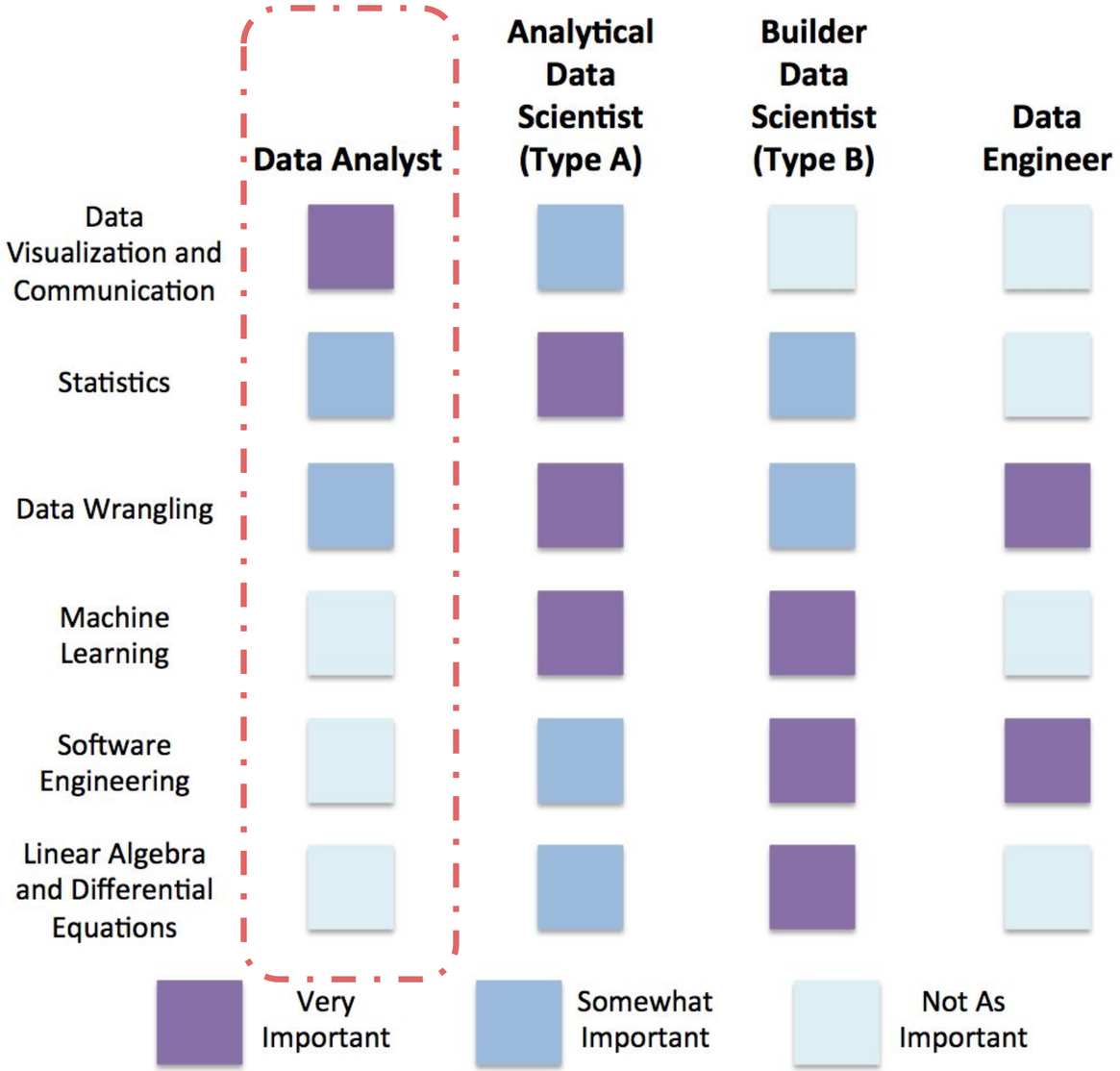
AI-school
Data Science - Big Data Analytics

Tendencias de búsqueda



AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

Diferentes expertises...



AISCHOOL y MARCUSRB is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Based on a work at www.marcusrb.com

AI-school
Data Science - Big Data Analytics

marcusRB

¿Qué es Data Science?

AI-school
Data Science - Big Data Analytics

¿Qué es Data Science?

TRANSFORMAR INFORMACIÓN DE LOS DATOS EN **CONOCIMIENTO**

NUEVOS **INSIGHTS** DE ANÁLISIS DE LOS DATOS

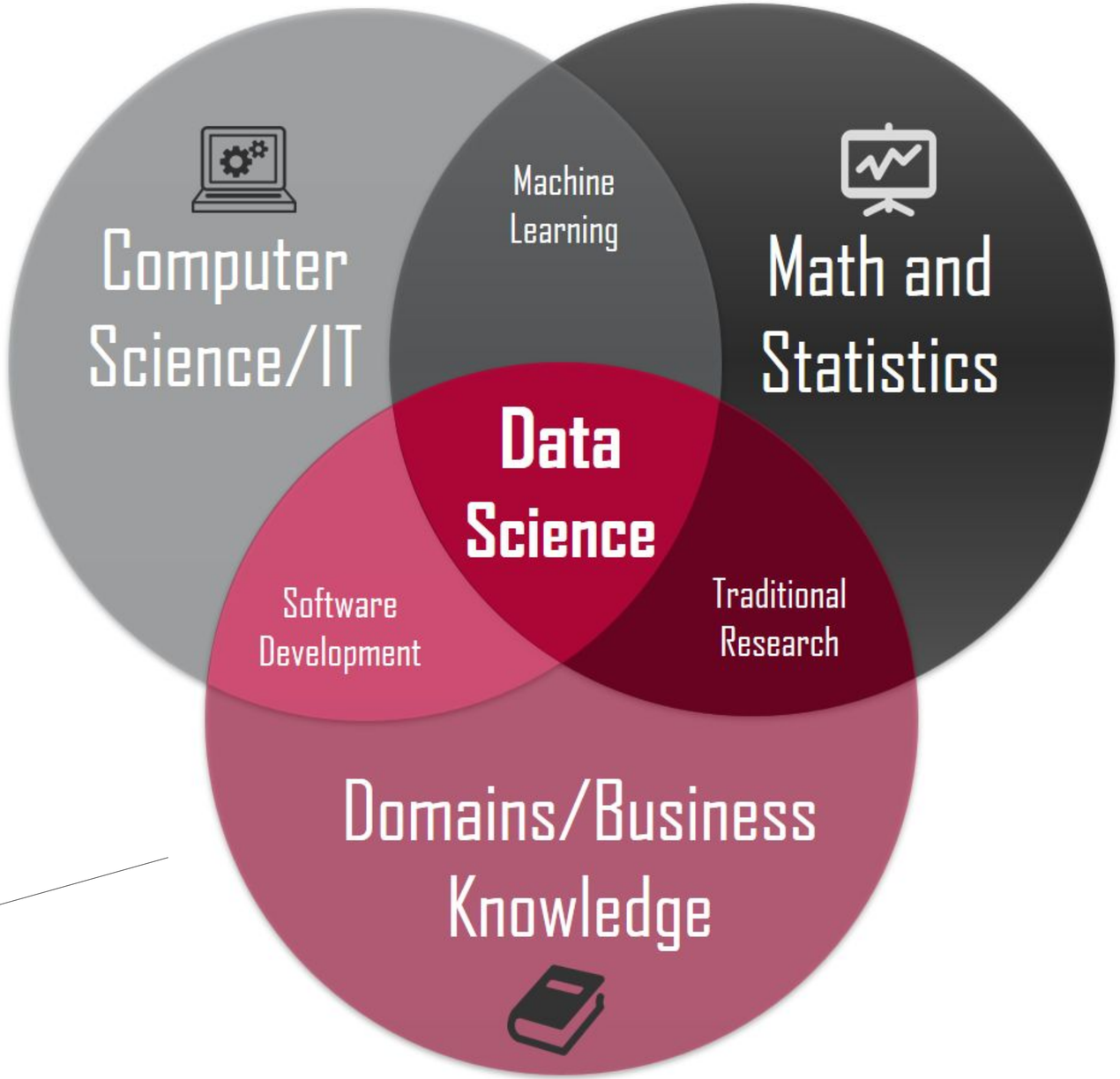
IDENTIFICAR **TENDENCIAS, PATRONES** Y **CORRELACIONES**

CONTEXTUALIZACIÓN

¿Qué es Data Science?

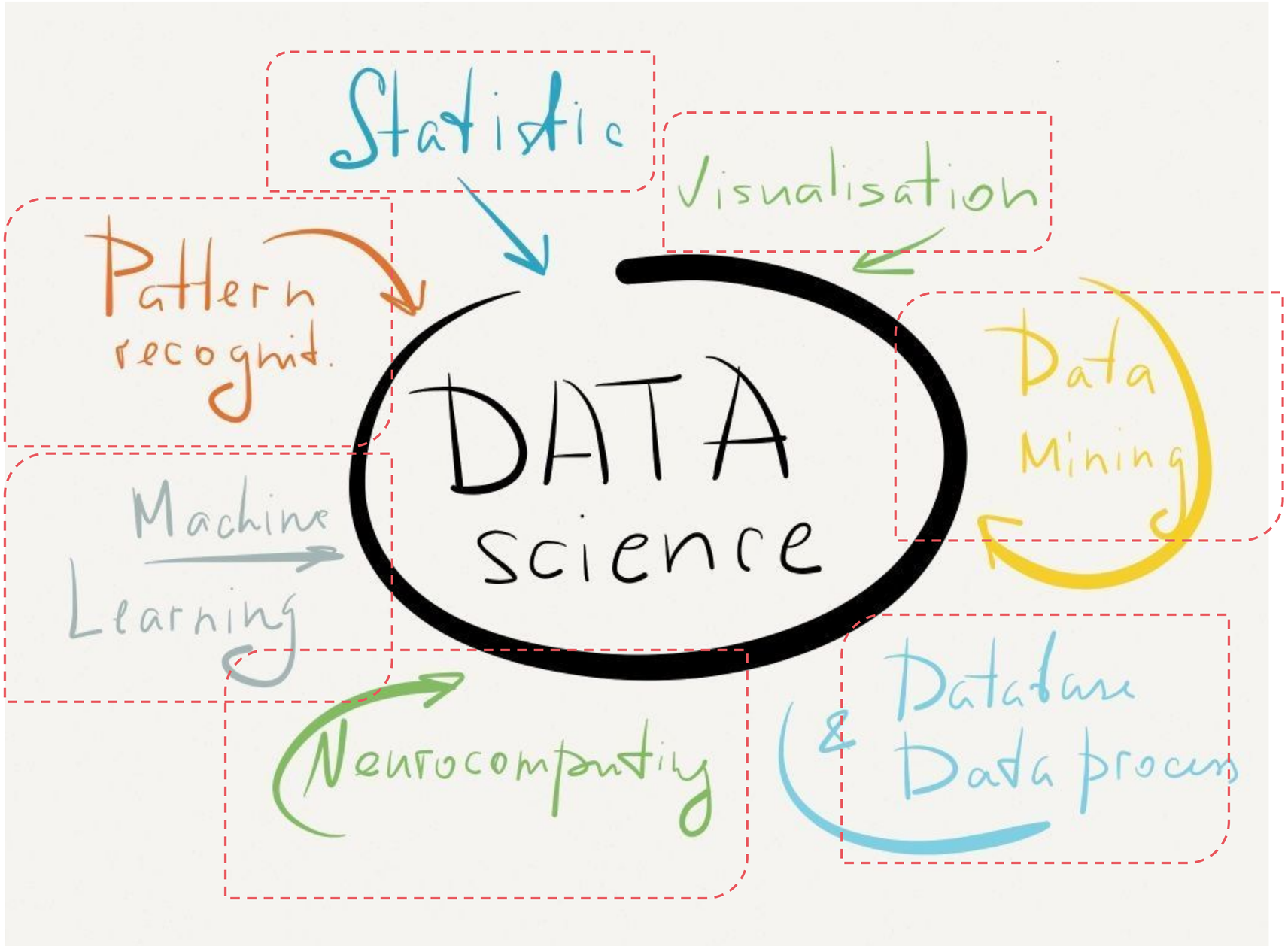
Pensamiento Computacional

Pensamiento analítico / Algoritmos



Business / Comunicación

¿Qué es Data Science?

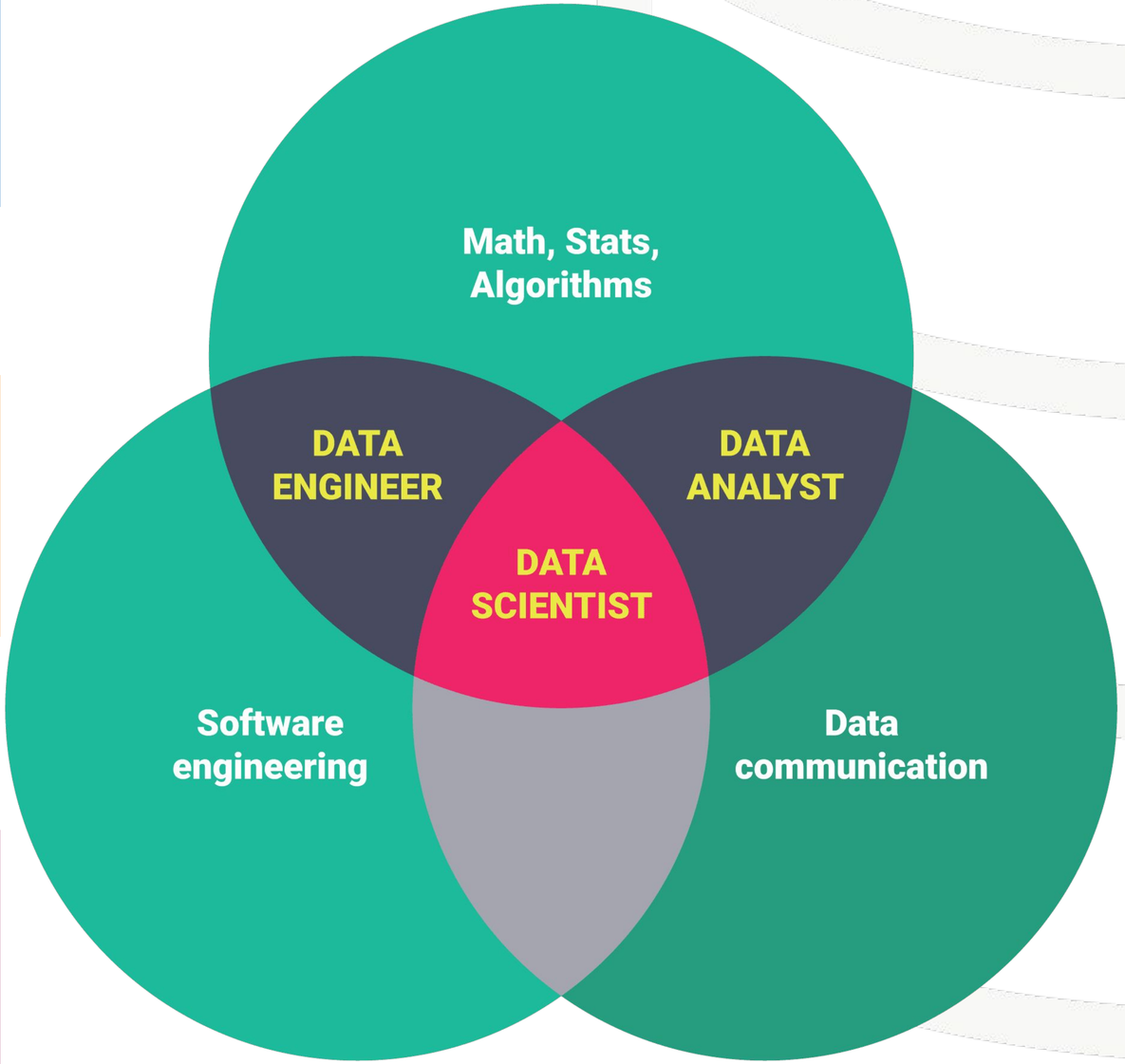


Diferentes Roles...

Creativity

Programming

Real-life Practice

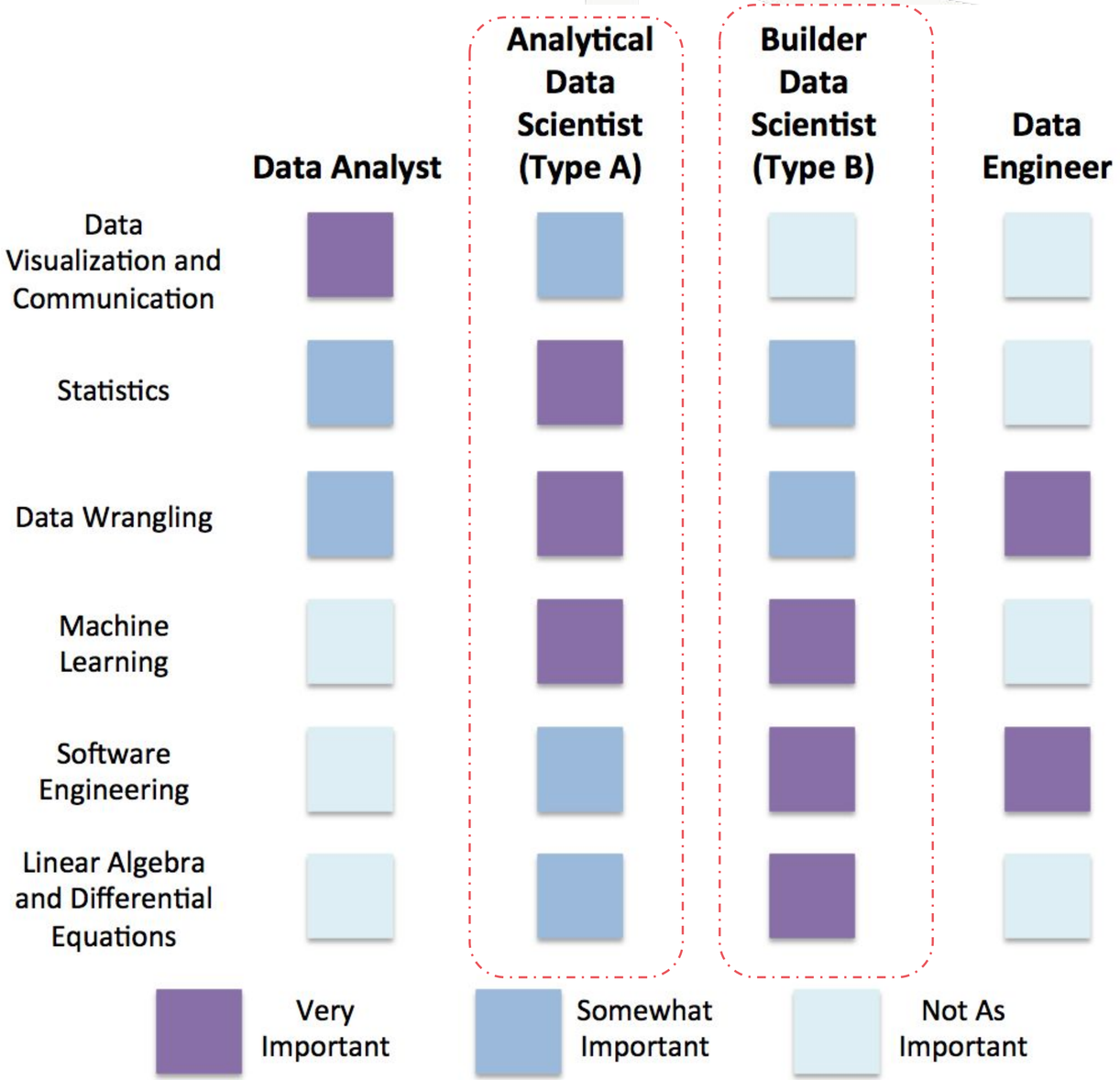


Communication

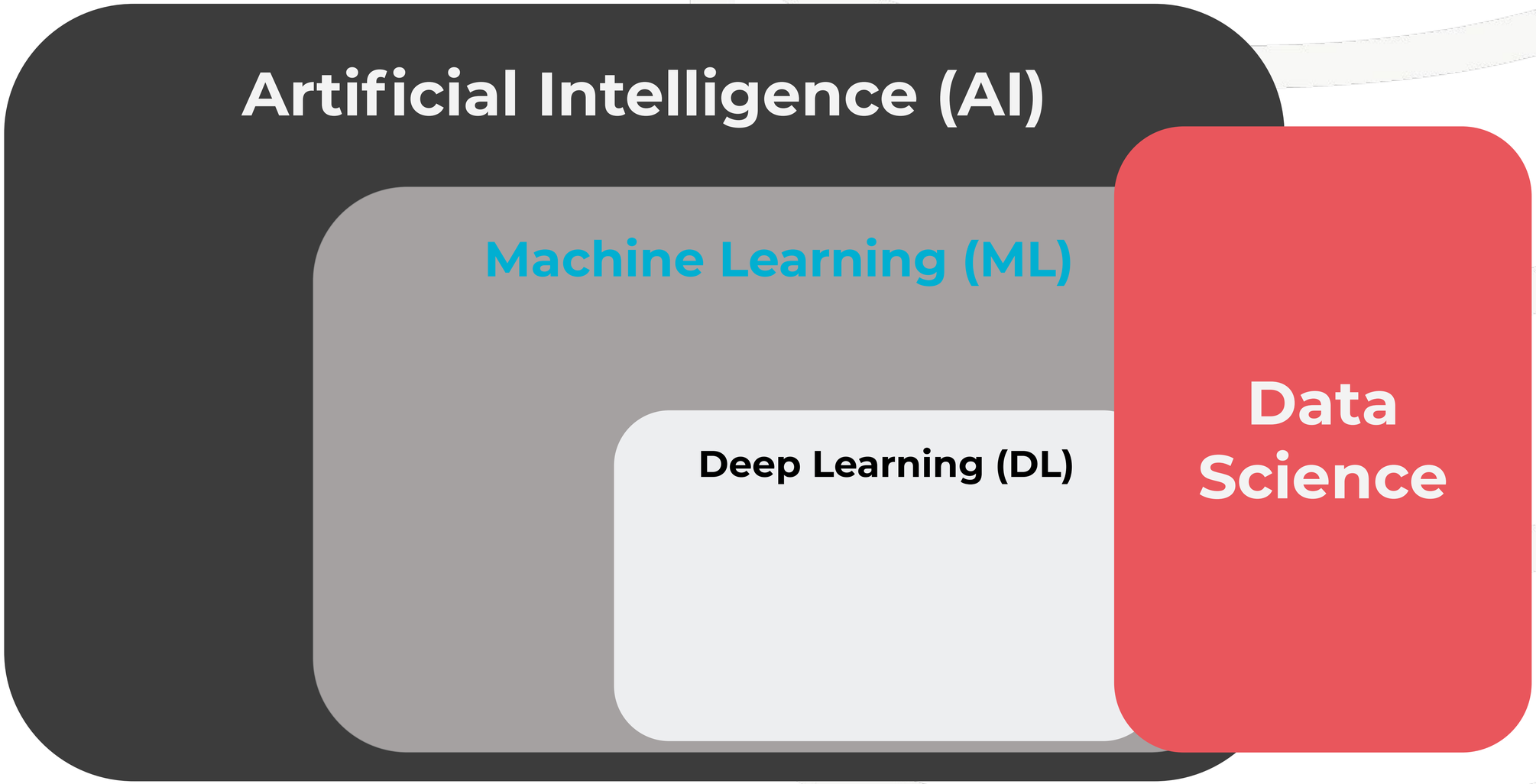
Presentation

Domain Knowledge

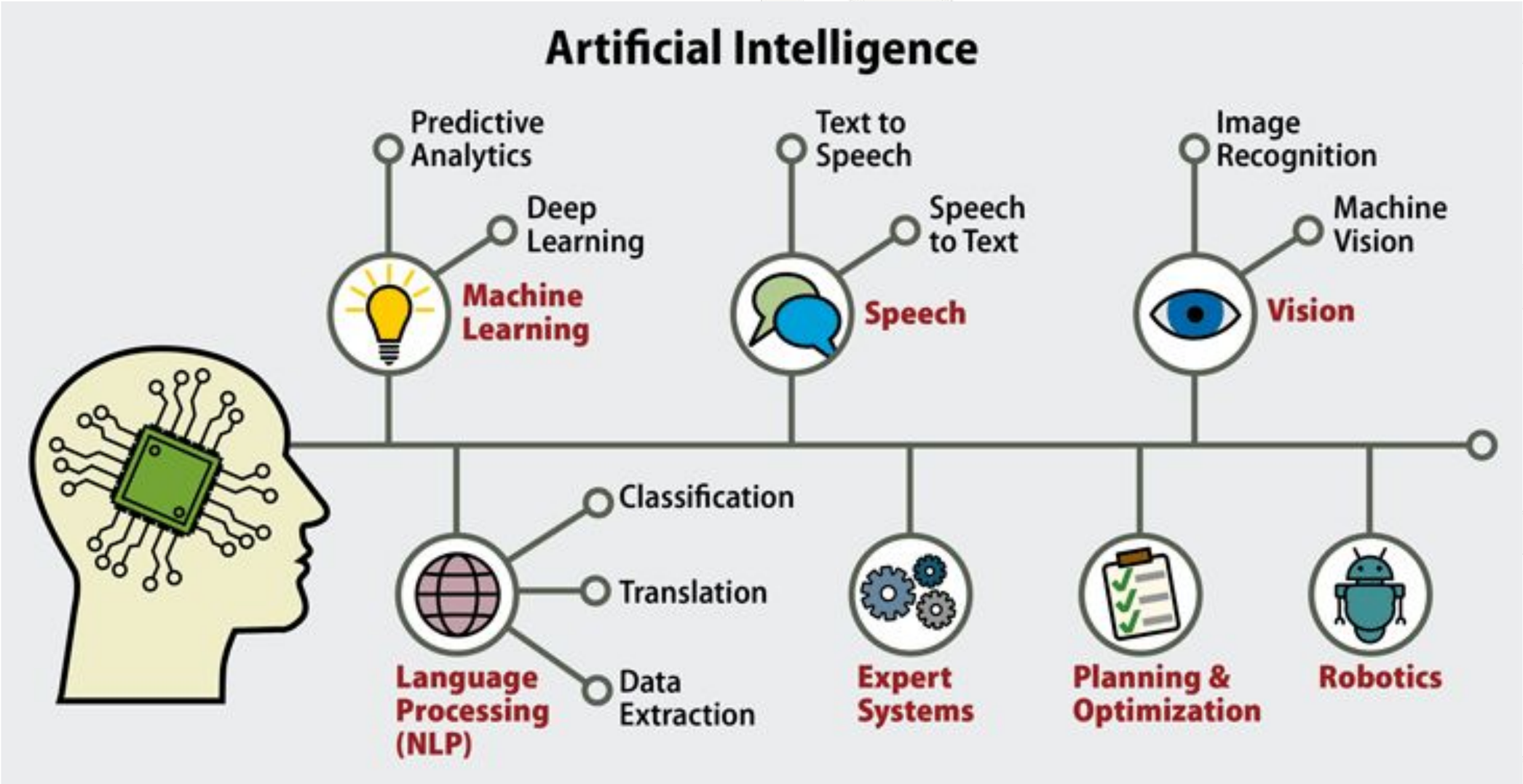
... diferentes Expertises



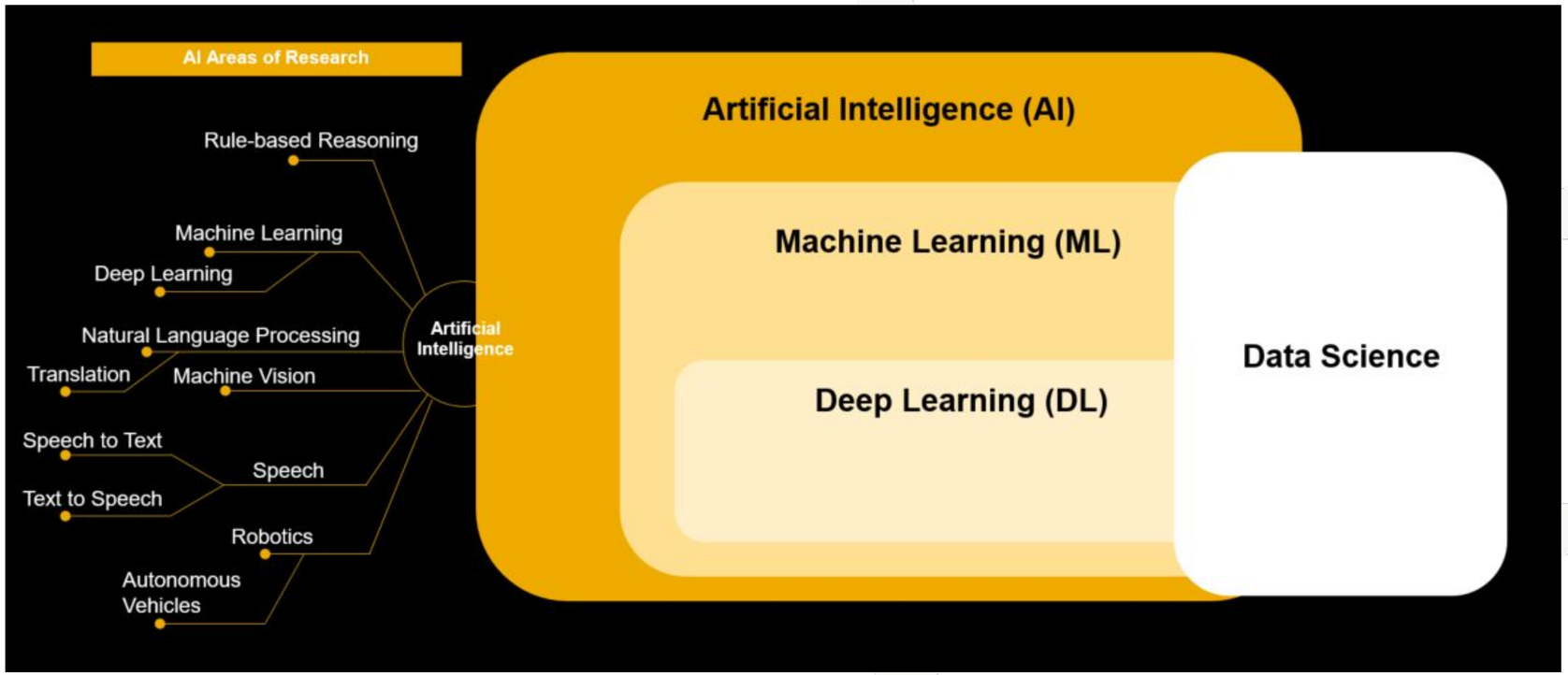
Áreas del Data Science



Áreas del Data Science



Áreas del Data Science



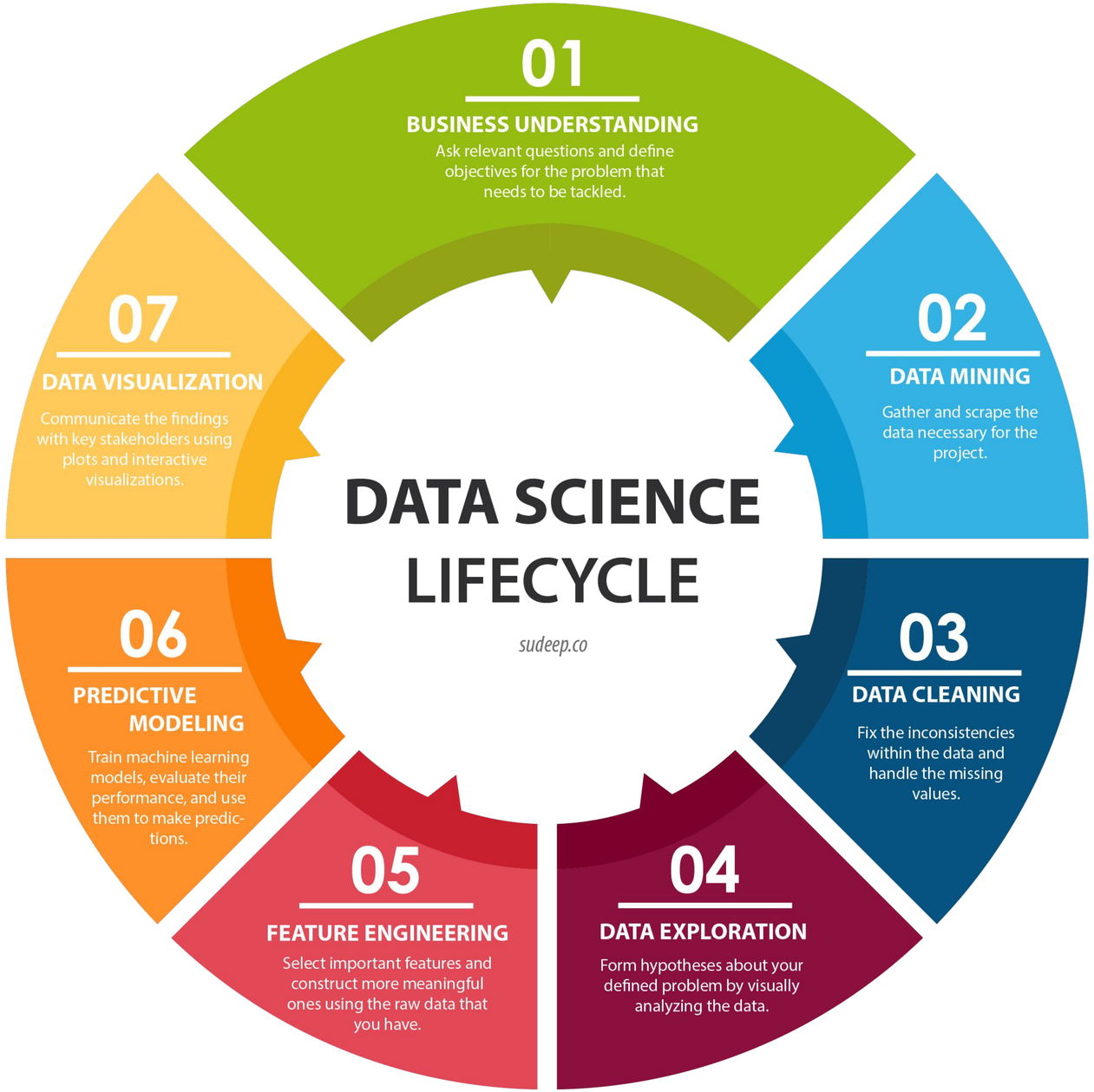
AI-school
Data Science - Big Data Analytics

marcusRB

Ciclo de vida de un proyecto

AI-school
Data Science - Big Data Analytics

Ciclo de vida del Data Scientist

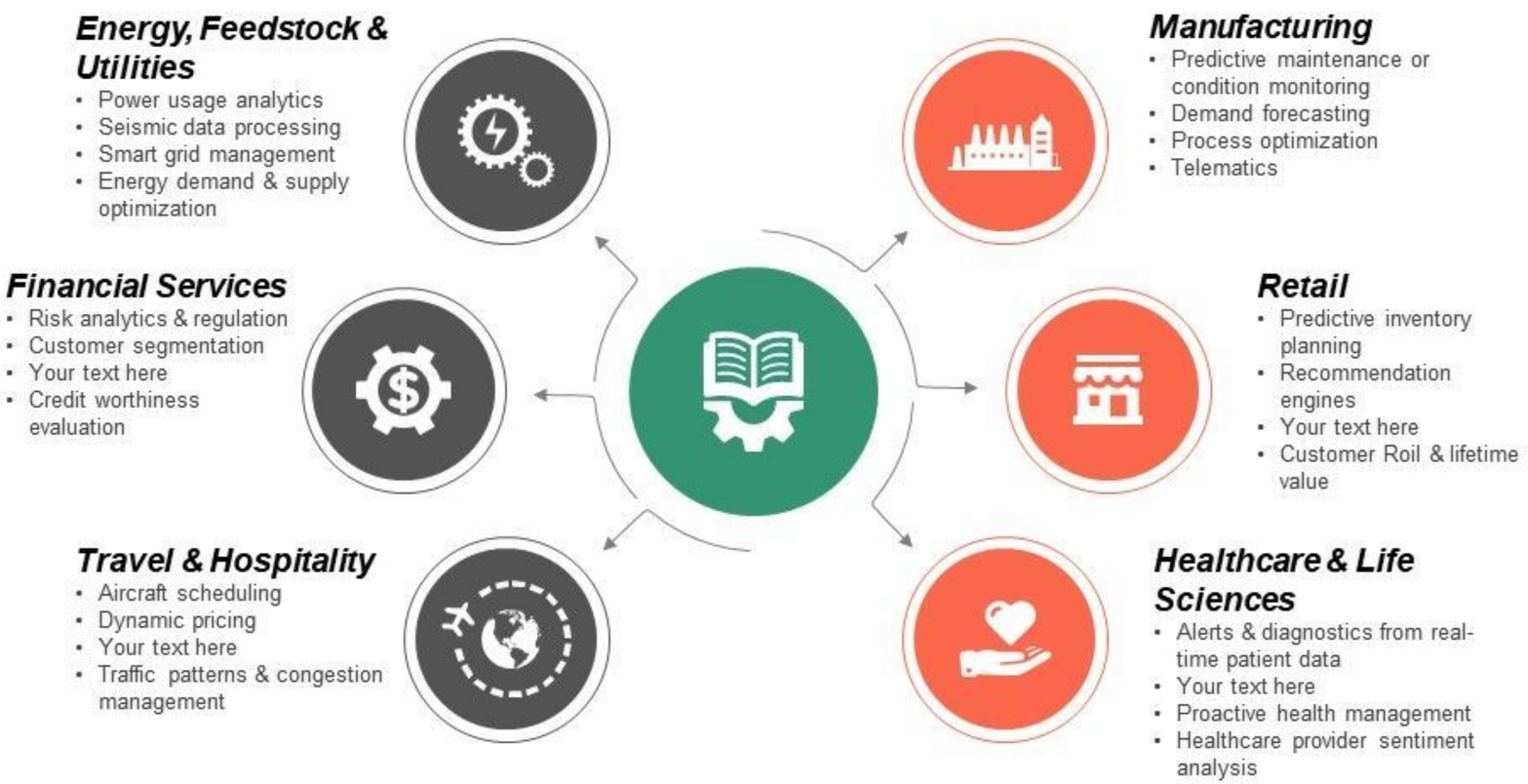


marcusRB

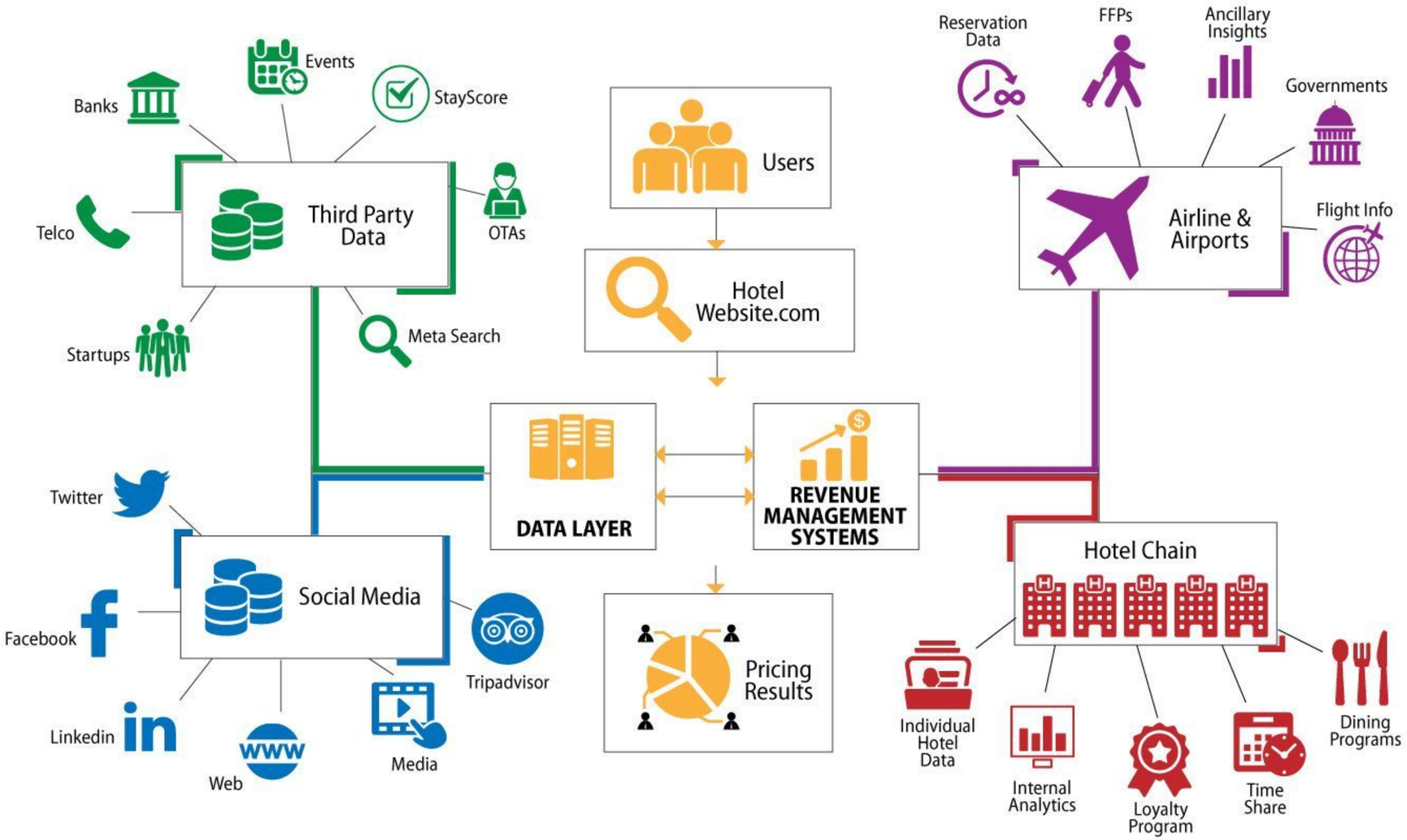
AI-school
Data Science - Big Data Analytics

Casos de uso

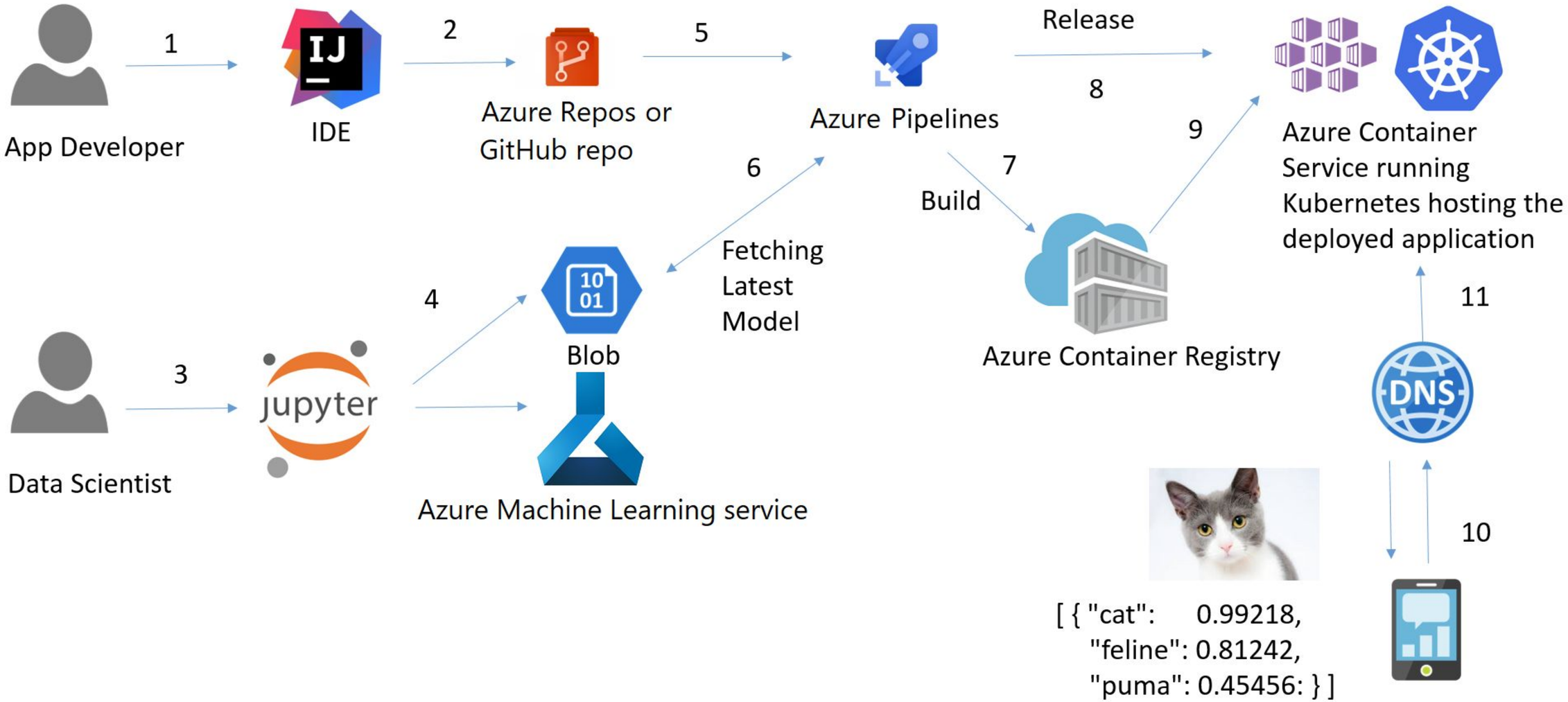
Machine Learning Use Cases



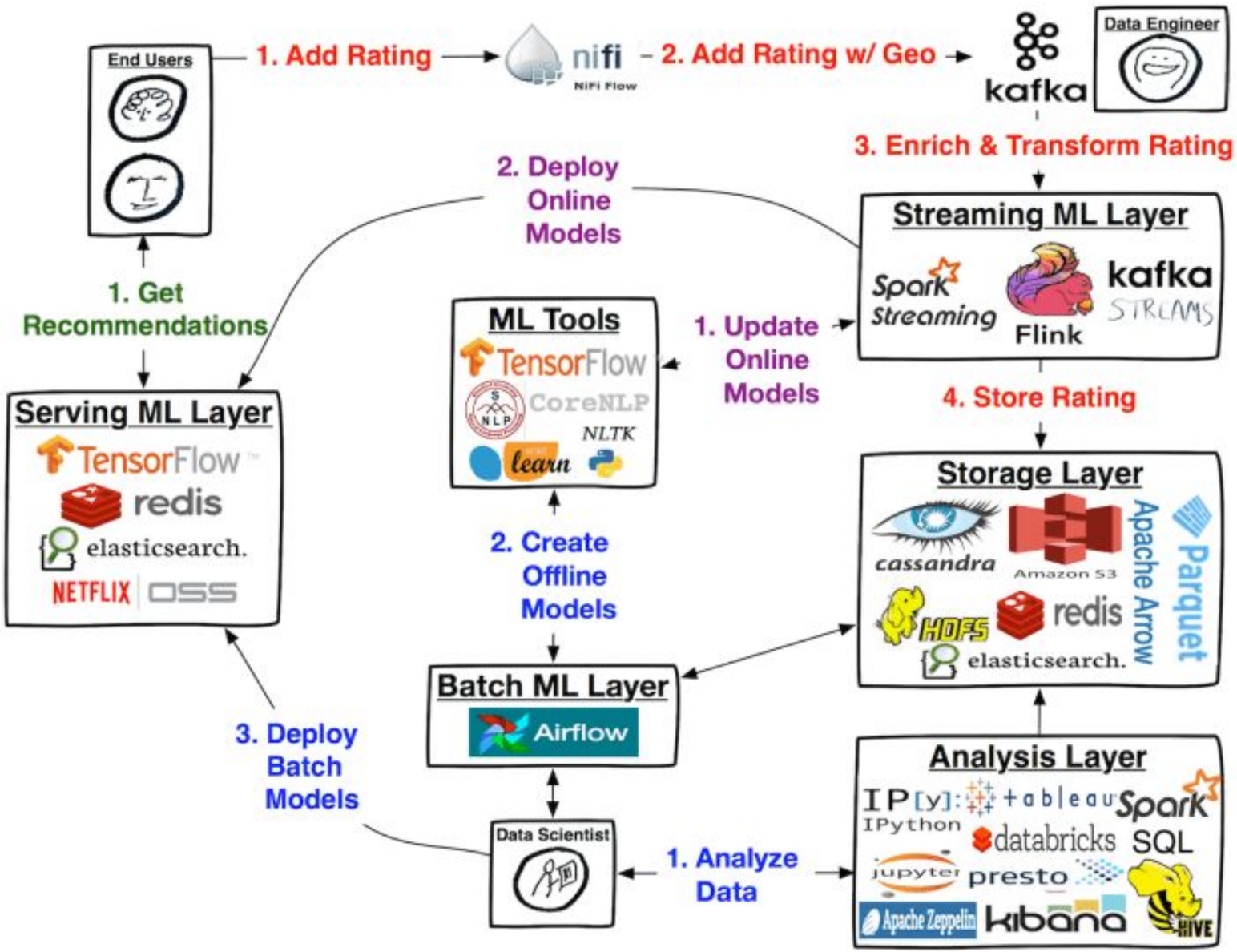
Casos de uso


Banks
Events
StayScore
Telco
Third Party Data
OTAs
Startups
Meta Search
Users
Hotel Website.com
Reservation Data
FFPs
Ancillary Insights
Governments
Airline & Airports
Flight Info
DATA LAYER
REVENUE MANAGEMENT SYSTEMS
Twitter
Social Media
Facebook
Tripadvisor
Linkedin
Web
Media
Pricing Results
Hotel Chain
Individual Hotel Data
Internal Analytics
Loyalty Program
Time Share
Dining Programs
© Copyright 2016 markrs.co

Ejemplo de un proyecto en Data Science



Ejemplo de un proyecto en Data Science



AI-school
Data Science - Big Data Analytics

marcusRB

Herramientas

AI-school
Data Science - Big Data Analytics

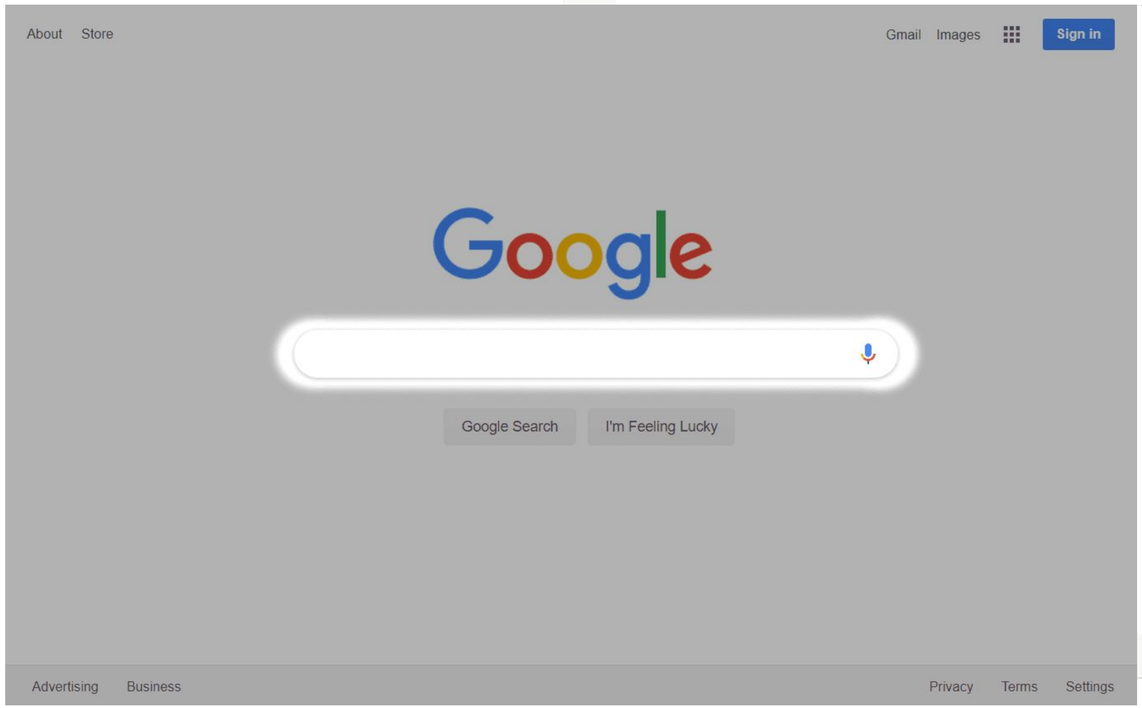
Tools utilizadas

Recursos

marcusRB

Data Science Central

stackoverflow

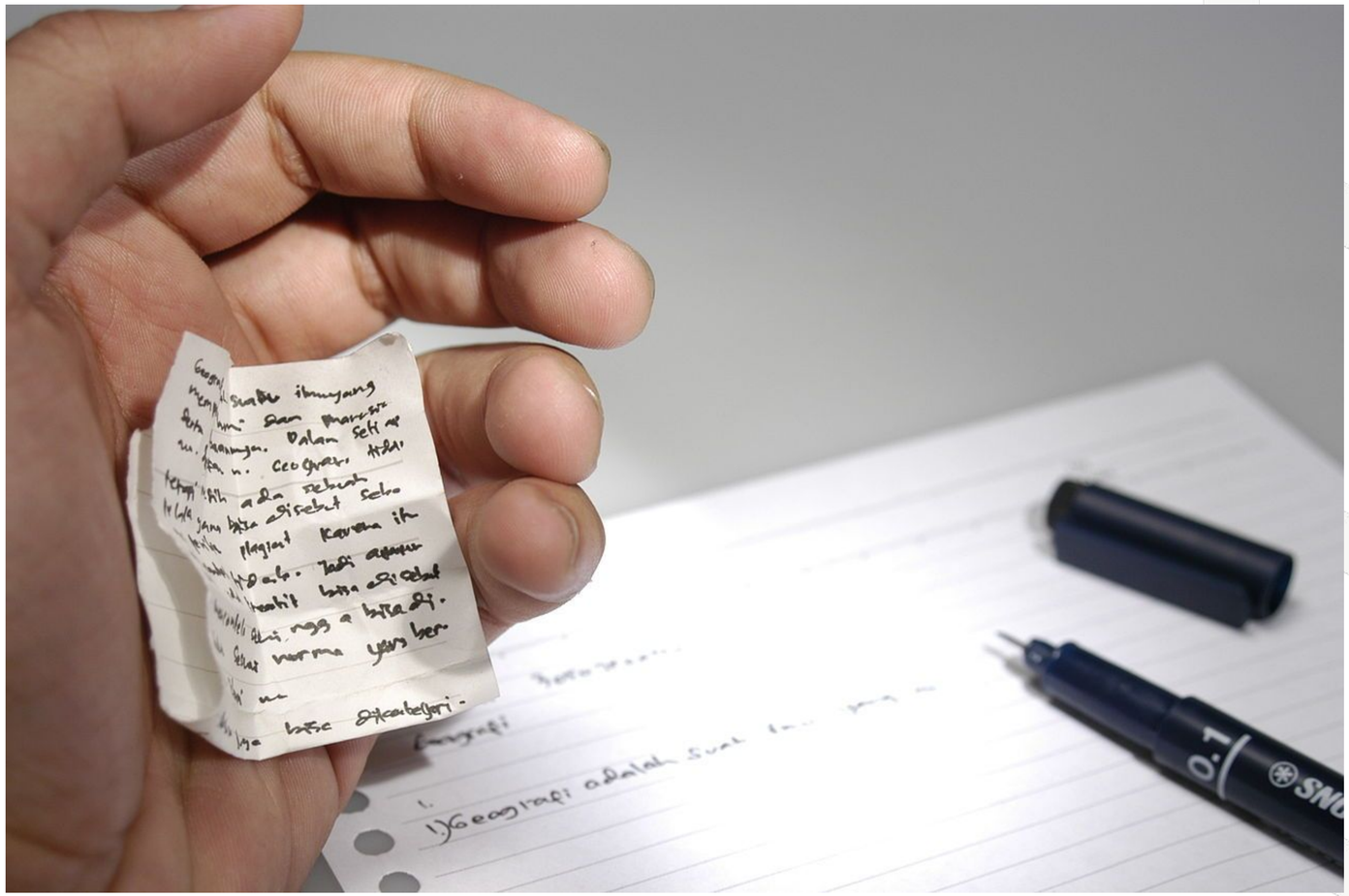


Towards Data Science

Medium

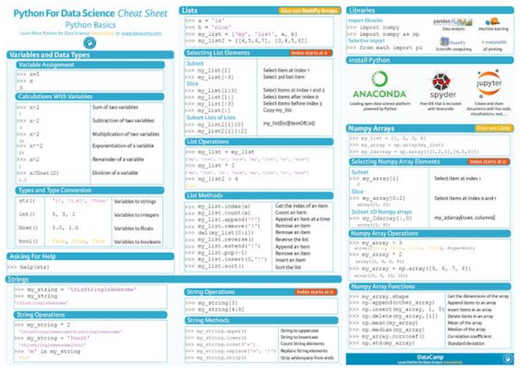
AI-school

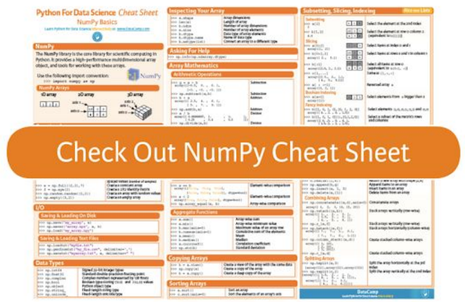
CHEAT SHEETS



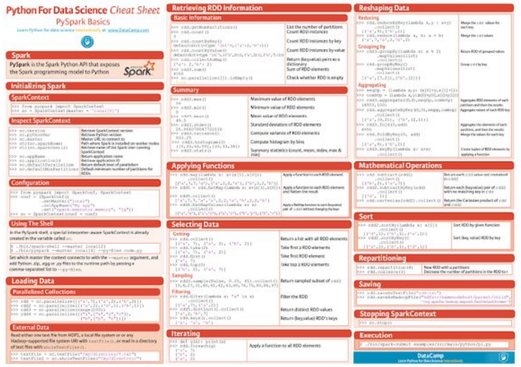
CHEAT SHEETS

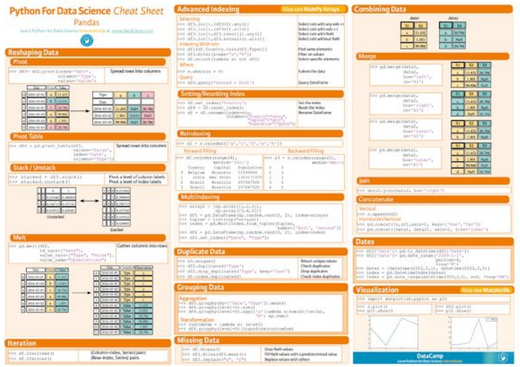
marcusRB

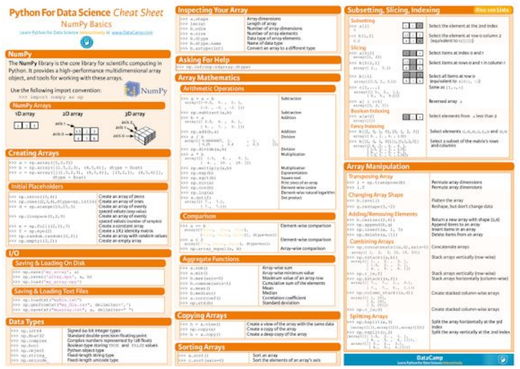


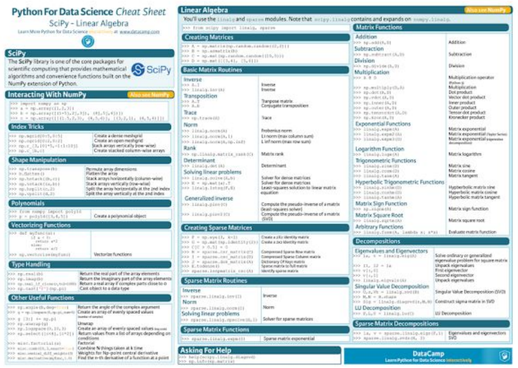


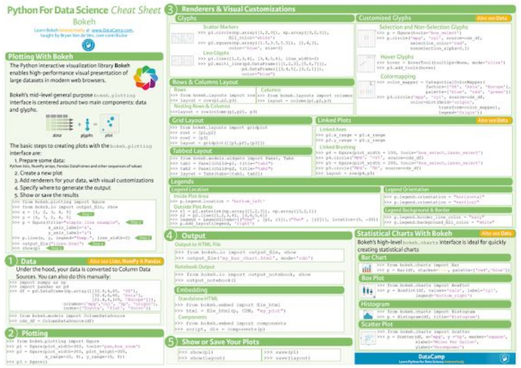
Check Out NumPy Cheat Sheet

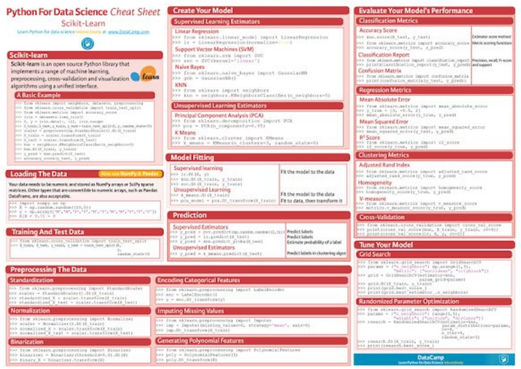


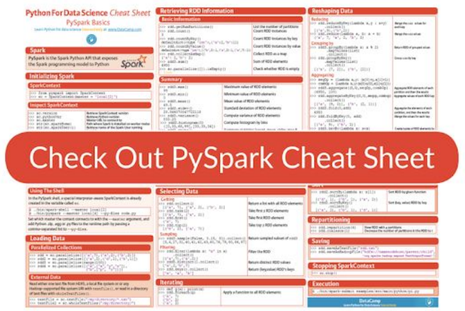




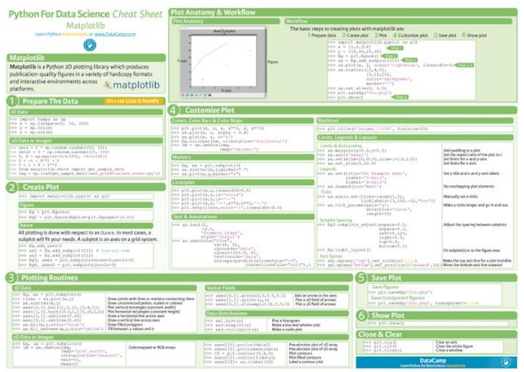


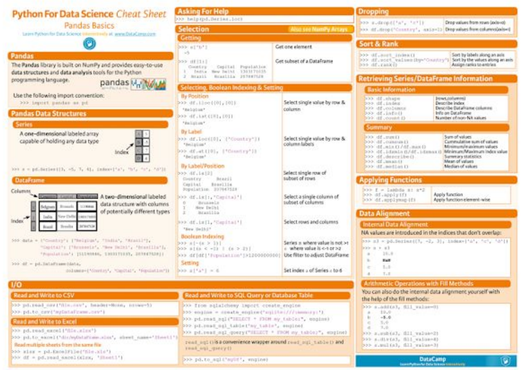


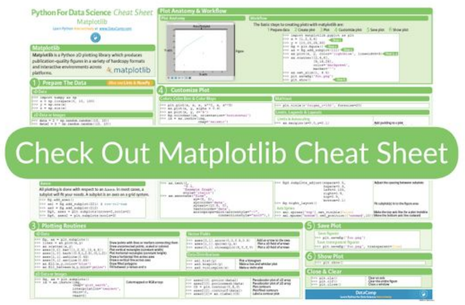




Check Out PySpark Cheat Sheet



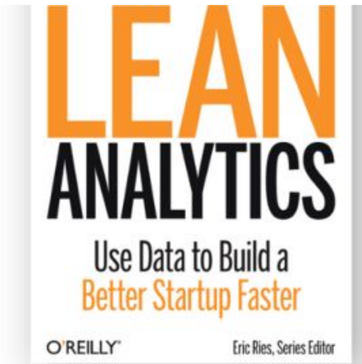


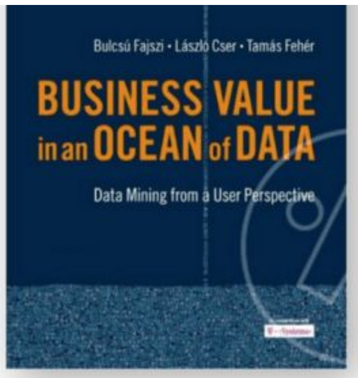


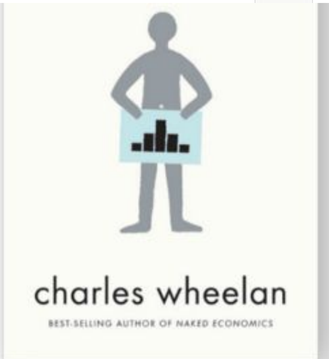
Check Out Matplotlib Cheat Sheet

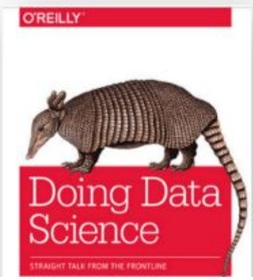
AI-school
Data Science - Big Data Analytics

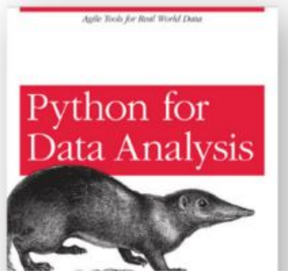
Bibliografía

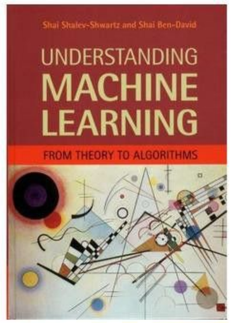


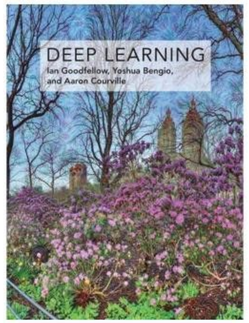


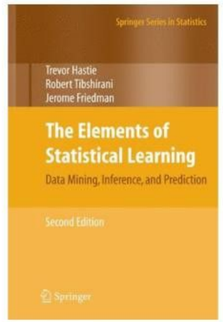


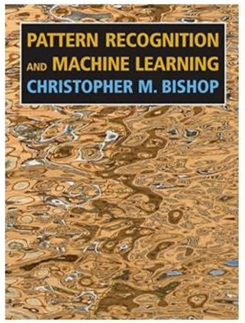


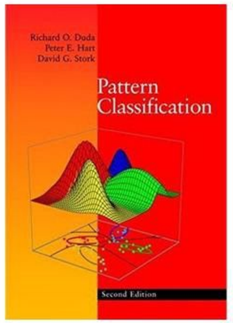


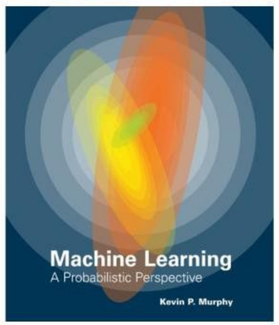


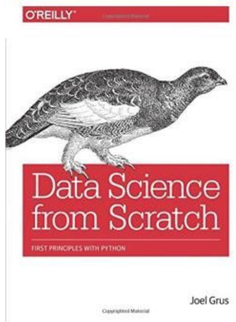


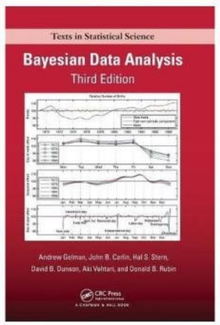












AI-school
Data Science - Big Data Analytics

marcusRB

¿Dudas - Preguntas?

AI-school
Data Science - Big Data Analytics