

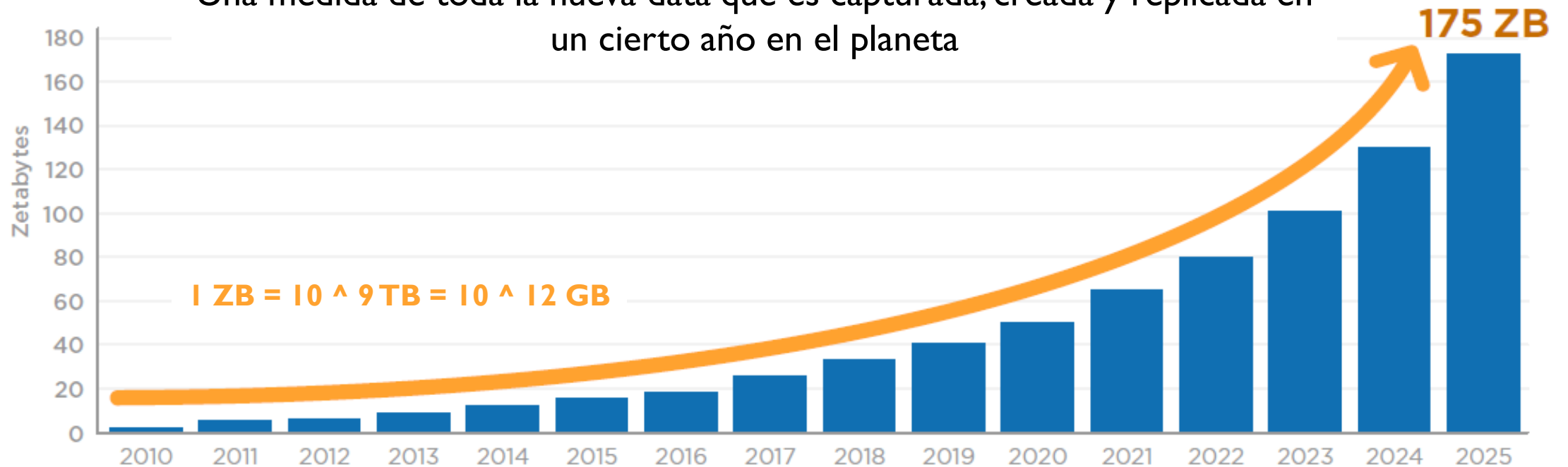
# ¿QUÉ ROL JUEGAN LOS DATOS E INFORMACIÓN EN LA ACTUALIDAD?



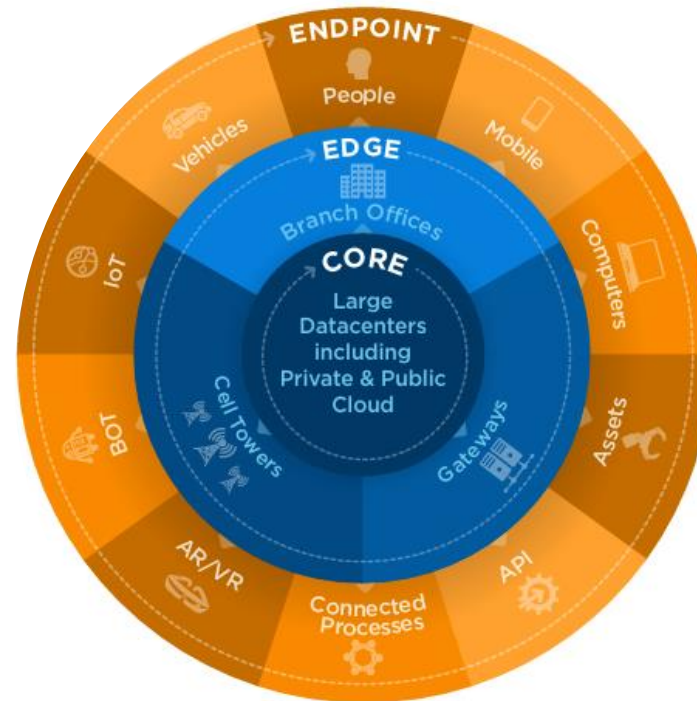
# DATA AGE: LA DIGITACIÓN DEL MUNDO

## Datósfera Global

Una medida de toda la nueva data que es capturada, creada y replicada en un cierto año en el planeta



# DATA AGE: LA DIGITACIÓN DEL MUNDO



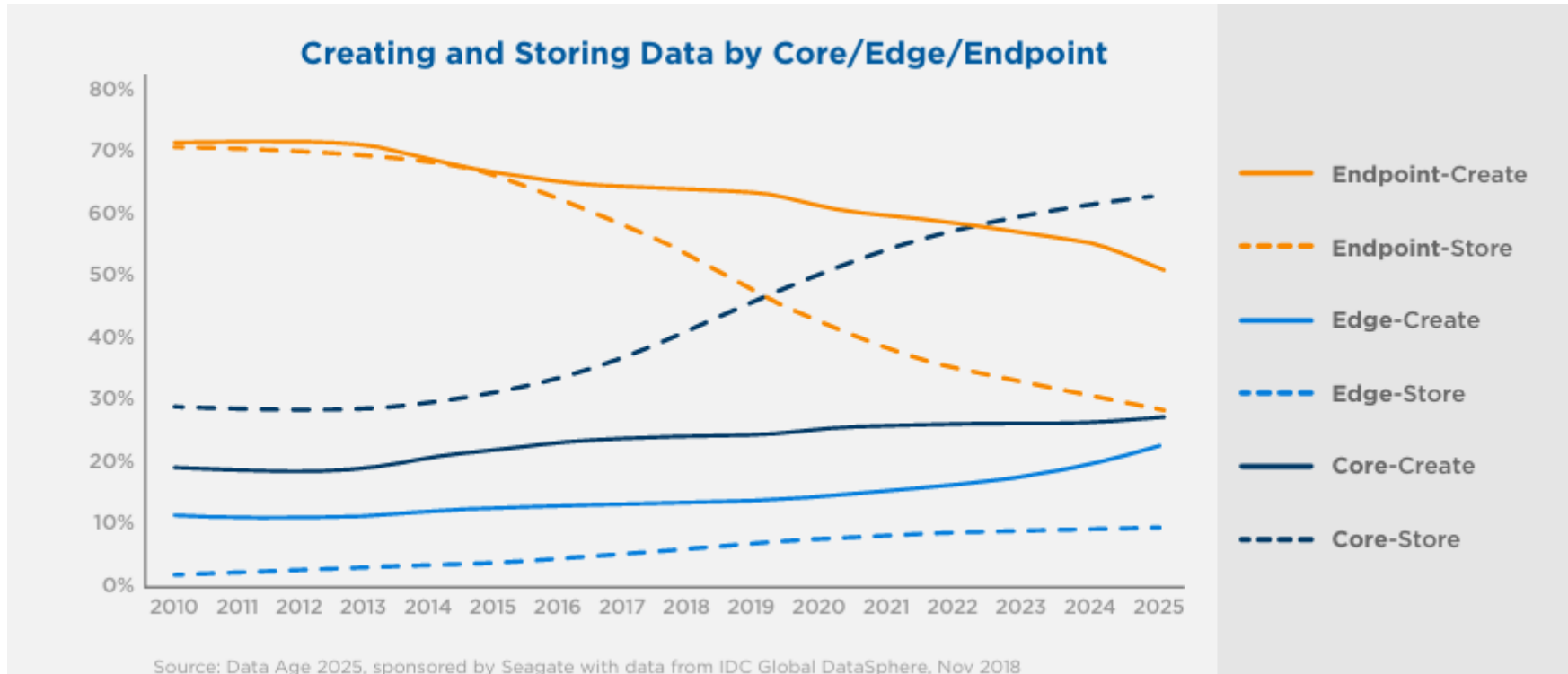
~ IDC's Data Age 2025 study ~

Mayor volumen, variedad, velocidad y acceso a datos

Fuente: The digitization of the world from Edge to Core, Disponible en:

<https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

# DATA AGE: LA DIGITACIÓN DEL MUNDO



Mayor volumen, variedad, velocidad y acceso a datos

Fuente: The digitization of the world from Edge to Core, Disponible en:

<https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

# DEFINICIONES

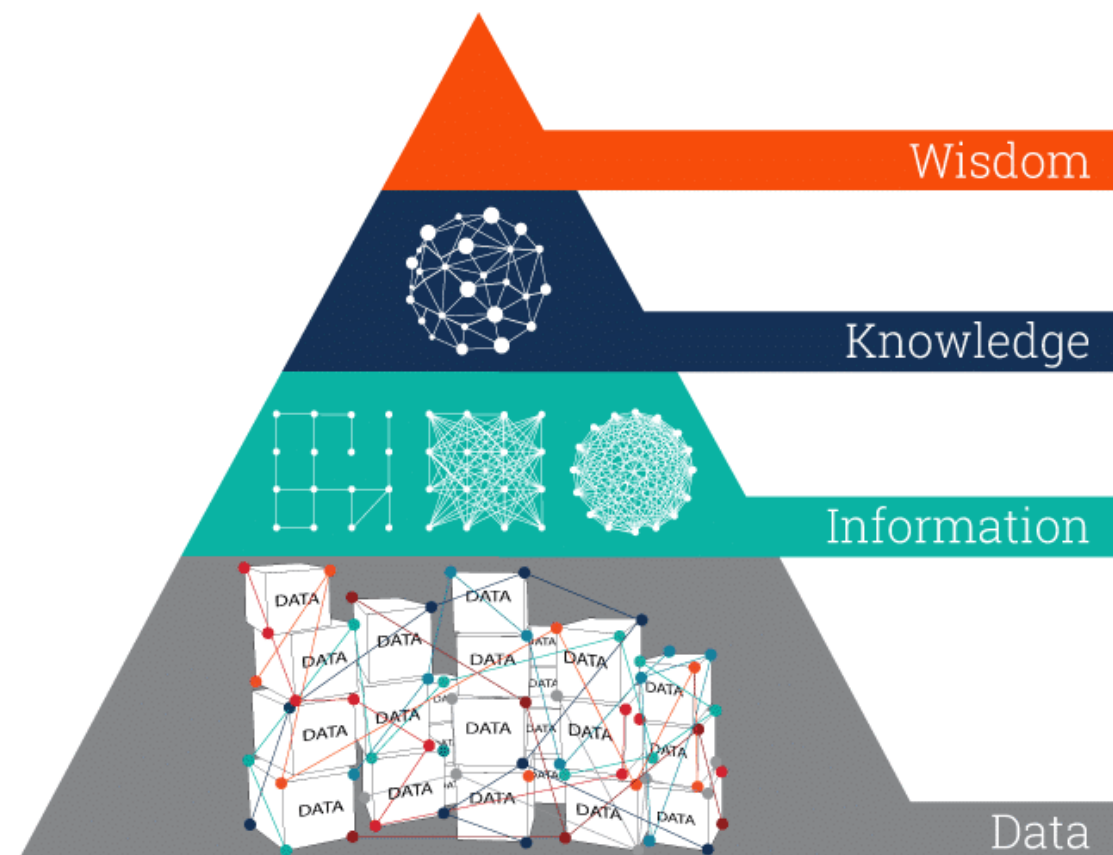
## DATOS VS INFORMACIÓN

**Datos** --> Abstracciones o mediciones obtenidas del mundo real

**Información** --> Datos que han sido procesados, estructurados o contextualizados de manera que son significativos para las personas.

**Conocimiento** --> Información que ha sido interpretada y comprendida por una persona, de manera que puede actuar a partir de ella.

**Sabiduría** --> Actuar de forma apropiada a partir del conocimiento



# Ciencia y Ciencia De Datos

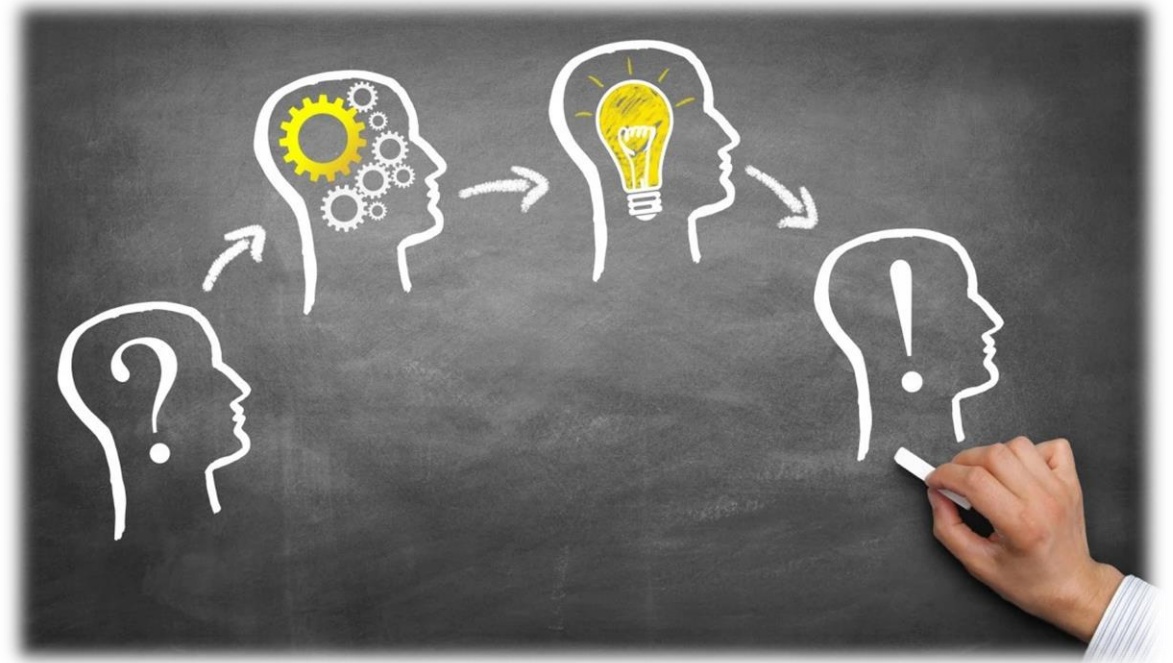
## MÉTODO CIENTÍFICO

Estudio sistemático de la estructura y comportamiento del mundo físico y natural, mediante la **observación y experimentación**.

Se basa en la construcción, testeo y comprobación de una **hipótesis**.

Busca identificar relaciones **causales**.

Fuente: <https://www.expil.com/t/scientific-method-steps-importance-11026>





# Ciencia de Datos: Definiciones

Es la ciencia de extraer **información significativa** de los datos.

Procesos y técnicas involucrados en la transformación de estos recursos (datos) en **información y conocimiento**.

Es el proceso de **formular una pregunta cuantitativa**, que puede ser respondida con datos,

- Recolectando y limpiando los datos
- Analizando los datos, y
- Comunicando la respuesta a la pregunta a una audiencia relevante.

Foco está en **correlaciones y patrones**, más que en causalidad.

Esto es un cambio de paradigma, respecto al método científico clásico.



# Ciencia de Datos e Inteligencia Artificial

---



ALGORITMOS

**ALGORITMOS**

**Conjunto de instrucciones  
automatizadas para llevar a cabo una  
tarea**



# Ciencia de Datos e Inteligencia Artificial



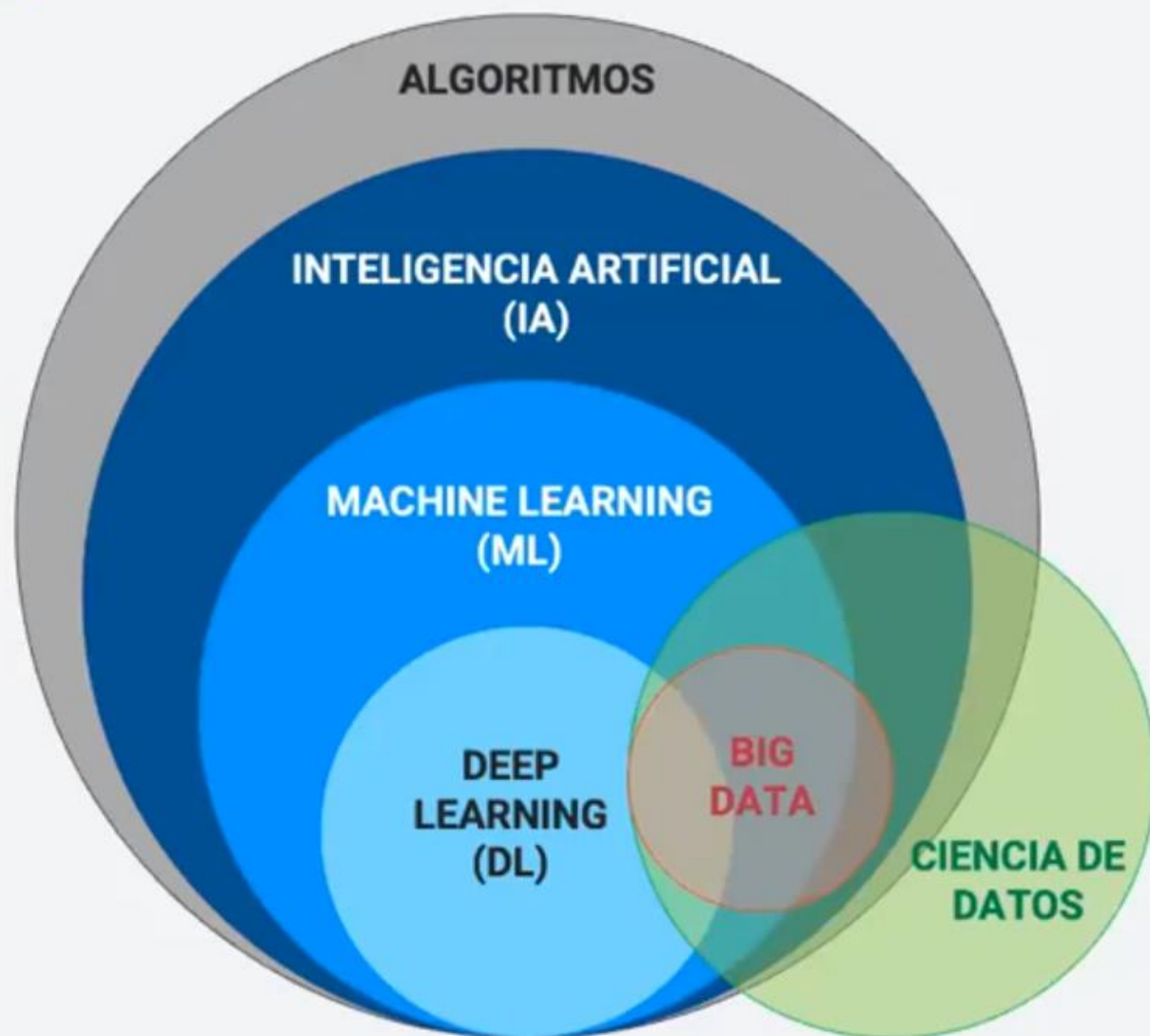
## ALGORITMOS

Conjunto de instrucciones automatizadas para llevar a cabo una tarea

## INTELIGENCIA ARTIFICIAL (IA)

Programas con la capacidad de imitar el comportamiento humano (sentir, razonar, actuar, adaptar)

# Ciencia de Datos e Inteligencia Artificial



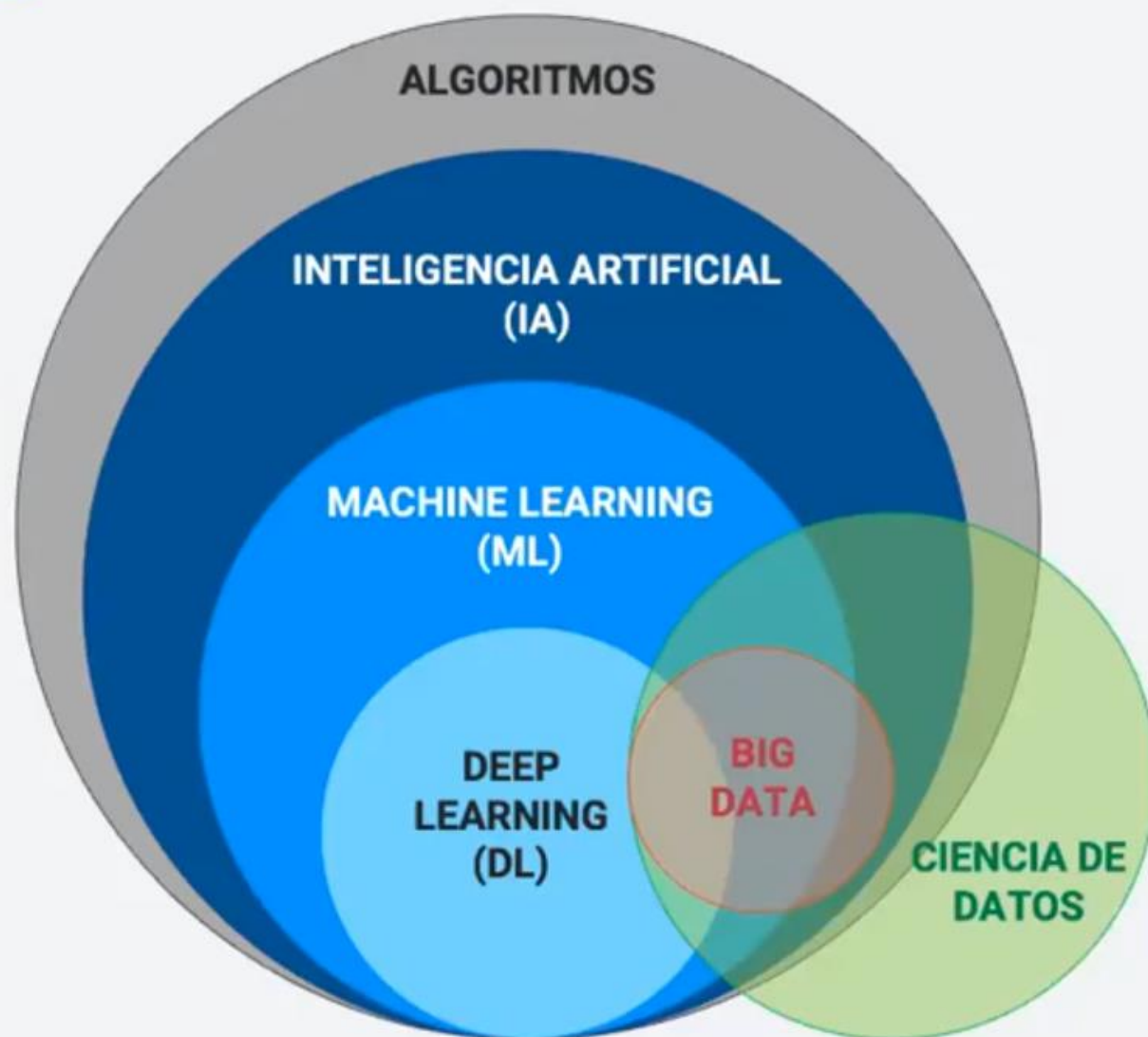
## MACHINE LEARNING (ML)

Algoritmos con la habilidad de aprender (mejorar) en base a la experiencia (datos)

## DEEP LEARNING (DL)

Subconjunto de ML en el que redes neuronales se adaptan y aprenden de grandes conjuntos de datos

# Ciencia de Datos e Inteligencia Artificial



## BIG DATA

Es una **definición relativa**.

“es Big Data cuando **no puede ser contenida en una sola unidad de cómputo**”

(David Cranshaw, Google)

Algunas propiedades fundamentales (VVV):

- Gran volumen
- Gran velocidad (ej: tiempo real)
- Gran variedad: estructurada y no estructurada
- Veracidad, valor....



# ¿Cómo se desarrolla la Ciencia de Datos?



# THE DATA SCIENCE PROCESS

## THE DATA SCIENCE PROCESS



Data Engineers

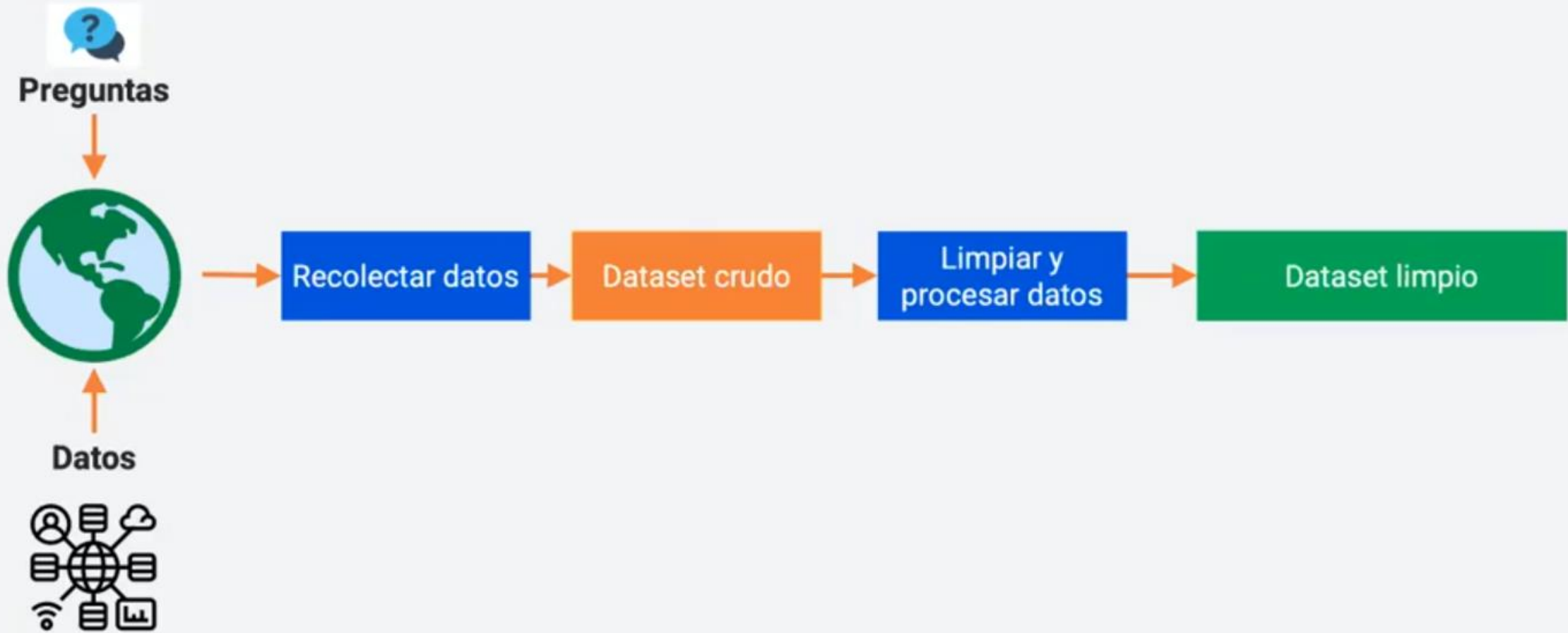
Data Analysts

Machine Learning Engineers

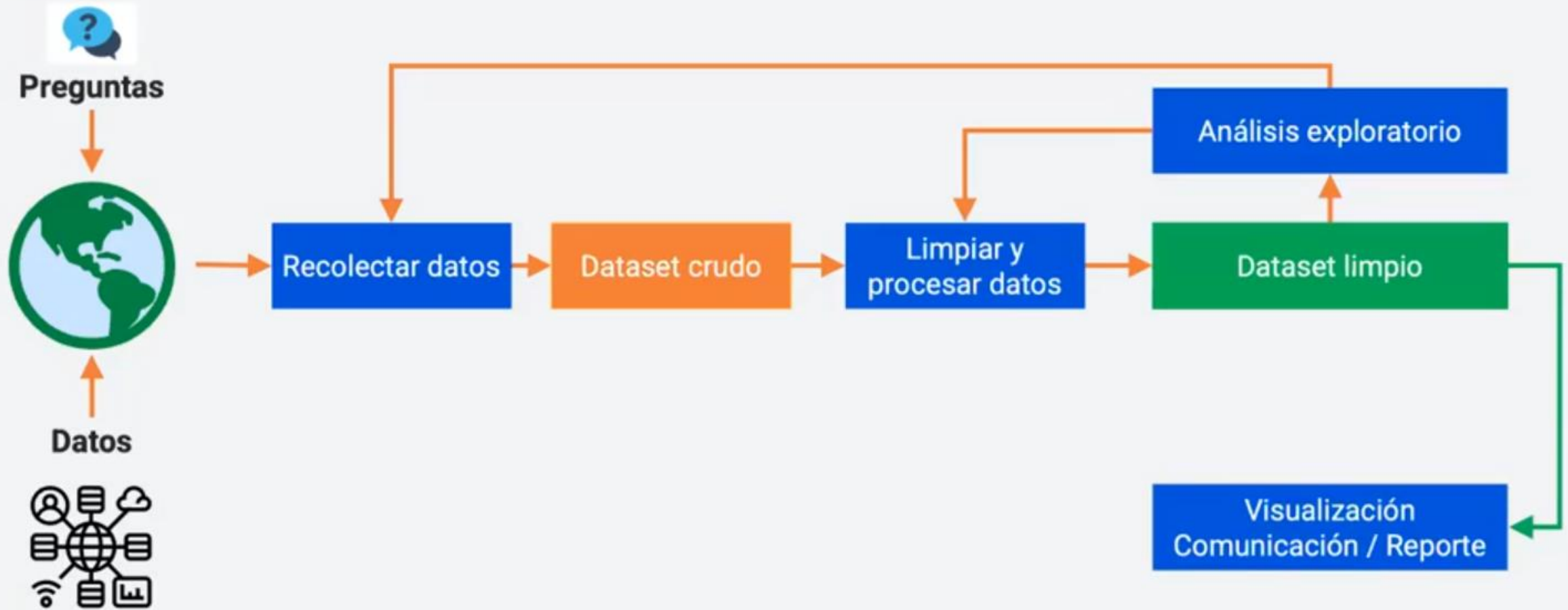
Data Scientists



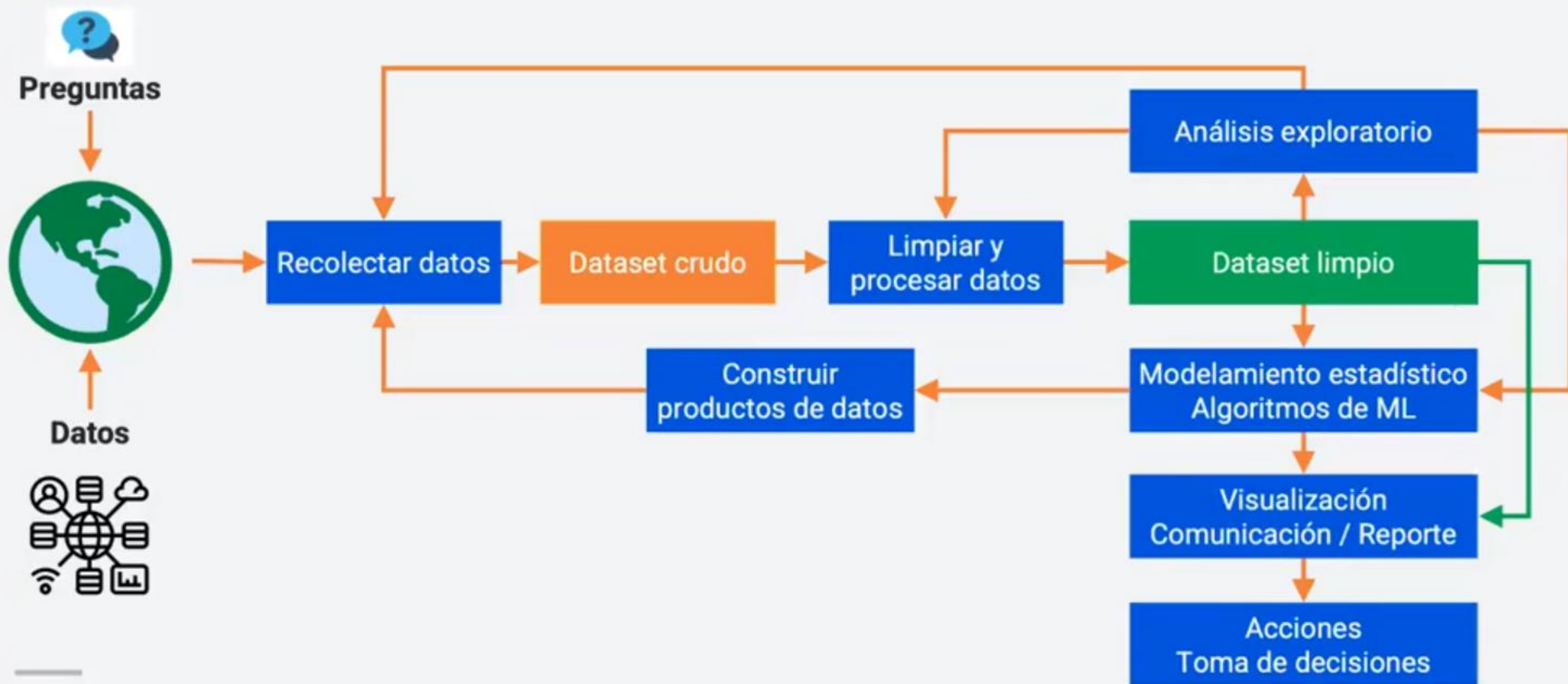
# Proceso de Ciencia de Datos



# Proceso de Ciencia de Datos



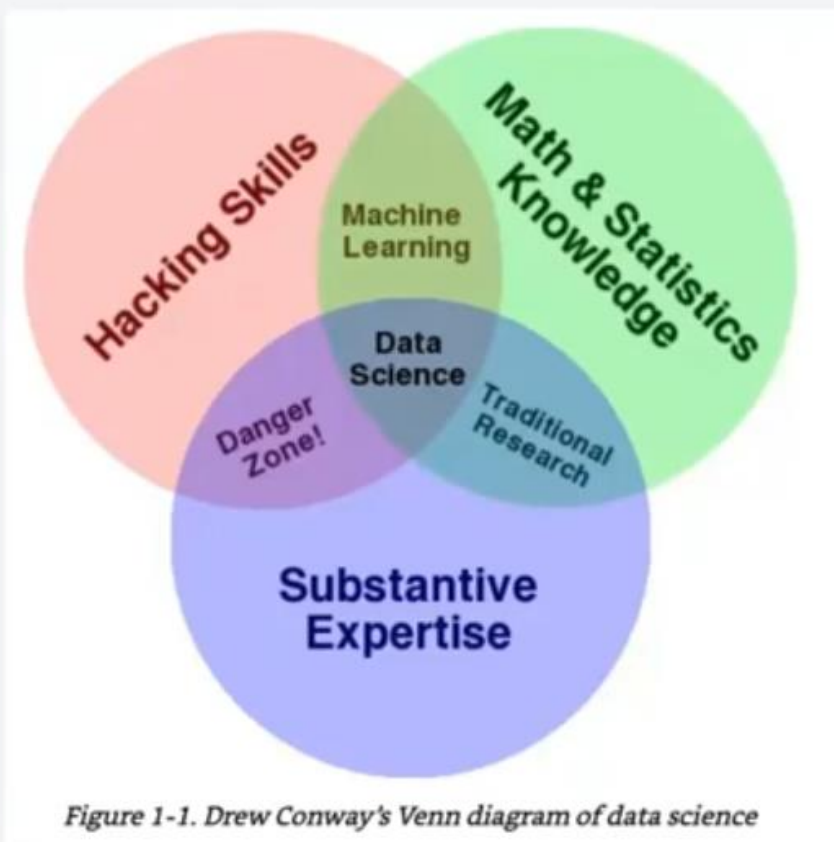
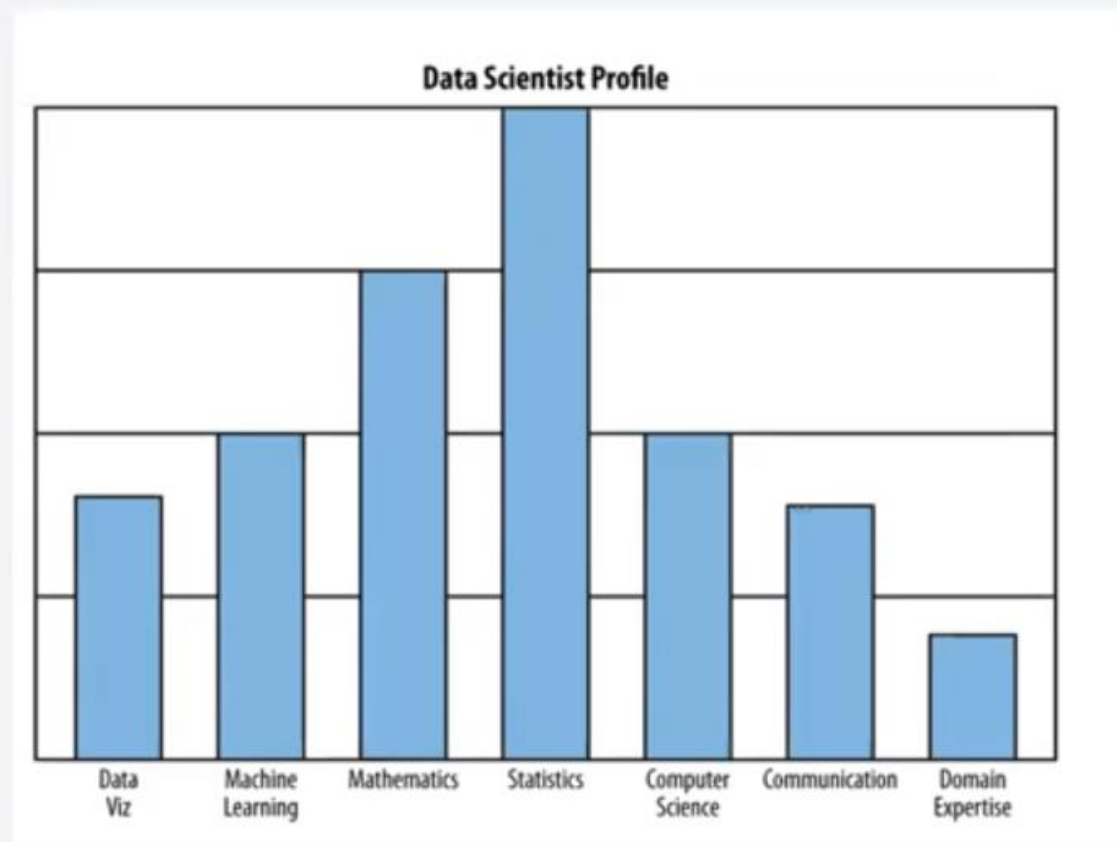
# Proceso de Ciencia de Datos



Fuente: Adaptado de O'Neil, Cathy, Schutt, Rachel. "Doing Data Science", O'Reilly Media.

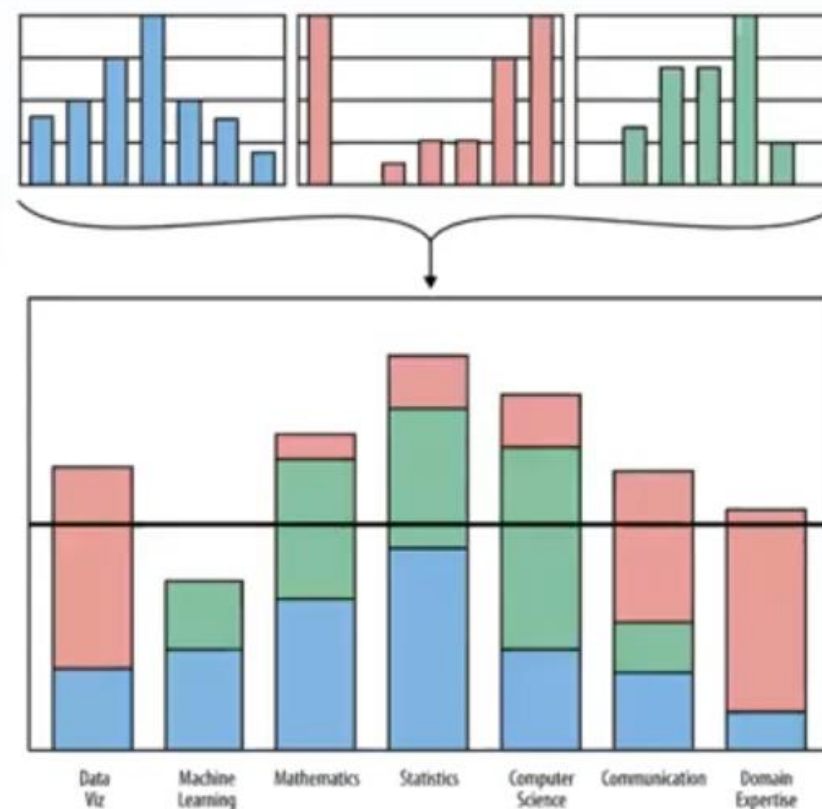
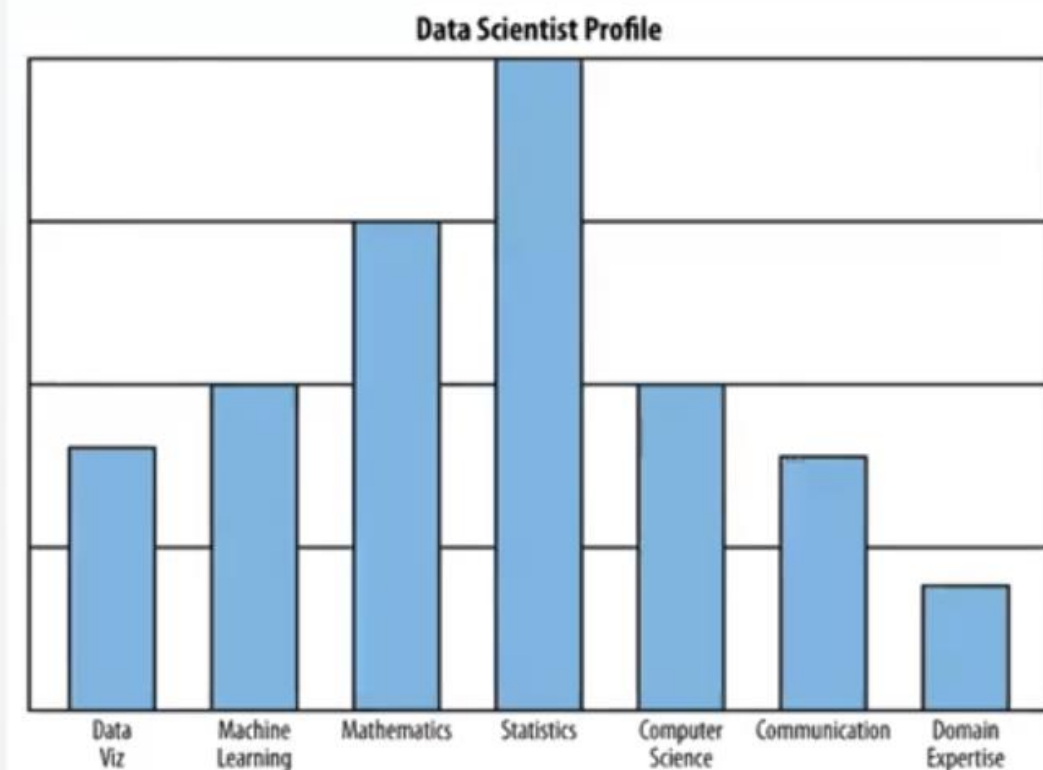
# Perfil de Cientistas de Datos

Se requiere una **combinación de competencias y conocimientos...**



# Perfil de Cientistas de Datos

...y trabajo **colaborativo**





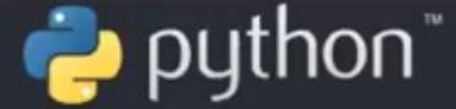
# **Herramientas computacionales para Ciencia de Datos**



# Python [www.python.org](http://www.python.org)

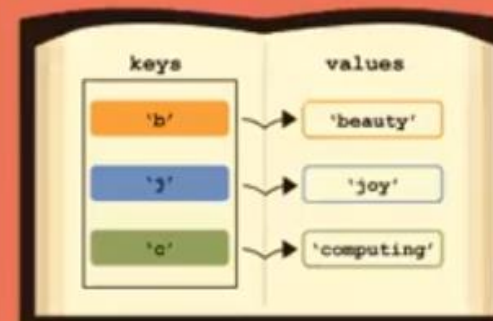
- Lenguaje **interpretado** de programación: los comandos o instrucciones se ejecutan paso por paso, sin compilación previa.
- Gran y activa **comunidad** de computación científica y ciencia de datos desarrollada en torno a Python.
- Integración de **múltiples librerías** con distintas funcionalidades y usos.

```
>>> import this
The Zen of Python, by Tim Peters
```

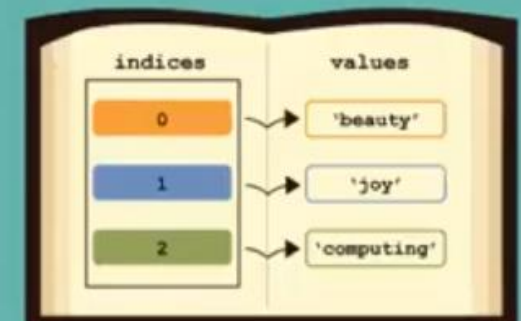


```
Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Flat is better than nested.
Sparse is better than dense.
Readability counts.
Special cases aren't special enough to break the rules.
Although practicality beats purity.
Errors should never pass silently.
Unless explicitly silenced.
In the face of ambiguity, refuse the temptation to guess.
There should be one— and preferably only one —obvious way to do it.
Although that way may not be obvious at first unless you're Dutch.
Now is better than never.
Although never is often better than *right* now.
If the implementation is hard to explain, it's a bad idea.
If the implementation is easy to explain, it may be a good idea.
Namespaces are one honking great idea — let's do more of those!
```

## dictionaries



## lists



# Librerías de Python para Ciencia de Datos



## ALGUNAS LECTURAS ...

- **A Survey on Bias and Fairness in Machine Learning**  
([Enlace](#))
- **Doing Data Science: A Framework and Case Study**  
([Enlace](#))

