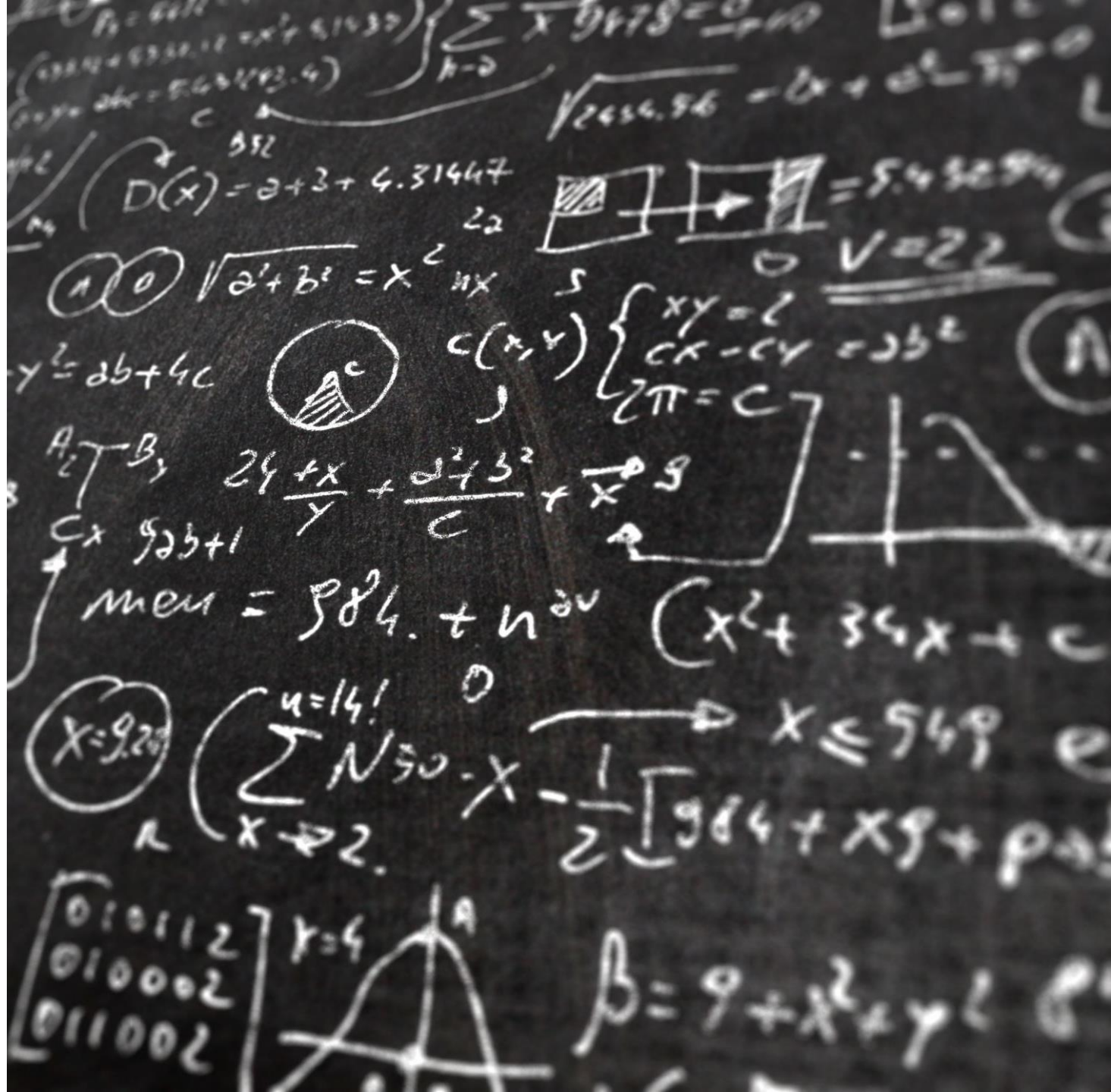


Ciencia de Datos



¿Qué es la Ciencia de Datos?

La **Ciencia de Datos** es un enfoque multidisciplinario para obtener perspectivas (insights) a partir de una cantidad cada vez mayor de datos.

La ciencia de datos combina las matemáticas, estadística, programación especializada, IA y el aprendizaje automático para descubrir estos insights y orientar la toma de decisiones.



¿Cómo crear valor con los datos?



DATOS



CLASIFICAR

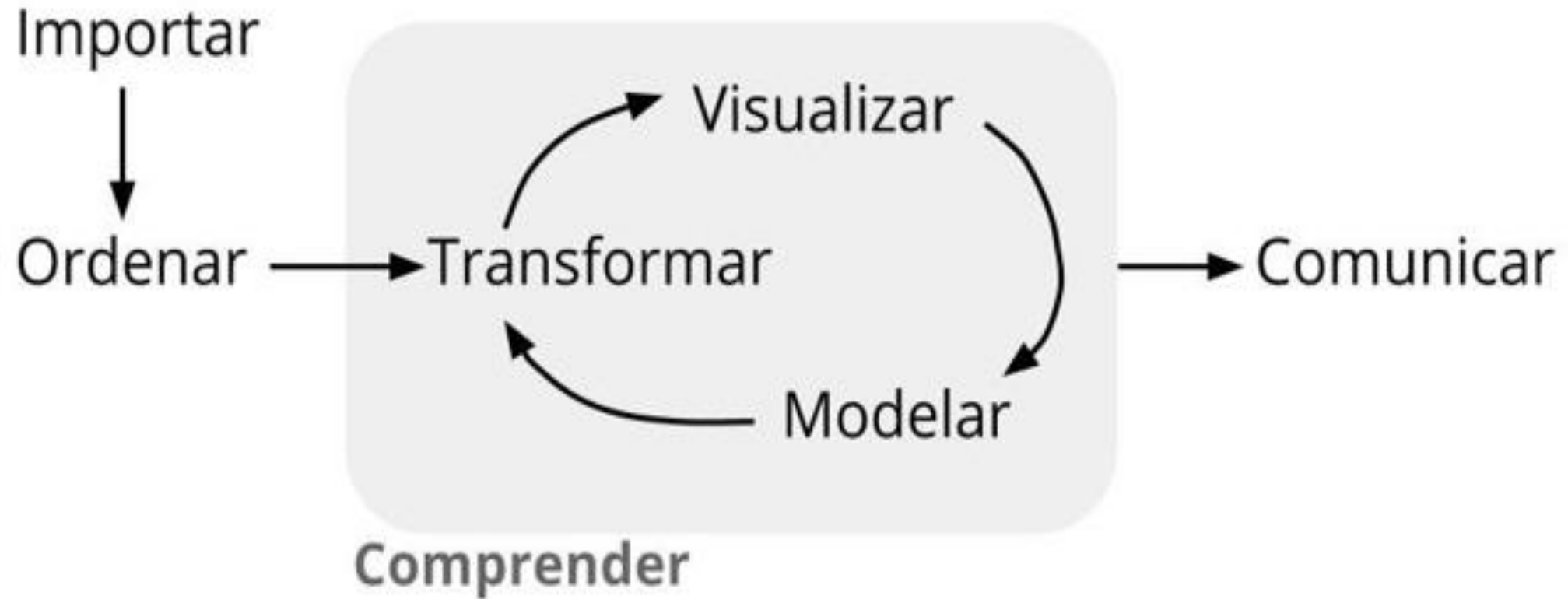


ORDENAR



VISUALIZAR Y ANALIZAR

El proceso en la Ciencia de Datos



Fuente: definido por Hadley Wickham

TIPOS DE DATOS

ESTRUCTURADOS

Datos que tienen un modelo definido o provienen de un campo determinado en un registro



NO ESTRUCTURADOS

Datos que no tienen un modelo predefinido o no están organizados de alguna manera



Tipos de datos

Datos estructurados Tablas, bases de datos.



Datos semi-estructurados

XML data



JSON data



Datos no estructurados

Textual



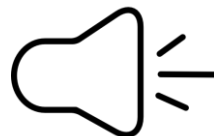
Imagen



Video



Audio



Datos acerca de datos

DATA

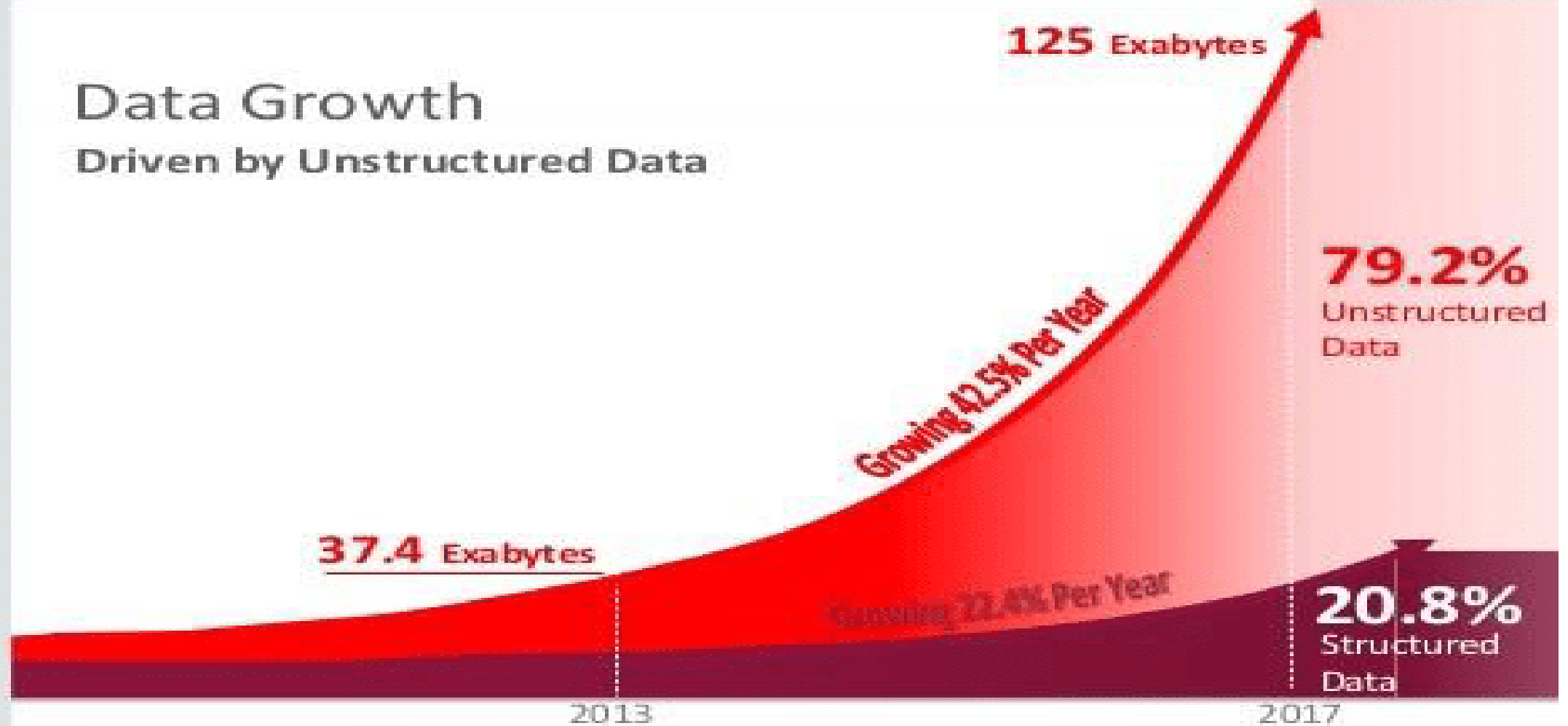


METADATA



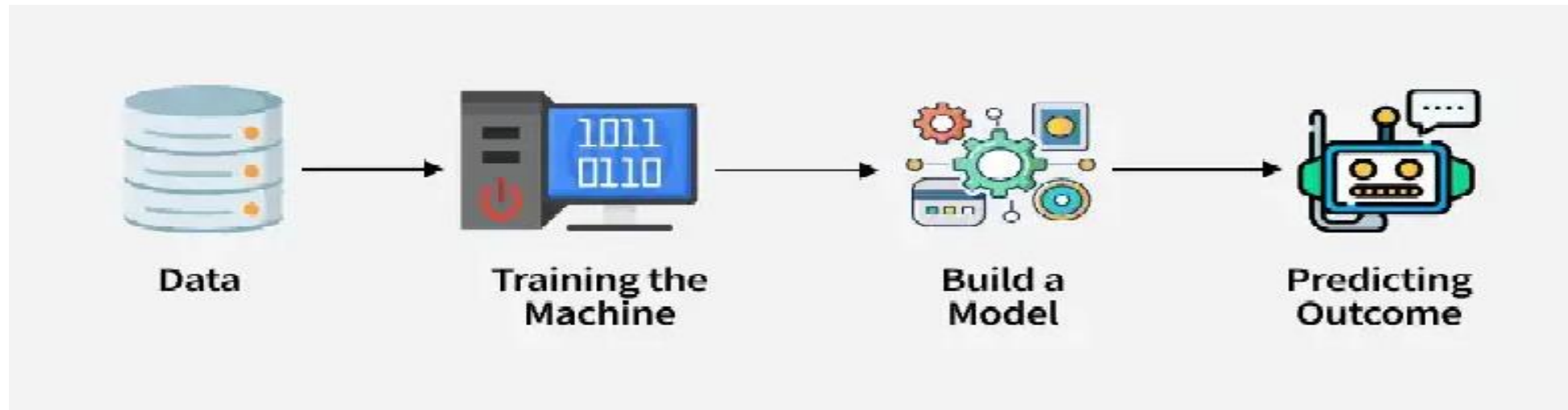
Data Growth

Driven by Unstructured Data

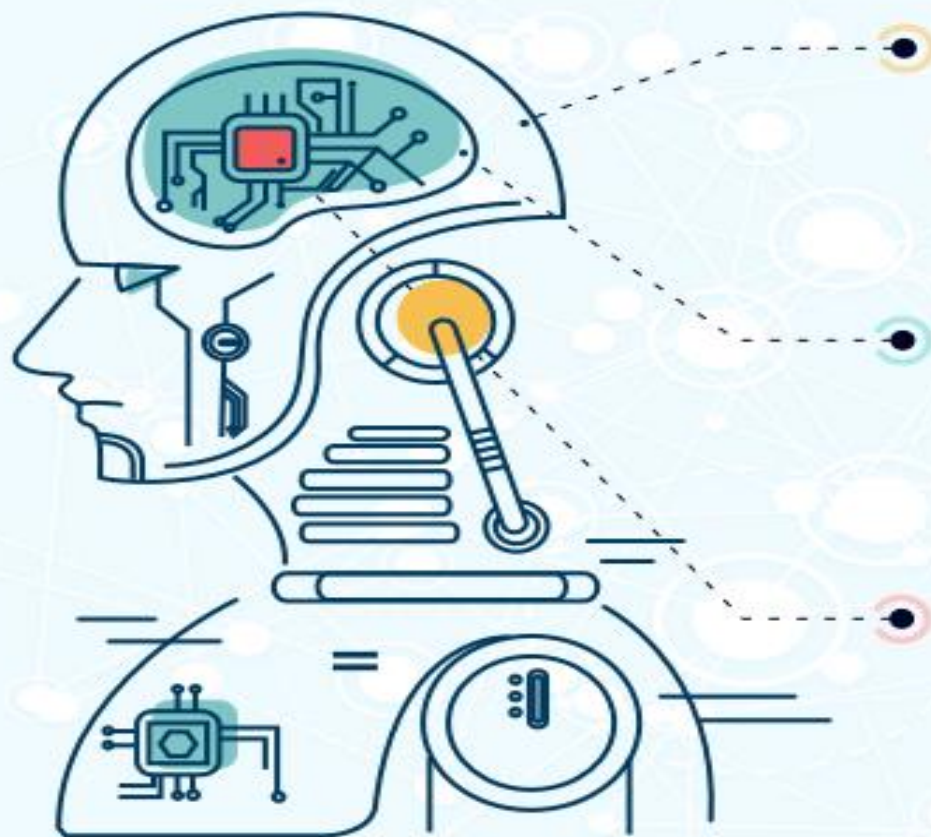


¿Qué es el Machine Learning?

- El aprendizaje automático (ML) permite a las computadoras aprender y tomar decisiones sin necesidad de programación explícita.
- Implica introducir datos en algoritmos para identificar patrones y realizar predicciones a partir de nuevos datos.
- El aprendizaje automático se utiliza en diversas aplicaciones, como el reconocimiento de imágenes y voz, el procesamiento del lenguaje natural y los sistemas de recomendación.



Las diferencias entre Inteligencia Artificial, Machine Learning y Deep Learning



Inteligencia Artificial (IA)

Campo que estudia cómo crear programas informáticos con **la habilidad de aprender y razonar como los humanos** para resolver problemas de forma creativa.

Machine Learning (ML)

Aplicación de la IA dedicada a la creación de algoritmos que permitan a los sistemas **aprender sin intervención humana**, es decir, sin necesidad de programarlos explícitamente.

Deep Learning (DL)

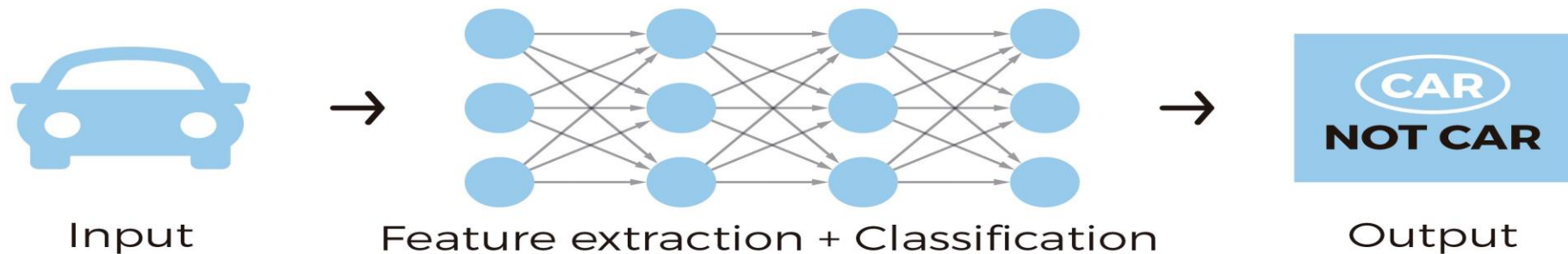
Subconjunto del ML enfocado a la **creación de redes neuronales artificiales**, es decir, sistemas que imitan al cerebro humano, adaptándose y aprendiendo a partir de grandes cantidades de datos.

Diferencias entre Machine Learning y Deep Learning

Machine Learning



Deep Learning



BIG DATA



Definición

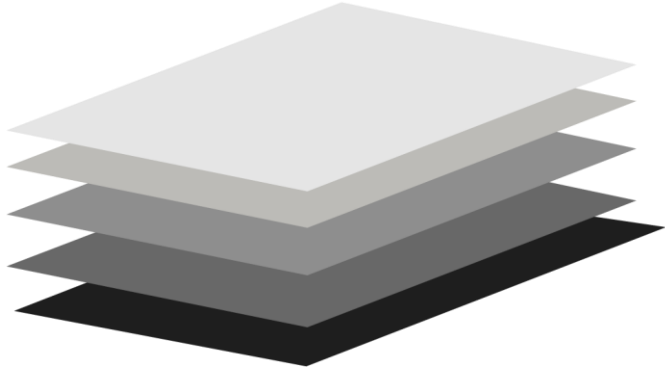
***Big data** se refiere a conjuntos de datos extremadamente grandes y complejos que no pueden gestionarse ni analizarse fácilmente con las herramientas tradicionales de procesamiento de datos*

***Big Data** cubre las estrategias no-tradicionales y tecnologías necesarias para recolectar, organizar, procesar y obtener información desde grandes conjuntos de datos (datasets).*



Las 5 "V" de Big Data

Volumen



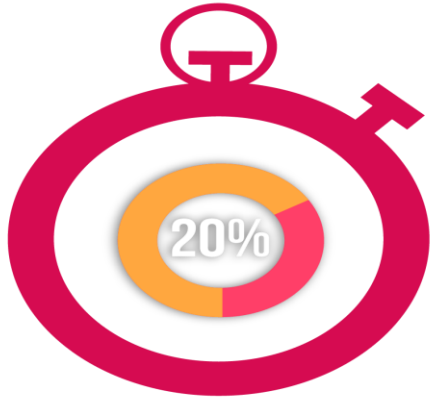
- Cantidades que exigen un almacenamiento mucho más complejo y con una capacidad de almacenamiento enorme a nivel terabyte.

Variedad

- Los datos provienen de diferentes fuentes
- Datos estructurados y no estructurados.



Velocidad



- Es necesario contar con capacidades robustas que hagan frente a la volatilidad de los datos.
- Muchos de ellos tienen una vida útil corta.
- Por lo que es necesario capturarlos y analizarlos en el momento oportuno para que no pierdan valor.

Valor

- Se refiere a la información que puedan tener los datos para alcanzar el objetivo que se está buscando.



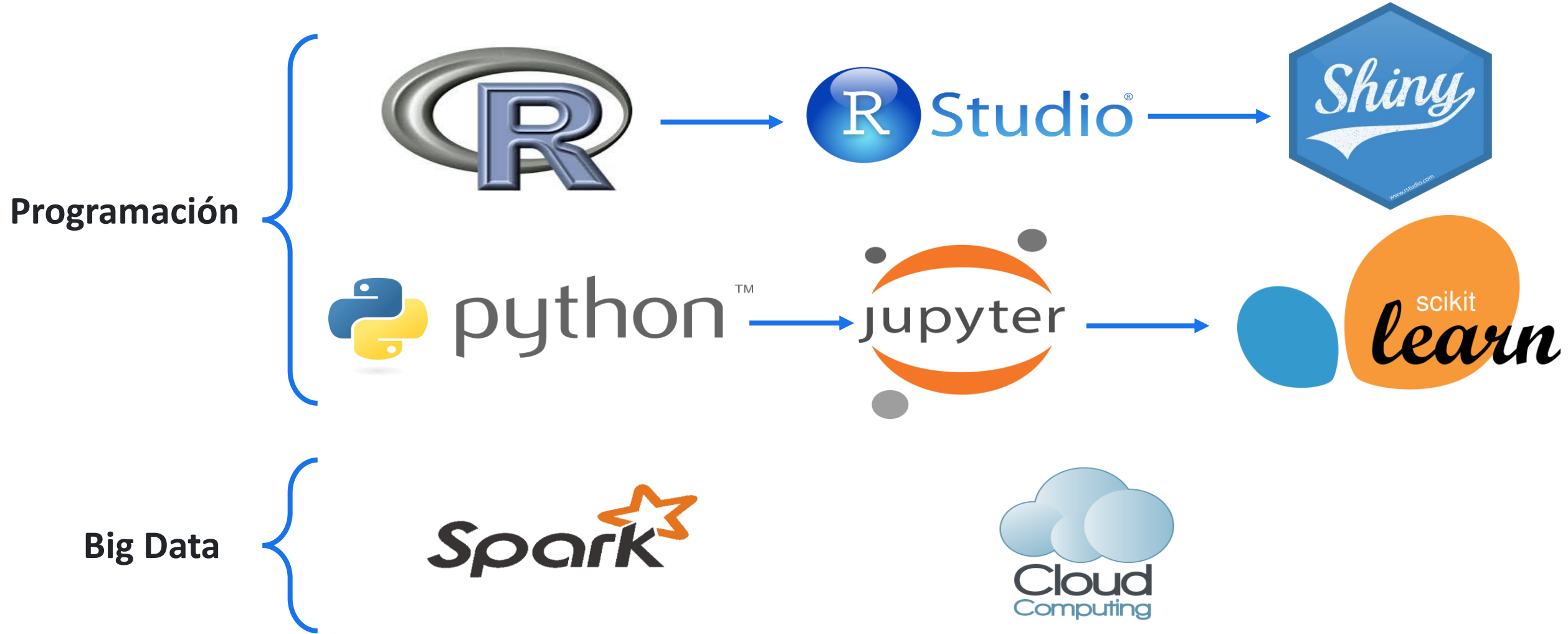
Veracidad

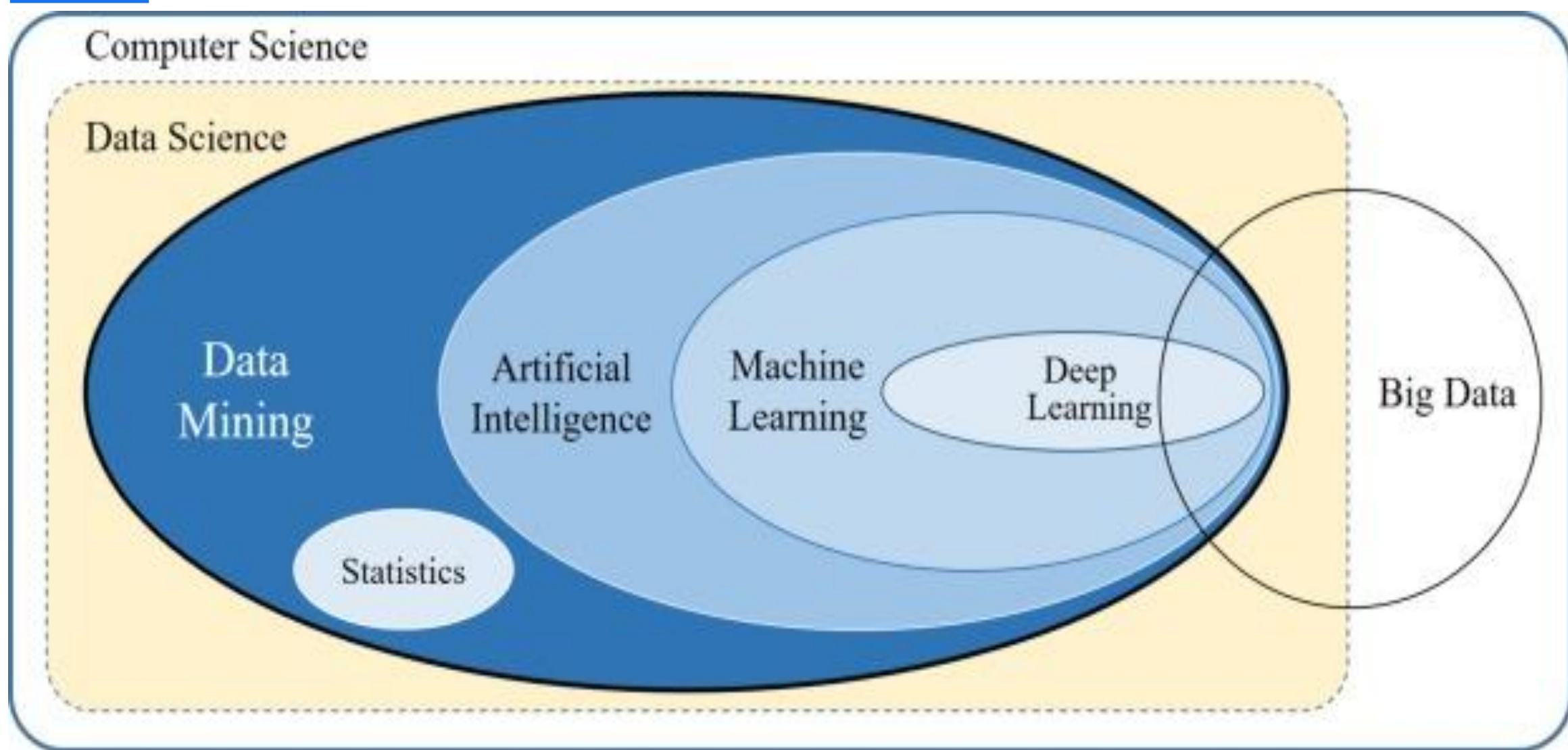


```
10101100101
10010101011
00110101001
11010110010
11010101111
```

- Calidad en cuanto al resguardo de los datos.
- Al trabajar con grandes volúmenes de información, se pueden presentar problemas como:
 - registros incompletos o erróneos,
 - datos faltantes
 - información que, proveniente de diferentes fuentes, son discrepantes.

Relación de la Ciencia de Datos y el Big Data





Ciencia de Datos en la vida cotidiana

Autos autónomos



Casas inteligentes



Gestión de tránsito



Ciudades inteligentes



Ofertas personalizadas



Ejemplos en el uso de Big Data



- Netflix utiliza Big Data para sus sistemas de recomendación y algunas estrategias/sugerencias comerciales y de marketing basadas en el análisis FODA y PESTLE.
- Hoy cuenta con más de 300 millones de suscriptores en todo el mundo

Fuente de datos

Para los sistemas de recomendación Netflix utiliza:



Calificaciones de sus usuarios.



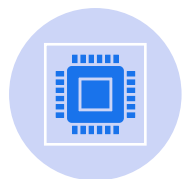
Ubicación, la hora y la fecha



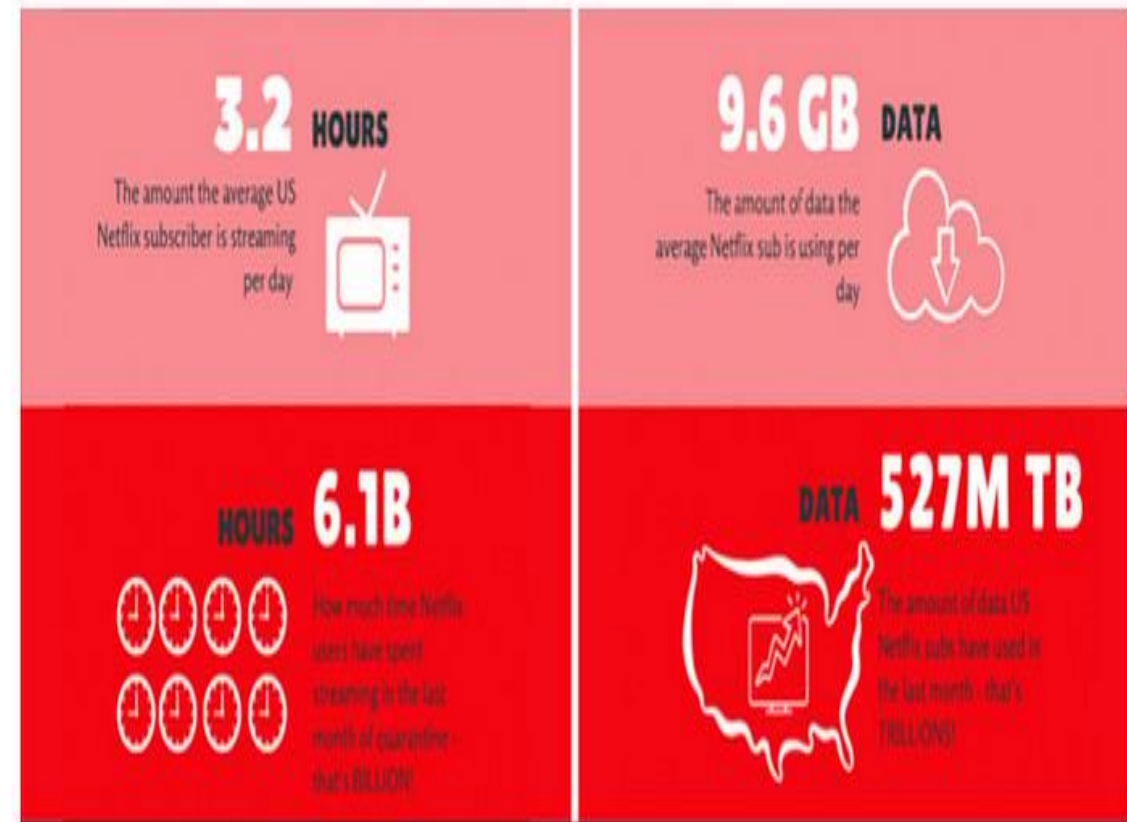
El dispositivo que están usando para transmitirlo



Los puntos en los que se pausó, reanudó o adelantó el programa, o si el programa se ve repetidamente



Los datos generados por Netflix han aumentado hasta el 1000% entre 2016 y 2019..



El éxito de amazon



- Gracias al Big Data es una de las empresas líderes del comercio online
- El uso del Big Data en Amazon está basado en el machine learning
- El Big Data lo utiliza para predecir el comportamiento de los usuarios y mejorar su experiencia en la plataforma.
- Filtración colaborativa ítem a ítem

El éxito de amazon

Big data en AWS

- Amazon incorpora las funciones Big Data en Amazon Web Services (o AWS, la plataforma de soluciones en la nube de la compañía)
- No incluye solo el almacenamiento y análisis de datos, sino también sus análisis predictivos y aprendizaje automático
- Integración con otras herramientas como Hadoop

