

Big Data y Visualización

Características del Big Data

Universidad Autónoma de Manizales

Docente: Alejandro Mora Rubio



Acreditación Institucional
DE ALTA CALIDAD
Resolución 009527 Mineducación Sep. 6 de 2019

Contenido

¿Qué es Big Data?

Las V's del Big Data
(Volumen, Variedad, Velocidad, Veracidad, Valencia)

Generación de Valor

Computación Escalable

¿Qué es Big Data?

Grandes volúmenes de datos



Imagen de macrovector en [Freepik](#)

Las V's del Big Data - Volumen

Cantidad de información.

Crecimiento exponencial de la generación y capacidad de almacenamiento.

Retos: almacenamiento, recuperación y distribución.

Unit	Value	Example
Kilobytes (KB)	1,000 bytes	a paragraph of a text document
Megabytes (MB)	1,000 Kilobytes	a small novel
Gigabytes (GB)	1,000 Megabytes	Beethoven's 5th Symphony
Terabytes (TB)	1,000 Gigabytes	all the X-rays in a large hospital
Petabytes (PB)	1,000 Terabytes	half the contents of all US academic research libraries
Exabytes (EB)	1,000 Petabytes	about one fifth of the words people have ever spoken
Zettabytes (ZB)	1,000 Exabytes	as much information as there are grains of sand on all the world's beaches
Yottabytes (YB)	1,000 Zettabytes	as much information as there are atoms in 7,000 human bodies

Imagen de [My NASA Data](#)

Las V's del Big Data - Volumen

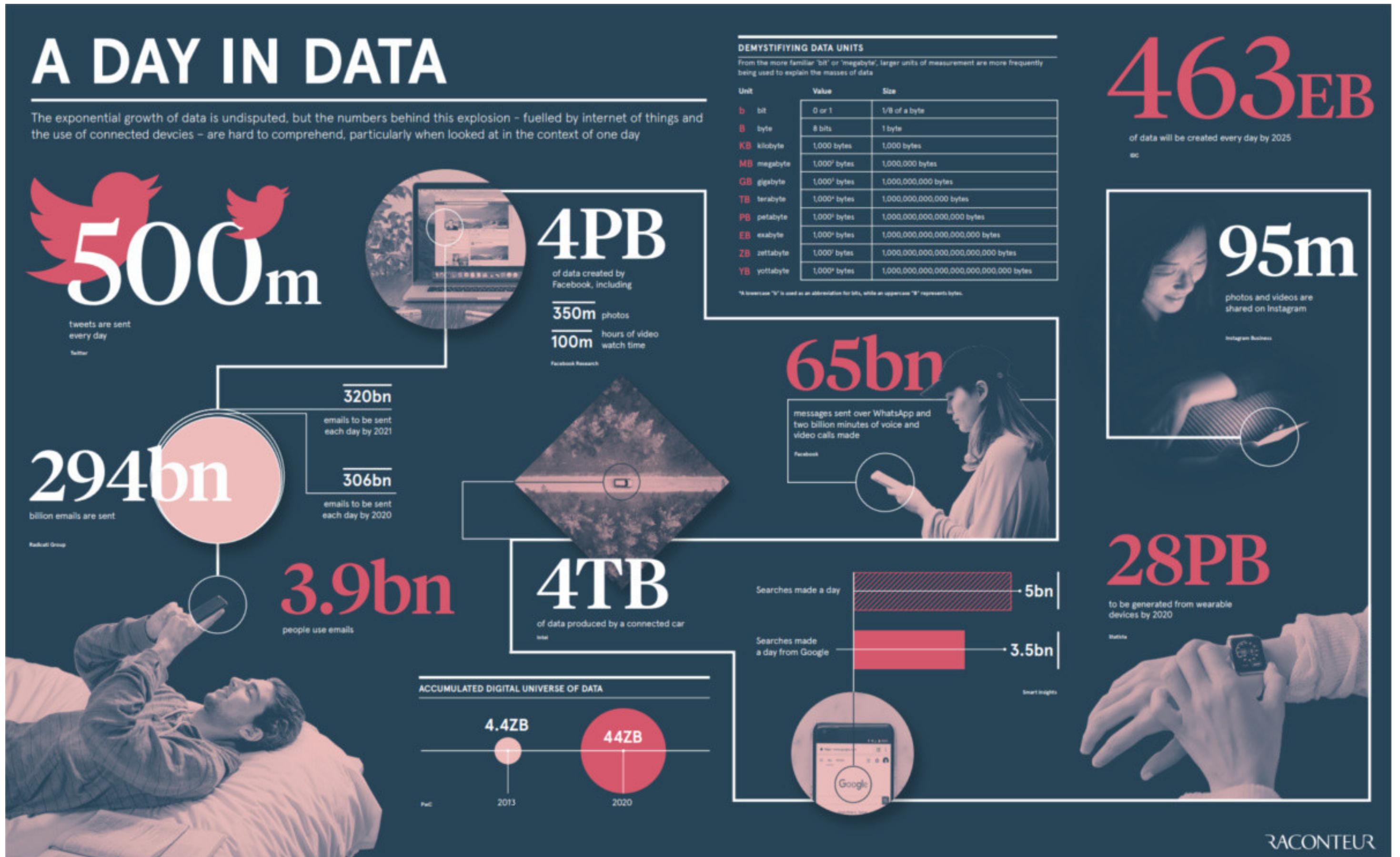


Imagen de World Economic Forum

2019 *This Is What Happens In An Internet Minute*



Imagen de Visual Capitalist

Las V's del Big Data - Variedad

Diferentes formas en la que se pueden presentar los datos:

Texto, imágenes, audio...

Retos: complejidad del procesamiento.

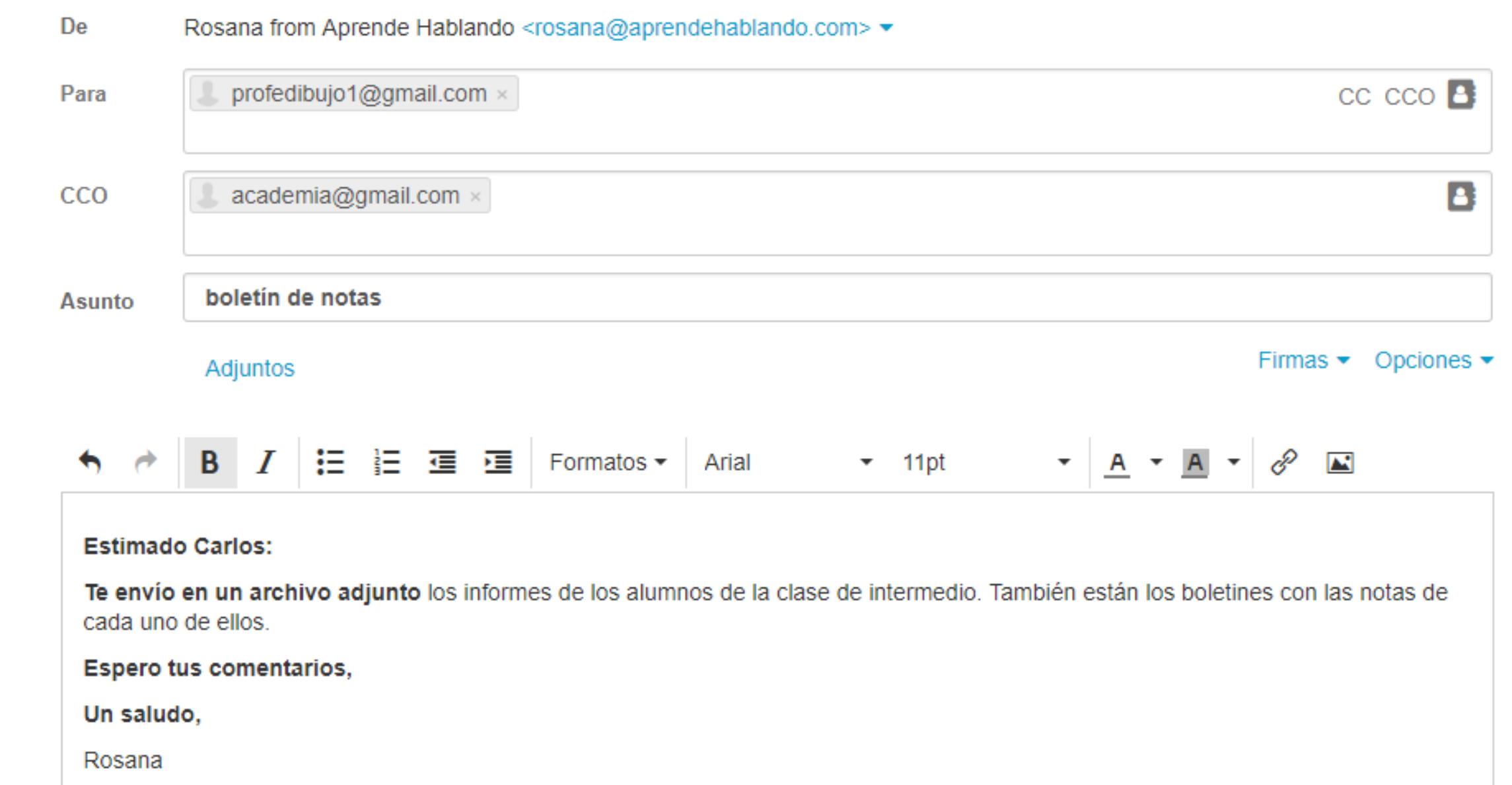


Imagen de [Aprende Hablando](#)

Las V's del Big Data - Variedad

DATOS ESTRUCTURADOS	DATOS NO ESTRUCTURADOS	DATOS SEMIESTRUCTURADOS
<p>Son los que tradicionalmente se han usado en el tratamiento de datos.</p> <p>Sus características principales son que se pueden almacenar en tablas y tienen una clara definición de longitud y formato.</p> <p>Entre ellos, se encuentran los números, cadenas de caracteres y las fechas.</p> <p>La irrupción gracias al Big Data de los otros tipos de datos no significa que estos no tengan su importancia.</p>	<p>Se trata de datos en su forma original, tal y como fueron recogidos.</p> <p>No poseen un formato específico que permita almacenarlos de forma tradicional, pues no se puede desglosar la información que facilitan a tipos de datos definidos en longitud y formato.</p> <p>Entre ellos son comunes, por ejemplo, los emails, las presentaciones multimedia como los PowerPoint, documentos de procesadores de textos o los archivos en formato PDF.</p>	<p>Siguen una especie de estructura, pero esta no es lo suficientemente regular como para gestionarla como datos estructurados.</p> <p>Posee ciertos patrones comunes que los describen y dan información sobre las relaciones entre los mismos.</p> <p>Como ejemplo, el HTML, lenguaje para la elaboración de páginas web, donde su sistema de etiquetas permite detectar esas pautas comunes.</p>

Imagen de Universidad Complutense de Madrid

Las V's del Big Data - Velocidad

Velocidad a la que se genera la información y a la que se desplaza entre dos puntos.

Retos: capacidad de computación, procesamiento en tiempo real.



Imagen de Tribu

Las V's del Big Data - Veracidad

Calidad de los datos en función de su precisión y confiabilidad.

Retos: calidad de los resultados.

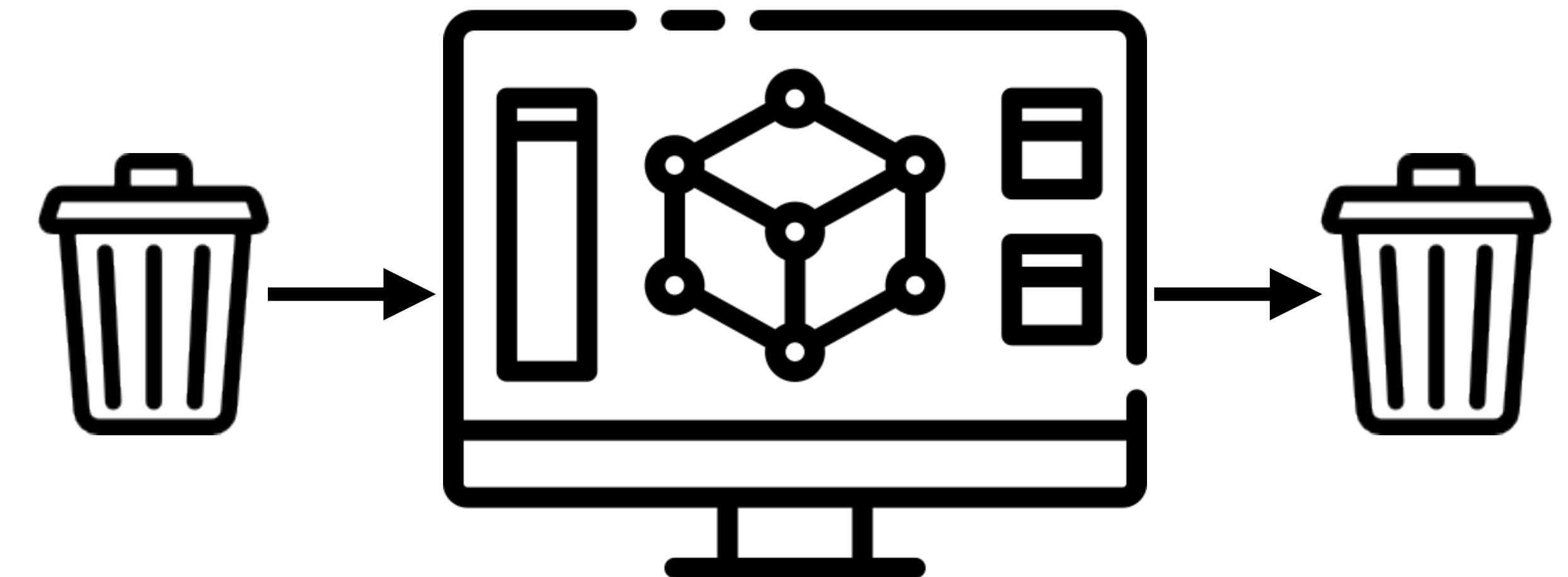


Imagen de [Freepik](#)

Las V's del Big Data - Valencia

Inspirado en el concepto químico, conexiones entre los datos o muestras.

Retos: complejidad en el análisis.

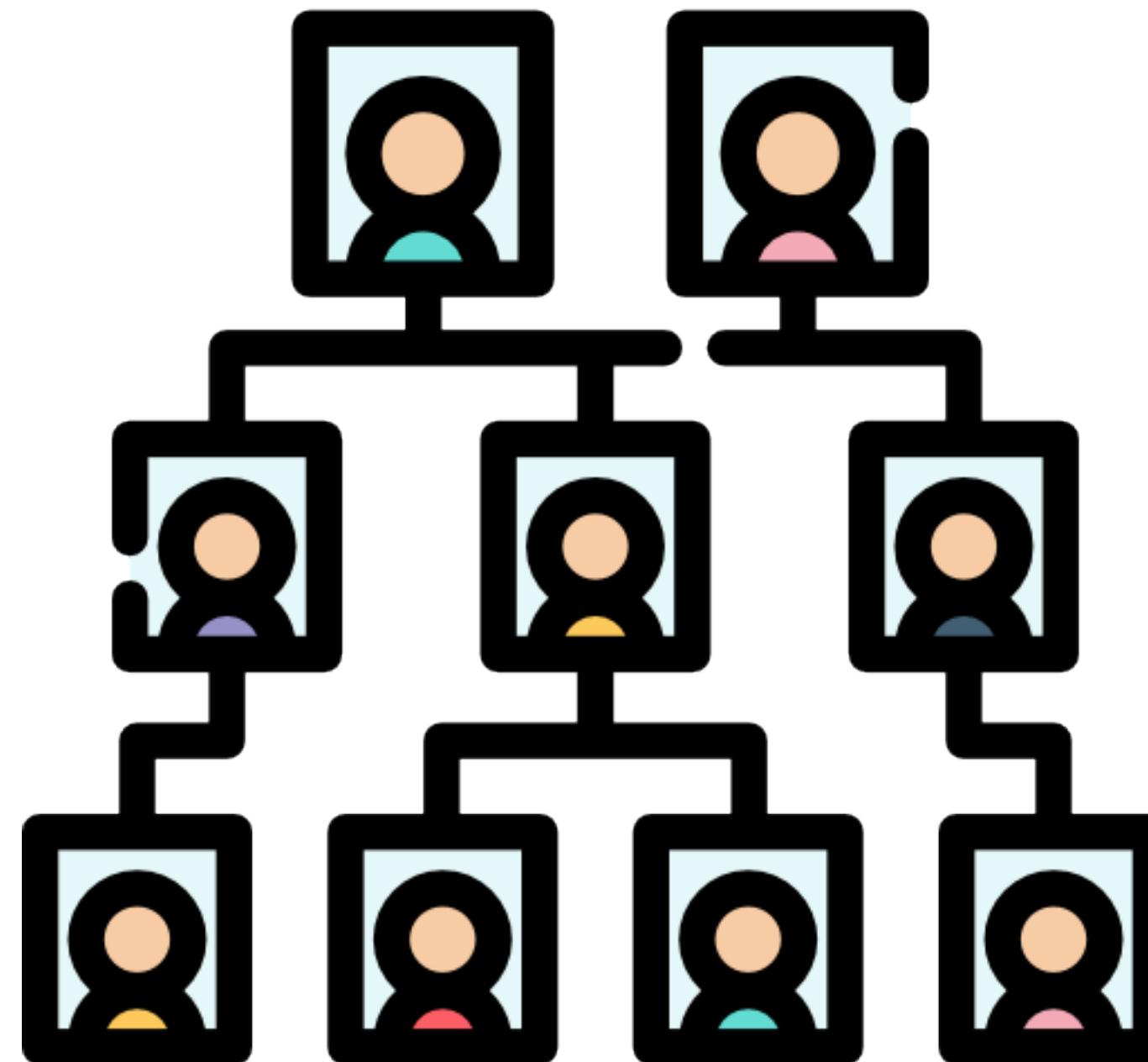


Imagen de [Freetipik](#)

Generación de Valor

Extraer información sobre la cuál tomar decisiones y obtener un beneficio.

Relacionado con el proceso de *Ciencia de Datos*.

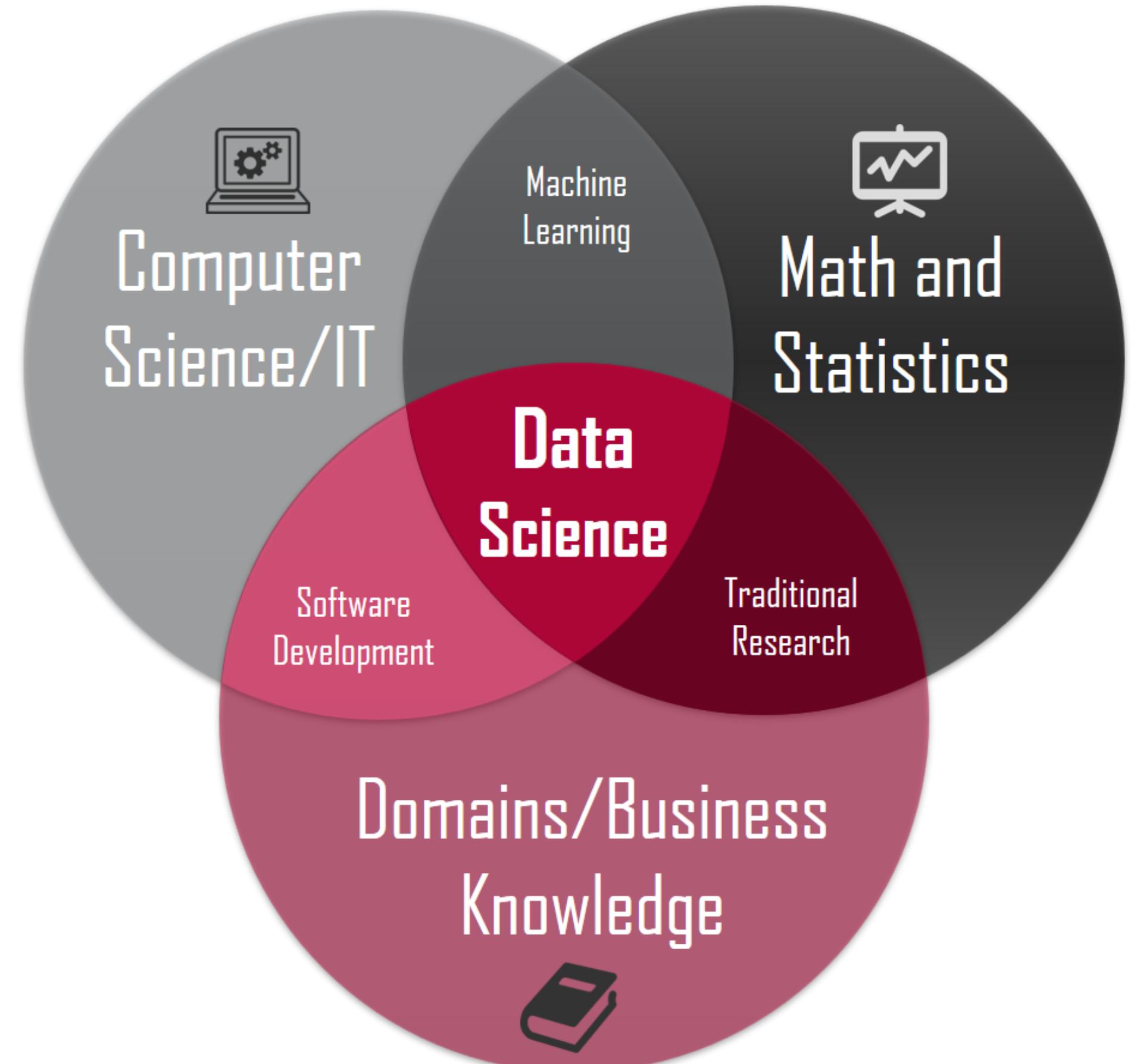


Imagen de [GitHub](#)

Generación de Valor



Etapas del proceso de Ciencia de Datos.

Computación Escalable

Sistemas de Archivos
Distribuidos
(Distributed File Systems, DFS)

Paralelización

Modelos de
Programación

Actividad propuesta

Consultar sobre Hadoop y MapReduce