

# Introducción

Big Data Aplicado

Dr. Francisco E. Cabrera

# ¿Qué es el Big Data?

Una gran cantidad de datos



# ¿Qué es el Big Data?

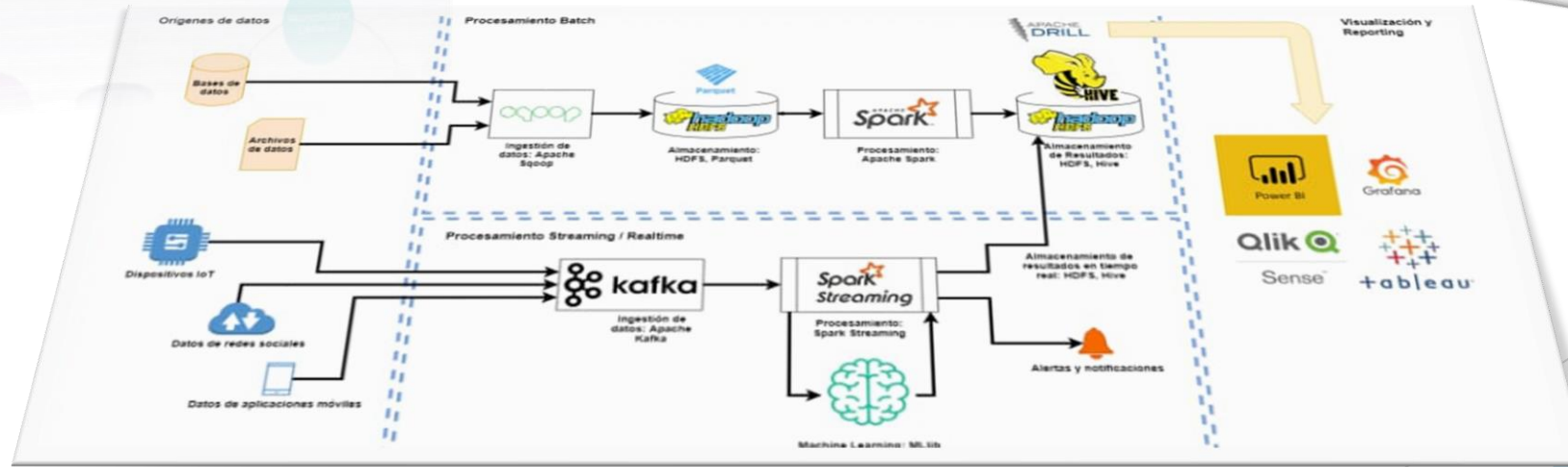
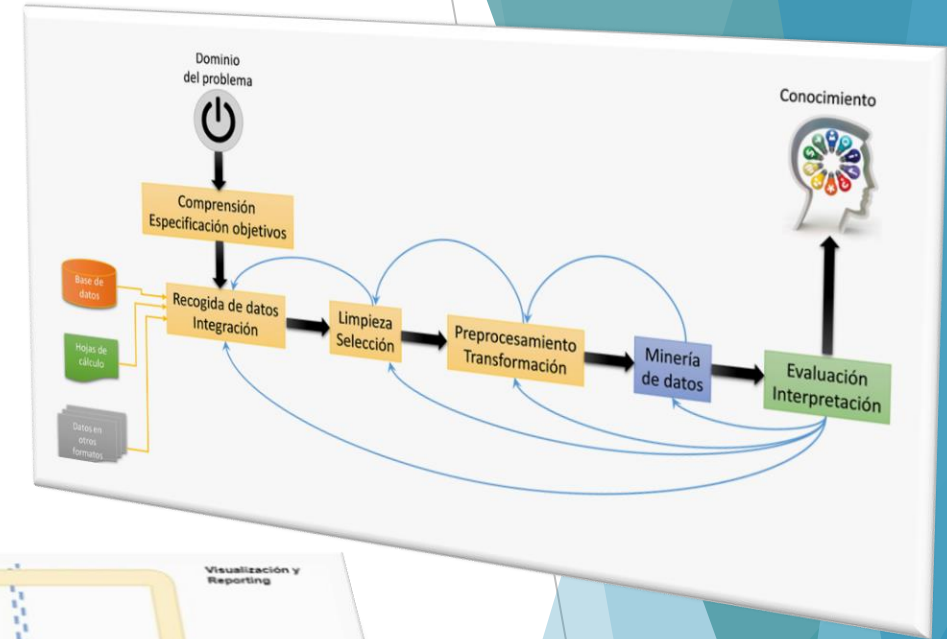
Es algo más que una gran cantidad de datos.

- ▶ Conjunto de datos extremadamente grandes y complejos que no pueden gestionarse ni analizarse de manera tradicional.
- ▶ No solo es importante el volumen de estos datos, también hay otras características que hacen que las técnicas tradicionales no sean las más adecuadas para tratar estos datos.
- ▶ El Big Data engloba también todo el conjunto de técnicas que se utilizan para tratar esos datos.

# ¿Qué es el Big Data?



**CONCEPTO PARAGUAS**



# Objetivo del Big Data

El Big Data sirve para extraer nueva información a partir de grandes conjuntos de datos.



# Aplicaciones

Obtener conocimiento a partir de los datos

- ▶ Estadísticas.
- ▶ Toma de decisiones.
- ▶ Personalizar la experiencia del usuario/cliente.
- ▶ Optimizar recursos.

# Aplicaciones



OpenAI

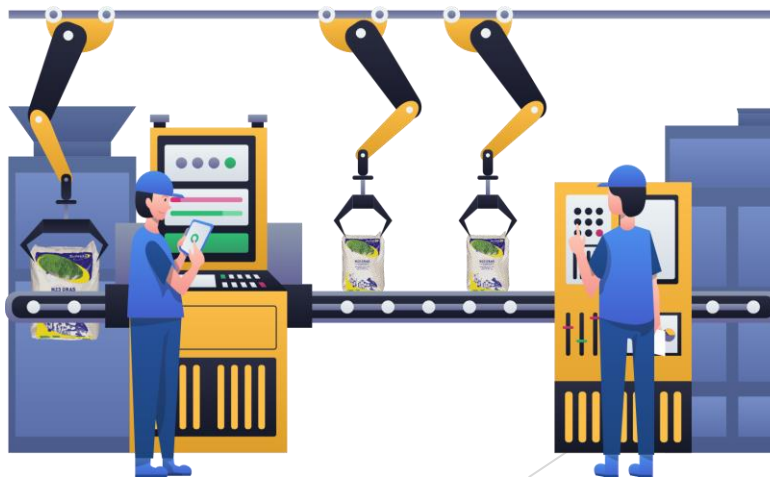
Google

TikTok

amazon



Microsoft



# ¿Qué vamos a aprender?

- ▶ ¿Qué características tienen los grandes conjuntos de datos?
- ▶ ¿Cómo obtengo los datos y que tipos de datos hay disponibles?
- ▶ ¿Dónde y cómo puedo almacenar los datos?
- ▶ ¿Cómo debo preparar estos datos?
- ▶ ¿Qué herramientas me pueden ayudar?
- ▶ ¿Qué clase de información puedo obtener?



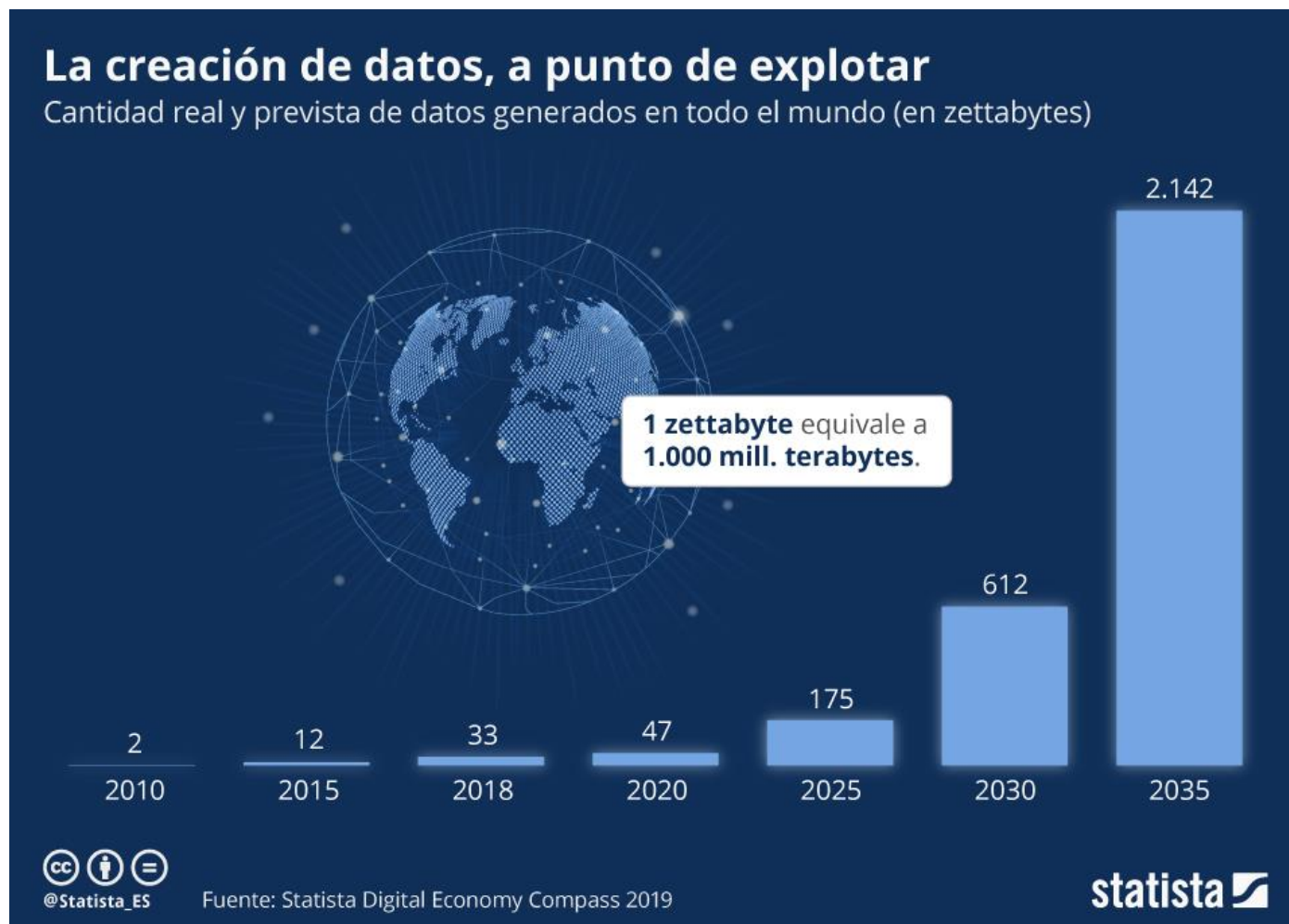
# Características del Big Data

Las 5 “V” del Big Data

# Las 5 “V” del Big Data



# Volumen



# Volumen



Google

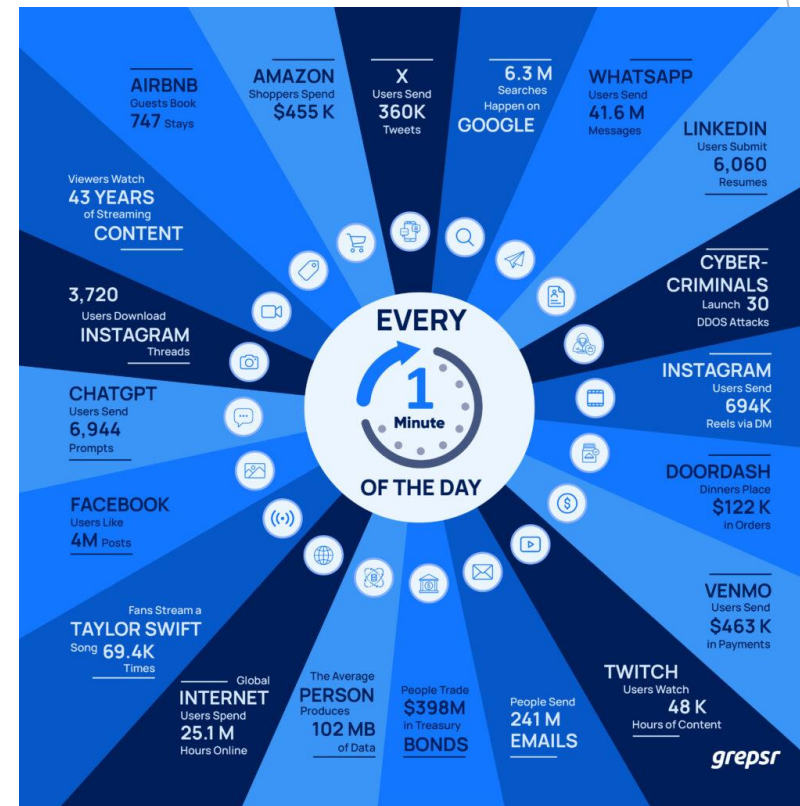


# Velocidad

Necesitamos respuesta a las preguntas en el tiempo adecuado.

Factores a tener en cuenta:

- ▶ Tiempos de respuesta.
  - ▶ Análisis en tiempo real.
- ▶ Datos generados en streaming.



# Variedad

Heterogeneidad de los datos.

- ▶ Distintas fuentes.
- ▶ Estructuras diferentes.
- ▶ Diferentes vías de acceso.



# Veracidad

No todas las fuentes son igual de fiables para todos los propósitos.

- ▶ Contexto.
- ▶ Calidad de la fuente.
- ▶ Consistencia.
- ▶ Relevancia.







# Implicaciones de las 5 Vs

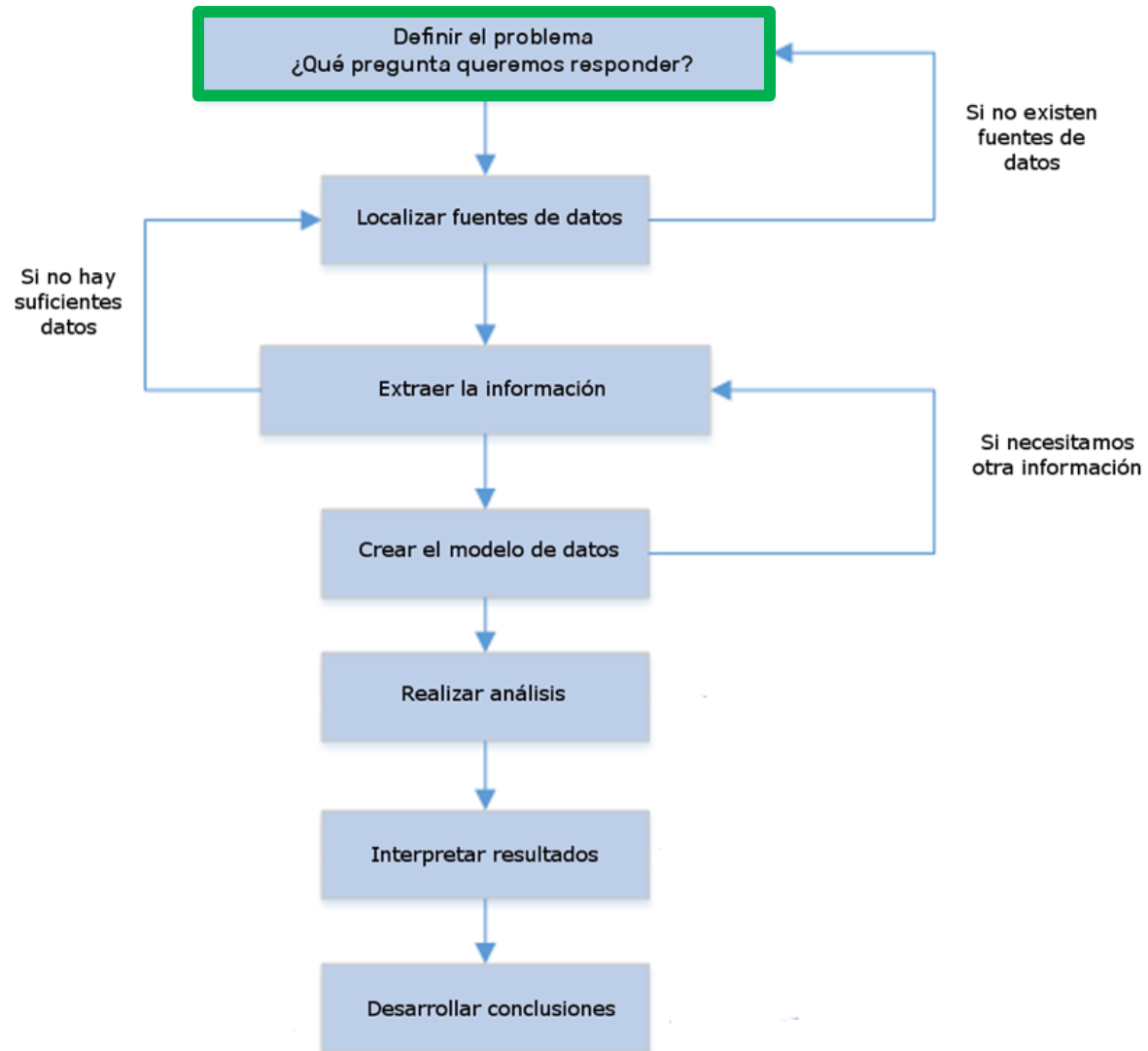
Afectan directamente a la calidad de los análisis

- ▶ Impactan la calidad del análisis.
  - ▶ Sin tener en cuenta la **Veracidad** ni evaluar correctamente el **Valor** podemos caer en sesgos o analizar información irrelevante.
- ▶ Determinan la eficiencia del procesamiento.
  - ▶ Si el sistema no está optimizado, la **Velocidad** y el **Volumen** pueden hacer que los datos lleguen muy tarde para ser útiles.
- ▶ Exigen manejar diferentes tipos de datos.
  - ▶ La **Variedad** implica que no solo trabajaremos con bases de datos estructuradas.
- ▶ Permiten tomar mejores decisiones.
  - ▶ Una decisión informada es la mejor que se puede tomar.



# El análisis de datos

# El proceso de un análisis de datos



# Definición del problema

*“Esa, definitivamente, es la respuesta. Creo que el problema, para ser sincero, es que en realidad nunca has sabido cuál es la pregunta.”*

- Deep Thought

# Definición del problema

Algunos ejemplos de problemas:

- ▶ Detección de fraude en transacciones financieras.
  - ▶ Identificar transacciones sospechosas en tiempo real para prevenir fraudes en una plataforma de pago.
- ▶ Mantenimiento predictivo en la industria.
  - ▶ Predecir fallos en máquinas y equipos para reducir tiempos de inactividad en una fábrica.
- ▶ Análisis de sentimiento en redes sociales para empresas.
  - ▶ Conocer la percepción pública sobre una marca o producto en redes sociales.
- ▶ Análisis de percepción política en elecciones.
  - ▶ Evaluar el sentimiento y las opiniones de la población sobre candidatos, partidos políticos o propuestas antes y durante una campaña electoral.