



---

# BIG DATA

---

Sebastián Humberto Sánchez Fuentes y Bernardo Rivera Ojeda



1 DE JULIO DE 2025  
INSTITUTO DE LA VERA CRUZ

## Introducción

El concepto de *Big Data* hace referencia al tratamiento y análisis de grandes volúmenes de datos generados a alta velocidad y con gran variedad. Su uso permite a las organizaciones obtener insights valiosos para la toma de decisiones estratégicas, operativas y predictivas.

## ¿Qué es Big Data?

Big Data es un término que describe conjuntos de datos tan grandes, complejos y cambiantes que superan la capacidad de las herramientas tradicionales de procesamiento.

## Evolución del Big Data

### ❖ Más de 7,000 años

Los primeros registros de uso de datos para rastrear y controlar negocios datan de esta época. Cuando se introdujo la contabilidad en Mesopotamia para registrar el crecimiento de cultivos y rebaños.

### ❖ Siglo XX

El primer gran proyecto de datos se creó en 1937 y fue encargado por la administración de Franklin D. Roosevelt en los Estados Unidos. Después de que la Ley del Seguro Social se hizo oficial, el gobierno tuvo que realizar un seguimiento de las contribuciones de 26 millones de estadounidenses y más de 3 millones de empleadores. IBM obtuvo el contrato para desarrollar una máquina de lectura de tarjetas perforadas para este enorme proyecto de contabilidad. La primera máquina de procesamiento de datos apareció en 1943. Fue desarrollada por los británicos para descifrar los códigos nazis durante la Segunda Guerra Mundial.

### ❖ Siglo XXI

En 2005, Roger Mougallas de O'Reilly Media acuñó el término Big Data por primera vez, solo un año después de que crearan el término Web 2.0. que se refiere a un gran conjunto de datos que es casi imposible de administrar y procesar con herramientas tradicionales de inteligencia empresarial. 2005 es también el año en que Yahoo! creó Hadoop, construido sobre MapReduce de Google. Su objetivo era indexar toda la World Wide Web y, hoy en día, muchas organizaciones utilizan Hadoop de código abierto para analizar grandes cantidades de datos.

## Desafíos del Big Data

- ❖ Muchas fuentes y Tipos de datos: Con tantas fuentes, tipos de datos y estructuras complejas, que hacen una dificultad de integración de datos aumenta.
- ❖ Volumen de Datos: El volumen de datos es enorme, y eso complica la ejecución de un proceso de calidad de datos dentro de un tiempo razonable. Y esto hace difícil recolectar, limpiar, integrar y obtener datos de alta calidad de forma rápida.
- ❖ Mucha volatilidad: Los datos cambian rápidamente y eso hace que tengan una validez muy corta. Para solucionarlo necesitamos un poder de procesamiento muy

alto. Si no lo hacemos bien, el procesamiento y análisis basado en estos datos puede producir conclusiones erróneas.

- ❖ No existen estándares de calidad de datos unificados: En 1987 la Organización Internacional de Normalización (ISO) publicó las normas ISO 9000 para garantizar la calidad de productos y servicios. Sin embargo, no fue hasta 2011 cuando ISO publicó las normas de calidad de datos ISO 8000.

<b>Ventajas</b>	<b>Desventajas</b>
Toma de decisiones basada en datos	Falta de personal
Avances en la salud	Altos costos
Mejoras en la seguridad	Sesgos en los datos
Optimización de procesos	Complejidad técnica
Innovación	Sobrecarga de información

## Referencias

¿Qué es el big data? Ventajas y Desventajas 2025. (s/f). Com.mx.  
<https://imepi.com.mx/que-es-el-big-data-ventajas-y-desventajas-2025/>

(S/f). Snaplogic.com. <https://www.snaplogic.com/glossary/big-data-maturity-model-bdmm>

*Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad.* (s/f). Powerdata.Es. <https://www.powerdata.es/big-data>