



Universidad Nacional
Federico Villarreal



FACULTAD DE INGENIERÍA
ELECTRÓNICA E INFORMÁTICA

INGENIERIA DE SISTEMAS DE INFORMACIÓN

Docente: Ing. Alejandro Rea
Lima Perú Junio 2025
Sesión 10





BIGDATA

LA INTELIGENCIA DE LOS DATOS

“La tecnología big data ya está aquí, ahora hay que aprender a usarla”



Big data in context

Chris Meyer. Fundador de Monitor Talent y Patrono de Fundación Innovación Bankinter

➔ **En la década de los 90**, cuando la informatización de la cada una de las funciones de negocio empezó a dar paso a los paquetes de software para empresas, los profesionales de la informática jerarquizaban la utilidad en "datos, información, conocimiento, sabiduría".

VOLUMEN
VELOCIDAD
VARIEDAD
VERACIDAD
VALOR



¿QUÉ ES EL BIG DATA?

Big Data es el océano de información en el que nadamos a diario (vastos zettabytes de datos que fluyen desde nuestras computadoras, dispositivos móviles y sensores de máquinas). Las organizaciones usan estos datos para tomar decisiones; mejorar procesos y políticas; y crear productos, servicios y experiencias centrados en el cliente. Big Data se llama "grande" no solo por su volumen, sino también por la variedad y complejidad de su naturaleza.

Normalmente, supera la capacidad de las bases de datos tradicionales para capturar, gestionar y procesar datos. Y Big Data puede provenir de cualquier lugar o cosa en el mundo que sea posible monitorear digitalmente. Satélites meteorológicos, dispositivos de internet de las cosas (IoT), cámaras de tránsito y tendencias en redes sociales son solo algunas de las fuentes de datos que se extraen y analizan para hacer que las empresas sean más resilientes y competitivas.

La importancia del análisis de Big Data

Mayor eficiencia

Mejora de la toma de
decisiones

Desafíos del análisis de Big
Data



Mejor experiencia del cliente

Ventaja competitiva

El verdadero valor de Big Data se mide por el grado en que usted puede analizar y comprender los datos. La inteligencia artificial (IA), el machine learning y las tecnologías modernas para base de datos permiten la visualización y análisis de Big Data para brindar información estratégica accionable –en tiempo real–.

Las analíticas de Big Data ayudan a las empresas a poner sus datos a trabajar para hacer realidad nuevas oportunidades y construir modelos de negocio. Como dijo acertadamente Geoffrey Moore, autor y analista de gestión, “sin analíticas de Big Data, las empresas son ciegas y sordas, entran a la web como animales distraídos a una autopista”.

LÍNEA DE TIEMPO BIG DATA

1960S - 1970S: INICIOS DE LA INFORMÁTICA Y LAS BASES DE DATOS

Se desarrollan las bases de datos relacionales y el lenguaje SQL (Structured Query Language) para gestionar datos estructurados

1960S



1970S

Se introducen las primeras computadoras y sistemas de almacenamiento de datos en empresas y organizaciones

1980S - 1990S: EVOLUCIÓN DE LAS BASES DE DATOS

La explosión de Internet y la web genera una gran cantidad de datos no estructurados, como texto e imágenes

1980S



1990S

Aparecen sistemas de gestión de bases de datos comerciales, como Oracle y Microsoft SQL Server, que permiten un mayor manejo de datos

El Big Data se convierte en un tema candente. Grandes empresas y organizaciones comienzan a adoptar soluciones de Big Data para analizar y aprovechar sus datos

2000S

2000S: SURGIMIENTO DE BIG DATA



2010S: ADOPCIÓN GENERALIZADA Y AVANCES TECNOLÓGICOS

2010S

Se acuña el término "Big Data" para describir conjuntos de datos extremadamente grandes y complejos. Tecnologías como Apache Hadoop y MapReduce se desarrollan para gestionar Big Data

El Big Data se combina con la inteligencia artificial y el aprendizaje automático para mejorar el análisis y la toma de decisiones

2010S



2020S - ACTUALIDAD: INTEGRACIÓN CON INTELIGENCIA ARTIFICIAL Y MÁS AVANCES

2020S

Aparecen plataformas de análisis en la nube, como AWS, Azure y Google Cloud, que facilitan el acceso y el procesamiento de Big Data.

ACTUALIDAD



La Internet de las cosas (IoT) y los dispositivos conectados generan grandes volúmenes de datos en tiempo real, lo que impulsa aún más la importancia del Big Data

Esta línea de tiempo destaca los hitos clave en la evolución del Big Data, desde sus inicios hasta su estado actual, en el que se ha convertido en una parte fundamental de la tecnología y los negocios.

Por inconcebible que parezca hoy, la computadora guía del Apolo llevó la primera nave espacial a la luna con menos de 80 kilobytes de memoria. Desde entonces, la tecnología informática ha crecido a un ritmo exponencial –y la generación de datos junto con ella–. De hecho, la capacidad tecnológica mundial para almacenar datos se ha duplicado cada tres años aproximadamente desde la década de 1980.

Hace poco más de 50 años, cuando despegó el Apolo 11, la cantidad de datos digitales generados en todo el mundo hubiera cabido en una laptop promedio. En 2020, Statista estima que se crearon o replicaron 64,2ZB y que "la cantidad de datos digitales creados durante los próximos cinco años será más del doble de los creados desde la llegada del almacenamiento digital".

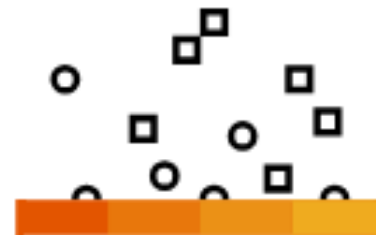
A medida que el software y la tecnología se vuelven más avanzados, menos viables son, en comparación, los sistemas no digitales. Los datos generados y recopilados digitalmente demandan sistemas de gestión más avanzados. Además, el crecimiento exponencial de plataformas de redes sociales, tecnologías de smartphones y dispositivos de IoT digitalmente conectados ha ayudado a crear la actual era de Big Data.

Tipos de Big Data: ¿qué son los datos estructurados y no estructurados?

Los data sets normalmente se clasifican en tres tipos según su estructura y cuán sencillo (o no) es indexarlos.



Structured data



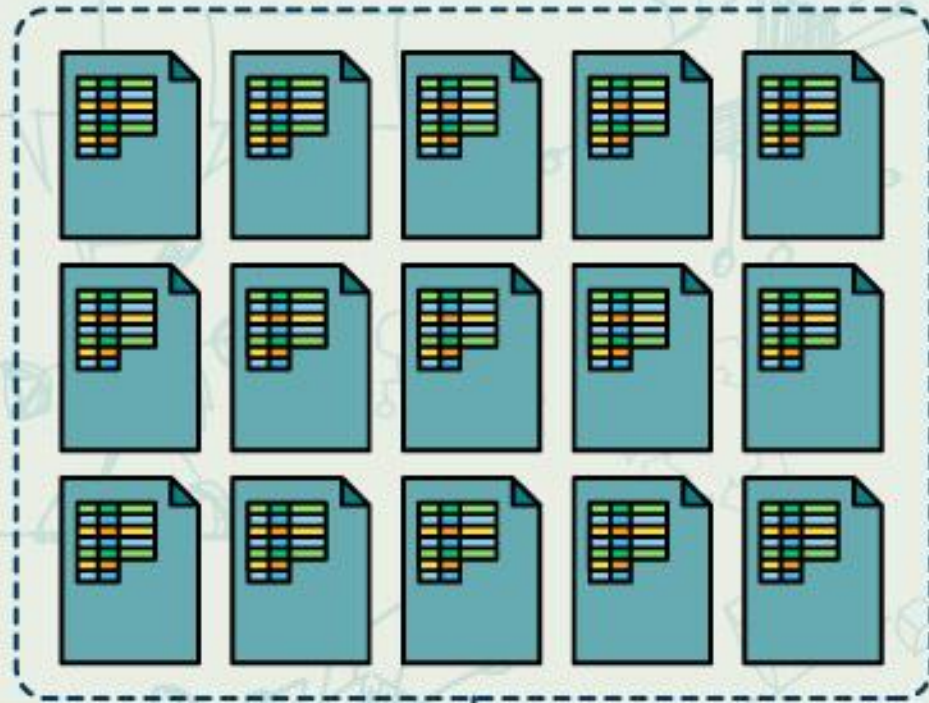
Unstructured data



Semi-structured data

Los tres tipos de Big Data

TIPOS DE BIG DATA



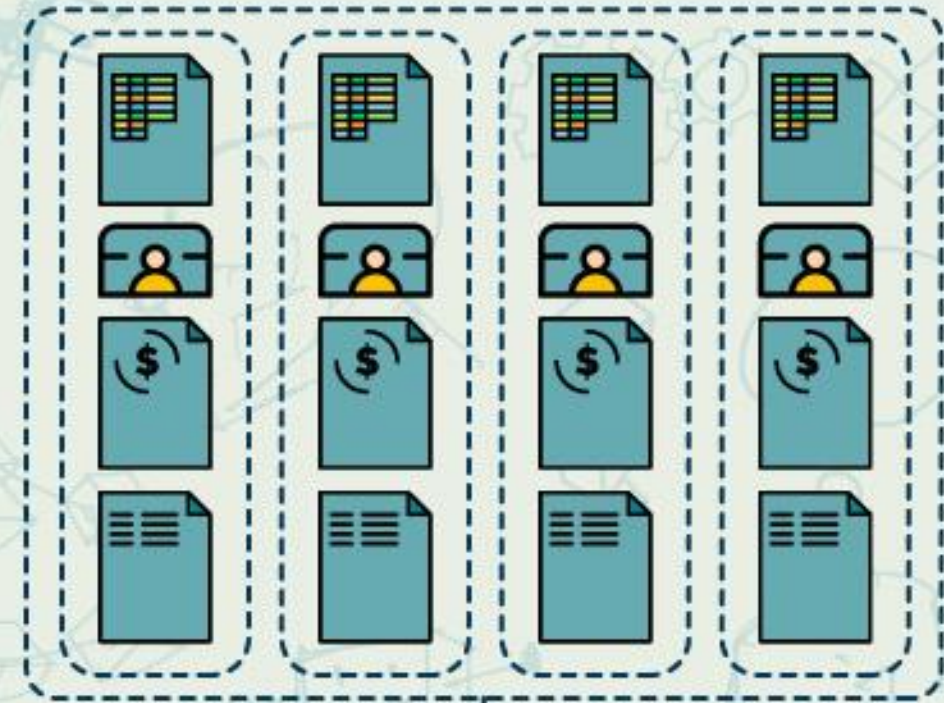
DATOS ESTRUCTURADOS

Datos cuidadosamente organizados que se guardan en bases de datos, conjuntos de datos y hojas de cálculo.



DATOS NO ESTRUCTURADOS

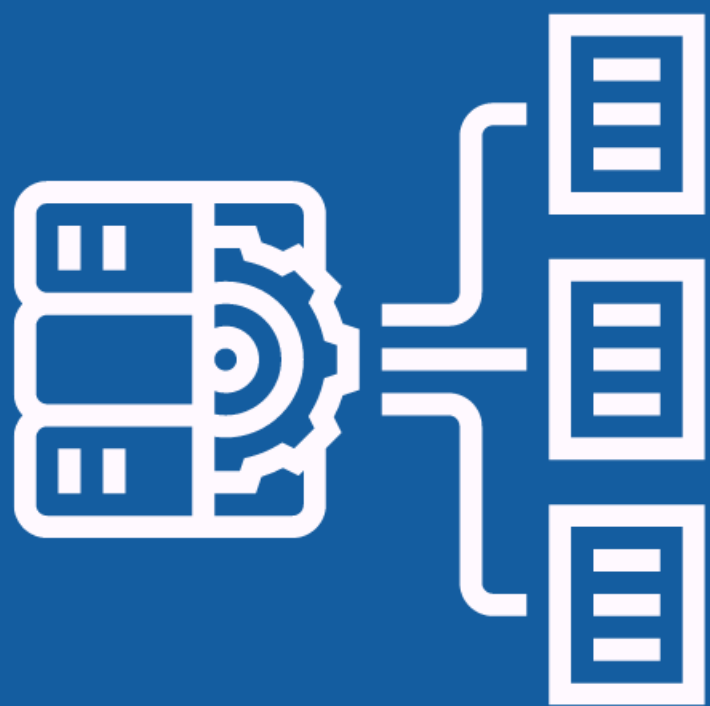
Datos sin procesar. Los datos no estructurados también pueden incluir datos de encuestas de clientes, notas y e-mails.



DATOS SEMIESTRUCTURADOS

Los datos semiestructurados tienen cierta estructura organizativa, pero no son fáciles de analizar tal como están.

¿Qué son datos estructurados?

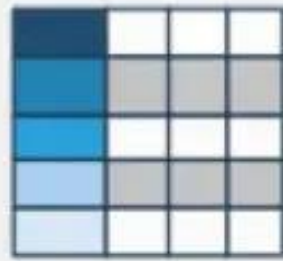


Datos estructurados: es el tipo más sencillo de organizar y buscar. Pueden incluir datos financieros, registros de máquinas, y detalles demográficos. Una hoja de cálculo de Excel, con su diseño de columnas y filas predefinidas, es una buena forma de visualizar datos estructurados.

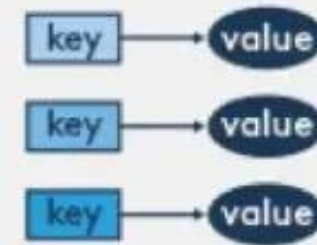
Sus componentes se categorizan fácilmente, lo cual permite a los diseñadores y administradores de bases de datos definir algoritmos simples para búsqueda y análisis. Incluso cuando existen grandes volúmenes de datos estructurados, no necesariamente califican como Big Data porque los datos estructurados por sí solos son relativamente simples de gestionar y, por lo tanto, no cumplen con los criterios de definición de Big Data.

Tradicionalmente, las bases de datos han usado un lenguaje de programación llamado lenguaje de consulta estructurado (SQL) para gestionar datos estructurados. El SQL fue desarrollado por IBM en la década de 1970 para permitirles a los desarrolladores construir y gestionar las bases de datos relacionales (al estilo de las hojas de cálculo) que comenzaban a prevalecer en ese momento.

Column-Family



Key-Value



NoSQL

Document



Graph



Datos no estructurados: esta categoría puede incluir cosas tales como posts en redes sociales, archivos de audio, imágenes, y comentarios abiertos de clientes. Este tipo de datos no son fáciles de capturar en bases de datos relacionales estándar de filas y columnas. Tradicionalmente, las empresas que querían buscar, gestionar o analizar grandes cantidades de datos no estructurados debían usar laboriosos procesos manuales.

Nunca hubo duda sobre el valor potencial de analizar y comprender tales datos, pero el costo de hacerlo a menudo era demasiado exorbitante como para que valiera la pena. Debido al tiempo que llevaba, los resultados solían quedar obsoletos incluso antes de ser entregados. En lugar de hojas de cálculo o bases de datos relacionales, los datos no estructurados se suelen almacenar en data lakes, almacenes de datos y bases de datos NoSQL.

Datos semiestructurados



Datos semiestructurados: tal como suena, los datos semiestructurados son un híbrido entre estructurados y no estructurados. Los correos electrónicos son un buen ejemplo, ya que incluyen datos no estructurados en el cuerpo del mensaje, al igual que propiedades organizativas tales como emisor, destinatario, asunto y fecha. Los dispositivos que usan etiquetado geográfico o semántico o marcas de tiempo también pueden brindar datos estructurados junto con contenido no estructurado.

Una imagen de smartphone no identificada, por ejemplo, puede indicar que es una selfie y el momento y lugar donde se tomó. Una base de datos moderna que opera tecnología de IA no solo puede identificar al instante diferentes tipos de datos, sino también generar algoritmos en tiempo real para gestionar y analizar eficazmente los dispares data sets involucrados.

Fuentes de Big Data

La gama de cosas que generan datos está creciendo a un ritmo fenomenal –desde satélites drones hasta tostadoras–. Pero a efectos de su categorización, las fuentes de datos generalmente se dividen en tres tipos:





BIG DATA



Volume



Value



Veracity



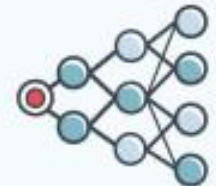
Visualization



Variety



Velocity



Virality