

FSID

*Fundamentos de los Sistemas de Información
Digitales*

Big Data y Business Intelligence

Año 2025



Definiciones de Big Data

Hay varias definiciones dependiendo de los autores:

- según Adrian Merv: Big Data excede el alcance de los entornos de hardware de uso común y las herramientas de software para gestionar los datos en un tiempo tolerable para una población de usuarios.
- según McKinsey: Big Data es un conjunto de datos más allá de la capacidad de herramientas típicas de Bases de datos.
- según IDC: Big Data es una nueva generación de tecnologías para capturar y analizar datos heterogéneos para sacar valor económico de ellos.
- según Deloitte: Big Data es un volumen que supera herramientas de uso común para analizar datos..
- según Gartner existen Las 3V: Volumen (cantidad), Velocidad(velocidad de creación y utilización), Variedad(fuentes de datos distintos)

Definiciones de Big Data

En definitiva la definición de Big Data puede variar según las características de las empresas.

Para unas empresas prima el volumen, para otras la velocidad, para otras la variabilidad de las fuentes.

Las empresas con mucho volumen o volumetría van a estar interesadas en capturar la información, guardarla, actualizarla e incorporarla en sus procesos de negocio; pero hay empresas que aunque tengan mucho volumen, no necesitan almacenar sino trabajar en tiempo real y a gran velocidad. Otras por el contrario, pueden estar interesadas en gestionar diferentes tipos de datos.

Entonces agrupando todas las definiciones podemos inferir que la definición podría ser: Big Data es el término que describe un gran volumen de datos, el cual crece de manera exponencial con el paso del tiempo. En pocas palabras, es un conjunto de datos tan grande y complejo que ninguna de las herramientas tradicionales de datos es capaz de almacenarlos o procesarlos de manera eficiente.

uso de Big data

Los sistemas de Big Data están brotando por todas partes y utilizándolos adecuadamente proporcionan una gran ventaja competitiva a las organizaciones y empresas. En cambio su ignorancia producirá grandes riesgos en las organizaciones y no las hará competitivas.

Es imperativo que las organizaciones persiguan agresivamente la captura y análisis de estas nuevas fuentes de datos para alcanzar los conocimientos y oportunidades que ellas ofrecen.

Tipos de datos en Big data

- Datos estructurados: Son la mayoría de de las fuentes de datos tradicionales, vienen en un formato o esquema fijo que poseen campos fijos. Por ejemplo bases de datos relacionales, hojas de cálculo y archivos.
- Datos semiestructurados: Son datos que no tienen formatos fijos, pero contienen etiquetas y otros marcadores que permiten separar los elementos dato. Por ejemplo XML, HTML, registros de logs.
- Datos no estructurados: Son datos sin tipos predefinidos, se almacenan en documentos u objetos. Por ejemplo audio, video, fotos, emails, SMS, redes sociales.
- Al menos 80% de la información de una organización es no estructurada.

Integración de datos y oportunidades de negocios

- El valor surge al combinar datos tradicionales con Big Data.
- Herramientas especializadas: Hadoop, MapReduce, MPP, Cloud.
- Ejemplo: preferencias de clientes en redes sociales + datos de ventas.
- Estrategia de Big Data debe integrarse con la estrategia global de la empresa.

Características de Big data (5V)

- Volumen: tamaño masivo de datos.
- Velocidad: rapidez en que fluyen los datos para generación y análisis.
- Variedad: diversidad de fuentes y formatos.
- Veracidad: fiabilidad y calidad de los datos.
- Valor: utilidad en la toma de decisiones.

Retos y Oportunidades de Big Data

- Retos: almacenamiento, privacidad, integración de fuentes heterogéneas, formación del personal.
- Oportunidades: ventaja competitiva, innovación, decisiones basadas en evidencia.
- Caso: Walmart procesa 1M+ transacciones/hora.
- Empresas que aplican Big Data son 5-6% más productivas (MIT).

Herramientas nombradas

- Hadoop

Es un marco de trabajo de software de código abierto, administrado por la Fundación de Software Apache, lo que permite el procesamiento paralelo distribuido de enormes cantidades de datos a través de computadoras económicas. Descompone un problema de Big Data en varios subproblemas, los distribuye entre miles de nodos de procesamiento de computadoras económicas y luego combina el resultado en un conjunto de datos de menor tamaño que es más fácil de analizar. Hadoop consta de varios servicios clave: el sistema de archivos distribuidos Hadoop (HDFS) para almacenamiento de datos y MapReduce para procesamiento de datos en paralelo de alto rendimiento. HDFS enlaza entre sí los sistemas de archivos en los numerosos nodos en un clúster Hadoop para convertirlos en un gran sistema de archivos. MapReduce de Hadoop se inspiró en el sistema MapReduce de Google para desglosar el procesamiento de enormes conjuntos de datos y asignar trabajo a los diversos nodos en un clúster. HBase, la base de datos no relacional de Hadoop, ofrece un acceso rápido a los datos almacenados en HDFS y una plataforma transaccional para ejecutar aplicaciones en tiempo real de alta escala.

Herramientas nombradas

- MapReduce es un modelo de programación distribuida para procesar grandes volúmenes de datos a través de un clúster de servidores, dividiendo las tareas en dos fases principales: Map, que transforma los datos de entrada en pares clave-valor, y Reduce, que agrupa los valores con la misma clave para producir un resultado resumido y más pequeño. Este enfoque permite la ejecución paralela de tareas, lo que acelera drásticamente el procesamiento de Big Data y se utiliza comúnmente dentro del ecosistema de [Apache Hadoop](#).
- MPP: utiliza las características de los equipos con múltiples procesadores y que procesan en forma paralela

Business Intelligence

El Business Intelligence o BI o Inteligencia de Negocios es el proceso de recopilar, almacenar, analizar y presentar datos empresariales para apoyar y mejorar la toma de decisiones.

Utiliza un conjunto de tecnologías, procesos y herramientas para transformar datos en información útil y conocimiento, permitiendo a las empresas identificar tendencias, mejorar el rendimiento y mantenerse competitivas.

El BI se enfoca en generar informes y visualizaciones, como paneles de control , que facilitan la comprensión de los datos históricos y actuales

Componentes principales de BI

- **Recopilación y limpieza de datos:** Se extraen y preparan datos de diversas fuentes internas y externas (como ventas, marketing, finanzas) utilizando técnicas como la extracción, transformación y carga (ETL).
- **Análisis de datos:** Se utilizan herramientas y metodologías para identificar patrones, tendencias y relaciones dentro de los datos.
- **Visualización de datos:** La información se presenta a través de paneles de control (dashboards), gráficos y otras representaciones visuales intuitivas para que sea fácil de entender.
- **Toma de decisiones:** La información generada proporciona el conocimiento necesario para que los responsables puedan tomar decisiones estratégicas y operativas más informadas.

Como funciona el BI

- **Identificación de fuentes de datos:** Se determinan los datos relevantes que provienen de sistemas internos (como ERP, CRM) y fuentes externas.
- **Procesamiento de datos:** Los datos se extraen, se limpian, se transforman y se cargan en un repositorio, como un almacén de datos (data warehouse).
- **Análisis:** Se exploran los datos para descubrir tendencias, identificar problemas y encontrar oportunidades.
- **Visualización:** Los resultados se presentan en informes y dashboards para una fácil comprensión y seguimiento de las métricas clave.
- **Acción:** La organización implementa cambios basados en la información obtenida para mejorar el rendimiento y alcanzar sus objetivos

Ejemplos de uso de BI

- **Identificar tendencias del mercado:** Permite comprender la dinámica del mercado y anticipar cambios.
- **Mejorar la eficiencia operativa:** Se optimizan procesos al detectar cuellos de botella o áreas de mejora.
- **Personalizar la experiencia del cliente:** Al comprender los datos de los clientes, se pueden ofrecer productos y servicios más personalizados.
- **Medir el impacto de iniciativas:** Se evalúa la efectividad de las campañas de marketing u otras acciones estratégicas.

FSID

GRACIAS

Año 2025

