

Tipos de datos

Índice

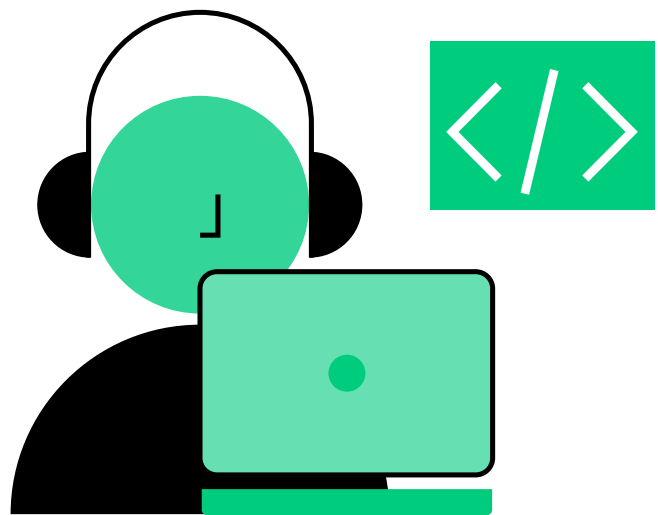
- 01 [Repasemos: ¿Qué son los datos?](#)
- 02 [Datos estructurados](#)
- 03 [Datos no estructurados](#)
- 04 [Datos semi estructurados](#)
- 05 [Conclusiones](#)



01

Repasemos:
¿Qué son los datos?

#Dato

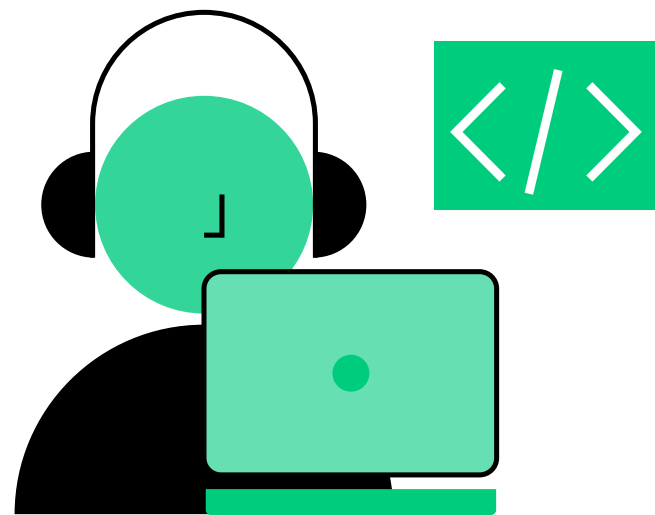


Entonces...

Un dato es una representación simbólica (numérica, alfabética, algorítmica, espacial, etc.) de un atributo o variable cuantitativa o cualitativa. Los datos describen hechos empíricos o eventos.

De manera que un dato es un valor o referencia que un software recibe por diferentes medios. Los datos representan una información que el programador manipula en la construcción de una solución o en el desarrollo de un algoritmo.

#Dato



Recuerden que...

Los datos por sí solos no contienen información relevante. Solamente cuando un conjunto de datos es examinado en conjunto a la luz de un abordaje, hipótesis o teoría, es ahí que cobra carácter informativo.

¿Cómo se muestran los datos?

Podemos encontrarlos como números, estadísticas o declaraciones descriptivas que informan estados de situación. A continuación, vamos a ver **tipos de datos**.

02

Datos estructurados

Fuentes relacionales y fuentes no relacionales

Para entender la aplicación y definición de cada etapa de ETL, necesitamos entender las fuentes y los tipos de datos.

Observemos a continuación la diferencia entre fuentes de datos relacionales y fuentes no relacionales.

- Las **fuentes relacionales** proveen **datos estructurados** (tablas).
- Las **fuentes no relacionales** proveen **datos no-estructurados o datos semi-estructurados**.

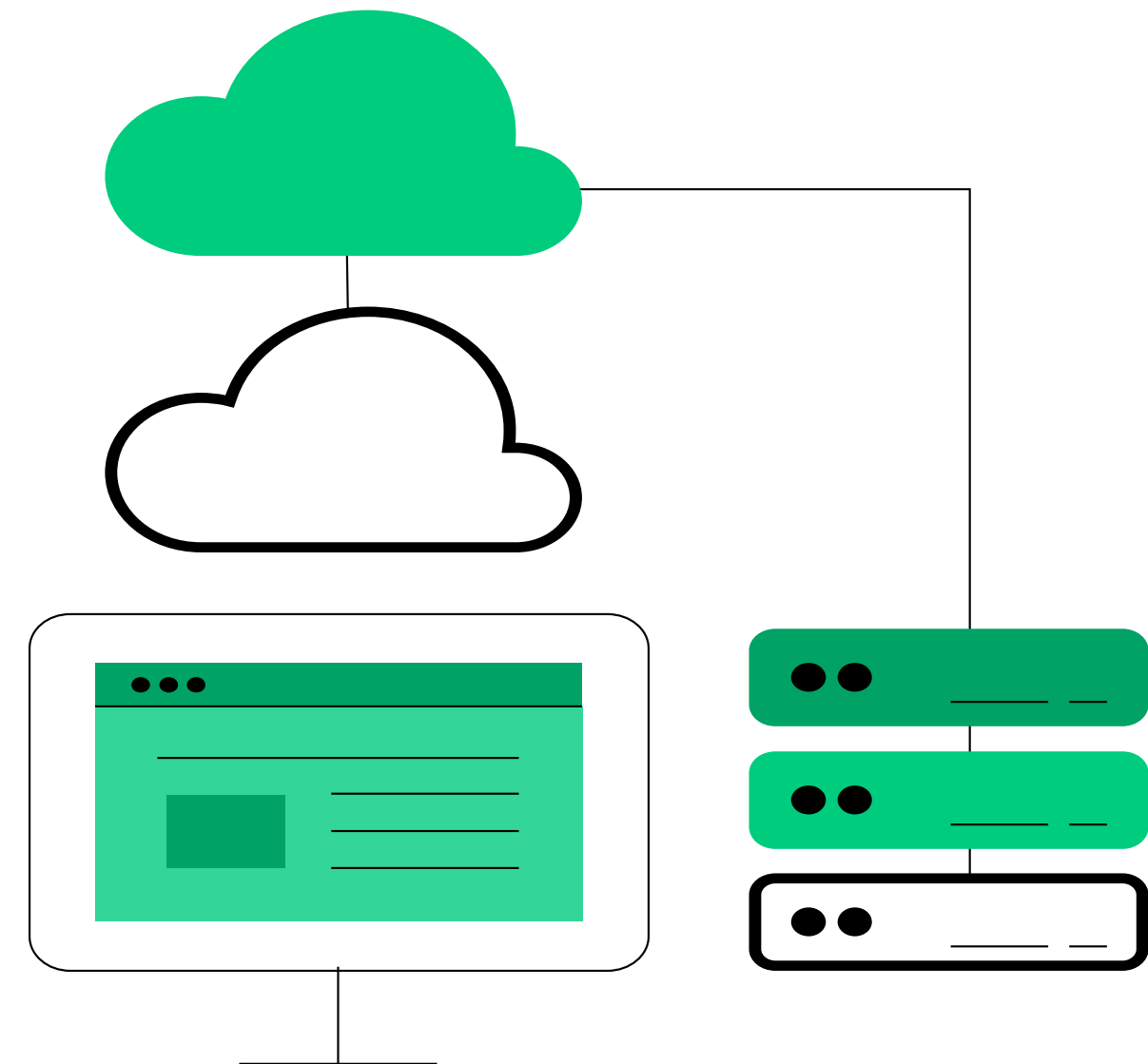
SQL

Existen varios **Sistemas de Administración de Bases de Datos Relacionales**, o en inglés, Relational Database Management System - RDBMS.

Estos sistemas permiten administrar los datos en **SQL** (Structured Query Language).

SQL es un lenguaje estructurado para consulta y manipulación en bases de datos relacionales.

En SQL, las bases de datos relacionales poseen la tarea de guardar. Es decir, mantener los datos y dejarlos en disponibilidad para otros sistemas y aplicaciones.

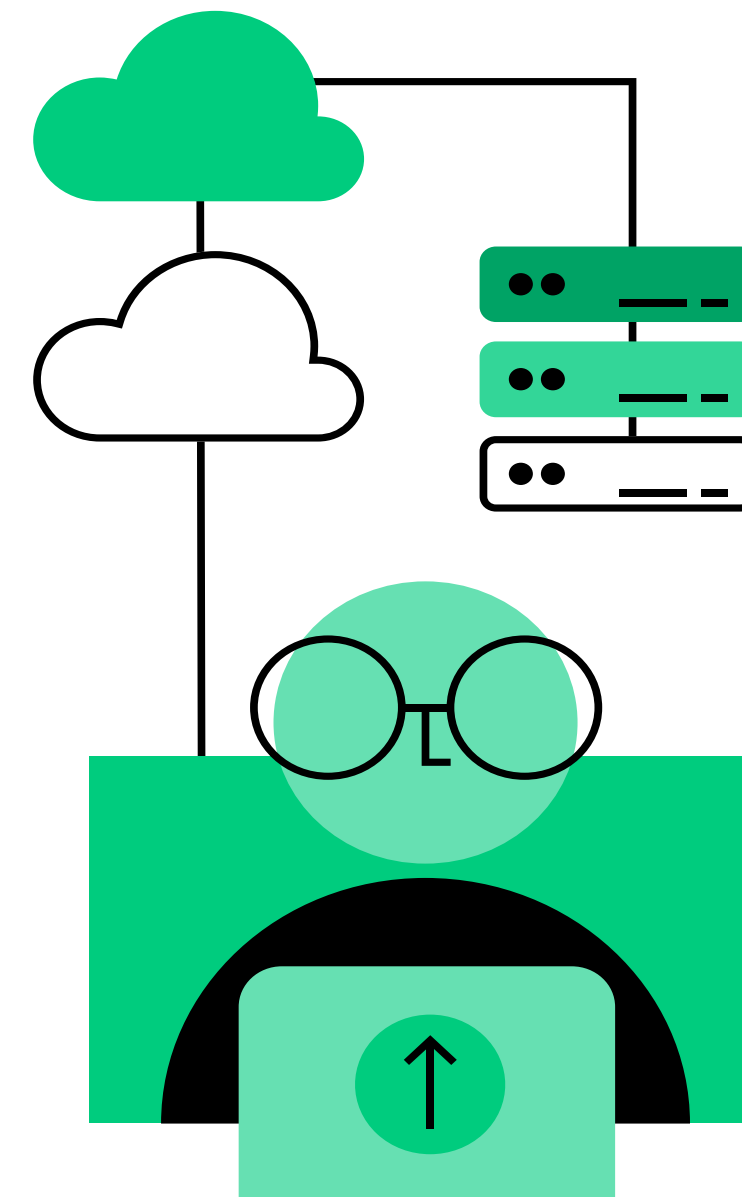


Modelado de datos + MER

Los Sistemas de administración de bases de datos más conocidos actualmente en el mercado son ORACLE, Microsoft SQL Server, MySQL y PostgreSQL.

La palabra “relacional” se debe al hecho de que el principal objetivo es relacionar los datos guardados. Esto es realizado mediante un proceso llamado modelado de datos, conjuntamente con MER, o más conocido como modelo ER (modelo de entidad y relación)

MER es la parte de las ciencias de la computación que se encarga de **construir estructuras que permitan almacenar y recuperar datos.**

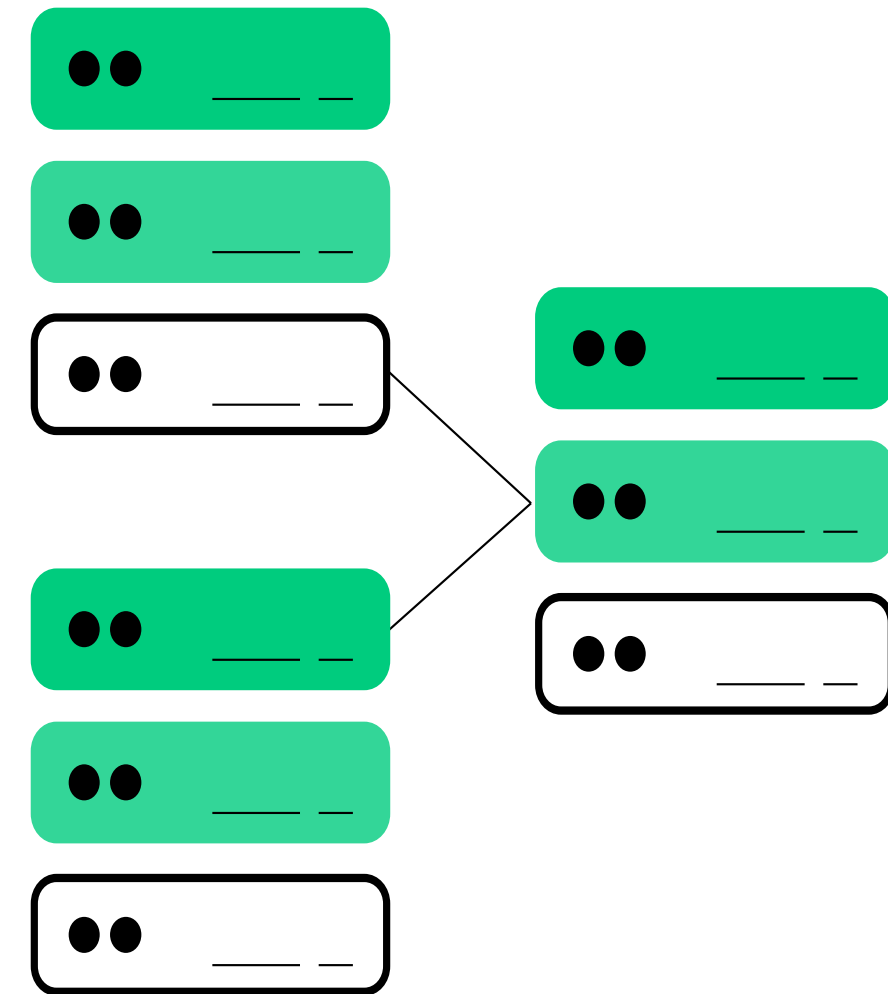


Modelado de datos + MER

Las estructuras construidas por MER guardan los datos como si fueran etiquetas, mediante los conceptos de líneas y columnas.

MER relaciona esos datos y simplifica el trabajo de almacenamiento y recuperación de los mismos.

Esta organización está hecha generalmente por líneas y columnas (parecidas a las planillas de Excel) pero puede variar de acuerdo a la fuente de datos.



Entonces, los datos estructurados son...

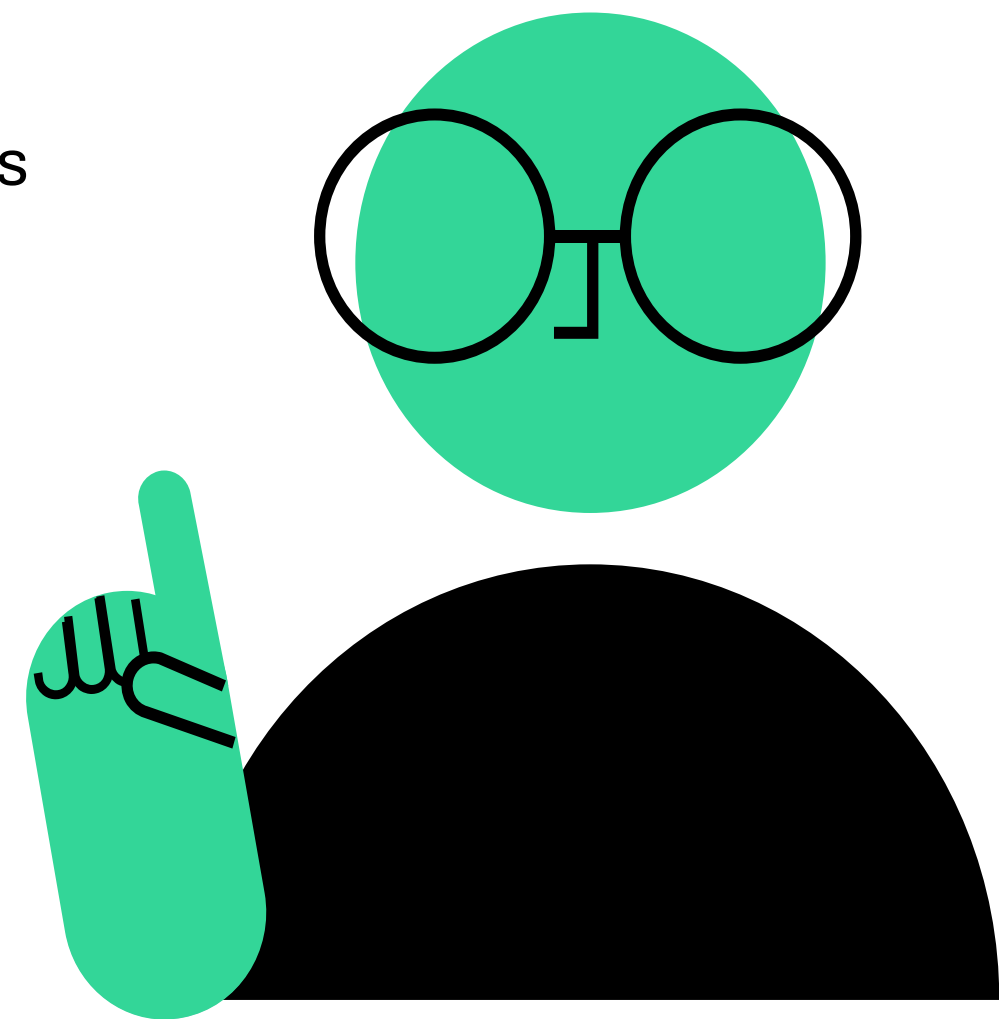
Podemos concluir que son datos agrupados, guardados y almacenados de una forma organizada para que puedan ser recuperados en algún momento.

El proceso para almacenar los datos es conocido como **persistencia**.

La unidad donde esos datos son almacenados se conoce como base de datos relacionales.

Ejemplos:

- Planillas de cálculo (Excel)
- •Bases de datos
- •Archivos XML
- •Archivos CSV



03

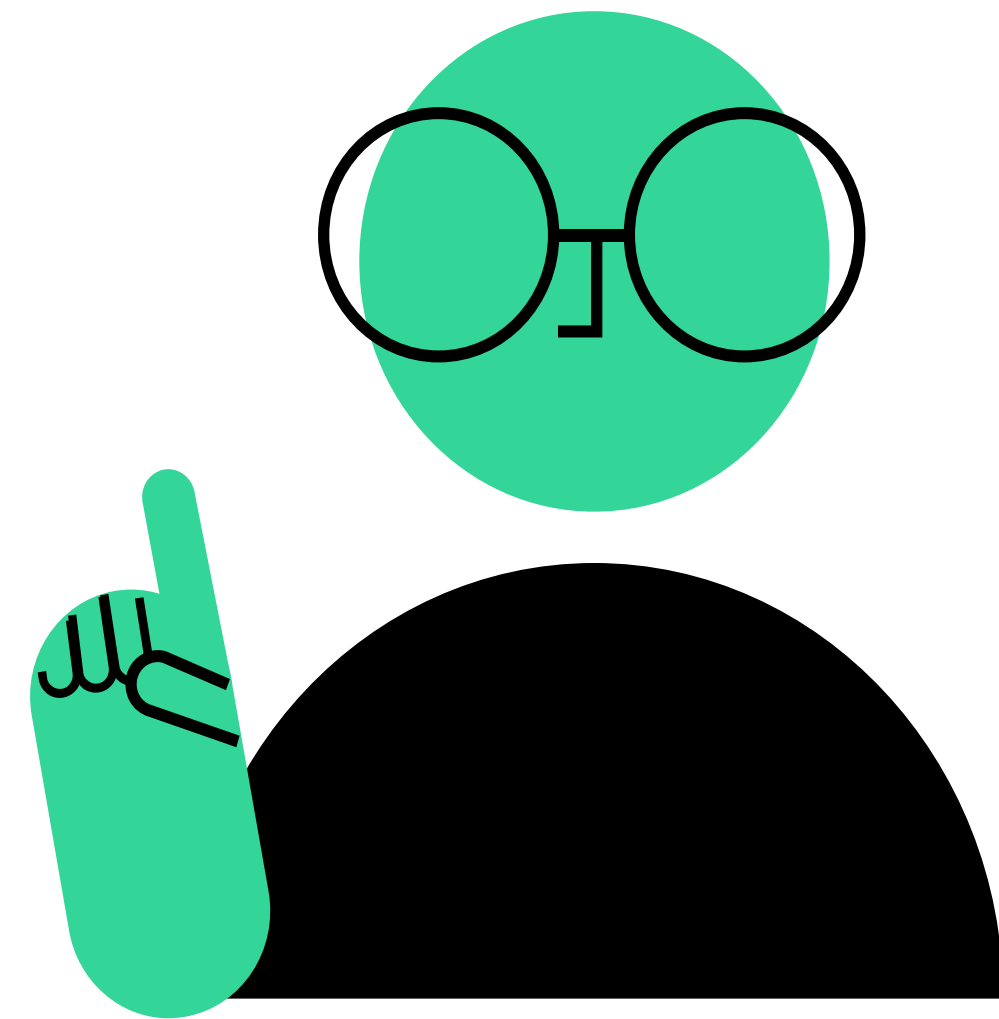
Datos no estructurados

Datos no estructurados

Las bases de datos en general no contienen toda la información posible sobre algo que tienen almacenado.

Como ya dijimos, un dato es una forma organizada de información. Pero la organización ocurre solamente si los campos específicos están llenos, para que la recuperación de datos se de en forma automatizada.

Entonces, cuando no conseguimos identificar una organización de los datos almacenados, concluimos que estos son datos no estructurados.



Entonces, datos no estructurados son...

Este tipo de datos no poseen estructuras bien definidas, alineadas o que respondan a un patrón. Pueden estar compuestos por diversos elementos dentro de un todo. Como por ejemplo, millares de pixeles únicos dentro del paisaje de una fotografía.

Ejemplos:

- ➔ Fotos
- ➔ Audios
- ➔ Vídeos
- ➔ Textos libres
- ➔ Emojis o stickers

04

Datos semi estructurados

Datos semi estructurados

Los datos semi estructurados poseen una **organización bastante heterogénea**. La distinción entre **estructura y valor es nebulosa**. Esto dificulta su consulta y clasificación.

De manera general, **los datos semi estructurados no son estrictamente clasificados**, no se mantienen en un Sistema de Administración de Base de Datos, pero sí en lo que se conoce como **Data Lake** (concepto de almacenamiento que veremos más adelante).

Ejemplos:

→ Archivos de Vídeo

→ Archivos de audio

→ Imágenes

→ Documentos de Texto

→ Información de las Redes Sociales

→ E-mails

05 Conclusiones

Conclusiones

Desde el punto de vista técnico, los datos son un conjunto de valores que surgen de variables cuantitativas y cualitativas sobre eventos, personas u objetos diversos.

Aunque los términos “datos” e “información” sean frecuentemente utilizados en forma intercambiable, estos conceptos tienen significados distintos.

