

Etapas de un proyecto de Ciencia de Datos

- Los **datos** son el registro de cosas que existen o han sucedido y tienen un significado implícito (Nombres, L.U., imágenes).
- **Base de Datos (DB)**: colección de datos ordenados y organizados.
 - ↳ Colección lógicamente coherente de datos con un significado que depende del dominio de la aplicación. (archivos en dispositivo)
 - ↳ Indep. de cómo se almacena
 - ↳ Se diseña con un propósito específico (para grupos o aplicaciones)
 - ↳ Representan algún aspecto del mundo real, enfocadas en el universo de discurso (abstracción).

Bases de datos digitales Sistema de admin. de DB (DBMS).

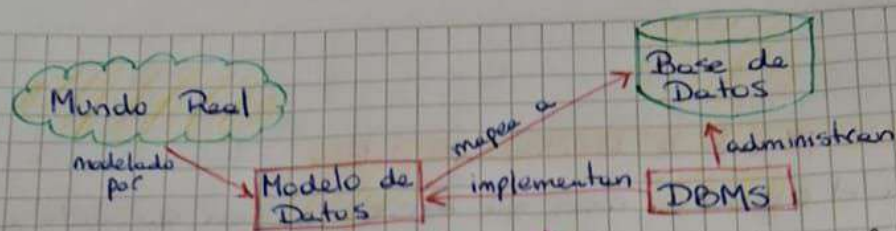
- ↳ Sistema de Software desarrollado para el manejo y administración de DB que administran datos persistentes y acceden de manera eficiente.
- ↳ Provee soporte al modelo de datos en lenguaje de alto nivel.
- ↳ Acceso correcto y concurrente (limita el acceso) y recuperación ante fallos.

- Características de un DBMS:

- Naturaleza descriptiva y autodescriptiva.
- Aislamiento entre programas y datos, abstracción de datos.
- Soporte de múltiples vistas de los datos.

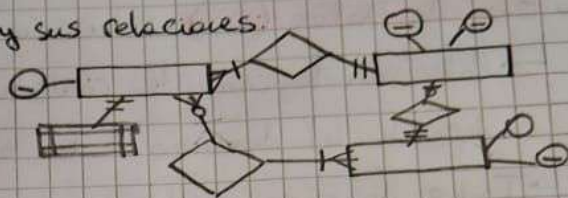
Modelo de Datos: representación abstracta que organiza elementos de datos y estandariza cómo se relacionan tanto entre sí como con las propiedades de las entidades del mundo real. (en alto nivel).

- ↳ Determina explícitamente la estructura (lógica) de los datos.
- ↳ Suelen usar notaciones en forma gráfica.



Modelo Entidad Relación (DER) herramienta gráfica que permite visualizar las entidades del modelo y sus relaciones.

→ Ayuda a explicar visualmente la estructura lógica de una BD



Etapas para la construcción de un modelo

- 1) **Identificar entidades**: algo del mundo real con existencia independiente.
 - Parte del negocio que necesita representarse en la BD (evento, persona, concepto).
 - Un conjunto de entidades forman una clase (personas).

•) En una universidad, los estudiantes se inscriben en materias.
Un estudiante debe estar inscripto en al menos una materia. Cada materia es dictada por un único profesor y, por cuestiones de carga horaria, cada profesor dicta una sola materia.

- 2) **Identificar atributos**: todas las características que se comporten entre elementos de una entidad (podrían diferir los valores).

•) Los estudiantes poseen un nombre y un número único de identificación (LU), dirección de mail y fecha de nacimiento.

• Los atributos pueden ser simples (unidades atómicas), compuestos o derivados. Los atributos compuestos se pueden descomponer en unidades atómicas y los derivados son combinación de simples, así que no se representan. → Evitar atributos multivaluados.

- 3) **Identificar claves** atributos cuyos valores son distintos para cada individuo (unívocamente) id, código (se notan subrayados).

4) Identificar las relaciones (los rombos)

- Estudiante se inscribe a materia, profesor dicta materia.

5) Identificar cardinalidades

- Relaciones uno a uno / uno a muchos / muchos a muchos.
- El tipo de participación de las entidades puede ser parcial (0) o total (1).

$\text{---} \text{---} \text{---}$

uno parcial

$\text{---} \text{---} \text{---}$

uno total

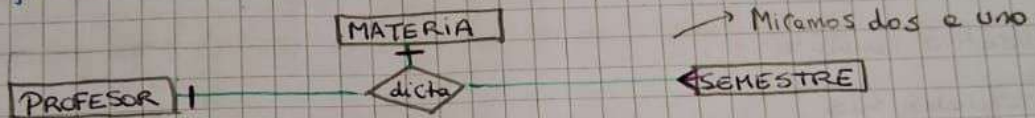
$\text{---} \text{---} \text{---}$

muchos parcial

$\text{---} \text{---} \text{---}$

muchos total.

6) Identificar relaciones no binarias (puede haber loops o ternarias).



Cada estudiante debe tener un único tutor (un estudiante). Un estudiante puede asesorar a varios estudiantes (o a ninguno).



7) Identificar entidades débiles no existen por sí mismas y su clave está construida total o parcialmente por la clave de la entidad que dependen.

