



# Datos 2.0

## ▼ Tareas

▼ 28/07

La carta de un rector de colegio a padres de familia que se hizo viral  
'Dígale que las notas que obtenga no son tan importantes', fueron algunas de sus palabras.

**ET** <https://www.eltiempo.com/cultura/gente/la-carta-de-un-director-de-colegio-a-los-padres-que-se-hizo-viral-580684>



## ETL:

Extracción, Transformación y Carga

Es un proceso de integración de datos que extrae datos de diversas fuentes, los transforma en un formato adecuado para el análisis y los carga en un sistema de destino.

1. **Extracción:** Los datos se obtienen de diferentes fuentes, como bases de datos, archivos planos, sistemas en la nube o aplicaciones web.
2. **Transformación:** Los datos extraídos suelen estar en diferentes formatos y estructuras, por lo que es necesario transformarlos para que sean consistentes y compatibles con el sistema destino.

Puede incluir limpieza de datos, filtrado, ordenación, agregación, validación y conversión de formatos.

3. **Carga:** Una vez transformados los datos, se cargan en el sistema de destino que puede ser un almacén de datos, un lago de datos o una base de datos.

## Metodologías Ágiles:

Son enfoques iterativos e incrementales para la gestión de proyectos, especialmente populares en el desarrollo de software, que priorizan la flexibilidad, la colaboración y la entrega continua de valor.

Se adaptan a los cambios y a las necesidades del cliente a lo largo del proyecto, mediante iteraciones cortas y regulares llamadas sprints.

### Tipos:

- Scrum: Basada en la creación de equipos autodesignados que trabajan en sprints para entregar incrementos de producto.
- Kanban: Utiliza un sistema visual de tarjetas para gestionar el flujo de trabajo y optimiza la eficiencia.
- XP (Extreme Programming): Enfocada en la programación en pareja, las pruebas automatizadas y la entrega continua.

### ▼ 30/07

Base de Datos de excel, más de 1.000 datos. (csv)

Proyecto Oficial.xlsx

### ▼ 1/08

Identificar qué datos tenemos en la base de datos.

¿Cómo se relacionan los datos?

Tipos de datos (Análisis primario)

¿Qué me sirve y qué no?

Definir el problema.

Datos:

- Identificación

- Identificación taxonómica
- Nombre científico
- Clasificación superior
- Reino
- Filio
- Clase
- Orden
- Familia
- Género
- Epiteto específico
- Epiteto infraespecífico
- Rango de taxón
- Nombre científico 1
- Nombre vernáculo
- Código nomenclatura
- Estado taxonómico
- Nombre original validado
- Estado de amenaza

Identifican una especie en específico, además de sus características y su estado crítico.

Fundamentalmente la base de datos está compuesta por caracteres (varchar).

Hay duplicados en los datos, además de urls que no tienen ningún propósito.

Identificar las especies con mayor riesgo de extinción.

#### ▼ 4/08

ETL base de datos

## Mapa de Stakeholders

Es una herramienta visual que representa a las personas y organizaciones afectadas por un proyecto (stakeholders), mostrando sus relaciones e intereses.

Permite identificar a los involucrados, comprender sus necesidades y definir estrategias para su participación.

## ¿Para qué sirve?

- *Identifica* a todas las partes interesadas relevantes.
- Facilita la *comprensión* de los intereses, necesidades y expectativas de cada stakeholder.
- *Planificación*: Ayuda a definir estrategias de comunicación y participación para cada stakeholder.
- Permite *gestionar* las relaciones con los stakeholders y mitigar posibles conflictos.
- Facilita la *toma de decisiones* informadas, considerando las perspectivas de todos los involucrados.

## ¿Cómo se crea?

1. *Identificación*: Enumera a todas las partes interesadas.
2. *Clasificación*: Agrupa a los stakeholders por su tipo.
3. *Análisis*: Determina el nivel de interés e influencia de cada stakeholder.
4. *Representación*: Utiliza un diagrama visual para mostrar las relaciones.
5. *Estrategia*: Define acciones para cada stakeholder.

## Ejemplo:

Usualmente se puede representar con un diagrama de dos ejes.

- *Eje Horizontal*: Representa el nivel de interés de cada stakeholder.
- *Eje Vertical*: Representa el nivel de influencia de cada stakeholder.

## Esto crea 4 cuadrantes:

- *Alta influencia, alto interés*: Stakeholders clave que deben ser gestionados cuidadosamente.
- *Alta influencia, bajo interés*: Stakeholders que deben ser mantenidos informados.
- *Baja influencia, alto interés*: Stakeholders que deben ser mantenidos informados y consultados.
- *Baja influencia, bajo interés*: Stakeholders que deben ser monitoreados.

## Beneficios:

- Reduce riesgos.
- Aumenta la satisfacción.

- Mejora la toma de decisiones.
- Optimiza la gestión del proyecto.

▼ 6/08

Pasar la base de datos por un software que realice la ETL.

```
import pandas as pd

# 1. Extract: Cargar datos
ruta_archivo = "Cobertura_m_vil_por_tecnolog_a_departamento_y_municipio_por_
proveedor_20250808.csv"
df = pd.read_csv(ruta_archivo)

# 2. Transform
# Renombrar columnas a formato snake_case y corregir error de escritura
df.columns = (
    df.columns.str.strip()
    .str.lower()
    .str.replace(" ", "_")
    .str.replace(",", "", regex=False)
    .str.replace("+", "plus", regex=False)
)

# Corregir error tipográfico en columna
df = df.rename(columns={"cobertuta_4g": "cobertura_4g"})

# Normalizar valores de cobertura: S → 1, N → 0
cols_cobertura = [
    "cobertura_2g",
    "cobertura_3g",
    "cobertura_hspaplus_hspaplusdc",
    "cobertura_4g",
    "cobertura_lte",
    "cobertura_5g"
]

for col in cols_cobertura:
    df[col] = df[col].map({"S": 1, "N": 0})

# Convertir códigos a enteros con soporte para valores nulos
```

```
df["cod_municipio"] = pd.to_numeric(df["cod_municipio"], errors="coerce").astype("Int64")
df["cod_centro_poblado"] = pd.to_numeric(df["cod_centro_poblado"], errors="coerce").astype("Int64")

# 3. Load: Guardar el CSV limpio
ruta_salida = "cobertura_movil_limpia.csv"
df.to_csv(ruta_salida, index=False)

print(f"Archivo limpio guardado en: {ruta_salida}")
```

Cobertura Móvil por Departamento.xlsx

▼ 8/08

## Técnica Camelcase

Es un estilo de escritura donde se combinan varias palabras en una sola, y la primera de cada palabra, excepto la primera se escribe en mayúscula.

Se llama así porque las letras mayúsculas en medio de la palabra se asemejan a las jorobas de un camello.

### Usos:

- Programación: Nombres de variables, funciones y métodos para mejorar la legibilidad.
- Nombres de dominio y URLs: Mejorar la apariencia y facilitar su lectura.

### Tipos:

- `lowerCamelCase`: La primera letra es minúscula. Ej: `miVariable` .
- `UpperCamelCase`: La primera letra es mayúscula. Ej: `MiVariable` .

### Ventajas:

- Facilita la lectura y comprensión de nombres compuestos sin espacios.
- Permite nombres más cortos que si se usaran espacios.

### Desventajas:

- Para quienes no están familiarizados con este estilo, puede resultar confuso al principio.

▼ 13/08

Métodos, técnicas e instrumentos del caso de estudio para resolver la necesidad:

- Método discreto
- Encuesta
- Cuestionario

¿Qué se requiere hacer para mejorar la resolución de dicho problema?

▼ 15/08

2 Redes Sociales: Definir 5 Requisitos funcionales y 3 Requisitos No funcionales.

- Instagram:
  - Requisitos Funcionales:
    - Crear una cuenta.
    - Crear una publicación (post, reels).
    - Grabar videos, tomar fotos y realizar en vivos.
    - Dar me gusta a una publicación y compartirla.
    - Editar su información y foto de perfil.
  - Requisitos No Funcionales:
    - Verificar una cuenta.
    - Restringir el uso inapropiado de una cuenta.
    - Mantener un tiempo de respuesta rápido en la carga de publicaciones.
- TikTok:
  - Requisitos Funcionales:
    - Crear una cuenta.
    - Dar me gusta a una publicación y compartirla.
    - Editar su información y foto de perfil.
    - Crear una publicación (video, historia).
    - Seguir ciertas cuentas de acuerdo a los gustos del usuario.
  - Requisitos No Funcionales:

- Verificar una cuenta.
- Restringir el uso inapropiado de una cuenta.
- Mantener un tiempo de carga óptimo para la reproducción de videos.

▼ 22/08

## Módulos de Green Bite

- Gestión de Clientes y Pedidos
  - Registro de pedidos.
  - Datos del cliente.
  - Medio de pago.
- Gestión de Productos
  - Tacos personalizados.
  - Pastor de carne marinero.
- Gestión de Domiciliarios
  - Asignación de pedidos.
  - Control de tiempos de entrega.
- Gestión de Comunicación
  - WhatsApp (pedidos y coordinación de domiciliarios).

## Clases (atributos y métodos)

- Cliente:
  - Atributos:
    - id Cliente
    - Nombre
    - Celular
    - Dirección
  - Métodos:
    - realizarPedido()
    - pagarPedido()
- Producto:

- Atributos:
  - Id Producto
  - Nombre
  - Tipo
  - Precio
- Métodos:
  - obtenerPrecio()
  - obtenerTipo()
- Pedido:
  - Atributos:
    - id Pedido
    - Fecha y hora.
    - Estado
    - Total
  - Métodos:
    - actualizarEstado()
    - calcularTotal()
- Domiciliario:
  - Atributos:
    - id Domiciliario
    - Nombre
    - Celular
    - Disponibilidad
  - Métodos:
    - asignarPedido()
    - confirmarEntrega()
- Pago:
  - Atributos:
    - id Pago

- Número de pago
- Tipo
- Monto
- Métodos:
  - procesarPago()

## Tipos de relaciones de clases

- Un cliente puede realizar muchos pedidos (1:M)
  - Un pedido puede ser realizado por un cliente (1:1).
- Un pedido puede tener muchos productos (1:M).
  - Muchos productos pueden tener muchos pedidos (M:M).
- Un pedido tiene un solo pago (1:1).
- Un pedido puede ser entregado por un domiciliario (1:1).
  - Muchos pedidos pueden ser entregados por un domiciliario (M:M).
  - Muchos domiciliarios pueden entregar muchos pedidos (M:M).

## Crear el diagrama de clases

```

Cliente -----1..* Pedido
+ nombre          + idPedido
+ teléfono       + fechaHora
+ dirección      + estado
+ realizarPedido() + total
                  + calcularTotal()
                  + actualizarEstado()

Pedido -----*..* Producto
Pedido -----1 Pago
Pedido -----1 Domiciliario

Producto
+ idProducto
+ nombre
+ precio
+ obtenerPrecio()

```

### Domiciliario

- + idDomiciliario
- + nombre
- + celular
- + disponibilidad
- + asignarPedido()
- + confirmarEntrega()

### Pago

- + idPago
- + tipo
- + monto
- + procesarPago()

▼ 5/09

Atributos y tipología.

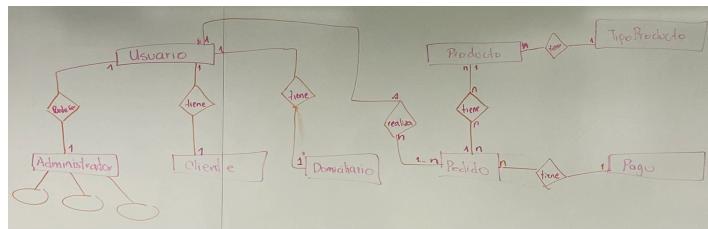
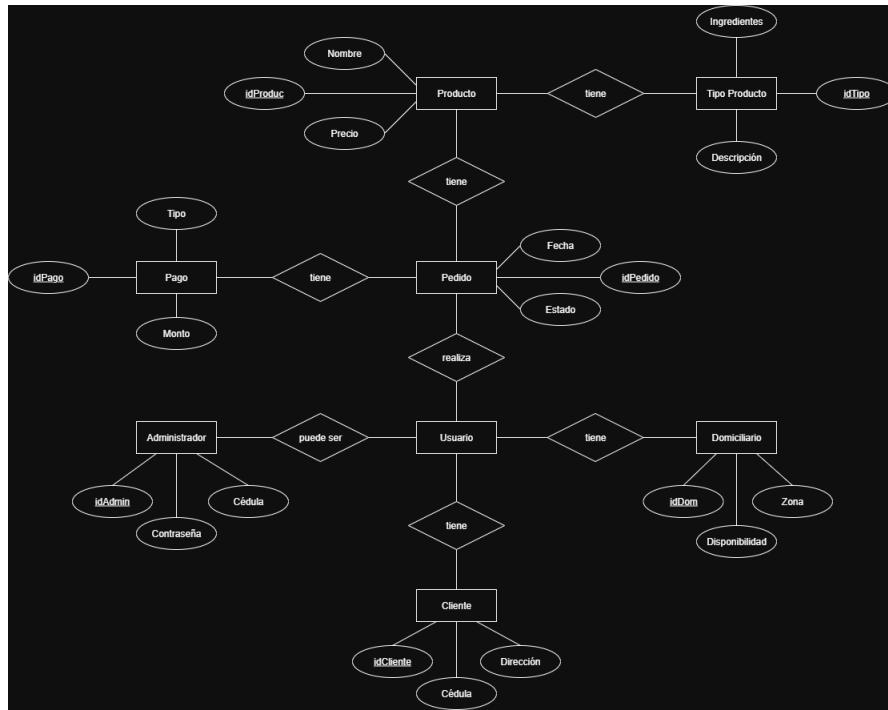
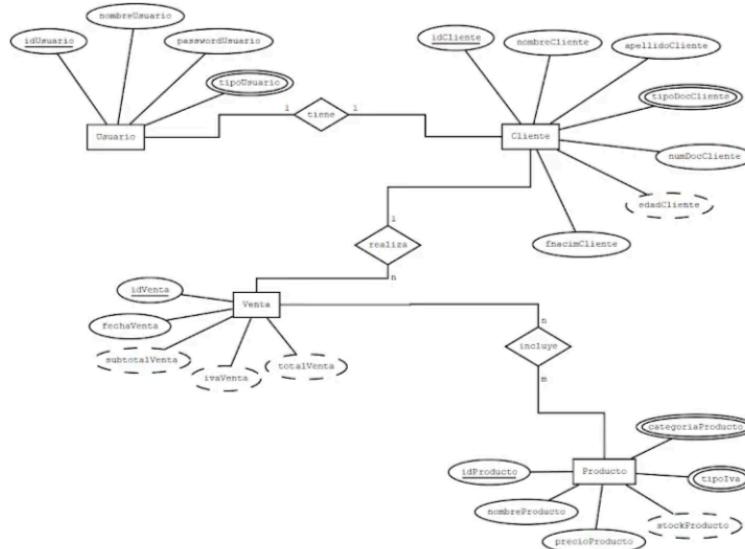


Diagrama de Clases y Modelo Conceptual del proyecto (a mano).

MER Green Bite:



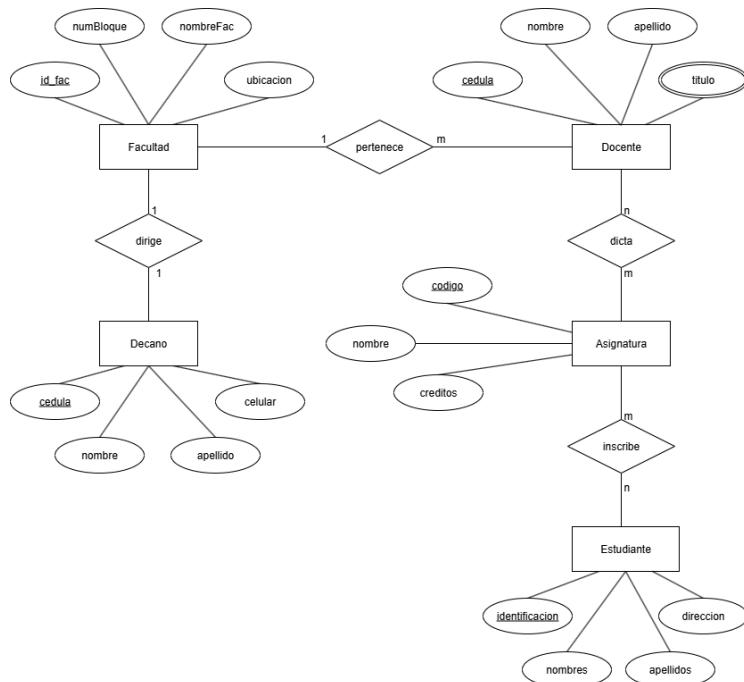
## Taller



1. ¿De qué tipo es cada uno de los atributos de la entidad Cliente?

- idCliente →
- nombreCliente → Simple
- apellidoCliente → Simple

- tipoDocCliente → Multivalorado
  - numDocCliente → Simple
  - edadCliente → Derivado
  - fnacimCliente → Simple
2. ¿Qué tipo de relación existe entre las entidades **Cliente - Venta** y entre las entidades **Venta - Producto**?
- Cliente - Venta:
    - Un cliente puede realizar muchas ventas.
    - Muchas ventas pueden ser realizadas pueden ser realizadas por un cliente.
  - Venta - Producto:
    - Muchas ventas pueden incluir muchos productos.
    - Muchos productos pueden estar incluidos muchas ventas.
3. Ahora vamos a elaborar un **Modelo Entidad Relación MER** desde ceros, preste atención al ejemplo:
- En una universidad las Facultades son dirigidas por un decano y a su vez, un decano dirige una facultad.
  - Cada facultad cuenta con una serie de docentes, pero cada docente solo puede pertenecer a una facultad.
  - Cada docente dicta varias asignaturas, así mismo una misma asignatura la pueden dictar docentes diferentes.
  - Los estudiantes inscriben las asignaturas que les corresponde cada semestre, en una materia puede haber varios estudiantes inscritos.
  - **De las facultades** interesa saber cuál es su nombre, ubicación y número de bloque.
  - **De los decanos** se requiere la cédula, nombres, apellidos y celular.
  - **De los docentes** se debe registrar la cédula, nombres y apellidos y título.
  - **Cada asignatura** tiene un código, nombre y número de créditos.
  - **De los estudiantes** se debe registrar identificación, nombres, apellidos y dirección de residencia.



▼ 8/09

Diagrama de Clases y Modelo Conceptual del proyecto (a mano).

### ¿Qué es un diccionario de datos?

Es un conjunto centralizado que contiene la descripción y características de los elementos de datos de una base de datos, funcionando como una guía para entender su estructura, significado y uso.

Incluye información técnica, lo cual ayuda a asegurar la coherencia, facilitar la interpretación y estandarizar el uso de los datos entre los diferentes usuarios de un sistema.

### ¿Qué información suele contener?

- *Nombres de los elementos de datos*: Cómo se llaman los campos o atributos.
- *Definiciones y descripciones*: El propósito de cada dato y su contexto.
- *Tipos de datos*: El tipo de información que contiene.
- *Longitud y formato*: El tamaño máximo de un dato y su formato específico.
- *Valores permitidos*: Los posibles valores que puede tomar un campo, incluyendo valores por defecto.
- *Restricciones y reglas*: Las validaciones que deben cumplir los datos para mantener la integridad de la información.

- *Metadatos*: Información adicional sobre los datos, como cuándo se actualizaron por última vez o quién es el responsable.

## ¿Por qué se llama modelo lógico?

Porque define los datos y las relaciones entre ellos de una manera abstracta y conceptual, independientemente de cómo se almacenarán físicamente o qué Sistema Gestor de Base de Datos (SGBD) se utilizará. Esta separación entre la visión lógica (cómo se ve y se entiende la información) y la visión física (cómo se implementa en el sistema) permite a los usuarios interactuar con los datos sin necesidad de conocer los detalles de la infraestructura de almacenamiento, lo que aporta flexibilidad e independencia.

### Naturaleza lógica del modelo:

- *Independencia tecnológica*: El MR está diseñado para ser independiente del hardware y del software específicos. Esto significa que las estructuras de las tablas, las relaciones y los atributos se definen sin depender de un SGBD particular.
- *Abstracción*: Representa los datos como un conjunto de "relaciones" o tablas, donde cada fila es un registro y cada columna es un atributo. Esta es una forma conceptual de organizar la información, que es fácil de entender y manipular, independientemente del mecanismo de almacenamiento subyacente.
- *Base matemática*: Se fundamenta en la lógica de predicados y la teoría de conjuntos, y utiliza conceptos como álgebra relacional para la manipulación de datos.

## Restricciones en bases de datos

Son reglas que aseguran la integridad, validez y consistencia de los datos almacenados, impidiendo la entrada de información incorrecta, incompleta o inconsistente.

### Tipos de restricciones:

- *Clave Primaria*: Garantiza que cada fila de una tabla sea identificada de forma única.
- *Clave Foránea*: Mantiene la integridad referencial entre tablas, asegurando que los valores de la tabla dependiente existan en la tabla principal.
- *Única*: Asegura que los valores en una columna(s) sean únicos, evitando duplicados.
- *No Nula*: Previene la inserción de valores NULL en una columna específica, requiriendo que siempre haya un valor presente.

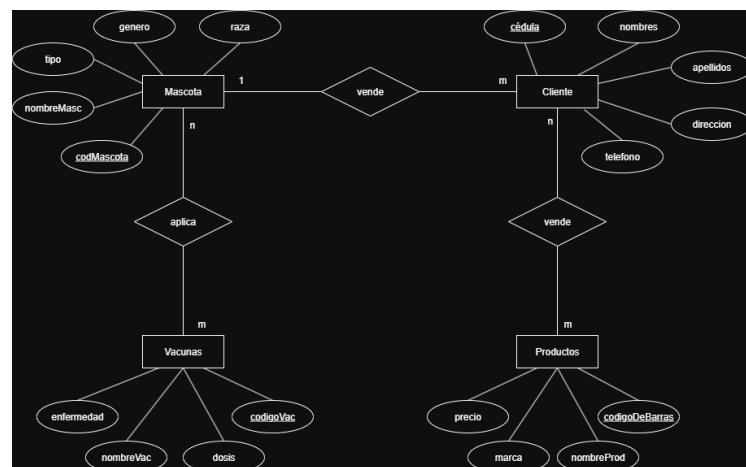
- *Comprobación*: Impone condiciones específicas sobre los datos que se pueden ingresar en una columna o tabla.
- *Dominio*: Definen el tipo de datos, rango y formato permitidos para un atributo específico, asegurando la precisión de los datos.

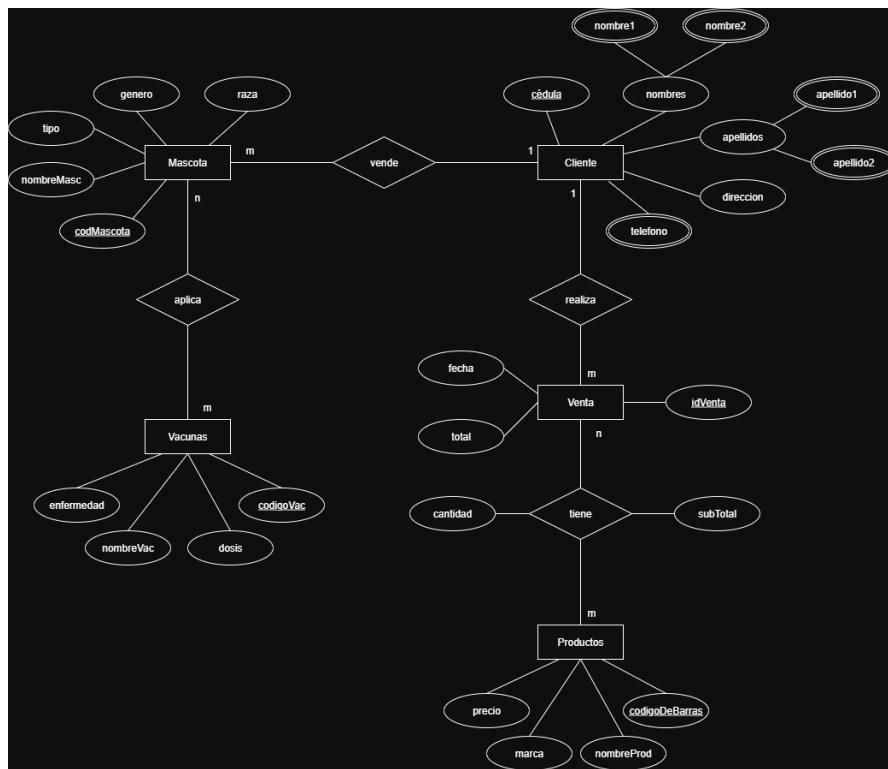
**Importancia:**

- Garantizan la calidad de los datos.
- Aseguran las reglas de negocio.
- Automatizan la validación.
- Mantienen la consistencia relacional.

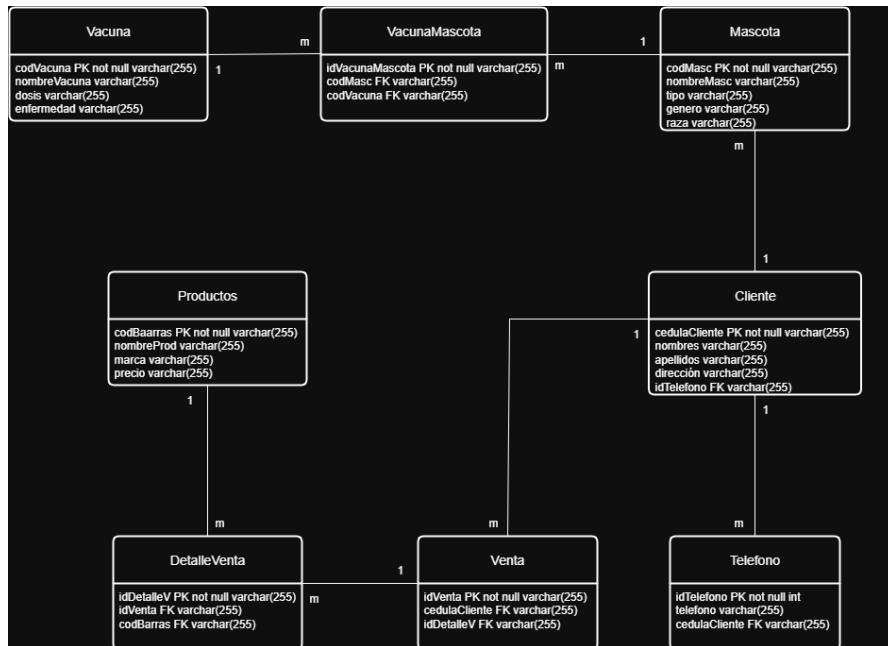
▼ 10/09

MER Ejercicio Mascotas





▼ 12/09



### Diccionario de datos de mascotas:

- Nombre Tabla: Vacuna.
- Descripción: Tabla que contiene los campos de la Vacuna.

Campo	Tipo de Dato	Tamaño	Restricción	Descripción
codVacuna	varchar	255	Primary Key Not null	Llave primaria de la vacuna.
nombreVacuna	varchar	255		Nombre de la vacuna.
dosis	varchar	255		Dosis de la vacuna.
enfermedad	varchar	255		Nombre de la enfermedad.

- Nombre Tabla: Mascota.
- Descripción: Tabla que contiene los campos de las Mascotas.

Campo	Tipo de Dato	Tamaño	Restricción	Descripción
codMasc	varchar	255	Primary Key Not null	Llave primaria de la mascota.
nombreMasc	varchar	255		Nombre de la mascota.
tipo	varchar	255		Tipo de mascota.
genero	varchar	255		Género de la mascota.
raza	varchar	255		Raza de la mascota.

- Nombre Tabla: Cliente.
- Descripción: Tabla que contiene los campos de los Clientes.

Campo	Tipo de Dato	Tamaño	Restricción	Descripción
cedulaCliente	varchar	255	Primary key Not null	Llave primaria del cliente.
nombres	varchar	255		Nombres del cliente.
apellidos	varchar	255		Apellidos del cliente.
dirección	varchar	255		Dirección del cliente.
idTelefono	varchar	255	Foreign key	Llave foránea del teléfono del cliente.

- Nombre Tabla: Productos.
- Descripción: Tabla que contiene los campos de los Productos.

Campo	Tipo de Dato	Tamaño	Restricción	Descripción
codBarras	varchar	255	Primary key Not null	Llave primaria del producto.
nombreProd	varchar	255		Nombre del producto.
marca	varchar	255		Marca del producto.
precio	varchar	255		Precio del producto.

Crear tablas de mascotas en base de datos.

```
CREATE TABLE Vacuna(
    codVacuna PRIMARY KEY NOT NULL VARCHAR(255),
    nombreVacuna VARCHAR(255),
    dosis VARCHAR(255),
    enfermedad VARCHAR(255)
);
```

```
CREATE TABLE Mascota(
    codMasc PRIMARY KEY NOT NULL VARCHAR(255),
    nombreMasc VARCHAR(255),
    tipo VARCHAR(255),
    genero VARCHAR(255),
    raza VARCHAR(255)
);
```

```
CREATE TABLE Cliente(
    cedulaCliente PRIMARY KEY NOT NULL VARCHAR(255),
    nombres VARCHAR(255),
    apellidos VARCHAR(255),
    idTelefono FOREIGN KEY VARCHAR(255)
);
```

```
CREATE TABLE Productos(
    codBarras PRIMARY KEY NOT NULL VARCHAR(255),
    nombreProd VARCHAR(255),
    marca VARCHAR(255),
    precio VARCHAR(255)
);
```

▼ 15/09

Terminar creación de las tablas para mascotas.

▼ 24/0

## Funciones calculadas en workbench:

Son expresiones que generan un nuevo valor utilizando datos existentes, en lugar de almacenar directamente el valor en la columna.

Se puede crear columnas calculadas que realicen operaciones matemáticas, lógica, de texto o de fecha/hora.

### Tipos:

- Matemáticas: Realizan operaciones aritméticas o trigonométricas.
- Lógicas: Permiten evaluar condiciones y devolver un resultado booleano o condicional.
- De texto: Manipulan cadenas de texto.
- De fecha y hora: Trabajan con fechas y horarios.
- De agregación: Calculan un valor único a partir de un conjunto de valores.

## ▼ Retos

▼ 03/10

The screenshot shows the MySQL Workbench interface. The top bar has tabs for 'BDMascotas', 'ventasTienda', 'BD Mascotas Actualizados', and 'Ejercicios Consultas'. The main area contains a code editor with the following SQL script:

```
286     Left Join
287     Right Join
288     Full Join - JOIN UNION, JOIN
289 */
290
291 /* Modificaciones update sintaxis:
292 UPDATE nombreTabla SET campovalor, campo1 = valor1 WHERE condicion */
293 •   SELECT * FROM Mascota;
294 •   UPDATE Mascota SET nombreMascota = "Milu" WHERE idMascota = 7;
295 •   UPDATE Mascota SET nombreMascota = "Maximo" WHERE idMascota = 9;
296 •   UPDATE Mascota SET nombreMascota = "Melody" WHERE idMascota = 3;
```

Below the code editor is a 'Result Grid' showing the data for the 'Mascota' table:

	idMascota	nombreMascota	raza	generMascota	tipoMascota	fechaCreacion	documentoClienteFK
▶	1	Firulais	Labrador	Macho	Perro	2025-09-24 07:35:55	C001
	2	Firulais	Labrador	Macho	Perro	2025-09-24 07:35:55	C001
	3	Melody	Siames	Hembra	Gato	2025-09-24 07:35:55	C002
	4	Misu	Siames	Hembra	Gato	2025-09-24 07:35:55	C002
	5	Rocky	Bulldog	Macho	Perro	2025-09-24 07:35:55	C003
	6	Rocky	Bulldog	Macho	Perro	2025-09-24 07:35:55	C003
	7	Milu	Persa	Hembra	Gato	2025-09-24 07:36:00	C004
	8	Luna	Persa	Hembra	Gato	2025-09-24 07:36:00	C004
	9	Maximo	Golden Retriever	Macho	Perro	2025-09-24 07:36:02	C005
	10	Max	Golden Retriever	Macho	Perro	2025-09-24 07:36:03	C005
	11	Max	Golden Retriever	Macho	Perro	2025-09-24 07:36:03	C005

```
/* Modificaciones update sintaxis:
UPDATE nombreTabla SET campovalor, campo1 = valor1 WHERE condicion */
```