

**ACTIVIDAD BASE DE DATOS**

**JEISON LEONARDO CRISTANCHO GARCÍA**

**J3**

**PEDRO FELIPE GÓMEZ BONILLA**

**CAMPUSLANDS**

**APOLO**

**INTRODUCCION AL BACKEND**

**FLORIDABLANCA**

**2025**

## **Tabla de Contenidos**

|  |    |
|--|----|
| Introducción .....                       | 3  |
| Caso de Estudio .....                    | 3  |
| Planificación .....                      | 3  |
| Construcción del Modelo Conceptual ..... | 3  |
| Construcción del Modelo Lógico .....     | 5  |
| Normalización del Modelo Lógico .....    | 6  |
| Construcción del Modelo Físico .....     | 7  |
| Diagrama E-R .....                       | 8  |
| Tablas y Relaciones .....                | 9  |
| Inserción de Datos .....                 | 9  |
| Referencias .....                        | 10 |

## Introducción

Este informe presenta el diseño completo de una base de datos orientada al registro y análisis de las muertes accidentales en Colombia durante el periodo 2015–2024. El trabajo se estructura en tres modelos de bases de datos; conceptual, lógico y físico, además, el proceso de normalización hasta la tercera forma normal (1FN, 2FN y 3FN). El objetivo es transformar un archivo de tablas de contenido en un sistema relacional claro y mantenible, garantizando integridad, coherencia y facilidad para la generación de consultas analíticas y estadísticas.

## Caso de Estudio

El conjunto de datos de referencia corresponde a registros de muertes accidentales en Colombia, con variables como: año del evento, departamento, causa, sexo y grupo de edad. El conjunto de datos original del DANE está en formato CSV y presenta redundancias y combinaciones de texto en celdas, lo cual dificulta la manipulación eficiente de la información. Por esta razón, se plantea un modelo relacional donde cada entidad y relación quede normalizada, evitando duplicidad y mejorando la consistencia de los datos.

## Planificación

Las etapas del desarrollo se definieron así: (a) análisis de la base de datos primarios y su estructura, (b) diseño del modelo conceptual para identificar entidades y relaciones, (c) construcción del modelo lógico con definición de claves primarias y foráneas, (d) normalización progresiva hasta la tercera forma normal con ejemplos prácticos aplicados al modelo, y (e) construcción del modelo físico mediante SQL estándar.

## Construcción del Modelo Conceptual

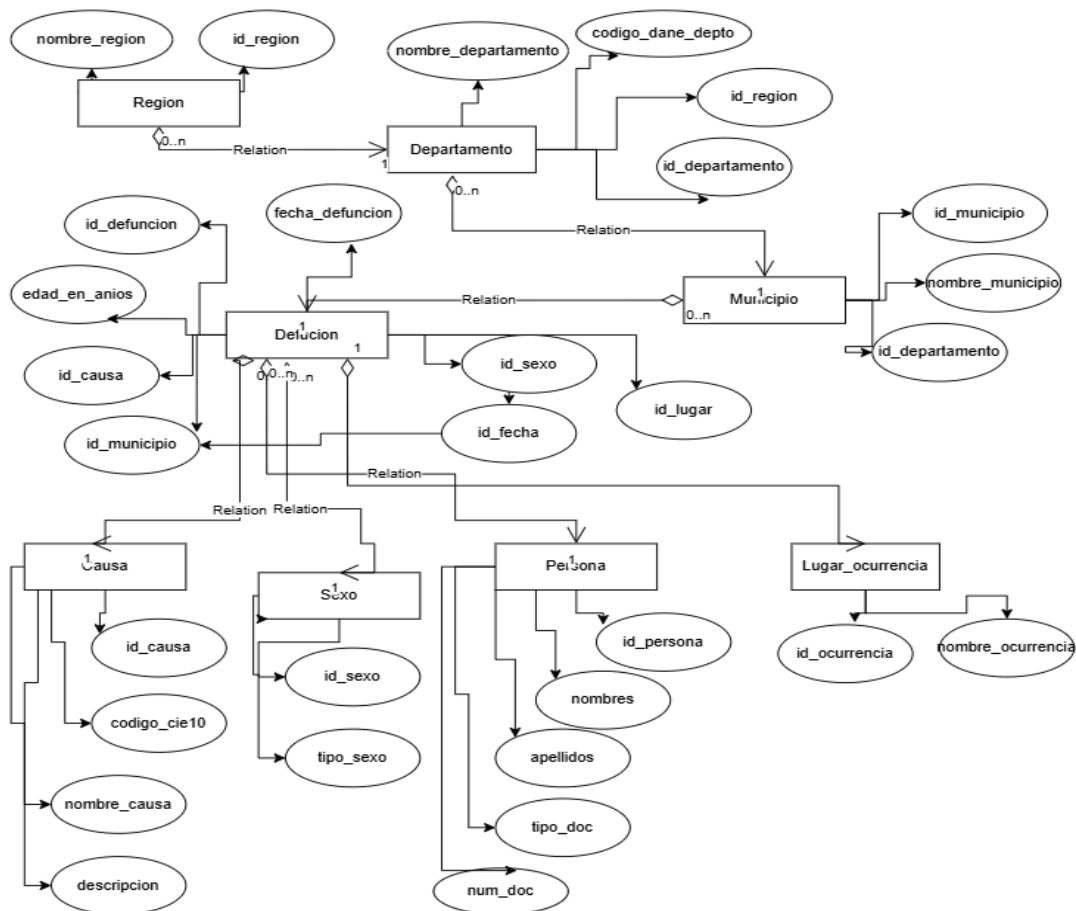
El modelo conceptual define las entidades fundamentales y sus relaciones. Cada entidad incluye atributos que representan las características clave de los datos. Este nivel de diseño se centra en el significado y estructura lógica de la información sin especificar implementaciones técnicas.

- REGIÓN (id\_region, nombre\_region) → Agrupa los departamentos del país.
- DEPARTAMENTO (id\_departamento, nombre\_departamento, id\_region) → Unidad territorial administrativa.

- CAUSA (id\_causa, nombre\_causa, descripcion\_causa) → Motivo de la muerte accidental.
- SEXO (id\_sexo, tipo\_sexo) → Clasificación biológica o administrativa.
- GRUPO\_EDAD (id\_grupo\_edad, descripcion\_grupo\_edad, edad\_min, edad\_max) → Rango etario estandarizado.
- AÑO (id\_año, año) → Periodo temporal de referencia.
- MUERTE (id\_muerte, cantidad, id\_causa, id\_departamento, id\_sexo, id\_grupo\_edad, id\_año) → Hecho principal del modelo.

Relaciones: REGIÓN 1—N DEPARTAMENTO; DEPARTAMENTO 1—N MUERTE; CAUSA 1—N MUERTE; SEXO 1—N MUERTE; GRUPO\_EDAD 1—N MUERTE; AÑO 1—N MUERTE.

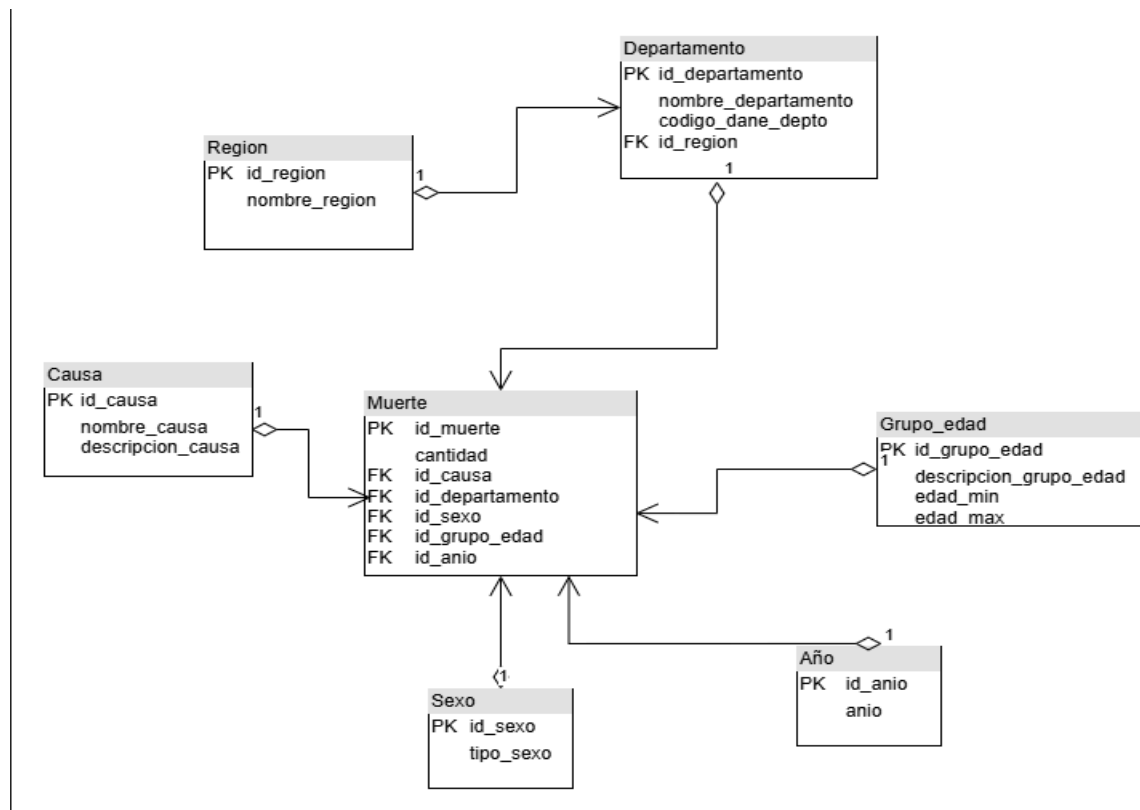
La entidad central ,muerte representa los registros agregados de muertes accidentales para una combinación específica de causa, sexo, grupo etario, año y departamento. La jerarquía REGIÓN → DEPARTAMENTO → MUERTE permite realizar análisis espaciales y demográficos, manteniendo independencia entre dimensiones.



## Construcción del Modelo Lógico

- REGION(id\_region PK, nombre\_region)
- DEPARTAMENTO(id\_departamento PK, nombre\_departamento, codigo\_dane\_depto, id\_region FK→REGIÓN)
- CAUSA(id\_causa PK, nombre\_causa, descripcion\_causa)
- SEXO(id\_sexo PK, tipo\_sexo)
- GRUPO\_EDAD(id\_grupo\_edad PK, descripcion\_grupo\_edad, edad\_min, edad\_max)
- ANIO(id\_año PK, año)
- MUERTES(id\_muerte PK, id\_causa FK, id\_departamento FK, id\_sexo FK, id\_grupo\_edad FK, id\_año FK, cantidad)

El modelo lógico define las claves primarias y foráneas, cumpliendo las tres primeras formas normales. Las relaciones 1:N se aplican entre las dimensiones y la tabla central de hechos (MUERTES).



## Entidades principales y relaciones

### Región

Atributos: *id\_region* (PK), *nombre\_region*. Define las grandes divisiones geográficas del país (por ejemplo: Caribe, Andina, Pacífica).

Se relaciona de forma 1:N con la tabla Departamento, ya que una región contiene varios departamentos.

### Departamento

Atributos: *id\_departamento* (PK), *nombre\_departamento*, *codigo\_dane\_depto*, *id\_region* (FK). Representa la unidad administrativa donde ocurrió el evento.

Su clave foránea *id\_region* lo vincula jerárquicamente con la entidad Región.

La relación es N:1 hacia región y 1:N hacia la tabla Muerte.

### Causa

Atributos: *id\_causa* (PK), *nombre\_causa*, *descripcion\_causa*. Contiene los tipos de accidente o motivo de la muerte (por ejemplo: caída, ahogamiento, intoxicación). Se relaciona 1:N con Muerte, ya que una causa puede aparecer en muchos registros de muertes.

### Sexo

Atributos: *id\_sexo* (PK), *tipo\_sexo*. Clasifica el sexo biológico de las víctimas (Masculino, Femenino, Noespecificado). Relación 1:N con Muerte.

### Grupo\_edad

Atributos: *id\_grupo\_edad* (PK), *descripcion\_grupo\_edad*, *edad\_min*, *edad\_max*. Define los rangos etarios (por ejemplo, 0–4, 5–9, 10–14 años). Relación 1:N con Muerte, ya que cada grupo agrupa muchos registros.

### Año

Atributos: *id\_anio* (PK), *anio*. Registra el periodo temporal de la observación.

Relación 1:N con Muerte, ya que cada año puede tener múltiples registros.

### Muerte

Atributos: *id\_muerte* (PK), *cantidad*, *id\_causa*, *id\_departamento*, *id\_sexo*, *id\_grupo\_edad*, *id\_anio*. Esta tabla integra todas las dimensiones del modelo: espacial (departamento/región), demográfica (sexo, grupo de edad), temporal (año) y causal (causa). El campo *cantidad* indica el número total de muertes para la combinación de factores referenciada por las claves foráneas.

## Normalización del Modelo Lógico

### Primera Forma Normal (1FN)

La Primera Forma Normal (1FN) busca que todos los atributos sean atómicos, es decir, que cada campo contenga un solo valor indivisible. También elimina columnas que contengan listas o valores repetidos dentro de una misma celda. En la base original MUERTES existían combinaciones como 'Caída/Ahogamiento' o '15-19 años / 20-24 años', lo que no cumple con esta

regla. Para cumplir la 1FN, se separaron estos datos en tablas individuales como CAUSA, SEXO y GRUPO\_EDAD. De esta manera, la tabla MUERTES contiene únicamente claves numéricas (FK) y el valor agregado de 'cantidad'. El resultado es una estructura más limpia y apta para consultas SQL.

## **Segunda Forma Normal (2FN)**

La Segunda Forma Normal (2FN) elimina dependencias parciales, que ocurren cuando un atributo depende solo de una parte de una clave compuesta. En la versión inicial, el campo 'nombre\_region' dependía únicamente del 'departamento', y no del resto de la clave compuesta (AÑO, DEPARTAMENTO, CAUSA, SEXO, GRUPO\_EDAD). Para resolverlo, se creó la tabla REGIÓN con una relación jerárquica 1:N hacia DEPARTAMENTO. Ahora, los atributos de cada tabla dependen completamente de su clave principal. Este ajuste simplifica las relaciones y mejora la consistencia de la información geográfica.

## **Tercera Forma Normal (3FN)**

La Tercera Forma Normal (3FN) elimina dependencias transitivas, es decir, aquellas en las que un atributo no clave depende de otro atributo no clave. En el modelo inicial, 'nombre\_region' podría inferirse a partir de 'nombre\_departamento', lo que generaba redundancia. En el nuevo modelo, los datos se separan correctamente: REGIÓN contiene la información regional, DEPARTAMENTO solo depende de su propia PK, y MUERTES mantiene referencias mediante claves foráneas.

## **Construcción del Modelo Físico**

```
CREATE TABLE region (  
    id_region INTEGER GENERATED ALWAYS AS IDENTITY PRIMARY KEY,  
    nombre_region VARCHAR(100) NOT NULL  
);
```

```
CREATE TABLE departamento (  
    id_departamento INTEGER GENERATED ALWAYS AS IDENTITY PRIMARY KEY,  
    nombre_departamento VARCHAR(120) NOT NULL,  
    codigo_dane_depto VARCHAR(10),  
    id_region INTEGER NOT NULL,
```

```
FOREIGN KEY (id_region) REFERENCES region(id_region)
);
```

```
CREATE TABLE causa (
    id_causa INTEGER GENERATED ALWAYS AS IDENTITY PRIMARY KEY,
    nombre_causa VARCHAR(150) NOT NULL,
    descripcion_causa VARCHAR(255)
);
```

```
CREATE TABLE sexo (
    id_sexo INTEGER GENERATED ALWAYS AS IDENTITY PRIMARY KEY,
    tipo_sexo VARCHAR(30) NOT NULL
);
```

```
CREATE TABLE grupo_edad (
    id_grupo_edad INTEGER GENERATED ALWAYS AS IDENTITY PRIMARY KEY,
    descripcion_grupo_edad VARCHAR(60),
    edad_min INTEGER,
    edad_max INTEGER
);
```

```
CREATE TABLE anio (
    id_anio INTEGER GENERATED ALWAYS AS IDENTITY PRIMARY KEY,
    anio INTEGER NOT NULL
);
```

```
CREATE TABLE muertes (
    id_muerte INTEGER GENERATED ALWAYS AS IDENTITY PRIMARY KEY,
    id_causa INTEGER NOT NULL,
    id_departamento INTEGER NOT NULL,
    id_sexo INTEGER NOT NULL,
    id_grupo_edad INTEGER NOT NULL,
    id_anio INTEGER NOT NULL,
    cantidad INTEGER NOT NULL,
    FOREIGN KEY (id_causa) REFERENCES causa(id_causa),
    FOREIGN KEY (id_departamento) REFERENCES departamento(id_departamento),
    FOREIGN KEY (id_sexo) REFERENCES sexo(id_sexo),
    FOREIGN KEY (id_grupo_edad) REFERENCES grupo_edad(id_grupo_edad),
    FOREIGN KEY (id_anio) REFERENCES anio(id_anio)
);
```



## Diagrama E-R

El diagrama entidad-relación posiciona a la tabla MUERTES como entidad central, conectada con las dimensiones CAUSA, SEXO, GRUPO\_EDAD, AÑO y DEPARTAMENTO. La entidad DEPARTAMENTO se enlaza con REGIÓN.

## Tablas y Relaciones

Las relaciones entre las tablas se definen mediante claves primarias y foráneas, garantizando la integridad referencial. El diseño propuesto permite analizar los datos desde múltiples perspectivas: por causa, género, edad, ubicación y año, sin redundancia ni pérdida de consistencia.

## Inserción de Datos

```
INSERT INTO region (nombre_region) VALUES ('Caribe');
INSERT INTO departamento (nombre_departamento, codigo_dane_depto, id_region) VALUES
('Atlántico', '08', 1);
INSERT INTO causa (nombre_causa, descripcion_causa) VALUES ('Caída', 'Caída desde altura o
superficie resbalosa');
INSERT INTO sexo (tipo_sexo) VALUES ('Femenino');
INSERT INTO grupo_edad (descripcion_grupo_edad, edad_min, edad_max) VALUES ('15-19',
15, 19);
INSERT INTO anio (anio) VALUES (2015);

INSERT INTO muertes (id_causa, id_departamento, id_sexo, id_grupo_edad, id_anio, cantidad)
VALUES (1);
```

## **Referencias**

DANE – Muertes accidentales en Colombia (2015–2024). Documentación y datos abiertos.  
Elmasri & Navathe. Sistemas de Bases de Datos. Pearson.  
Documentación SQL Estándar (ANSI/ISO).