

# Detección de niveles de inseguridad en el sector Colombiano

Saudith, Amy, Abril Jostin, Gómez Sebastian

<sup>1-2</sup>Dpto. de Periodismo, <sup>3</sup>Dpto. de Matemáticas,  
Universidad Externado de Colombia  
Pregrado Ciencia de Datos  
Curso de Bases de Datos  
Bogotá, Colombia  
{<sup>1</sup>autor1,<sup>2</sup>autor2}@correo1.com, <sup>3</sup>autor2@correo2.com

November 22, 2023

## Contents

<b>1</b>	<b>Introducción</b>	<b>2</b>
<b>2</b>	<b>Características del proyecto de investigación que hace uso de Bases de Datos</b>	<b>3</b>
2.1	Titulo del proyecto de investigación . . . . .	3
2.2	Objetivo general . . . . .	3
2.2.1	Objetivos especificos . . . . .	3
2.3	Alcance . . . . .	3
2.4	Pregunta de investigación . . . . .	4
2.5	Hipotesis . . . . .	4
<b>3</b>	<b>Reflexiones sobre el origen de datos e información</b>	<b>5</b>
3.1	¿Cual es el origen de los datos e información ? . . . . .	5
3.2	¿Cuales son las consideraciones legales o eticas del uso de la información? . . . . .	5
3.3	¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación? . . . . .	5
3.4	¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? . . . . .	6
<b>4</b>	<b>Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)</b>	<b>7</b>

4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto . . . . .	7
4.2	Diagrama modelo de datos . . . . .	7
4.3	Imágenes de la Base de Datos . . . . .	10
4.4	Código SQL - lenguaje de definición de datos (DDL) . . . . .	11
4.5	Código SQL - Manipulación de datos (DML) . . . . .	15
4.6	Código SQL + Resultados: Vistas . . . . .	15
4.7	Código SQL + Resultados: Triggers . . . . .	16
4.8	Código SQL + Resultados: Funciones . . . . .	17
4.9	Código SQL + Resultados: procedimientos almacenados . . . . .	21
<b>5</b>	<b>Bases de Datos No-SQL (<i>Segunda entrega</i>)</b>	<b>22</b>
5.1	Diagrama Bases de Datos No-SQL ( <i>Segunda entrega</i> ) . . . . .	22
5.2	Modelos conceptuales . . . . .	23
5.3	Modelos Lógicos . . . . .	24
5.4	Modelos físicos . . . . .	25
5.5	SMBD utilizado para la Base de Datos No-SQL ( <i>Segunda entrega</i> )	27
<b>6</b>	<b>Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos</b>	<b>29</b>
6.1	Ejemplo de aplicación de ETL y Bodega de Datos . . . . .	29
6.2	Automatización de Datos ( <i>Tercera entrega</i> ) . . . . .	30
6.3	Integración de Datos . . . . .	31
<b>7</b>	<b>Proximos pasos</b>	<b>32</b>
<b>8</b>	<b>Lecciones aprendidas</b>	<b>33</b>
<b>9</b>	<b>Bibliografía</b>	<b>34</b>

## 1 Introducción

La violencia es un término que abarca abundantes áreas de interpretabilidad, pues, esta puede verse desde el ámbito de medicina, política, economía, género, etc. De igual forma, cada uno de estos ámbitos puede tener diferentes formas o definiciones del término "violencia", ya sea por ejemplo en la medicina, violencia de género, violencia sexual o violencia psicológica (entre otros).

Así, si se le preguntase a cada persona cuál es su definición de "violencia", muy seguramente obtendríamos un sin fin de ideas diferentes sobre el mismo concepto, algunas relacionadas y otras no tanto. La forma en la que nosotros podríamos definir la violencia sería "todo proceso conductual que genere afectaciones negativas tanto así mismo como a su entorno", vease a su entorno como un espacio físico y/o las personas que rodean al agresor.

A su vez, cada acto considerado como violento puede tener diferentes razones que lo produzcan, por ejemplo, la inestabilidad emocional, los trastornos mentales

que induzcan sobre la conducta y/o percepción de su entorno, la economía de un país, el nivel educativo del agresor, etc.

Al poder identificar los sectores geográficos donde más se presentan casos de violencia y sus correspondientes tipos más frecuentes, se pueden generar diferentes estrategias que permitan la disminución de la presencia de estos casos en los diferentes sectores geográficos analizados.

## **2 Características del proyecto de investigación que hace uso de Bases de Datos**

### **2.1 Título del proyecto de investigación**

Recolección e inferencia de datos para realizar análisis de causalidad sobre la violencia en Bogotá, Colombia.

### **2.2 Objetivo general**

Determinar cuales son los posibles detonantes del incremento de los niveles de violencia en el sector Colombiano, para poder brindar posibles soluciones a los factores de violencia más comunes y que generen mayor afectación a nivel social dentro del país.

#### **2.2.1 Objetivos específicos**

- Identificar los principales tipos de violencia en Colombia.
- Generar diferentes clasificaciones sobre los datos con lo cual se pueda abordar el tema desde diferentes problemáticas de violencia.
- Encontrar relaciones entre las posibles causas que generen un incremento en los diferentes índices de violencia.

### **2.3 Alcance**

El proyecto se centra en la recopilación y estructuración de datos relacionados con actos violentos ocurridos en el país de Colombia durante un periodo específico de seis años, desde 2018 hasta 2023. La información se extraerá exclusivamente de fuentes oficiales del país, lo que garantiza la autenticidad y fiabilidad de los datos.

En cuanto a la naturaleza de los diferentes incidentes, se abordarán eventos específicos como homicidios, hurtos, secuestros, violaciones, acoso sexual y violencia intrafamiliar.

Cada incidente registrado en la base de datos contará con detalles específicos como el municipio donde tuvo lugar, la fecha en que ocurrió y el número de personas afectadas. Sin embargo, se ha decidido omitir cualquier dato

relacionado con intervenciones posteriores, como acciones legales, arrestos o juicios, debido a que no es el enfoque de este proyecto.

La delimitación geográfica del proyecto, como se mencionó anteriormente, se restringirá únicamente al sector Colombiano, excluyendo incidentes que puedan haber ocurrido fuera de esta área.

## **2.4 Pregunta de investigación**

¿Cuál es el comportamiento de los diferentes tipos de violencia en el territorio Colombiano?

## **2.5 Hipotesis**

En la capital, los tipos de violencia que más afectan socialmente son la violencia doméstica, la violencia de género y la violencia juvenil, debido a la prevalencia de estas en áreas urbanas, la cobertura mediática que reciben, y el impacto directo que tienen en las estructuras familiares y comunitarias.

### **3 Reflexiones sobre el origen de datos e información**

Al ubicarnos en el contexto de Bogotá, la capital de Colombia, nuestro proyecto está guiado por bases de datos e información gubernamental como fuente de información primordial. Debido a que la violencia es un problema que todos los días enfrentamos, esta se ve de manera más pronunciada en ciudades con mayor población y capitales. En este proyecto gracias a la concurrencia de violencia pudimos encontrar información en plataformas de

#### **3.1 ¿Cuál es el origen de los datos e información ?**

Nuestro proyecto se basa en datos gubernamentales de Bogotá, obtenidos de fuentes como el sitio web de la Policía Nacional, la Secretaría de Educación y el Instituto de Medicina Legal. A través de la página de datos abiertos de la capital, accedemos a diversas bases registradas por entidades públicas. Esta variedad de fuentes nos permite obtener una visión completa de las causas de la violencia en Bogotá y realizar análisis sólidos basados en datos confiables.

#### **3.2 ¿Cuales son las consideraciones legales o éticas del uso de la información?**

El uso de la información en nuestro proyecto conlleva importantes consideraciones legales y éticas. En el ámbito legal, debemos cumplir con las leyes de protección de datos y privacidad vigentes, garantizando la anonimización de información sensible. También debemos respetar los derechos de propiedad intelectual al utilizar datos gubernamentales. Desde una perspectiva ética, nos comprometemos a asegurar la confidencialidad de los datos y a utilizarlos exclusivamente para fines de investigación. Además, trabajamos para minimizar cualquier sesgo en el análisis y garantizar la transparencia en nuestros hallazgos. Nuestra prioridad es contribuir a la comprensión de las causas de la violencia de manera ética y legal.

#### **3.3 ¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación?**

Los datos que utilizaremos en nuestra base de datos enfrentan retos significativos en términos de calidad y consolidación. Primero, la heterogeneidad de las fuentes gubernamentales puede generar inconsistencias y errores. La veracidad y confiabilidad de los datos son esenciales y requerirán una exhaustiva verificación y limpieza. Además, la estandarización de la información proveniente de diversas entidades es un desafío para su consolidación efectiva. Garantizar la integridad y consistencia de los datos será fundamental para obtener resultados precisos. También, la actualización constante de los datos es crucial para mantener su relevancia a lo largo del tiempo. Nuestra labor incluye abordar estos retos para construir una base de datos sólida y confiable.

### **3.4 ¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto?**

Esperamos que la utilización de un sistema de Bases de Datos para nuestro proyecto aporte una gestión eficiente de la información, permitiendo un almacenamiento ordenado y seguro de los datos gubernamentales. Buscamos que facilite la recuperación ágil de información relevante, lo que agilizará nuestros análisis y la toma de decisiones. Asimismo, esperamos que el sistema permita facilitar la utilización de las bases de datos, para identificar patrones y causas subyacentes de la violencia en Bogotá. En resumen, nuestro objetivo es aprovechar al máximo la potencia de las bases de datos para lograr una investigación eficaz y una comprensión profunda de las causas de la violencia en la ciudad.

## 4 Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)

### 4.1 Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto

Oracle es un Sistema Gestor de Bases de Datos con características objeto-relacionales, que pertenece al modelo evolutivo de SGBD. Se considera a Oracle Database como uno de los sistemas de bases de datos más completos, destacando:

- Soporte de transacciones y multiplataforma
- Estabilidad y escalabilidad
- Oracle soporta todas las características que se esperan de un servidor, admite un lenguaje de diseño de bases de datos muy completo (PL/SQL) que permite implementar diseños “activos”. Con triggers y procedimientos almacenados, con una integridad referencial declarativa bastante sólida.
- Posibilita el uso de participaciones para la mejora de la eficiencia, de replicación e incluso ciertas versiones admiten la administración de bases de datos distribuidas.

### 4.2 Diagrama modelo de datos

El modelado general de nuestra base de datos se puede observar en el siguiente diagrama

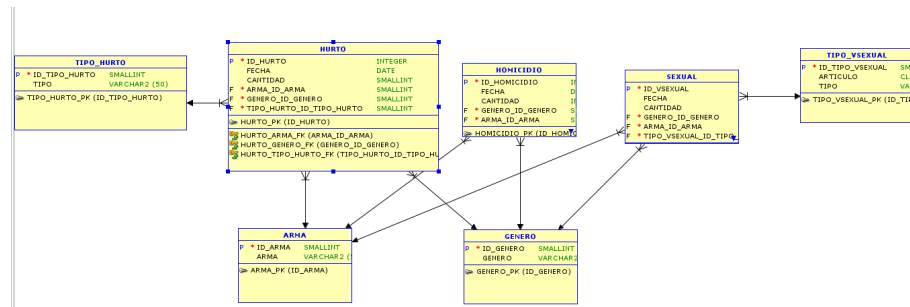


Figure 1: Modelo general

Sin embargo, como podemos observar que la imagen se encuentra borrosa, describiremos cada uno de las entidades de manera breve.

Tenemos dos entidades con información básica sobre variables incluidas en cualquier tipo de violencia. Cabe aclarar que la entidad "GENERO" es en referencia a la víctima y no al victimario.

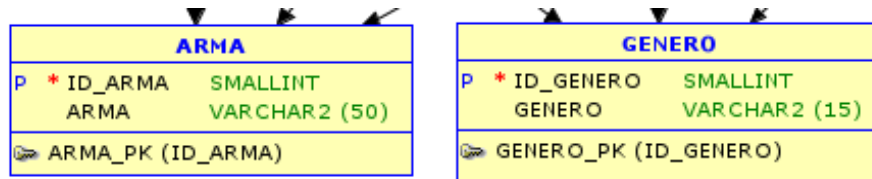


Figure 2: Entidades básicas

También, contamos con dos tablas que definen el tipo de abuso y hurto:

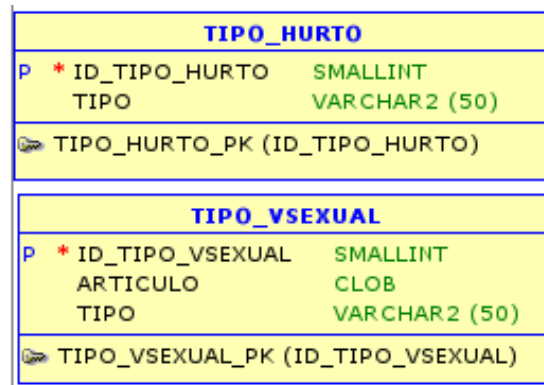


Figure 3: Entidades de tipo

Por último contamos con tres entidades correspondientes a diferentes tipos de expresión de violencia.



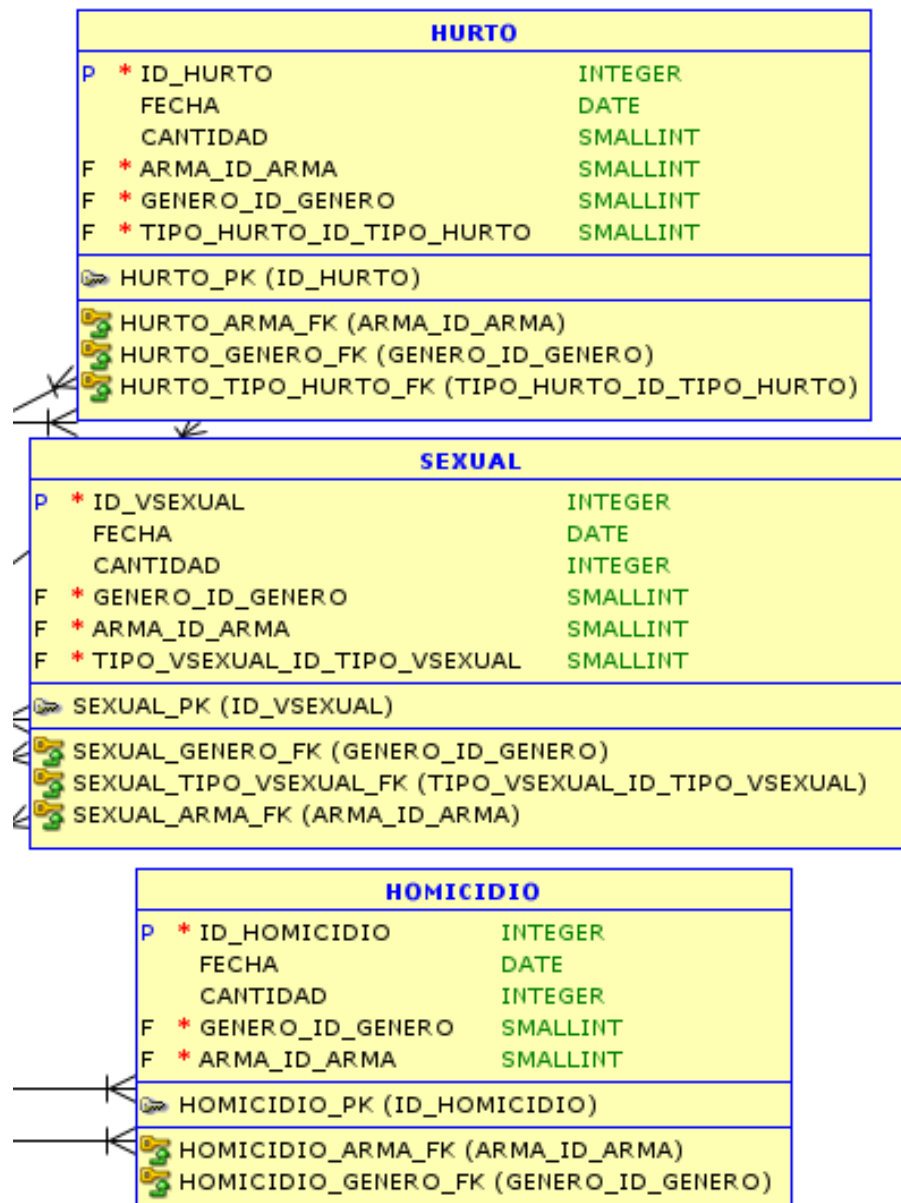


Figure 4: Principales

### 4.3 Imágenes de la Base de Datos

❖	COLUMN_NAME	❖	DATA_TYPE	❖	NULLABLE	DATA_DEFAULT	❖	COLUMN_ID	❖	COMMENTS
1	EMPLEADO_ID		NUMBER(5,0)		No	(null)		1		(null)
2	NOMBRE		VARCHAR2(50 BYTE)		Yes	(null)		2		(null)
3	APELLIDO		VARCHAR2(50 BYTE)		Yes	(null)		3		(null)
4	SALARIO		NUMBER(10,2)		Yes	(null)		4		(null)

Figure 5: Arma BD

❖	COLUMN_NAME	❖	DATA_TYPE	❖	NULLABLE	DATA_DEFAULT	❖	COLUMN_ID	❖	COMMENTS
1	ID_HURTO		NUMBER(38,0)		No	(null)		1		(null)
2	FECHA		DATE		Yes	(null)		2		(null)
3	CANTIDAD		NUMBER(38,0)		Yes	(null)		3		(null)
4	ARMA_ID_ARMA		NUMBER(38,0)		No	(null)		4		(null)
5	GENERO_ID_GENERO		NUMBER(38,0)		No	(null)		5		(null)
6	TIPO_HURTO_ID_TIPO_HURTO		NUMBER(38,0)		No	(null)		6		(null)

Figure 6: Hurto BD

❖	COLUMN_NAME	❖	DATA_TYPE	❖	NULLABLE	DATA_DEFAULT	❖	COLUMN_ID	❖	COMMENTS
1	ID_GENERO		NUMBER(38,0)		No	(null)		1		(null)
2	GENERO		VARCHAR2(15 BYTE)		Yes	(null)		2		(null)

Figure 7: Género BD

❖	COLUMN_NAME	❖	DATA_TYPE	❖	NULLABLE	DATA_DEFAULT	❖	COLUMN_ID	❖	COMMENTS
1	ID_VSEXUAL		NUMBER(38,0)		No	(null)		1		(null)
2	FECHA		DATE		Yes	(null)		2		(null)
3	CANTIDAD		NUMBER(38,0)		Yes	(null)		3		(null)
4	GENERO_ID_GENERO		NUMBER(38,0)		No	(null)		4		(null)
5	ARMA_ID_ARMA		NUMBER(38,0)		No	(null)		5		(null)
6	TIPO_VSEXUAL_ID_TIPO_VSEXUAL		NUMBER(38,0)		No	(null)		6		(null)

Figure 8: Sexual BD

	❖ COLUMN_NAME	❖ DATA_TYPE	❖ NULLABLE	DATA_DEFAULT	❖ COLUMN_ID	❖ COMMENTS
1	ID_HOMICIDIO	NUMBER(38,0)	No	(null)	1	(null)
2	FECHA	DATE	Yes	(null)	2	(null)
3	CANTIDAD	NUMBER(38,0)	Yes	(null)	3	(null)
4	GENERO_ID_GENERO	NUMBER(38,0)	No	(null)	4	(null)
5	ARMA_ID_ARMA	NUMBER(38,0)	No	(null)	5	(null)

Figure 9: Homicidio BD

	❖ COLUMN_NAME	❖ DATA_TYPE	❖ NULLABLE	DATA_DEFAULT	❖ COLUMN_ID	❖ COMMENTS
1	ID_TIPO_HURTO	NUMBER(38,0)	No	(null)	1	(null)
2	TIPO	VARCHAR2(50 BYTE)	Yes	(null)	2	(null)

Figure 10: Tipo hurto BD

	❖ COLUMN_NAME	❖ DATA_TYPE	❖ NULLABLE	DATA_DEFAULT	❖ COLUMN_ID	❖ COMMENTS
1	ID_TIPO_VSEXUAL	NUMBER(38,0)	No	(null)	1	(null)
2	ARTICULO	CLOB	Yes	(null)	2	(null)
3	TIPO	VARCHAR2(50 BYTE)	Yes	(null)	3	(null)

Figure 11: Tipo violencia sexual BD

#### 4.4 Código SQL - lenguaje de definición de datos (DDL)

Nuestro código DDL para la creación de las tablas de la base de datos es el siguiente, cabe aclarar que, en esta parte solo se definen el código DDL de la tabla y bajo la misma la llave primaria correspondiente.

```

CREATE TABLE arma (
    id_arma SMALLINT NOT NULL,
    arma VARCHAR2(50)
);

ALTER TABLE arma ADD CONSTRAINT arma_pk PRIMARY KEY ( id_arma );

CREATE TABLE genero (
    id_genero SMALLINT NOT NULL,
    genero VARCHAR2(15)
);

ALTER TABLE genero ADD CONSTRAINT genero_pk PRIMARY KEY ( id_genero );

CREATE TABLE homicidio (
    id_homicidio INTEGER NOT NULL,
    fecha DATE,
    cantidad INTEGER,
    genero_id_genero SMALLINT NOT NULL,
    arma_id_arma SMALLINT NOT NULL
);

ALTER TABLE homicidio ADD CONSTRAINT homicidio_pk PRIMARY KEY ( id_homicidio );

CREATE TABLE hurto (
    id_hurto INTEGER NOT NULL,
    fecha DATE,
    cantidad SMALLINT,
    arma_id_arma SMALLINT NOT NULL,
    genero_id_genero SMALLINT NOT NULL,
    tipo_hurto_id_tipo_hurto SMALLINT NOT NULL
);

```

Figure 12: Parte 1 de la creación

```

CREATE TABLE sexual (
    id_vsexual          INTEGER NOT NULL,
    fecha              DATE,
    cantidad            INTEGER,
    genero_id_genero    SMALLINT NOT NULL,
    arma_id_arma        SMALLINT NOT NULL,
    tipo_vsexual_id_tipo_vsexual SMALLINT NOT NULL
);

ALTER TABLE sexual ADD CONSTRAINT sexual_pk PRIMARY KEY ( id_vsexual );

CREATE TABLE tipo_hurto (
    id_tipo_hurto SMALLINT NOT NULL,
    tipo          VARCHAR2(50)
);

ALTER TABLE tipo_hurto ADD CONSTRAINT tipo_hurto_pk PRIMARY KEY ( id_tipo_hurto );

CREATE TABLE tipo_vsexual (
    id_tipo_vsexual SMALLINT NOT NULL,
    articulo        CLOB,
    tipo            VARCHAR2(50)
);

ALTER TABLE tipo_vsexual ADD CONSTRAINT tipo_vsexual_pk PRIMARY KEY ( id_tipo_vsexual );

```

Figure 13: Parte 2 de la creación

En la siguiente imagen se hacen las diferentes relaciones entre las entidades.

```

ALTER TABLE homicidio
    ADD CONSTRAINT homicidio_arma_fk FOREIGN KEY ( arma_id_arma )
        REFERENCES arma ( id_arma );

ALTER TABLE homicidio
    ADD CONSTRAINT homicidio_genero_fk FOREIGN KEY ( genero_id_genero )
        REFERENCES genero ( id_genero );

ALTER TABLE hurto
    ADD CONSTRAINT hurto_arma_fk FOREIGN KEY ( arma_id_arma )
        REFERENCES arma ( id_arma );

ALTER TABLE hurto
    ADD CONSTRAINT hurto_genero_fk FOREIGN KEY ( genero_id_genero )
        REFERENCES genero ( id_genero );

ALTER TABLE hurto
    ADD CONSTRAINT hurto_tipo_hurto_fk FOREIGN KEY ( tipo_hurto_id_tipo_hurto )
        REFERENCES tipo_hurto ( id_tipo_hurto );

ALTER TABLE sexual
    ADD CONSTRAINT sexual_arma_fk FOREIGN KEY ( arma_id_arma )
        REFERENCES arma ( id_arma );

ALTER TABLE sexual
    ADD CONSTRAINT sexual_genero_fk FOREIGN KEY ( genero_id_genero )
        REFERENCES genero ( id_genero );

ALTER TABLE sexual
    ADD CONSTRAINT sexual_tipo_vsexual_fk FOREIGN KEY ( tipo_vsexual_id_tipo_vsexual )
        REFERENCES tipo_vsexual ( id_tipo_vsexual );

```

Figure 14: Creación de relaciones

## 4.5 Código SQL - Manipulación de datos (DML)

```
-- Insertar diferentes tipos de violencia sexual
INSERT INTO TIPO_VSEXUAL VALUES(1, 206, 'Acceso sexual violento');
INSERT INTO TIPO_VSEXUAL VALUES(2, 205, 'Acceso carnal violento');

--Insertar los diferentes tipos de hurtos
INSERT INTO TIPO_HURTO VALUES(1, 'Comercio');
INSERT INTO TIPO_HURTO VALUES(2, 'Residencias');
INSERT INTO TIPO_HURTO VALUES(3, 'Personas');
INSERT INTO TIPO_HURTO VALUES(4, 'Motocicletas');

-- Insertar generos
INSERT INTO GENERO VALUES (1,'Femenino');
INSERT INTO GENERO VALUES (2,'Masculino');
INSERT INTO GENERO VALUES (3, 'No reportado');

--Insertar los diferentes tipos de armas
INSERT INTO ARMA VALUES(1, 'Arma blanca');
INSERT INTO ARMA VALUES(2, 'Arma de fuego');
INSERT INTO ARMA VALUES(3, 'Sin empleo de arma');
INSERT INTO ARMA VALUES(4, 'No reportado');

-- Insertar datos a la tabla de hurtos
INSERT INTO HURTO VALUES(1, '15-JUL-2023', 8, 4, 2, 3);
INSERT INTO HURTO VALUES(2, '02-JAN-2023', 1, 1, 2, 2);
INSERT INTO HURTO VALUES(3, '18-FEB-2023', 1, 1, 3, 2);

-- Insertar datos a la tabla de homicidios
INSERT INTO HOMICIDIO VALUES(1, '28-FEB-2023', 2, 1, 3);
INSERT INTO HOMICIDIO VALUES(2, '08-SEP-2023', 1, 2, 1);

-- Insertar datos a la tabla de sexual
INSERT INTO SEXUAL VALUES(1, '17-DEC-2023', 2, 1, 3, 2);
INSERT INTO SEXUAL VALUES(2, '07-JAN-2023', 1, 2, 1, 1);
INSERT INTO SEXUAL VALUES(3, '07-JUL-2023', 10, 2, 1, 1);

COMMIT;
```

Figure 15: Inserción de datos

## 4.6 Código SQL + Resultados: Vistas

Primero identificamos de las diferentes tablas cuales datos no son necesarios o relevantes al momento de realizar una visualización sobre la tabla, sin embargo, llegamos a la conclusión que todos los datos pueden ser relevantes según la variable de interés al momento de analizar, por ende, lo único que no incluimos en las vistas de las tablas fue el campo ID.

```

-- Vistas para tabla hurto
CREATE VIEW view_hurto AS
SELECT FECHA, CANTIDAD, ARMA_ID_ARMA, GENERO_ID_GENERO, TIPO_HURTO_ID_TIPO_HURTO FROM HURTO;

-- Vistas para tabla homicidio
CREATE VIEW view_homicidio AS
SELECT FECHA, CANTIDAD, ARMA_ID_ARMA, GENERO_ID_GENERO FROM HOMICIDIO;

-- Vistas para tabla sexual
CREATE VIEW view_vsexual AS
SELECT FECHA, CANTIDAD, GENERO_ID_GENERO, ARMA_ID_ARMA, TIPO_VSEXUAL_ID_TIPO_VSEXUAL FROM SEXUAL;

-- Vistas para tabla sexual
CREATE VIEW view_violencia AS
SELECT TIPO, FECHA_ACTUALIZACION FROM VIOLENCIA;

COMMIT;

```

Figure 16: DDL Vistas

Las siguientes imagenes muestran los resultados al hacer la consulta con SELECT, cabe aclarar que, los indices numéricos son los que por defecto pone el DBMS, más no, los del ID de cada tabla.

	FECHA	CANTIDAD	ARMA_ID_ARMA	GENERO_ID_GENERO
1	28-FEB-23	2	3	1
2	08-SEP-23	1	1	2

Figure 17: Vista de homicidio

	FECHA	ARMA_ID_ARMA	GENERO_ID_GENERO	TIPO_HURTO_ID_TIPO_HURTO
1	15-JUL-23	8	2	3
2	02-JAN-23	1	2	2
3	18-FEB-23	1	3	2

Figure 18: Vista de hurto

	FECHA	CANTIDAD	GENERO_ID_GENERO	ARMA_ID_ARMA	TIPO_VSEXUAL_ID_TIPO_VSEXUAL
1	17-DEC-23	2	1	3	2
2	07-JAN-23	1	2	1	1
3	07-JUL-23	10	2	1	1

Figure 19: Vista de abuso sexual

## 4.7 Código SQL + Resultados: Triggers

La primera imagen muestra la definición del código DDL de los triggers, la que le sigue muestra el resultado.



```

-- Actualización fechas para SEXUAL
CREATE OR REPLACE TRIGGER actualizar_fecha_S
BEFORE INSERT ON SEXUAL

BEGIN
    UPDATE VIOLENCIA
    SET FECHA_ACTUALIZACION = SYSDATE
    WHERE ID = 1;
END;

-- Actualización fechas para HURTO
CREATE OR REPLACE TRIGGER actualizar_fecha_HU
BEFORE INSERT ON HURTO

BEGIN
    UPDATE VIOLENCIA
    SET FECHA_ACTUALIZACION = SYSDATE
    WHERE ID = 3;
END;

-- Actualización fechas para HOMICIDIO
CREATE OR REPLACE TRIGGER actualizar_fecha_HO
BEFORE INSERT ON HOMICIDIO

BEGIN
    UPDATE VIOLENCIA
    SET FECHA_ACTUALIZACION = SYSDATE
    WHERE ID = 2;
END;

COMMIT;

```

Figure 20:

#### 4.8 Código SQL + Resultados: Funciones

Las siguientes funciones harán lo mismo pero enfocado para tablas diferentes, sé que existe algo llamado funciones dinámicas pero, a la fecha no entiendo como funcionan.

El funcionamiento de la función es cuantificar por géneros los cada tipo de

violencia

```
-- Homicidio
CREATE OR REPLACE FUNCTION proporciones_generoHOMICIDIO RETURN VARCHAR2 IS
    freq_m NUMBER;
    freq_h NUMBER;
    freq_n NUMBER;
    v_genero_id_genero NUMBER;
    result VARCHAR2(100);
BEGIN
    FOR registro IN (SELECT GENERO_ID_GENERO, SUM(CANTIDAD) AS SUM_CANTIDAD
                     FROM HOMICIDIO
                     GROUP BY GENERO_ID_GENERO) LOOP

        v_genero_id_genero := registro.GENERO_ID_GENERO;
        IF v_genero_id_genero = 1 THEN
            freq_m := registro.SUM_CANTIDAD;
        ELSIF v_genero_id_genero = 2 THEN
            freq_h := registro.SUM_CANTIDAD;
        ELSIF v_genero_id_genero = 3 THEN
            freq_n := registro.SUM_CANTIDAD;
        END IF;

    END LOOP;

    result := 'Proporción de homicidios por género' || CHR(10) || CHR(9) ||
    'Masculino --> ' || freq_h || CHR(10) || CHR(9) || 'Femenino --> ' ||
    freq_m || CHR(10) || CHR(9) || 'No especificado --> ' || freq_n;
    RETURN result;
END proporciones_generoHOMICIDIO;
/

DECLARE
    v_result VARCHAR2(100);
BEGIN
    v_result := proporciones_generoHOMICIDIO();
    DBMS_OUTPUT.PUT_LINE(v_result);
END;
```

Figure 21: Declarando FHO

```

-- Hurto
CREATE OR REPLACE FUNCTION proporciones_generoHURTO RETURN VARCHAR2 IS
    freq_m NUMBER;
    freq_h NUMBER;
    freq_n NUMBER;
    v_genero_id_genero NUMBER;
    result VARCHAR2(100);
BEGIN
    FOR registro IN (SELECT GENERO_ID_GENERO, SUM(CANTIDAD) AS SUM_CANTIDAD
                     FROM HURTO
                     GROUP BY GENERO_ID_GENERO) LOOP

        v_genero_id_genero := registro.GENERO_ID_GENERO;
        IF v_genero_id_genero = 1 THEN
            freq_m := registro.SUM_CANTIDAD;
        ELSIF v_genero_id_genero = 2 THEN
            freq_h := registro.SUM_CANTIDAD;
        ELSIF v_genero_id_genero = 3 THEN
            freq_n := registro.SUM_CANTIDAD;
        END IF;

    END LOOP;

    result := 'Proporción de hurtos por género' || CHR(10) || CHR(9) ||
    'Masculino --> ' || freq_h || CHR(10) || CHR(9) || 'Femenino --> ' ||
    freq_m || CHR(10) || CHR(9) || 'No especificado --> ' || freq_n;
    RETURN result;
END proporciones_generoHURTO;
/

DECLARE
    v_result VARCHAR2(100);
BEGIN
    v_result := proporciones_generoHURTO();
    DBMS_OUTPUT.PUT_LINE(v_result);
END;

```

Figure 22: Declarando FH

```

-- SEXUAL
CREATE OR REPLACE FUNCTION proporciones_generoSEXUAL RETURN VARCHAR2 IS
    freq_m NUMBER;
    freq_h NUMBER;
    freq_n NUMBER;
    v_genero_id_genero NUMBER;
    result VARCHAR2(100);
BEGIN
    FOR registro IN (SELECT GENERO_ID_GENERO, SUM(CANTIDAD) AS SUM_CANTIDAD
                     FROM SEXUAL
                     GROUP BY GENERO_ID_GENERO) LOOP

        v_genero_id_genero := registro.GENERO_ID_GENERO;
        IF v_genero_id_genero = 1 THEN
            freq_m := registro.SUM_CANTIDAD;
        ELSIF v_genero_id_genero = 2 THEN
            freq_h := registro.SUM_CANTIDAD;
        ELSIF v_genero_id_genero = 3 THEN
            freq_n := registro.SUM_CANTIDAD;
        END IF;

    END LOOP;

    result := 'Proporción de violencia sexual por género' || CHR(10) || CHR(9)
    || 'Masculino --> ' || freq_h || CHR(10) || CHR(9) || 'Femenino --> ' ||
    freq_m || CHR(10) || CHR(9) || 'No especificado --> ' || freq_n;
    RETURN result;
END proporciones_generoSEXUAL;
/

DECLARE
    v_result VARCHAR2(100);
BEGIN
    v_result := proporciones_generoSEXUAL();
    DBMS_OUTPUT.PUT_LINE(v_result);
END;

```

Figure 23: Declarando FS

```

Proporción de homicidios por género
Masculino --> 1
Femenino --> 2
No especificado -->

Proporción de hurtos por género
Masculino --> 9
Femenino -->
No especificado --> 1

Proporción de violencia sexual por género
Masculino --> 11
Femenino --> 2
No especificado -->

```

Figure 24: Resultados de las funciones

## 4.9 Código SQL + Resultados: procedimientos almacenados

Este procedimiento tiene como propósito actualizar el identificador de un tipo de arma en las tablas homicidio, hurto y sexual. Si, por ejemplo, tienes registros que usaron un identificador incorrecto para un tipo de arma, puedes corregir estos registros llamando a este procedimiento.

```
CREATE PROCEDURE actualizar_tipo_arma(p_arma_id_actual SMALLINT, p_arma_id_nuevo SMALLINT) AS
BEGIN
    UPDATE homicidio SET arma_id_arma = p_arma_id_nuevo WHERE arma_id_arma = p_arma_id_actual;
    UPDATE hurto SET arma_id_arma = p_arma_id_nuevo WHERE arma_id_arma = p_arma_id_actual;
    UPDATE sexual SET arma_id_arma = p_arma_id_nuevo WHERE arma_id_arma = p_arma_id_actual;
    COMMIT;
END actualizar_tipo_arma;

EXECUTE actualizar_tipo_arma(1,3);
```

Figure 25: DDL del proceso

ID_VSE...	FECHA	CANTID...	GENERO_ID_GENERO	ARMA_ID_ARMA	TIPO_VSEXUAL_ID_TIPO_VSEXUAL
1	1 17-DEC-23	2	1	3	2
2	2 07-JAN-23	1	2	1	1
3	3 07-JUL-23	10	2	1	1

Figure 26: Antes del proceso

ID_VSE...	FECHA	CANTID...	GENERO_ID_GENERO	ARMA_ID_ARMA	TIPO_VSEXUAL_ID_TIPO_VSEXUAL
1	1 17-DEC-23	2	1	3	2
2	2 07-JAN-23	1	2	3	1
3	3 07-JUL-23	10	2	3	1

Figure 27: Después del proceso

## 5 Bases de Datos No-SQL (*Segunda entrega*)

### 5.1 Diagrama Bases de Datos No-SQL (*Segunda entrega*)

Empezamos definiendo nuestro modelo de datos en un nivel conceptual, esto para ofrecer un vistazo general del modelamiento.

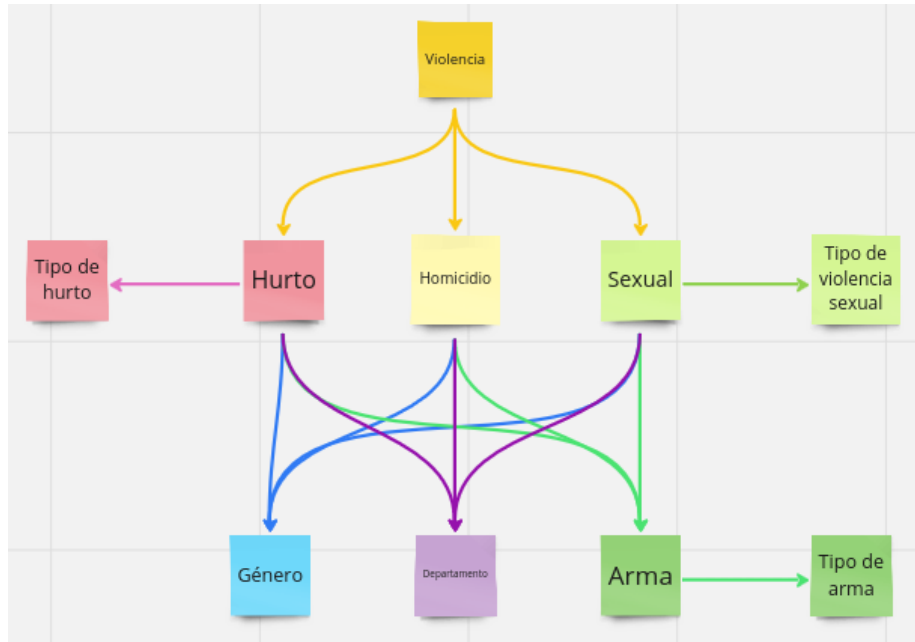


Figure 28: Diagrama de modelo NoSQL

## 5.2 Modelos conceptuales

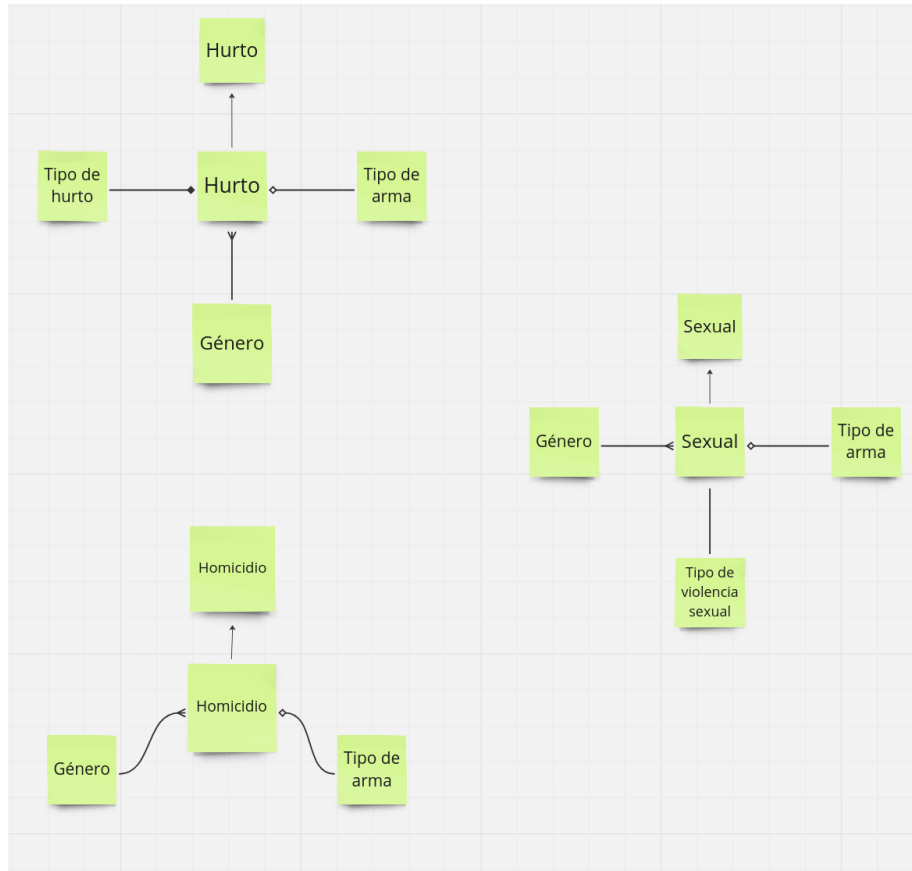


Figure 29: Tres modelos de conceptualización

### 5.3 Modelos Lógicos

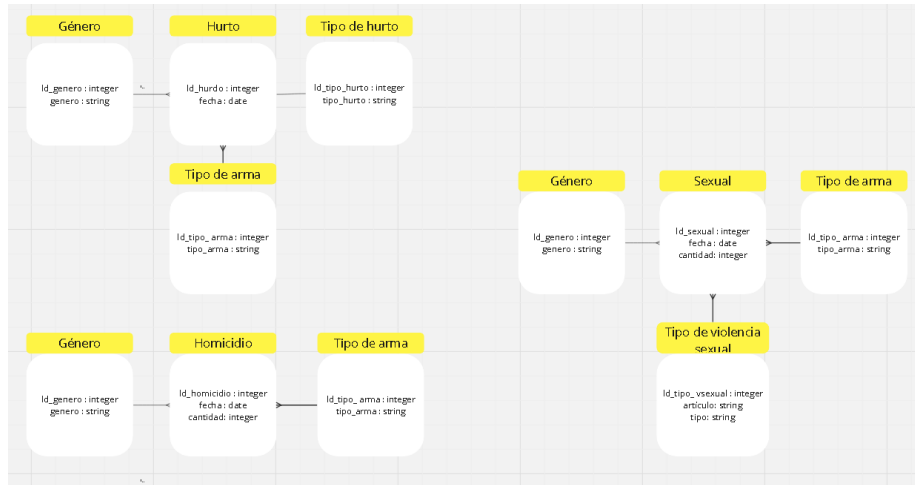


Figure 30: Modelos de logicos



## 5.4 Modelos físicos

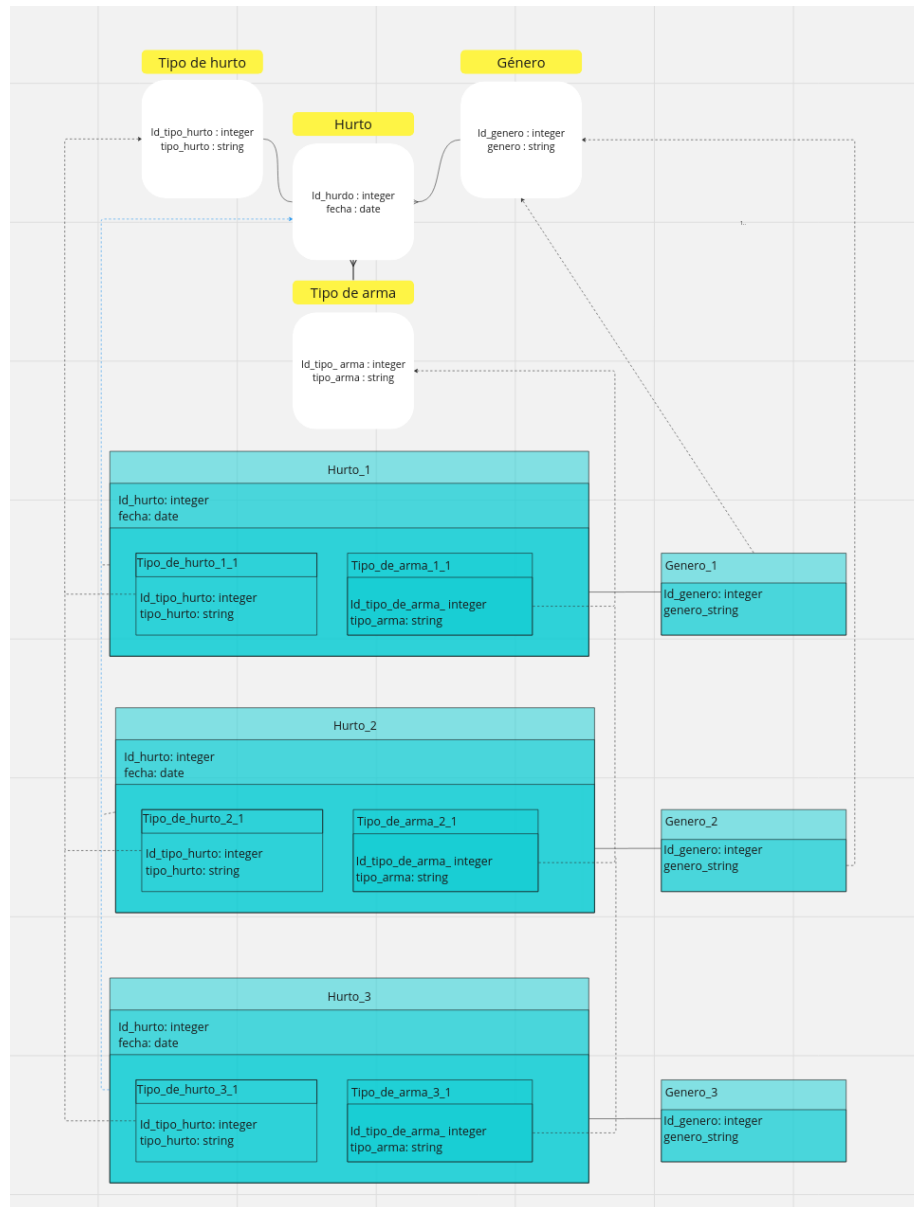


Figure 31: Modelo físico Hurto

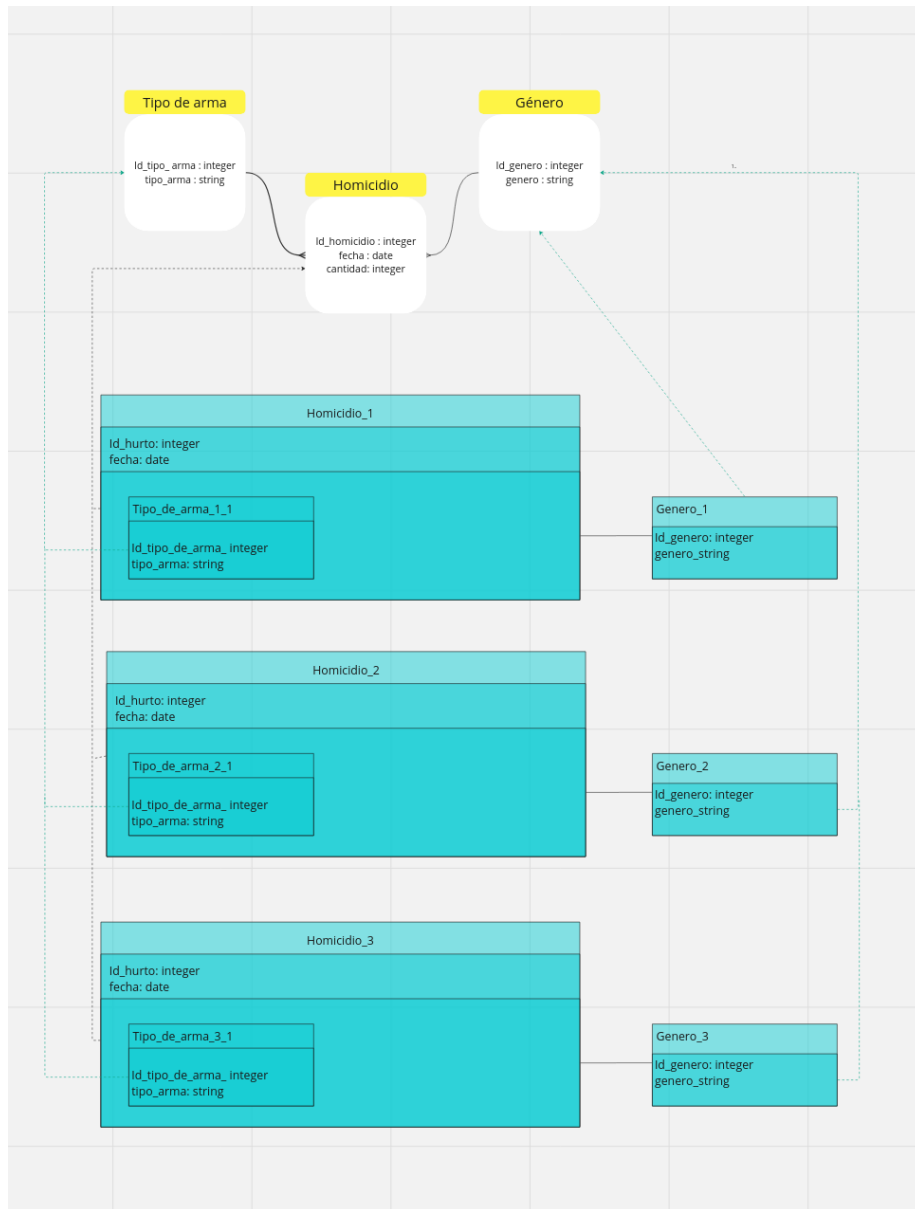


Figure 32: Modelo físico Homicidio

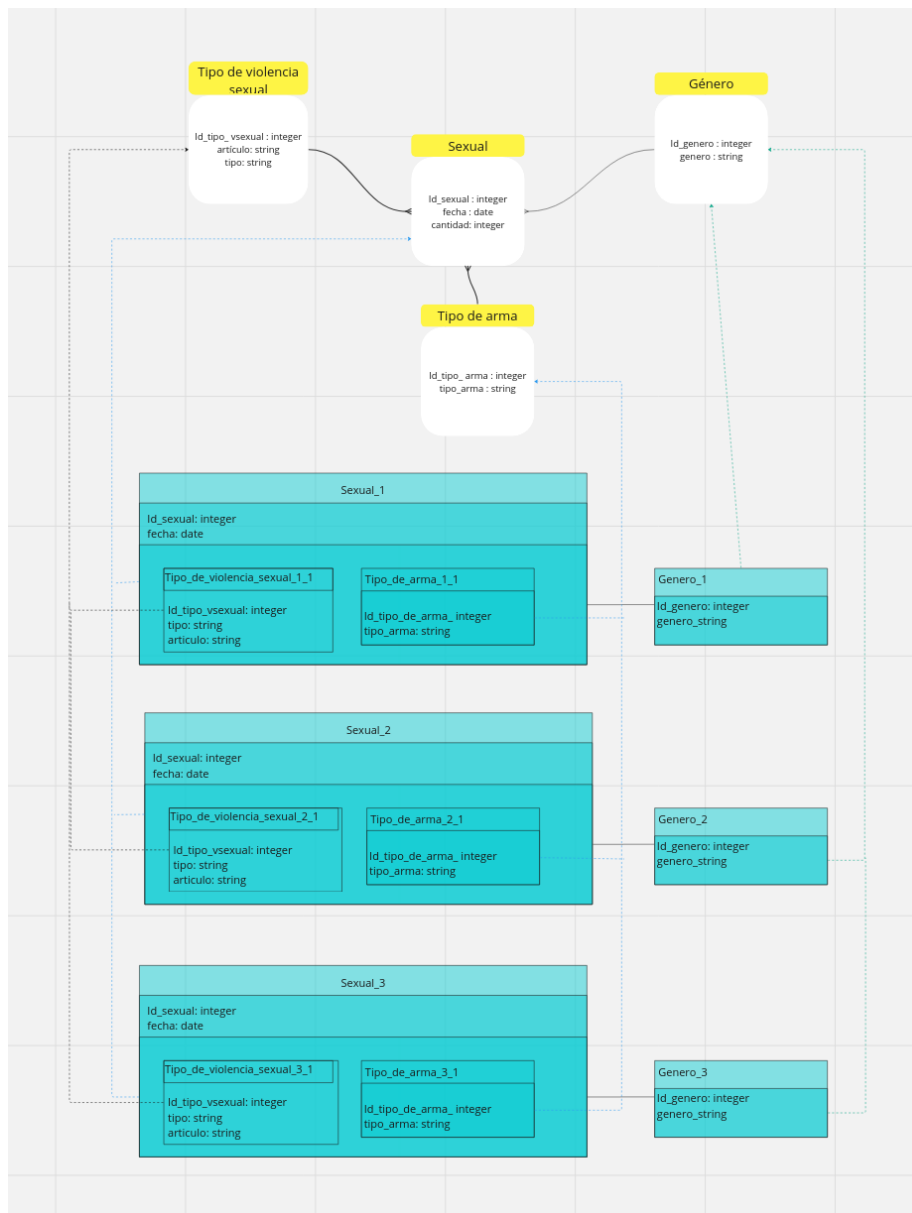


Figure 33: Modelo físico Sexual

## 5.5 SMBD utilizado para la Base de Datos No-SQL (*Segunda entrega*)

El DBMS utilizado para la implementación NoSQL fue MongoDB, esto debido a que nuestros datos son de tipo documentales y MongoDB permite el almacenamiento

de datos en este tipo de formato, adicional a lo anterior, este DBMS facilita la replicación de nuestros datos.

## 6 Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos

### 6.1 Ejemplo de aplicación de ETL y Bodega de Datos

- **Extracción**

Para el desarrollo de la extracción de los datos se tomaron diferentes fuentes, más sin embargo, para ejemplificar este punto se escogió la base de datos de hurto a personas en el año 2020 recolectada del sitio web de la policía nacional de Colombia **Estadística delictiva**



The screenshot shows a web interface with two dropdown menus: 'Delito de impacto' set to 'Hurto a personas' and 'Año' set to '2020'. An 'Aplicar' button is to the right. Below, a table lists the selected filters and a download link: 'hurto\_personas\_2020.xlsx'.

Delito de impacto	Año	Descargable
Hurto a personas	2020	hurto_personas_2020.xlsx

Figure 34: Base de datos de hurtos a personas en el año 2020

- **Transformación de datos**

Esta tabla contiene un total de 103.306 registros y 8 campos, para nuestro modelo de base de datos poseemos dos campos que no son de interés. El campo "Código DANE" actúa como meta variable, por lo que no es de utilidad basándose en nuestro modelo ER, el campo "Municipio" aunque brinda una información más detallada por sectores con una mayor delimitación, consideramos que por motivos prácticos es mejor solo delimitar la información por el campo "Departamento".

	Municipio	Código DANE
9	Leticia (CT)	91001000
10	Leticia (CT)	91001000
11	Leticia (CT)	91001000
12	Leticia (CT)	91001000
13	Leticia (CT)	91001000

Figure 35: Variables excluidas del conjunto de datos

Este dataset provee información sobre entidades diferentes dentro de nuestro modelo ER, el campo "Arma empleada" proporciona información relevante para poder agregar registros a nuestra entidad "ARMA", de manera análoga los campos "GÉNERO" y "Departamento" proporcionan información con su respectiva entidad en el modelo propuesto.

Para concretar, se realizaron cambios en el nombre de la categorías con el fin de obtener una mejor optimización al momento de realizar el paso de carga de datos en ETL y que de esta forma nuestro modelo cumpla con las reglas de normalización, este proceso se realizó para cada campo mencionado en el párrafo anterior, la siguiente tabla ilustra lo anteriormente dicho.

ID	Género
0	MASCULINO
1	FEMENINO
2	NO REPORTADO

Table 1: Normalización de los datos correspondientes al género

- **Carga de datos**

Primero decidimos cargar los datos conectados mediante llaves foráneas a nuestra entidad "HURTO", por ejemplo, fueron subidos a la entidad Departamento de nuestro modelo los Departamentos únicos presentes en la base de datos obtenida.

Los diferentes departamentos contenidos en el dataset fueron identificados y almacenados en la entidad "Departamento", esto para poder iniciar con la normalización de los datos y generar una mejor estructura en la carga de datos al servidor. De manera análoga se realizó el mismo ejercicio con los datos relacionados con el tipo de arma y el género de la víctima.

ID	Arma
0	ARMA BLANCA / CORTOPUNZANTE
1	ARMA DE FUEGO
2	CONTUNDENTES
3	ESCOPOLAMINA
4	LLAVE MAESTRA
5	NO REPORTADO
6	SIN EMPLEO DE ARMAS

Table 2: Normalización de los datos recolectados para la entidad Arma

## 6.2 Automatización de Datos (*Tercera entrega*)

Para poder hacer un proceso óptimo de automatización de datos se implementó el lenguaje de programación python debido a su alta eficiencia y simplicidad a la hora de escribir código. La estructura que manejamos se separó en tres partes:

- Separar información: se separó de la base de datos extraída los campos correspondientes a las diferentes entidades relacionadas a nuestra entidad principal (en este caso Hurto).
- Reconocer los datos: Se estructuraron los datos según nuestro modelo propuesto para tener simplicidad e información normalizada.
- Despliegue: Una vez los datos llevan una estructura según el modelo propuesto, se realiza el proceso de carga a cada entidad correspondiente.

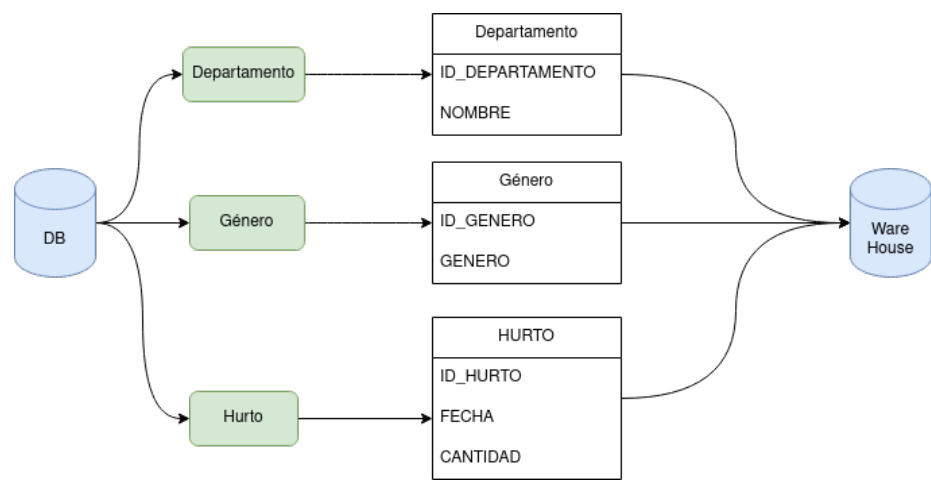


Figure 36: Modelo de procedimiento automatización

6.3 Integración de Datos

La mayoría de base de datos fueron obtenidas del sitio oficial de estadísticas delictivas ofrecida por la policía nacional, lo que permitió considerar diferentes bases de datos con campos similares y esto genera mayor uniformidad y facilidad para el proceso de automatización.

ARMAS/VEHICULO	DEPARTAMENTO	VAHICULO	FECHA	ID_CODIGO_USAR	CANTIDAD
ARMA BLANCA / CORTOPUJAMAZONAS	Leticia (CT)		24/01/2023	91001000	1
ARMA BLANCA / CORTOPUJANTIOQUIA	Andes		16/05/2023	920034000	1
ARMA BLANCA / CORTOPUJANTIOQUIA	Barbosa		31/01/2023	96079000	1
ARMA BLANCA / CORTOPUJANTIOQUIA	Barbosa		27/03/2023	96079000	1

Figure 37: Campos de la base de datos Hurto a motocicletas 2023. Fuente: Estadística delictiva

ARMAS/VEHICULO	DEPARTAMENTO	VEHICULO	FECHA	SEXO	TIPO DE VEHICULO	CANTIDAD
ARMA BLANCA / AMAZONAS	Leticia (CT)		24/03/2023	MASCULINO	ADULTOS	91001000
ARMA BLANCA / AMAZONAS	Leticia (CT)		16/04/2023	MASCULINO	ADULTOS	91001000
ARMA BLANCA / AMAZONAS	Leticia (CT)		16/05/2023	FEMENINO	ADULTOS	91001000
ARMA BLANCA / AMAZONAS	Leticia (CT)		30/06/2023	MASCULINO	ADULTOS	91001000

Figure 38: Campos de la base de datos Homicidios 2023. Fuente: Estadística delictiva

## 7 Proximos pasos

Implementar la aplicación de técnicas analíticas más avanzadas, como modelos de aprendizaje automático, para predecir patrones y tendencias en los datos recopilados. Esto podría proporcionar una visión más profunda y predictiva sobre la violencia en Colombia. Además, podrías explorar la posibilidad de realizar análisis geoespaciales para identificar áreas específicas con mayor incidencia de violencia, lo que podría ser crucial para el diseño de estrategias de intervención y prevención.

Podría considerarse la implementación de indicadores de desempeño y realizar un seguimiento continuo para medir el impacto de las estrategias sugeridas (futuras). Además, podrías investigar y discutir en detalle las implicaciones éticas y legales asociadas con el uso de los datos recopilados, brindando a los lectores una comprensión completa de los desafíos y consideraciones éticas en la gestión de la información relacionada con la violencia.



## 8 Lecciones aprendidas

- Fue fundamental saber buscar información que tuviese relación con el modelo propuesto.
- El mal modelamiento de una base de datos puede influir en procesos futuros sobre el tratamiento de datos e inclusive sobre la recolección de los mismos.
- La implementación del proceso ETL nos brindó un entendimiento más amplio sobre técnicas de recolección de datos y su tratamiento para poder ser desplegadas en un almacén de datos.
- Ver la importancia de todo el tema de PL/SQL y su correcta implementación para obtener resultados óptimos

## 9 Bibliografía

Datos.gov.co. (s.f.). Datos Abiertos Colombia. Recuperado de <https://www.datos.gov.co/> (Accedido el 22 de noviembre de 2023).

Policía Nacional de Colombia. (s.f.). *Estadística delictiva*. Recuperado de <https://www.policia.gov.co/grupo-informacion-criminalidad/estadistica-delictiva> (Accedido el 22 de noviembre de 2023).