Universidad del Rosario
Escuela de Ingeniería, Ciencia y Tecnología
Matemáticas Aplicadas y Ciencias de la Computación
Ingeniería de datos
Claudia Hernández
Sofia Ochoa Gutiérrez, Valentina Pabón León, Mariana Cadena Moreno
Semestre 1, 2022



### Segunda Entrega

Repositorio Github: <a href="https://github.com/Meiiyuu/proyectoinge">https://github.com/Meiiyuu/proyectoinge</a>

### 1. Diagrama Relacional Normalizado

Para normalizar las tablas, con respecto a la primera forma normal se presentaba redundancia horizontal en la tabla de caso con el atributo de sexo por lo que creamos la tabla de género solucionado esto, al igual que con la unidad\_medida\_edad. En la tabla estado también se presentaba redundancia horizontal en todos los atributos por lo que creamos una tabla para cada atributo y relacionamos las llaves primarias como foráneas en esta tabla. En la tabla etnia, se presentaba redundancia horizontal con los grupos de las etnias, por lo que creamos la tabla pertenecia\_etnia para luego asociar la llave primaria como foránea en la tabla etnia. Al hacer estas tablas se satisfacen la segunda forma normal ya que cada tabla solo tiene una llave primaria y todos los atributos dependen de esta, y la tercera forma normal también se cumple ya que no se presentan dependencias transitivas. El diagrama se encuentra en el repositorio como diagramarelacional.jpeg

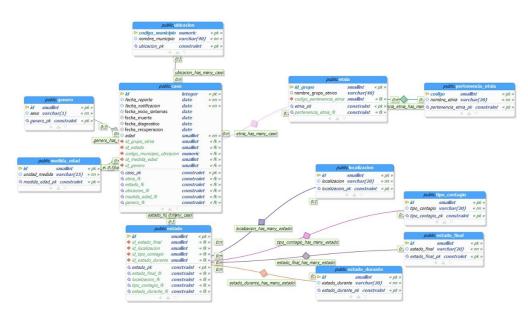


Figura 1) Diagrama relacional normalizado en 3NF

# 2. Implementación de la base de datos normalizada en tercera forma normal en PostgreSQL

Al tener el Diagrama Relacional en PGMoldeler se exporta como un archivo sql, que nos brinda el BDM para ejecutarlo en PostgreSQL con la creación de todas las tablas de la BD. El archivo sql se encuentra en el repositorio como **sqlproyecto.sql** 

## 3. Carga de todos los datos en la base de datos previamente creada.

Se tenían todos los datos en un solo archivo de Excel, se separa en once archivos para insertar los datos en cada tabla. Mediante la función COPY en PostgreSQL se carga a cada tabla sus respectivos datos. Para cargar los datos se debe ubicar la carpeta csv en C:\ y ejecutar cada COPY. Para visualizar la correcta inserción de los datos hacemos consultas con select \* from de cada una de las tablas. Nos decidimos por la carga masiva con la función COPY porque era la forma más eficiente de insertar los datos, solamente tocaba ajustar el Excel. Los archivos de Excel se encuentran en el repositorio en la carpeta csv y cada archivo se llama como cada una de las tablas y tienen sus respectivos datos ajustados. La carga de los datos se encuentra en el repositorio como sqlproyecto.sql.

```
COPY pertenencia_etnia FROM 'C:\csv\pertenencia_etnia.csv' WITH delimiter ';' CSV HEADER;
COPY etnia FROM 'C:\csv\etnia.csv' WITH delimiter ';' CSV HEADER;
COPY localizacion FROM 'C:\csv\localizacion.csv' WITH delimiter ';' CSV HEADER;
COPY tipo_contagio FROM 'C:\csv\tipo_contagio.csv' WITH delimiter ';' CSV HEADER;
COPY estado_final FROM 'C:\csv\estado_final.csv' WITH delimiter ';' CSV HEADER;
COPY estado_durante FROM 'C:\csv\estado_durante.csv' WITH delimiter ';' CSV HEADER;
COPY estado FROM 'C:\csv\estado.csv' WITH delimiter ';' CSV HEADER;
COPY medida_edad FROM 'C:\csv\medida_edad.csv' WITH delimiter ';' CSV HEADER;
COPY genero FROM 'C:\csv\medida_edad.csv' WITH delimiter ';' CSV HEADER;
COPY ubicacion FROM 'C:\csv\ubicacion.csv' WITH delimiter ';' CSV HEADER;
COPY caso FROM 'C:\csv\caso.csv' WITH delimiter ';' CSV HEADER;
```

Figura 2) Funciones COPY para la carga masiva

| 246<br>247 | FROM ubicacion                   |  |  |  |
|------------|----------------------------------|--|--|--|
| Data       | Output Explain                   | Messages Notifications                     |  |  |
| 4          | codigo_municipio<br>[PK] numeric | nombre_municipio<br>character varying (40) |  |  |
| 1          | 70771                            | SUCRE                                      |  |  |
| 2          | 7082                             | TOLU                                       |  |  |
| 3          | 70221                            | COVENIAS                                   |  |  |
| 4          | 70215                            | COROZAL                                    |  |  |
| 5          | 70001                            | SINCELEJO                                  |  |  |
| 6          | 7067                             | SAMPUES                                    |  |  |
| 7          | 70742                            | SINCE                                      |  |  |
| 8          | 70523                            | PALMITO                                    |  |  |
| 9          | 70418                            | 70418 LOS PALMITOS                         |  |  |
| 10         | 70124                            | CAIMITO                                    |  |  |
| 11         | 70508                            | OVEJAS                                     |  |  |
| 12         | 70713                            | SAN ONOFRE                                 |  |  |
| 13         | 70473                            | MORROA                                     |  |  |
| 14         | 70235                            | GALERAS (NUEVA GRANADA)                    |  |  |
| 15         | 70233                            | EL ROBLE                                   |  |  |
| 16         | 70708                            | SAN MARCOS                                 |  |  |
| 17         | 7011                             | BUENAVISTA                                 |  |  |
| 18         | 7023                             | CHALAN                                     |  |  |
| 19         | 70429                            | MAJAGUAL                                   |  |  |
| 20         | 70823                            | TOLUVIEJO                                  |  |  |

Figura 3) Consulta de la tabla ubicacion

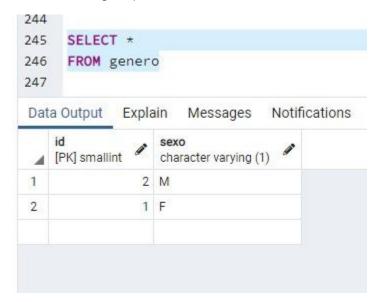


Figura 4) Consulta de la tabla genero

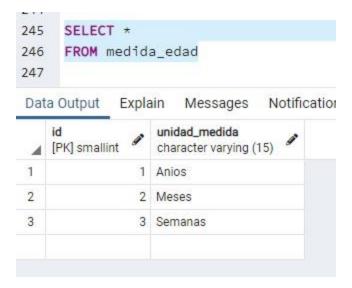


Figura 5) Consulta de la tabla medida\_edad

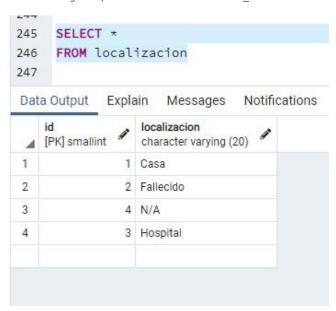


Figura 6) Consulta de la tabla localizacion

| 47   |                     |                          |                             |                              |                               |  |  |  |
|------|---------------------|--------------------------|-----------------------------|------------------------------|-------------------------------|--|--|--|
| Data | Output Explain      | n Messages N             | lotifications               |                              |                               |  |  |  |
| 4    | id<br>[PK] smallint | id_estado_final smallint | id_localizacion<br>smallint | id_tipo_contagio<br>smallint | id_estado_durante<br>smallint |  |  |  |
| 1    | 4112                | 4                        | 1                           | 1                            | 2                             |  |  |  |
| 2    | 4142                | 4                        | 1                           | 4                            | 2                             |  |  |  |
| 3    | 2211                | 2                        | 2                           | 1                            | 1                             |  |  |  |
| 4    | 2241                | 2                        | 2                           | 4                            | 1                             |  |  |  |
| 5    | 4414                | 4                        | 4                           | 1                            | 4                             |  |  |  |
| 6    | 3414                | 3                        | 4                           | 1                            | 4                             |  |  |  |
| 7    | 3444                | 3                        | 4                           | 4                            | 4                             |  |  |  |
| 8    | 4444                | 4                        | 4                           | 4                            | 4                             |  |  |  |
| 9    | 4132                | 4                        | 1                           | 3                            | 2                             |  |  |  |
| 10   | 1112                | 1                        | 1                           | 1                            | 2                             |  |  |  |
| 11   | 1142                | 1                        | 1                           | 4                            | 2                             |  |  |  |
| 12   | 1313                | 1                        | 3                           | 1                            | 3                             |  |  |  |
| 13   | 4122                | 4                        | 1                           | 2                            | 2                             |  |  |  |
| 14   | 1323                | 1                        | 3                           | 2                            | 3                             |  |  |  |
| 15   | 1343                | 1                        | 3                           | 4                            | 3                             |  |  |  |
| 16   | 2221                | 2                        | 2                           | 2                            | 1                             |  |  |  |

Figura 7) Consulta de la tabla estado

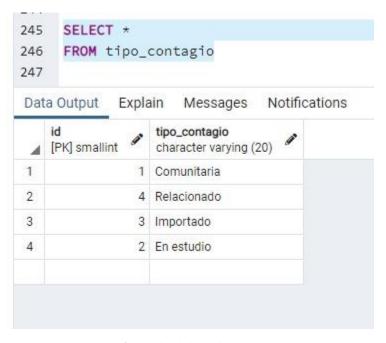


Figura 8) Consulta de la tabla tipo\_contagio



Figura 9) Consulta de la tabla estado\_final

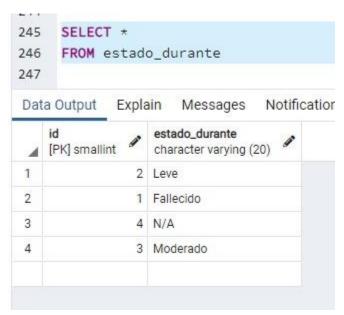


Figura 10) Consulta de la tabla estado\_durante



Figura 11) Consulta de la tabla etnia

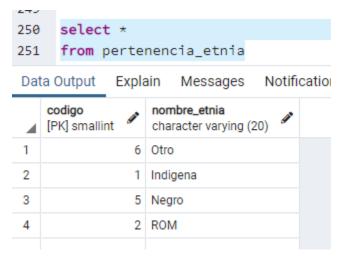


Figura 12) Consulta de la tabla pertenencia\_etnia

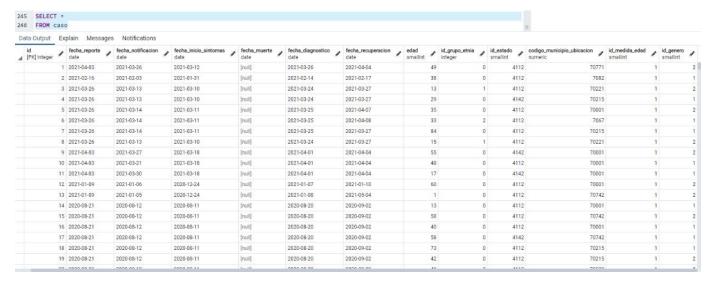


Figura 13) Consulta de la tabla caso

#### 4. Posibles escenarios de análisis

- Se analizará si la etnia influye en los casos de contagio, para ello se tendrá en cuenta la información de codigo\_etnia que corresponde a indígenas, ROM, raizal, palenquero, negro u otros; y estos se compararán con el comportamiento del contagio teniendo en cuenta las variables de inicio de síntomas y recuperación.
- Se observará en que edades se vuelve más probable que los pacientes se recuperen o no del COVID-19 y que fallezcan o no por causa del COVID-19, esto se realizará mediante los registros de fecha\_inicio\_sintomas, fecha\_diagnostico y fecha\_recuperacion o fecha\_muerte.
- Compararemos los casos de contagio con cada municipio para así identificar en cuál de los 26 municipios se presentaron más cantidad de contagios, cantidad de pacientes recuperados y cantidad de pacientes que fallecieron.
- Analizaremos la cantidad de casos que se presentaron, cuantos pacientes se recuperaron y cuantos fallecieron en los años 2020 y 2021, con esta información analizaremos los picos de la pandemia y como las medidas de bioseguridad y vacunas han influido.

## 5. Consulta de los datos de todas las tablas mediante Python.

Combinando las librerías psycopg2 y pandas de Python, se realizan las consultas a la base de datos por medio de la función read\_sql\_query, la cual a partir de un query permite

almacenar la tabla resultante como un dataframe de pandas, facilitando el manejo de los datos, y mostrar su visualización en la consola. Se adjuntan las consultas parciales de todas las tablas implementadas en el modelo relacional. Para ejecutar el modulo, solamente se debe ubicar en C:\ y ejecutar la siguiente línea de código y se conectará y mostrará los datos de cada tabla:

C:\python proyecto.py

El módulo de python se encuentra en el repositorio como **proyecto.py** 

```
:\Users\Laboratorios EICT\AppData\Local\Packages\Py
 nnection) ordatabase string URI or sqlite3 DBAPI2 c
 warnings.warn(
   codigo_município
                              nombre municipio
             70771.0
                                          SLICRE
             7082.0
70221.0
                                           TOLU
                                       COVENIAS
             70215.0
                                       COROZAL
                                      SINCELEJO
             70001.0
                                        SAMPUES
             70742.0
                                         SINCE
                                        PALMITO
             70523.8
                                  LOS PALMITOS
             70124.0
                                        CAIMITO
             70508.0
                                         OVEJAS
             70473.0
                                         MORROA
                      GALERAS (NUEVA GRANADA)
             70233.0
                                       EL ROBLE
                                    SAN MARCOS
             70708.0
              7011.0
                                    BUENAVISTA
16
17
18
19
20
              7023.0
                                         CHALAN
                                      MAJAGUAL
             70429.0
             70823.0
                                      TOLUVIEJO
             70717.0
                                      SAN PEDRO
21
             70678.0
                               SAN BENITO ABAD
             70204.0
                             COLOSO (RICAURTE)
               784.8
                                      LA UNION
             70265.0
                                       GUARANDA
                           SAN JUAN DE BETULIA
             70702.0
```

Figura 14) Visualización de la tabla ubicacion

```
genero
C:\Users\Laboratorios EICT\AppD
onnection) ordatabase string UR
warnings.warn(
id sexo
0 2 M
1 1 F
```

Figura 15) Visualización de la tabla genero

```
medida_edad
C:\Users\Laboratorios EICT\AppData\Loca
onnection) ordatabase string URI or sql
warnings.warn(
id unidad_medida
0 1 Anios
1 2 Meses
2 3 Semanas
```

Figura 16) Visualización de la tabla medida\_edad

Figura 17) Visualización de la tabla localizacion

Figura 18) Visualización de la tabla estado

Figura 19) Visualización de la tabla tipo\_contagio

Figura 20) Visualización de la tabla estado\_final

```
estado_durante
C:\Users\Laboratorios EICT\Ap
onnection) ordatabase string
warnings.warn(
   id estado_durante
8 2 Leve
1 1 Fallecido
2 4 N/A
3 3 Moderado
```

Figura 21) Visualización de la tabla estado\_durante

Figura 22) Visualización de la tabla etnia

Figura 23) Visualización de la tabla pertenencia\_etnia

Figura 24) Visualización de la tabla caso