# Portada.

Creación de una base de datos analítica Sistema de Gestión de Base de Datos de Casos Positivos de COVID-19 en Antioquia

Elaborado por: Carlos Dueiner Castaño Rodríguez

Facultad de Ingeniería y ciencias agropecuarias

Big Data

Presentado a: Andrés Felipe Callejas Jaramillo

Ingeniería de Software y Datos

Institución Universitaria Digital de Antioquia

Mayo 2025

Contenido

[Portada. 1](#_Toc197585804)

[Introducción. 2](#_Toc197585805)

[Descripción del problema. 2](#_Toc197585806)

[Objetivos. 3](#_Toc197585807)

[**Objetivo General:** 3](#_Toc197585808)

[**Objetivos Específicos:** 3](#_Toc197585809)

[Descripción de los datos disponibles. 3](#_Toc197585810)

[Solución propuesta (Elección del SGBD y esquema diseñado). 4](#_Toc197585811)

[Metodología empleada. 5](#_Toc197585812)

[Resultados y Conclusiones. 5](#_Toc197585813)

[Bibliografía. 6](#_Toc197585814)

# Introducción.

El COVID-19 representó uno de los mayores retos de salud pública en la historia reciente. En Colombia, y particularmente en Antioquia, la pandemia evidenció profundas falencias en la gestión y manejo de la información sanitaria. La dispersión de los datos, la falta de un sistema centralizado y la ausencia de herramientas robustas limitaron la capacidad de respuesta oportuna y efectiva de las autoridades durante los momentos más críticos de la emergencia.

Este documento aborda el diseño e implementación de un sistema de gestión de base de datos que centralice y organice la información de casos positivos de COVID-19 reportados en Antioquia entre junio y diciembre de 2020. Este proyecto busca ser una herramienta clave para reforzar la preparación ante eventuales crisis sanitarias futuras.

# Descripción del problema.

En el contexto de la pandemia de COVID-19, Antioquia, al igual que muchas regiones del mundo, enfrentó múltiples desafíos asociados a la gestión de información. Entre junio y diciembre de 2020, el número de casos positivos creció de manera exponencial. Sin embargo, la falta de una base de datos centralizada dificultó la recopilación, el almacenamiento y la integración de la información sanitaria, generando inconsistencias y retrasos en la implementación de medidas de control y mitigación.

Cinco años después, estas falencias resaltan la necesidad urgente de desarrollar herramientas tecnológicas que permitan la gestión eficaz de datos en tiempo real. Este sistema representa un esfuerzo por cerrar esa brecha, mejorando la capacidad de respuesta de la región frente a emergencias sanitarias futuras.

# Objetivos.

## Objetivo General:

Concebir e implementar un sistema centralizado de base de datos destinado a la gestión integral de los casos positivos de COVID-19 reportados en la geografía antioqueña entre junio y diciembre de 2020, con el firme propósito de optimizar los procesos de análisis epidemiológico y robustecer la toma de decisiones estratégicas en el ámbito de la salud pública.

## Objetivos Específicos:

* Identificar meticulosamente los datos que revisten relevancia para el análisis epidemiológico, asegurando en cada etapa su calidad e integridad.
* Diseñar un modelo de datos relacional que, con su lógica inherente, facilite la consulta ágil y el análisis eficiente de la vasta información contenida.
* Seleccionar un Sistema de Gestión de Bases de Datos (SGBD) que se ajuste a las necesidades del proyecto, priorizando criterios de escalabilidad que permitan su crecimiento futuro y una interfaz de usuario que favorezca su adopción y usabilidad.
* Implementar de manera precisa el modelo de datos previamente diseñado en el SGBD escogido, garantizando una migración de la información que preserve su fidelidad y exactitud.
* Validar exhaustivamente el sistema a través de pruebas funcionales rigurosas que certifiquen su operatividad sin fisuras y la precisión de los datos que alberga.

# Descripción de los datos disponibles.

La materia prima fundamental para este proyecto reside en los reportes oficiales emanados del Gobierno Nacional de Colombia, en el apartado de Salud y Protección Social, documentos que atesoran información de vital importancia para la comprensión epidemiológica del fenómeno. Entre los campos de datos que emergen como pilares del análisis, se destacan:

* **Identificadores:** La huella digital del reporte web y el ID único asignado a cada caso, elementos esenciales para la trazabilidad.
* **Información Temporal:** Las fechas cruciales de notificación, inicio de los síntomas, confirmación diagnóstica, recuperación (cuando acontece) y el luctuoso deceso (en los casos correspondientes).
* **Datos Demográficos:** La edad, el sexo biológico y la autoidentificación de la pertenencia étnica, variables fundamentales para identificar patrones y desigualdades.
* **Localización Geográfica:** El nombre del municipio donde se reporta el caso y su respectivo código DIVIPOLA, permitiendo análisis espaciales detallados.
* **Estado del Caso:** La tipología del contagio (importado, comunitario, etc.), la ubicación del paciente durante la enfermedad (hospitalización, domicilio, etc.), la gravedad del cuadro clínico (leve, grave, fallecido) y la condición final de recuperación.

Este conjunto de datos, en su integralidad, dibuja un panorama detallado de la dinámica con la que el virus se desenvolvió en nuestra región, constituyéndose en la piedra angular sobre la cual se erige la construcción del sistema propuesto.

# Solución propuesta (Elección del SGBD y esquema diseñado).

**Elección del SGBD:**

Se seleccionó SQLite como sistema de gestión de bases de datos debido a su simplicidad, portabilidad y capacidad para manejar bases de datos locales de forma eficiente. Esta elección fue estratégica para un sistema que requería una implementación rápida, con recursos mínimos, pero que pudiera garantizar la integridad y accesibilidad de los datos.

SQLite ofrece un modelo relacional que asegura la compatibilidad con las mejores prácticas en diseño de bases de datos. Además, al operar sin necesidad de un servidor, su instalación y mantenimiento son sencillos, lo que resulta ideal para proyectos pequeños o medianos.

**Esquema Diseñado:**

El modelo relacional propuesto consta de las siguientes tablas:

1. **Casos**:
   * ID\_Caso (PK)
   * Fecha\_Reporte
   * Fecha\_Notificacion
   * Estado
   * Recuperado
   * Tipo\_Recuperacion
   * Fecha\_Inicio\_Sintomas
   * Fecha\_Diagnostico
   * Fecha\_Recuperacion
   * Fecha\_Muerte
2. **Pacientes**:
   * ID\_Paciente (PK)
   * Edad
   * Sexo
   * Pertenencia\_Etnica
3. **Municipios**:
   * ID\_Municipio (PK)
   * Nombre
   * Codigo\_DIVIPOLA
4. **Contagios**:
   * ID\_Contagio (PK)
   * ID\_Caso (FK)
   * ID\_Paciente (FK)
   * ID\_Municipio (FK)
   * Tipo\_Contagio

# Metodología empleada.

1. **Identificación de Datos:** Se inició con un análisis exhaustivo de los datos disponibles para seleccionar los campos clave que garantizarían la relevancia y calidad de la información almacenada. Se priorizó la integridad de los datos, asegurando que cada campo tuviera una función clara y definida dentro del modelo.
2. **Diseño del Modelo de Datos:** Basándonos en los principios de normalización, se diseñó un modelo relacional que minimizara las redundancias y asegurara la integridad referencial entre las tablas. Este diseño consideró tanto la consulta eficiente como la escalabilidad futura del sistema.
3. **Implementación del Sistema:**
   * Se instaló SQLite como el SGBD seleccionado.
   * Se definieron las tablas y relaciones conforme al esquema diseñado.
   * Utilizando scripts automatizados, se procedió a la carga de los datos al sistema, validando cada paso para garantizar su precisión.
4. **Validación:**
   * Se ejecutaron pruebas funcionales que incluyeron consultas básicas y complejas, asegurando que los datos se recuperaran y presentaran de manera correcta.
   * También se evaluaron las relaciones entre tablas para garantizar su coherencia.

# Resultados y Conclusiones.

El sistema implementado logró centralizar los datos de casos positivos de COVID-19 en Antioquia, transformando un conjunto de datos dispersos en una base de datos organizada y de fácil acceso. Esta herramienta no solo permite consultas eficientes, sino que también establece un precedente para la gestión de información sanitaria en la región.

La elección de SQLite demostró ser acertada, proporcionando una solución económica, simple y eficaz. Aunque el sistema fue diseñado inicialmente para un conjunto de datos específico, su estructura es adaptable a otros contextos y periodos.

Finalmente, este proyecto destaca la importancia de la tecnología en la preparación y respuesta ante crisis sanitarias, subrayando la necesidad de inversiones sostenidas en infraestructura de datos para fortalecer los sistemas de salud pública.

# Bibliografía.

Python Software Foundation. (2025). *Python.org*. Obtenido de Python: https://www.python.org/

GitHub, Inc. (2025). *github.com*. Obtenido de Github: https://github.com/

Ministerio de Tecnologías de la Información y las Comunicaciones . (2025). *datos.gov.co*. Obtenido de Datos Abiertos de Colombia: https://www.datos.gov.co/