

## **Fragmentación Horizontal Primaria y Derivada de la base datos CovidHistorico.**

### **Prompt utilizado en Deep research:**

con este prompt que había hecho para una fragmentación horizontal de otra base de datos :  
prompt : Genera una fragmentación de forma horizontal primaria de acuerdo con la teoría del autor tamer orzu, considerando las tablas Productsubcategory, products, salesorderdetail. Esas tablas provienen de la base de datos de ejemplo de SQL server adventureworks.

toma en cuenta el algoritmo com min y las siguientes consultas para poder desarrollar la fragmentación:

- a. Listar el producto más vendido de cada una de las categorías registradas en la base de datos.
- b. Listar el nombre de los clientes con más ordenes por cada uno de los territorios registrados en la base de datos.
- c. Listar los datos generales de las ordenes que tengan al menos los mismos productos de la orden con salesorderid = 43676

Te adjunto el esquema de las tablas en una foto.

Ahora bien quiero hacer otra descripción y otra fragmentacion horizontal con la teoría del autor TAMER M. OZSUy con el algoritmo CON MIN para otra base de datos, que es CovidHistorico, Solo hay dos tablas en la base de datos: datoscovid y cat\_entidades. se había propuesto una fragmentación por entidades, entonces cada nodo se dividía en regiones como "oeste,suroeste, noroeste,noreste" y demás, pero las regiones pues no proponen una mejora para las consultas que te anexo. entonces quiero que me propongas la mejor manera o el criterio por el cual seria mejor fragmentar.

estas son las consultas a analizar para poder proponer el esquema de fragmentación :

Número 1. Listar el top 5 de las entidades con más casos confirmados por cada uno de los años registrados en la base de datos.

Número 2. Listar el municipio con más casos confirmados recuperados por estado y por año.

Número 3. Listar el porcentaje de casos confirmados en cada una de las siguientes morbilidades a nivel nacional: diabetes, obesidad e hipertensión.

Número 4. Listar los municipios que no tengan casos confirmados en todas las morbilidades: hipertensión, obesidad, diabetes, tabaquismo.

Número 4. Listar los municipios que no tengan casos confirmados en todas las morbilidades: hipertensión, obesidad, diabetes, tabaquismo.

Número 5. Listar los estados con más casos recuperados con neumonía.

Número 6. Listar el total de casos confirmados/sospechosos por estado en cada uno de los años registrados en la base de datos.

Número 7. Para el año 2020 y 2021 cuál fue el mes con más casos registrados, confirmados, sospechosos, por estado registrado en la base de datos.

Número 8. Listar el municipio con menos defunciones en el mes con más casos confirmados con neumonía en los años 2020 y 2021.

Número 9. Listar el top 3 de municipios con menos casos recuperados en el año 2021.

Número 10. Listar el porcentaje de casos confirmado por género en los años 2020 y 2021.

Número 11. Listar el porcentaje de casos hospitalizados por estado en el año 2020.

Requisitos: Crear una consulta que calcule el porcentaje de pacientes hospitalizados por estado en el año 2020, considerando únicamente aquellos registros con ingreso en dicho año.

Número 12. Listar total de casos negativos por estado en los años 2020 y 2021.

Número 13. Listar porcentajes de casos confirmados por género en el rango de edades de 20 a 30 años, de 31 a 40 años, de 41 a 50 años, de 51 a 60 años y mayores a 60 años a nivel nacional.

Número 14. Listar el rango de edad con más casos confirmados y que fallecieron en los años 2020 y 2021.

Requisitos:

Se utilizó la tabla principal de la base de datos denominada datoscovid, aplicando un CASE para agrupar las edades en rangos específicos, lo cual permitió analizar de forma detallada la distribución de los casos.

## 1. Análisis del Esquema Relacional Objetivo

En esta sección, realizamos un análisis de las tablas principales involucradas en la base de datos CovidHistorico, que son:

- **DatosCovid:** Contiene los registros de los casos de COVID-19, incluyendo información sobre la clasificación de los casos (confirmados, sospechosos) y las morbilidades asociadas (diabetes, obesidad, hipertensión, tabaquismo).
- **cat\_entidades:** Contiene información sobre las entidades geográficas (estados,

municipios) en las que se registran los casos de COVID-19.

Las consultas que vamos a analizar dependen principalmente de filtros por morbilidades (diabetes, obesidad, hipertensión, tabaquismo) y clasificación final de los casos (confirmados, comité de dictaminación, asociación clínica).

## **2. Deconstrucción de las Consultas de Alta Frecuencia**

Las consultas más relevantes que vamos a considerar son las siguientes:

- Consulta 3: Listar el porcentaje de casos confirmados en cada una de las siguientes morbilidades a nivel nacional: diabetes, obesidad e hipertensión.
- Consulta 4: Listar los municipios que no tengan casos confirmados en todas las morbilidades: hipertensión, obesidad, diabetes, tabaquismo.
- Consulta 5: Listar los estados con más casos recuperados con neumonía.
- Consulta 7: Para el año 2020 y 2021, ¿cuál fue el mes con más casos registrados, confirmados, sospechosos, por estado registrado en la base de datos?

### **Implicación para la Fragmentación:**

1. Consulta 3: Requiere filtrar los casos por morbilidades específicas (diabetes, obesidad e hipertensión) y su clasificación final.
2. Consulta 4: Necesita identificar municipios sin casos de morbilidades específicas como diabetes, obesidad, hipertensión y tabaquismo, lo que implica un filtrado por varias morbilidades.
3. Consulta 5: Se basa en los casos de neumonía, que deben estar filtrados por clasificación final y, posiblemente, por morbilidad.
4. Consulta 7: Necesita agrupar los datos por estado y mes. Aunque no filtra por morbilidades directamente, sí requiere una buena segmentación por clasificación final.

## **Definición del Conjunto Inicial de Predicados Simples (Pr).**

Basado en el análisis de las consultas anteriores, los predicados simples para la fragmentación de la tabla DatosCovid son los siguientes:

- Morbilidades:
  - p1: SI\_NO = 1 (Diabetes)
  - p2: SI\_NO = 2 (Obesidad)
  - p3: SI\_NO = 3 (Hipertensión)
  - p4: SI\_NO = 4 (Tabaquismo)
- Clasificación de los Casos:
  - p5: CLASIFICACION\_FINAL = 1 (Casos confirmados)
  - p6: CLASIFICACION\_FINAL = 2 (Comité de dictaminación)
  - p7: CLASIFICACION\_FINAL = 3 (Asociación clínica-epidemiológica)

## **Construcción de Predicados Minterm.**

Los minterms se generan combinando los predicados simples, asegurando que los fragmentos sean mutuamente excluyentes. A continuación, se definen los minterms para la base de datos CovidHistorico, considerando las morbilidades y la clasificación final de los casos.

- m1: SI\_NO = 1 AND CLASIFICACION\_FINAL = 1 (Diabetes y Casos Confirmados)
- m2: SI\_NO = 2 AND CLASIFICACION\_FINAL = 2 (Obesidad y Comité de Dictaminación)
- m3: SI\_NO = 3 AND CLASIFICACION\_FINAL = 3 (Hipertensión y Asociación Clínica)
- m4: SI\_NO = 4 AND CLASIFICACION\_FINAL = 1 (Tabaquismo y Casos Confirmados)
- m5: SI\_NO = 1 AND CLASIFICACION\_FINAL = 2 (Diabetes y Comité de Dictaminación)
- m6: SI\_NO = 2 AND CLASIFICACION\_FINAL = 3 (Obesidad y Asociación Clínica)
- m7: SI\_NO = 3 AND CLASIFICACION\_FINAL = 1 (Hipertensión y Casos Confirmados)
- m8: SI\_NO = 4 AND CLASIFICACION\_FINAL = 2 (Tabaquismo y Comité de Dictaminación)
- m9: SI\_NO = 1 AND CLASIFICACION\_FINAL = 3 (Diabetes y Asociación Clínica)
- m10: SI\_NO = 2 AND CLASIFICACION\_FINAL = 1 (Obesidad y Casos Confirmados)
- m11: SI\_NO = 3 AND CLASIFICACION\_FINAL = 2 (Hipertensión y Comité de Dictaminación)
- m12: SI\_NO = 4 AND CLASIFICACION\_FINAL = 3 (Tabaquismo y Asociación Clínica)

## **Definición de los Fragmentos Horizontales Primarios.**

Los fragmentos horizontales se definen aplicando los predicados minterm correspondientes a la tabla DatosCovid. A continuación, se definen los fragmentos:

- **Fragmento 1 (Diabetes y Casos Confirmados):**
  - Datos: Seleccionar aquellos donde SI\_NO = 1 y CLASIFICACION\_FINAL = 1.
- **Fragmento 2 (Obesidad y Comité de Dictaminación):**
  - Datos: Seleccionar aquellos donde SI\_NO = 2 y CLASIFICACION\_FINAL = 2.
- **Fragmento 3 (Hipertensión y Asociación Clínica):**
  - Datos: Seleccionar aquellos donde SI\_NO = 3 y CLASIFICACION\_FINAL = 3.
- **Fragmento 4 (Tabaquismo y Casos Confirmados):**
  - Datos: Seleccionar aquellos donde SI\_NO = 4 y CLASIFICACION\_FINAL = 1.
- **Fragmento 5 (Diabetes y Comité de Dictaminación):**
  - Datos: Seleccionar aquellos donde SI\_NO = 1 y CLASIFICACION\_FINAL = 2.
- **Fragmento 6 (Obesidad y Asociación Clínica):**
  - Datos: Seleccionar aquellos donde SI\_NO = 2 y CLASIFICACION\_FINAL = 3.
- **Fragmento 7 (Hipertensión y Casos Confirmados):**
  - Datos: Seleccionar aquellos donde SI\_NO = 3 y CLASIFICACION\_FINAL = 1.
- **Fragmento 8 (Tabaquismo y Comité de Dictaminación):**

- Datos: Seleccionar aquellos donde SI\_NO = 4 y CLASIFICACION\_FINAL = 2.
- **Fragmento 9 (Diabetes y Asociación Clínica):**
  - Datos: Seleccionar aquellos donde SI\_NO = 1 y CLASIFICACION\_FINAL = 3.
- **Fragmento 10 (Obesidad y Casos Confirmados):**
  - Datos: Seleccionar aquellos donde SI\_NO = 2 y CLASIFICACION\_FINAL = 1.
- **Fragmento 11 (Hipertensión y Comité de Dictaminación):**
  - Datos: Seleccionar aquellos donde SI\_NO = 3 y CLASIFICACION\_FINAL = 2.
- **Fragmento 12 (Tabaquismo y Asociación Clínica):**
  - Datos: Seleccionar aquellos donde SI\_NO = 4 y CLASIFICACION\_FINAL = 3.

## **Fragmentación Derivada.**

La fragmentación derivada se aplica a las tablas relacionadas como cat\_entidades, para asegurar que las relaciones entre las tablas dependientes también se mantengan correctamente fragmentadas. Dado que cat\_entidades contiene información sobre las entidades donde se registran los casos de COVID-19, se fragmenta según los mismos predicados que se aplican a DatosCovid. De esta manera, los fragmentos de cat\_entidades estarán alineados con los fragmentos de DatosCovid, asegurando una fragmentación coherente entre las tablas.

## **Justificación de la Fragmentación Derivada vs. Primaria.**

Dado que las tablas como cat\_entidades no requieren fragmentación propia por sus atributos (ya que su relación con DatosCovid es directa), la fragmentación derivada es la opción correcta. Esta asegura que cat\_entidades se fragmenta siguiendo los mismos predicados que DatosCovid, garantizando la coherencia entre las tablas relacionadas.

## **Esquema de Fragmentación Completo.**

El diseño final resulta en un esquema donde las tablas DatosCovid y cat\_entidades están particionadas horizontalmente en fragmentos de acuerdo con los predicados definidos:

- **Fragmentos de DatosCovid (Primarios):**
  - **Fragmento 1:** SI\_NO = 1 AND CLASIFICACION\_FINAL = 1 (Diabetes y Casos Confirmados)
  - **Fragmento 2:** SI\_NO = 2 AND CLASIFICACION\_FINAL = 2 (Obesidad y Comité de Dictaminación)
  - **Fragmento 3:** SI\_NO = 3 AND CLASIFICACION\_FINAL = 3 (Hipertensión y Asociación Clínica)
  - **Fragmento 4:** SI\_NO = 4 AND CLASIFICACION\_FINAL = 1 (Tabaquismo y Casos Confirmados)
  - **Fragmento 5:** SI\_NO = 1 AND CLASIFICACION\_FINAL = 2 (Diabetes y Comité de Dictaminación)
  - **Fragmento 6:** SI\_NO = 2 AND CLASIFICACION\_FINAL = 3 (Obesidad y Asociación Clínica)

- **Fragmento 7:** SI\_NO = 3 AND CLASIFICACION\_FINAL = 1 (Hipertensión y Casos Confirmados)
- **Fragmento 8:** SI\_NO = 4 AND CLASIFICACION\_FINAL = 2 (Tabaquismo y Comité de Dictaminación)
- **Fragmento 9:** SI\_NO = 1 AND CLASIFICACION\_FINAL = 3 (Diabetes y Asociación Clínica)
- **Fragmento 10:** SI\_NO = 2 AND CLASIFICACION\_FINAL = 1 (Obesidad y Casos Confirmados)
- **Fragmento 11:** SI\_NO = 3 AND CLASIFICACION\_FINAL = 2 (Hipertensión y Comité de Dictaminación)
- **Fragmento 12:** SI\_NO = 4 AND CLASIFICACION\_FINAL = 3 (Tabaquismo y Asociación Clínica)
- 
- **Fragmentos Derivados de cat\_entidades:**
  - Los fragmentos de **cat\_entidades** se derivan de los fragmentos de **DatosCovid**, asegurando que las consultas sobre las entidades y municipios se beneficien de la misma fragmentación.

#### **Fragmentación combinada (Por Clasificación Final y Morbilidades):**

1. Mejor adaptación a las consultas: Dado que las consultas que me proporcionaste requieren un análisis detallado tanto de morbilidades como de clasificación final, una fragmentación que combine ambos atributos sería la más adecuada para optimizar el rendimiento de todas las consultas. Esto permitirá una segmentación más fina de los datos, mejorando la eficiencia de las consultas que filtran por morbilidad y las que filtran por clasificación final.
2. Optimización del rendimiento: Aunque la fragmentación combinada implica más fragmentos, las consultas específicas sobre morbilidades (por ejemplo, porcentaje de casos por morbilidad, listado de municipios sin casos de determinadas morbilidades) se beneficiarán de esta fragmentación, ya que solo accederán a los datos relacionados con la morbilidad deseada.
3. Flexibilidad: Si bien algunas consultas solo filtrarán por clasificación final, otras necesitarán el desglose por morbilidad. La fragmentación combinada ofrece la flexibilidad para manejar ambos tipos de consulta de manera eficiente.

## **Conclusión.**

En el estudio de la fragmentación horizontal de la base de datos CovidHistorico, se evaluaron dos enfoques: uno basado en la fragmentación por regiones (que engloban varias entidades por cada región), y otro basado en la fragmentación combinada por clasificación final y morbilidades. Tras un análisis detallado de las consultas, se vio que la fragmentación por

regiones no era la adecuada para las consultas requeridas ya que los resultados que solicitaban las consultas no necesitaban ser filtradas por alguna entidad en específico solo requerían resultados que englobaran a toda la población de los estados de la república, mientras que la fragmentación combinada por clasificación final y morbilidades es más adecuada para poder hacer más eficientes las consultas.

La fragmentación combinada garantiza que los fragmentos contengan solo los datos relevantes según el tipo de caso y morbilidad deseados. Al hacerlo, se optimiza el procesamiento de las consultas, ya que cada fragmento estará más alineado con los filtros aplicados, reduciendo el volumen de datos a procesar y mejorando la rapidez en la ejecución de las consultas.