



INSTITUTO POLITECNICO NACIONAL

**UNIDAD PROFESIONAL INTERDISCIPLINARIA EN
INGENIERIA Y TECNOLOGIAS AVANZADAS**

**Tarea 4: Propuesta de Fragmentación Primaria y
Derivada para AdventureWorks (ChatGPT)**

**INGENIERIA TELEMÁTICA
BASE DE DATOS DISTRIBUIDAS**

PROFESOR: De la Cruz Sosa Carlos

ALUMNOS:

Legorreta Rodríguez María Fernanda

Macías Galván Arturo Daniel

Palacios Reyes Leslie Noemi

GRUPO: 3TM3

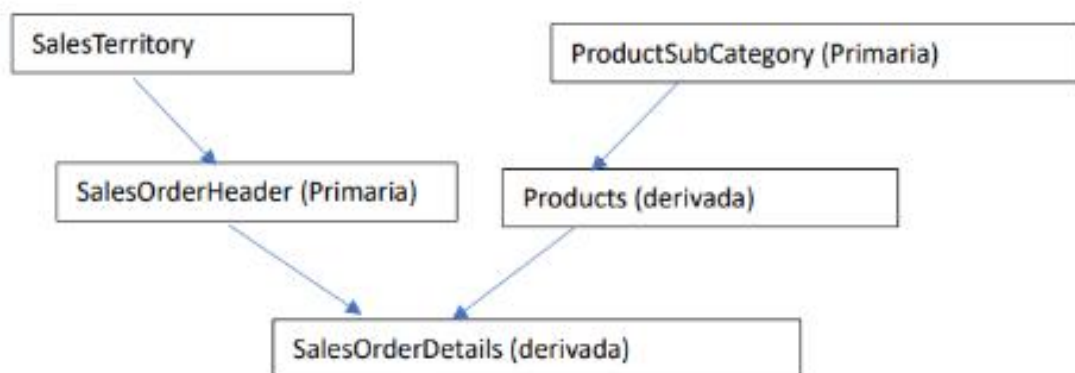
SEMESTRE 2025/2

Consulta inicial a ChatGPT

Para comprender y aplicar correctamente el algoritmo conmin, realizamos una consulta a ChatGPT solicitando una explicación detallada y los pasos para diseñar una fragmentación primaria basada en esta metodología. También preguntamos cómo identificar los predicados mínimos para garantizar que los fragmentos, basado en la siguiente figura:

FRAGMENTACIÓN HORIZONTAL PRIMARIA

Ejemplo de fragmentación horizontal primaria basada en el análisis de las consultas que se ejecutan con frecuencia en un segmento de tablas de la base de datos AdventureWorks



Para iniciar el proceso de fragmentación se requiere: modelo relacional a particionar, conjunto de consultas que se ejecutan con mayor frecuencia y se accede por grupos de usuarios.

1. Consultas

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

Figura 1: Diagrama con tablas de datos de AdventureWorks

Lo que le preguntamos a ChatGPT, fue lo siguiente, se muestra en la siguiente figura:

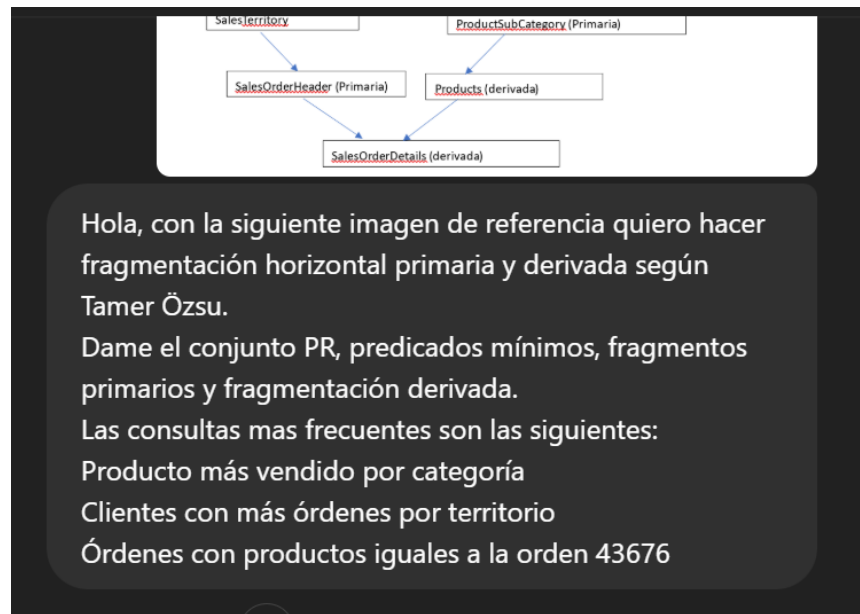


Figura 2: Prompt de la consulta a ChatGPT

Respuesta recibida de ChatGPT

ChatGPT nos proporcionó una guía teórica clara, donde destacó los siguientes puntos clave:

1. Identificación de Tablas Clave

Según el diagrama y las consultas:

- **Primarias candidatas** (por criterios de segmentación directa):
 - ProductSubCategory (por categorías → Consulta 1)
 - SalesOrderHeader (por territorio → Consulta 2)
- **Derivadas** (dependen de las primarias por FK o JOIN):
 - Products (depende de ProductSubCategory)
 - SalesOrderDetails (depende de SalesOrderHeader y Products)
 - SalesTerritory (se consulta indirectamente desde SalesOrderHeader)

2. Conjunto PR: Predicados de fragmentación

Fragmentación primaria de SalesOrderHeader basada en territorio:

Supongamos 3 territorios frecuentes:

- Northwest
- Northeast
- Central

Entonces el **PR para SalesOrderHeader** será:

plaintext

CopiarEditar

```
PR_SalesOrderHeader = {  
  p1: TerritoryID = 1 (Northwest),  
  p2: TerritoryID = 2 (Northeast),  
  p3: TerritoryID = 3 (Central)  
}
```

Fragmentación primaria de ProductSubCategory basada en categoría:

Supongamos las categorías:

- Bikes
- Accessories
- Clothing

Entonces el **PR para ProductSubCategory** será:

plaintext

CopiarEditar

```
PR_ProductSubCategory = {  
  q1: CategoryID = 1 (Bikes),  
  q2: CategoryID = 2 (Accessories),  
  q3: CategoryID = 3 (Clothing)  
}
```

3. Fragmentos Primarios

Fragmentos de SalesOrderHeader (Primaria):

plaintext

CopiarEditar

F1: $\sigma_{\text{TerritoryID}=1}$ (Northwest)
F2: $\sigma_{\text{TerritoryID}=2}$ (Northeast)
F3: $\sigma_{\text{TerritoryID}=3}$ (Central)

Fragmentos de ProductSubCategory (Primaria):

plaintext

CopiarEditar

G1: $\sigma_{\text{CategoryID}=1}$ (Bikes)
G2: $\sigma_{\text{CategoryID}=2}$ (Accessories)
G3: $\sigma_{\text{CategoryID}=3}$ (Clothing)

4. Fragmentación Derivada

Para Products (depende de ProductSubCategory)

plaintext

CopiarEditar

Derivada de PR_ProductSubCategory:

Products_G1: Productos donde ProductSubCategoryID \in G1
Products_G2: Productos donde ProductSubCategoryID \in G2
Products_G3: Productos donde ProductSubCategoryID \in G3

Para SalesOrderDetails (depende tanto de SalesOrderHeader como de Products)

Se hace **derivación compuesta**:

plaintext

CopiarEditar

SalesOrderDetails_F1_G1: Detalles donde OrderID \in F1 y ProductID \in Products_G1

SalesOrderDetails_F1_G2: ...

SalesOrderDetails_F2_G1: ...

...

SalesOrderDetails_F3_G3

Resumen

► PR (Predicados Mínimos):

Tabla	Predicados
SalesOrderHeader	TerritoryID = 1, 2, 3
ProductSubCategory	CategoryID = 1, 2, 3

► Fragmentos Primarios:

Fragmento Definición

F1	SalesOrderHeader donde TerritoryID = 1
F2	SalesOrderHeader donde TerritoryID = 2
F3	SalesOrderHeader donde TerritoryID = 3
G1	ProductSubCategory donde CategoryID = 1
G2	ProductSubCategory donde CategoryID = 2
G3	ProductSubCategory donde CategoryID = 3

► Fragmentación Derivada:

Tabla	Fragmentos Derivados
Products	Products_G1, Products_G2, Products_G3
SalesOrderDetails	SalesOrderDetails_Fx_Gy (combinaciones)

Conclusiones

A partir del análisis de las consultas más frecuentes sobre la base de datos AdventureWorks, se identificaron las tablas ProductSubCategory y SalesOrderHeader como candidatas para una fragmentación horizontal primaria, ya que contienen los atributos clave: productcategoryid y territoryid. A estas tablas se les aplicó un conjunto de predicados simples (PR), que luego fueron refinados para obtener un conjunto completo y mínimo de predicados mutuamente excluyentes (M), garantizando una fragmentación correcta y no redundante.

La fragmentación primaria se realizó sobre ProductSubCategory, por ser la que genera menos fragmentos y agrupa más tuplas, considerando que la mayoría de las consultas acceden a los productos por categoría. A partir de estos fragmentos primarios, se desarrollaron fragmentaciones derivadas en las tablas relacionadas: Product, SalesOrderDetail y SalesOrderHeader, utilizando semijoins para mantener la coherencia entre los fragmentos.

Esta estrategia de fragmentación horizontal permite una mejor localización de datos, reducción de volumen de procesamiento y mejora del rendimiento al ejecutar consultas específicas por categoría o territorio. Además, cumple con los criterios propuestos por Tamer Özsu para una fragmentación eficiente en sistemas distribuidos