

Consultas en la Base de datos no fragmentada de AdventureWorks

1. La Información de los clientes se debe almacenar por región, considerando las regiones de acuerdo con el atributo Group de SalesTerritory

Predicados de la tabla Customer en base al atributo TerritoryID

```
PrCustomer{  
  P1: TerritoryID = 1  
  P2: TerritoryID = 2 :  
  P10: TerritoryID = 10  
}
```

Predicados minitermino

```
MrCustomer{  
  M1: P1 ∧ P2 ∧ P3 ∧ P4 ∧ P5 ∧ P6 ∧ ¬P7 ∧ ¬P8 ∧ ¬P9 ∧ ¬P10 → Región North America  
  M2: (¬P1) ∧ (¬P2) ∧ (¬P3) ∧ (¬P4) ∧ (¬P5) ∧ (¬P6) ∧ P7 ∧ P8 ∧ (¬P9) ∧ P10 → Región Europe  
  M3: (¬P1) ∧ (¬P2) ∧ (¬P3) ∧ (¬P4) ∧ (¬P5) ∧ (¬P6) ∧ (¬P7) ∧ (¬P8) ∧ P9 ∧ (¬P10) → Region Pacific  
  M4: (¬P1) ∧ (¬P2) ∧ (¬P3) ∧ (¬P4) ∧ (¬P5) ∧ (¬P6) ∧ (¬P7) ∧ (¬P8) ∧ (¬P9) ∧ (¬P10) → Resto  
}
```

$Sales.Customer_1 = \sigma_{1 \leq TerritoryID \leq 6}(Sales.Customer)$

$Sales.Customer_2 = \sigma_{7 \leq TerritoryID \leq 8 || TerritoryID = 10}(Sales.Customer)$

$Sales.Customer_3 = \sigma_{TerritoryID = 9}(Sales.Customer)$

$Sales.Customer_4 = \sigma_{1 > TerritoryID > 10}(Sales.Customer)$

2. Listar datos del empleado que atendió más ordenes por territorio.

```
PrSalesOrderHeader{  
  P1: TerritoryID = 1  
  P2: TerritoryID = 2  
  :  
  P10: TerritoryID = 10  
}
```

MrSalesOrderHeader{

$M1: P1 \wedge \neg P2 \wedge \neg P3 \wedge \neg P4 \wedge \neg P5 \wedge \neg P6 \wedge \neg P7 \wedge \neg P8 \wedge \neg P9 \wedge \neg P10$

$M2: \neg P1 \wedge P2 \wedge \neg P3 \wedge \neg P4 \wedge \neg P5 \wedge \neg P6 \wedge \neg P7 \wedge \neg P8 \wedge \neg P9 \wedge \neg P10$

$M3: \neg P1 \wedge \neg P2 \wedge P3 \wedge \neg P4 \wedge \neg P5 \wedge \neg P6 \wedge \neg P7 \wedge \neg P8 \wedge \neg P9 \wedge \neg P10$

$M4: \neg P1 \wedge \neg P2 \wedge \neg P3 \wedge P4 \wedge \neg P5 \wedge \neg P6 \wedge \neg P7 \wedge \neg P8 \wedge \neg P9 \wedge \neg P10$

$M5: \neg P1 \wedge \neg P2 \wedge \neg P3 \wedge \neg P4 \wedge P5 \wedge \neg P6 \wedge \neg P7 \wedge \neg P8 \wedge \neg P9 \wedge \neg P10$

$M6: \neg P1 \wedge \neg P2 \wedge \neg P3 \wedge \neg P4 \wedge \neg P5 \wedge P6 \wedge \neg P7 \wedge \neg P8 \wedge \neg P9 \wedge \neg P10$

$M7: \neg P1 \wedge \neg P2 \wedge \neg P3 \wedge \neg P4 \wedge \neg P5 \wedge \neg P6 \wedge P7 \wedge \neg P8 \wedge \neg P9 \wedge \neg P10$

$M8: \neg P1 \wedge \neg P2 \wedge \neg P3 \wedge \neg P4 \wedge \neg P5 \wedge \neg P6 \wedge \neg P7 \wedge P8 \wedge \neg P9 \wedge \neg P10$

$M9: \neg P1 \wedge \neg P2 \wedge \neg P3 \wedge \neg P4 \wedge \neg P5 \wedge \neg P6 \wedge \neg P7 \wedge \neg P8 \wedge P9 \wedge \neg P10$

$M10: \neg P1 \wedge \neg P2 \wedge \neg P3 \wedge \neg P4 \wedge \neg P5 \wedge \neg P6 \wedge \neg P7 \wedge \neg P8 \wedge \neg P9 \wedge P10$

}

Se fragmentaramos por territorio tendríamos 10 fragmentos para contabilizar más fácil las ventas por territorio.

Si listáramos los datos de los empleados y tomáramos ese fragmento tendríamos dos fragmentos en SalesPerson.

3. Listar los datos del cliente con más ordenes solicitadas en la región "north america"

PrSalesOrderHeader{

$P1: TerritoryID = 1$

$P2: TerritoryID = 2$

⋮

$P6: TerritoryID = 6$

}

MrSalesOrderHeader{

$M1: P1 \wedge P2 \wedge P3 \wedge P4 \wedge P5 \wedge P6 \rightarrow \text{Región North America}$

$M2: (\neg P1) \wedge (\neg P2) \wedge (\neg P3) \wedge (\neg P4) \wedge (\neg P5) \wedge (\neg P6) \rightarrow \text{Resto de regiones}$

}

$Sales.OrderHeader_1 = \sigma_{1 \leq TerritoryID \leq 6}(Sales.OrderHeader)$

$Sales.OrderHeader_2 = \sigma_{TerritoryID > 6}(Sales.OrderHeader)$

4. Listar el producto más solicitado en la región "europe"

Predicados de la tabla SalesOrderHeader en base a TerritoryID

PrSalesOrderHeader{

$P1: TerritoryID = 7$

$P2: TerritoryID = 8$

$P3: TerritoryID = 10$

}

Predicados Minitermino

MSalesOrderHeader{

$M1: (P1) \wedge (P2) \wedge (P3) \rightarrow \text{Región Europa}$

$M2: (\neg P1) \wedge (\neg P2) \wedge (\neg P3) \rightarrow \text{Demás regiones}$

}

$Sales.SalesOrderHeader = \sigma_{7 \leq TerritoryID \leq 8 || TerritoryID = 10}(Sales.SalesOrderHeader)$

5. Listar las ofertas que tienen los productos de la categoría "bikes"

PrProduct{

$P1: ProductSubcategoryID = 1 // \text{Subcategoría "Bikes"}$

$P2: ProductSubcategoryID = 2 // \text{Subcategoría "Bikes"}$

$P3: ProductSubcategoryID = 3 // \text{Subcategoría "Bikes"}$

}

MrProductSubcategory{

$M1: P1 // \text{Productos Categorías Bike}$

$M2: \neg P1 // \text{Demás categorías}$

}

MrProduct{

$M1: P1 \wedge P2 \wedge P3 // \text{Productos Subcategorías Bike}$

$M2: \neg P1 \wedge \neg P2 \wedge \neg P3 // \text{Demás categorías}$

}

$FSOP = SpecialOfferProduct \bowtie (\sigma_{1 \leq ProductSubcategoryID \leq 3}(Product))$

6. Listar los 3 productos menos solicitados en la región "pacific"

PrSalesOrderOrderHeader{

$P1: TerritoryID = 9$

}

Predicados miniterminos

MSalesOrderHeader{

$M1 = P1 \rightarrow \text{Región Pacific}$

$M2 = \neg P2 \rightarrow \text{Demás regiones}$

}

$\text{Sales.SalesOrderHeader} = TerritoryID = 9(\text{Sales.SalesOrderHeader})$

7. Actualizar la subcategoría de los productos con productID del 1 al 4 a la subcategoría válida para el tipo de producto.

PrProducto{

$P1: ProductID = 1$

$P2: ProductID = 2$

$P3: ProductID = 3$

$P4: ProductID = 4$

}

MProduct{

$M1: \neg P1 \wedge \neg P2 \wedge \neg P3 \wedge P4 // \text{Actualizar ID del 1 al 4}$

$M1: \neg P1 \wedge \neg P2 \wedge \neg P3 \wedge P4 // \text{Demás productos}$

}

$FP1 = \sigma_{1 \leq ProductID \leq 4}(\text{Product})$

8. Listar los productos que no estén disponibles a la venta

PrProduct{

$P1: SellEndDate! = NULL$

}

MProduct{

$M1: P1 \rightarrow \text{Productos No disponibles}$

$M2: \neg P1 \rightarrow \text{Productos disponibles}$

}

$FP_1 = \sigma_{SellEndDate \neq NULL}(\text{Product})$

$FP_2 = \sigma_{SellEndDate = NULL}(\text{Product})$

9. Listar los clientes del territorio 1 y 4 que no tengan asociado un valor en personID

PCustomer{

$P1: \text{TerritoryID} = 1$

$P2: \text{TerritoryID} = 4$

$P3: \text{PersonID} = NULL$

}

MCustomer{

$M1: P1 \wedge P2 \wedge P3 \rightarrow \text{Clientes del territorio 1 y 4 sin PersonID}$

$M2: \neg P1 \wedge \neg P2 \wedge \neg P3 \rightarrow \text{Todo lo demás}$

}

$FP_1 = \sigma_{\text{TerritoryID}=1 \ \&\& \ \text{TerritoryID}=4 \ \&\& \ \text{PersonID}=NULL}(\text{Customer})$

10. Listar los clientes del territorio 1 que tengan ordenes en otro territorio.

PrCustomer{

$P1: \text{Territory} = 1$

$P2: \text{TerritoryID} \neq \text{SalesOrderHeader.TerritoryID}$

}

MCustomer{

$M1: P1 \wedge P2 \rightarrow \text{Clientes del territorio 1 con ordenes en otro territorio.}$

$M2: \neg P1 \wedge \neg P2 \rightarrow \text{Todo lo demas.}$

}

$FP_1 = \sigma_{\text{TerritoryID}=1}(\text{Customer}) \ \&\& \ \sigma_{\text{TerritoryID} \neq 1}(\text{SalesOrderHeader})$

Propuesta de Fragmentación

Haciendo un análisis de los 10 siguientes enunciados:

1. La Información de los clientes se debe almacenar por región, considerando las regiones de acuerdo con el atributo Group de SalesTerritory
2. Listar datos del empleado que atendió más ordenes por territorio.
3. Listar los datos del cliente con más ordenes solicitadas en la región "north america"
4. Listar el producto más solicitado en la región "europe"
5. Listar las ofertas que tienen los productos de la categoría "bikes"
6. Listar los 3 productos menos solicitados en la región "pacific"
7. Actualizar la subcategoría de los productos con productID del 1 al 4 a la subcategoría válida para el tipo de producto.
8. Listar los productos que no estén disponibles a la venta
9. Listar los clientes del territorio 1 y 4 que no tengan asociado un valor en personID
10. Listar los clientes del territorio 1 que tengan ordenes en otro territorio.

Echando un vistazo en las consultas y en la selectividad de estas podemos descartar los siguientes predicados de los siguientes enunciados:

- Enunciado 6: Es descartado en el caso de que se pretenda fragmentar products con solamente 3 productos en un fragmento.
- Enunciado 7: Este lo descartamos ya que tiene una selectividad de 4 filas en una tabla de 504 registros.
- Enunciado 9: Es descartado ya que tiene una selectividad de 223 filas, mientras que el complemento tendría una selectividad de 15,597 filas.
- Enunciado 10: Es descartado ya que su selectividad es de 0 filas.

Analizando el enunciado 1, si la información se almacena por región se podrían hacer los fragmentos bien ya sea a través de una fragmentación horizontal primaria en customer, o bien a través de una fragmentación derivada a través de sales territory. No vemos factible que sea a través de salesTerritory, ya que esta tabla es referenciada hacia otras muchas más tablas lo que haría que se generen muchos fragmentos derivados, pero el principal problema sería en SalesOrderHeader, ya que esta tabla se referencia a SalesTerritory y Customer.

Del enunciado 2, si lo que se quiere es los empleados con más ventas por territorio esto nos daría un total de 10 empleados máximo. Si este es el caso serían muy pocas filas para una tabla y estos registros podrían variar cada vez que se quieran los empleados con más ventas.

Del enunciado 3, al igual que el enunciado 2, no tendría sentido fragmentar Customer solo con un registro que pueda variar cada vez que se haga la consulta.

De los 10 enunciados, las tablas a las que más se hacen consultas son Customer y SalesOrderHeader en base a sus regiones o territorios.

Los enunciados que hacen referencia a customer son el 1,3,9 y 10, y solamente el enunciado agrupa los resultados por region.

Los enunciados que hacen referencia a SalesOrderHeader son 2,3,4 y 6.

Como comentamos anteriormente, no es conveniente hacer la fragmentación por SalesTerritory para fragmentar Customer y SalesOrderHeader, ya que esta última tiene llaves foráneas a SalesTerritory y Customer lo que generaría fragmentos diferentes y necesarios para una misma tabla.

Otra opción es hacer la fragmentación por regiones en customer, en base al enunciado 1, pero esto haría que sea más complicado hacer las consultas 2,3,4 y 6, ya que el territorio de la venta es distinto al territorio del cliente y OrderHeader se fragmentaría en base a los clientes. Se podrían hacer las consultas obteniendo la tabla original al hacer la unión de sus fragmentos cada que vez que se necesite.

Por último, podemos fragmentar SalesOrderHeader para hacer más fácil las consultas 3,4 y 6, ya que las ordenes serían almacenadas por región. Para la consulta 2 tampoco habría mucha más complicación, ya que los territorios forman parte de las regiones y no necesitaríamos unir los fragmentos.

Al final solo se lleva a cabo una consulta en clientes en base a su región. Por esta razón decidimos fragmentar la tabla SalesOrderHeader por regiones, para llevar a cabo de manera más rápida y fácil la mayoría de las consultas, en especial la 2,3,4 y 6 que serían las más complicadas.