

Consultas

1. Listar el número de ventas por territorio
2. Listar el número de ventas por tienda
3. Listar el total de clientes que pertenecen a cada territorio
4. Actualizar la oferta de llantas de montaña con un descuento del 40%
5. Listar las ordenes realizadas debidas a anuncio de revista
6. Listar el total de ordenes hechas por cada representante de ventas
7. Agregar el producto "HL Road Frame - Black, 58" a la oferta "Descuento por volumen 11 a 14"
8. La suma de total de venta por PersonID
9. Listar el producto más vendido por cada categoría
10. Listar los productos con ofertas en el territorio 5
11. Listar los cinco productos más solicitados en cada uno de los territorios
12. Listar el producto menos solicitado por cada territorio
13. Actualizar nombre de tarjeta de crédito SuperiorCard a SCard
14. Listado de registros donde el ID de territorio de clientes es 1 y 3
15. Listado de ventas con costo mayor a 2000 y menor a 4000
16. Listado de cambio de moneda de USD a MXN

Predicados

Letras rojas: se elimina

Letras naranjas: se le preguntara a profesor

Letras amarillas: hacer el análisis

```
Sales.SalesOrderHeader{
```

```
    P1 => TerritoryID = 1
```

```
    P2 => TerritoryID = 2
```

```
    ...
```

```
    P10 => TerritoryID = 10
```

```
    P1 => SalesPersonID = 274
```

```
    P2 => SalesPersonID = 275
```

```
    ...
```

```
    P17 => SalesPersonID = 290
```

P18 => SalesPersonID = NULL

P1 => CustomerID = 11000

...

P19,119 => CustomerID = 30118

P19,120 => CustomerID = NULL

P1 = TotalDue < 2000

P2 = TotalDue > 2000

P3 = TotalDue < 4000

P4 = TotalDue > 4000

}

Frecuencia de acceso Sales.SalesOrderHeader

Predicado	Consultas relacionadas	Frecuencia de acceso	Importancia
TerritoryID (10)	1	1	Relevante
SalesPersonID (25)	2	1	Irrelevante
CustomerID (>20000)	8, 6	1	Irrelevante
TotalDue (4)	15	1	Irrelevante

Para la tabla Sales.SalesOrderHeader, consideramos que es relevante el apartado TerritoryID ya que a pesar de que CustomerID tiene una mayor frecuencia de acceso tenemos que una gran cantidad de predicados, lo que nos costaría mucho trabajo hacer una clasificación para repartirlo entre fragmentos, ahora en el caso de SalesPersonID y TotalDue nos encontramos con una cantidad igual de frecuencia de acceso y podemos creer que no son muchos predicados, pero entre todos es mucho más fácil lograr hacer una clasificación por territorios.

Sales.Customer {

P1 => TerritoryID = 1

P2 => TerritoryID = 2

...

P10 => TerritoryID = 10

}

Frecuencia de acceso Sales.Customer

Predicado	Consultas relacionadas	Frecuencia de acceso	Importancia
TerritoryID (10)	3, 14	2	Irrelevante

A pesar de tener una frecuencia de acceso más en el TerritoryID que el caso de Sales.SalesOrderHeader aquí no es tan relevante como en este caso ya que solo son dos consultas y que esto no podría ser un gran impedimento y puede llegar a generarse una mayor carga de trabajo en Sales.SalesOrderHeader por la cantidad de consultas que hay.

Sales.SpecialOfferProduct {

P1 = ProductID = [Al id del las llantas de montaña (buscar cual es)]

P1 => SpecialOfferID = 2}

Frecuencia de acceso Sales.SpecialOfferProduct

Predicado	Consultas relacionadas	Frecuencia de acceso	Importancia
ProductID (1)	4	1	Irrelevante
SpecialOfferID (1)	7	1	Irrelevante

Al igual que en el caso anterior aquí solamente tenemos dos consultas, además de que si se le aplicará el algoritmo COM_MIN llegaríamos a tener fragmentos muy dispares por el tipo de consultas que son, ya que por ser tan específicas o dirigidas nos encontramos con un complemento muy grande en comparación.

Sales.SalesOrderHeaderSalesReason {

P1 => SalesReasonID = 28

P2 => SalesReasonID != 28

}

Frecuencia de acceso Sales.SalesOrderHeaderSalesReason

Predicado	Consultas relacionadas	Frecuencia de acceso	Importancia
SalesReasonID (2)	5	1	Irrelevante

Nos encontramos con la situación de que solamente tenemos una consulta relacionada con esta tabla, lo que nos lleva a descartarla por ser muy poca y además de que nos generaría fragmentos disparejos.

Sales.SalesOrderDetail {

P1 => ProductCategoryID = 1

P2 => ProductCategoryID = 2

P3 => ProductCategoryID = 3

P4 => ProductCategoryID = 4

P1 => TerritoryID = 5

P2 => TerritoryID != 5

P1 => TerritoryID = 1

P2 => TerritoryID = 2

...

P10 => TerritoryID = 10

}

Frecuencia de acceso Sales.SalesOrderDetail

Predicado	Consultas relacionadas	Frecuencia de acceso	Importancia
TerritoryID (10)	10, 11, 12	3	Irrelevante

ProductCategoryID (4)	9	1	Irrelevante
-----------------------	---	---	-------------

En este caso a diferencia del caso de la tabla Sales.SalesOrderHeader nos encontramos con la consulta 10 la cual solamente hace una búsqueda sobre el territoryId = 5 o diferente a 5 que para este caso no creemos que nos convenga debido a que habría una desigualdad en el sentido de que no se recuperaremos la misma cantidad de tuplas, lo que nos podría generar fragmentos dispares. Por otro lado el predicado ProductCategoryID lo encontramos

Sales.CreditCard {

P1 => CardType = SuperiorCard

P2 => CardType != SuperiorCard

}

Frecuencia de acceso Sales.CreditCard

Predicado	Consultas relacionadas	Frecuencia de acceso	Importancia
CardType(2)	13	1	Irrelevante

Nos encontramos con la situación de que solamente tenemos una consulta relacionada con esta tabla, lo que nos lleva a descartarla por ser muy poca y además de que nos generaría fragmentos dispares.

Sales.CurrencyRate {

P1 => ToCurrencyCode = MXN

P2 => ToCurrencyCode != MXN

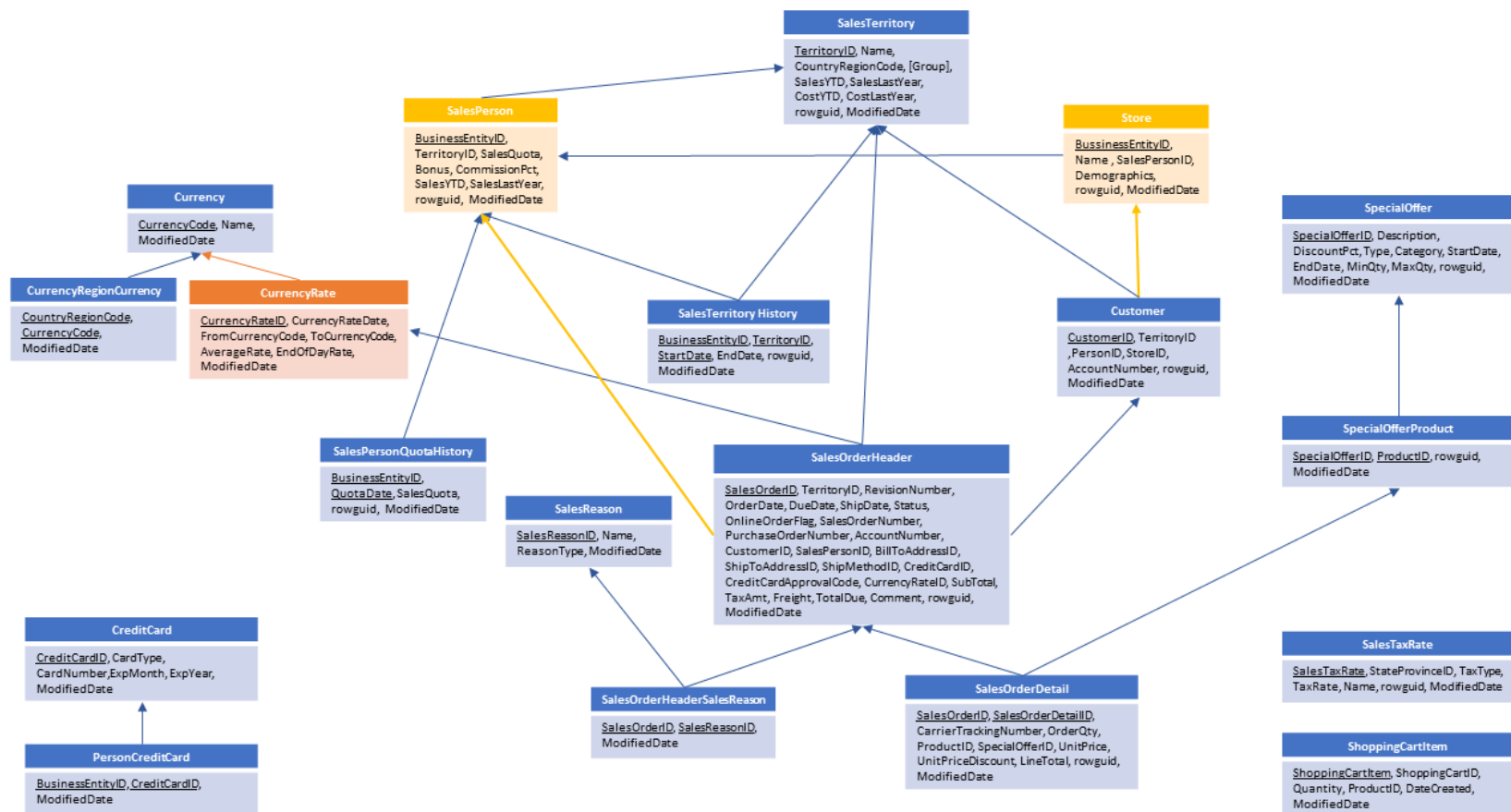
}

Frecuencia de acceso Sales.CurrencyRate

Predicado	Consultas relacionadas	Frecuencia de acceso	Importancia
ToCurrencyCode(2)	16	1	Irrelevante

Nos encontramos con la situación de que solamente tenemos una consulta relacionada con esta tabla, lo que nos lleva a descartarla por ser muy poca y además de que nos generaría fragmentos dispares.

Grafo



Comentarios de la revisión de su propuesta.

1. Analizando las consultas planteadas, se observa que al menos 5, requieren procesar datos por la columna TerritoryId de la tabla SalesOrderHeader y puede plantearse una propuesta de fragmentación por territorio y para ello deben detallar los valores que proponen para determinar la relevancia de cada predicado, es decir, se requiere especificar datos para calcular $\text{acc}(q_i \text{ o } m_i \text{ o } f_i)$ y $\text{card}(q_i \text{ o } m_i \text{ o } f_i)$, además, deben aplicar el algoritmo correspondiente para describir su fragmentación.
2. Por otro lado, dependiendo de la selectividad que calculen en cuanto a productos con descuento y sin descuento, se puede desarrollar una fragmentación en SpecialOfferProduct y si se cumple la regla 1.
3. Para las consultas que requieren datos de SalesOrderDetails se ejecutan a través de datos de SalesOrderHeader por lo que la fragmentación en esta tabla debe ser derivada a partir de la fragmentación primaria en SalesOrderHeader.
4. La consulta 3 opera sobre la tabla Customer en la columna TerritoryId, pero al ser una única consulta y no tener datos de la frecuencia de acceso, al momento se observa que no es necesario particionar esta tabla.
5. En resumen, pueden proponer una fragmentación horizontal primaria por territorios en SalesOrderHeader. Inclusive pueden agrupar territorios para generar un menor número de fragmentos. La fragmentación horizontal derivada se debe propagar a cada una de las tablas en donde exista una relación directa o indirecta en una jerarquía inferior con SalesOrderHeader, por ejemplo, en SalesOrderDetail, SalesOrderHeaderSalesReason, las demás tablas quedan sin fragmentar, excepto la propuesta de fragmentación en SalesOfferProduct.

Quedo al pendiente de dudas.

ALGORITMO COM_MIN

```
Pr'Sales.SalesOrderHeader{  
    P1 => TerritoryID = 1  
    P2 => TerritoryID=2  
    P3 => TerritoryID=3  
    ...  
    P9 => TerritoryID=9  
    P10 => TerritoryID=10  
}
```

```
F{  
    F1:TerritoryID = 1  
    F2: TerritoryID =2  
    F3:TerritoryID=3  
    ...  
    F10:TerritorioID=10  
}
```

A partir de esto checamos relevancia de acuerdo con los miniterminos siguientes:

```
Miniterminos{  
    M1: P1 and not P2 and not P3 and not P4 ...  
    M2: not P1 and P2 and not P3 and not P4 ...  
    M3: not P1 and not P2 and P3 and not P4 ...  
    M4: ... not P2 and not P3 and P4 and not P5 ....  
    M5: ... not P3 and not P4 and P5 and not P6 ...  
    M6: ... not P4 and not P5 and P6 and not P7 ...  
    M7: ... not P5 and not P6 and P7 and not P8 ...  
    M8: ... not P6 and not P7 and P8 and not P9 ...  
    M9: ... not P7 and not P8 and P9 and not P10
```


M10: ... not P8 and not P9 and P10

}

Obteniendo la siguiente tabla, considerando que las consultas listadas representan el 100%

Minitermino	Accesibilidad		Cardinalidad	Acc(mi)/Car(mi)
	Consultas	%		
M1	1, 2, 3, 5, 6, 8, 9, 11, 12, 14, 15	0.687	4594	149.54x10 ⁻⁶
M2	1, 2, 3, 5, 6, 8, 9, 11, 12, 15	0.625	352	1.77x10 ⁻³
M3	1, 2, 3, 5, 6, 8, 9, 11, 12, 14, 15	0.687	385	1.78x10 ⁻³
M4	1, 2, 3, 5, 6, 8, 9, 11, 12, 15	0.625	6224	100.41x10 ⁻⁶
M5	1, 2, 3, 5, 6, 8, 9, 10, 11, 12, 15	0.687	486	1.41x10 ⁻³
M6	1, 2, 3, 5, 6, 8, 9, 11, 12, 15	0.625	4067	153.67x10 ⁻⁶
M7	1, 2, 3, 5, 6, 8, 9, 11, 12, 15	0.625	2672	233.90x10 ⁻⁶
M8	1, 2, 3, 5, 6, 8, 9, 11, 12, 15	0.625	2623	238.27x10 ⁻⁶
M9	1, 2, 3, 5, 6, 8, 9, 11, 12, 15	0.625	6843	91.33x10 ⁻⁶
M10	1, 2, 3, 5, 6, 8, 9, 11, 12, 15	0.625	3219	194.15x10 ⁻⁶

FRAGMENTACIÓN

La fragmentación propuesta se hará por territorio, para lo cual se usará la columna de TerritoryID de la tabla SalesOrderHeader, en la cual se identifican 10 diferentes territorios, que se dividen en tres diferente grupos o regiones que son: Norteamérica, Europa y Pacifico, los cuales serán distribuidos en 2 diferentes instancias como se muestra en la siguiente tabla:

TerritoryID	Name	Group	Instancia
1	Northwest	North America	1
2	Northeast	North America	1
3	Central	North America	1
4	Southwest	North America	1
5	Southeast	North America	1
6	Canada	North America	1
7	France	Europe	2
8	Germany	Europe	2
9	Australia	Pacific	2
10	United Kingdom	Europe	2