SQL - DML Data Manipulation Language Subqueries - UNION - INTERSECT - EXCEPT SQL DDL Índices

UTN - FRBA
Ing. en Sistemas de Información *Gestión de Datos*

Prof.: Ing. Juan Zaffaroni

SQL – Hoy Veremos

•DDL – Data Definition Language

-INDICES

- •DML Data Manipulation Language
 - -SUBQUERIES en
 - SELECT
 - DELETE
 - UPDATE
 - INSERT
 - -Operadores Multi Select
 - UNION
 - INTERSECT
 - EXCEPT

Indices

Los índices son estructuras opcionales asociadas a una tabla.

La función de los índices es la de permitir un acceso más rápido a los datos de una tabla, se pueden crear distintos tipos de índices sobre uno o más campos.

Los índices son lógica y físicamente independientes de los datos en la tabla asociada. Se puede crear o borrar un índice en cualquier momento sin afectar a las tablas base o a otros índices.

Indices (Cont.) TIPOS DE INDICES

Btree Index

Estructura de índice estándar y más utilizada.

Btree Cluster Index

Este tipo de índice provoca al momento de su creación que físicamente los datos de la tabla sean ordenados por el mismo. (Informix / SQLServer / DB2)

Bitmap Index (Oracle)

Son utilizados para pocas claves con muchas repeticiones

Cada bit en el Bitmap corresponde a una fila en particular.

Si el bit esta en on significa que la fila con el correspondiente rowid tiene el valor de la clave.

Hash Index (MySql)

Están implementados en tablas de hash y se basan en otros indices Btree existentes para una tabla. Si una tabla entra integramente en memoria, la manera más rápida de ejecutar consultas sobre ella es usando un hash index.

TIPOS DE INDICES (Cont.)

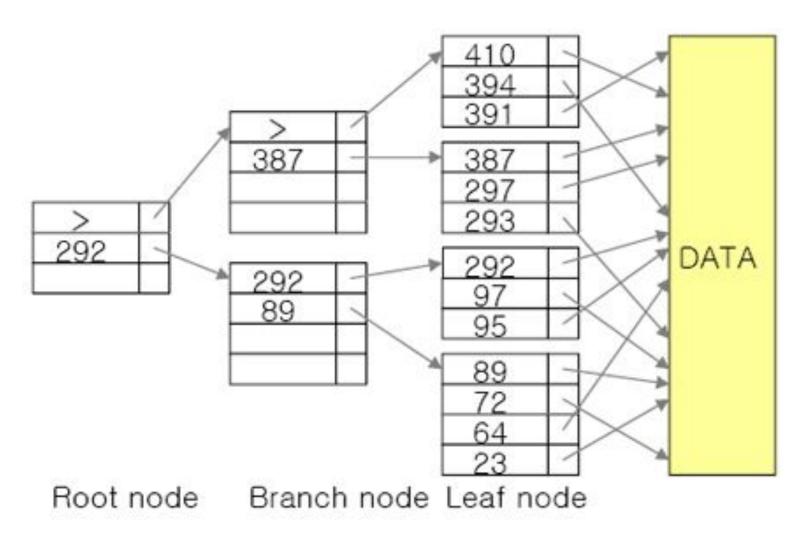
Functional Index / Function based Index

Son indices cuya clave deriva del resultado de una función. En general las funiciones deben ser funciones definidas por un usuario.

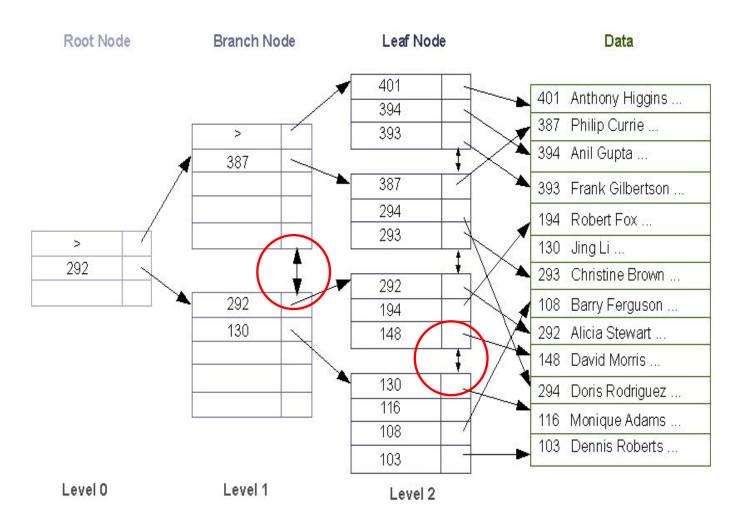
Reverse Key Index (Oracle)

Invierte los bytes de la clave a indexar. Esto sirve para los índices cuyas claves son una serie constante con por ej. Crecimiento ascendente. para que las inserciones se distribuyan por todas las hojas del árbol de índice.

Indices B Tree



Indices B Tree +



CARACTERÍSTICAS DIFERENCIADORAS PARA LOS ÍNDICES

Unique Índice de clave única. Sólo admite una fila por clave.

Duplicado Permite múltiples filas para una misma clave.

Simple La clave está integrada por una sola columna.

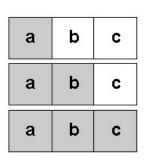
Compuesto La clave se compone de varias columnas.

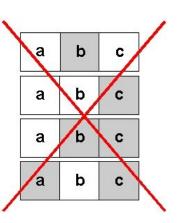
Las principales funciones de un índice compuesto son:

Facilitar múltiples joins entre columnas

Incrementar la unicidad del valor de los índices

CREATE INDEX ix_sample
ON sample table (a, b, c);





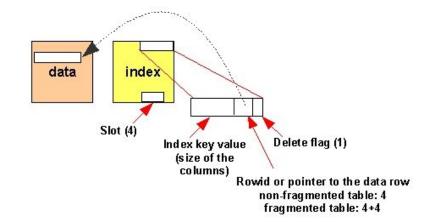
BENEFICIOS DE LA UTILIZACIÓN DE INDICES:

- •Se le provee al sistema mejor perfomance al equipo ya que no debe hacer lecturas secuenciales sino accede a través de los índices, sólo en los casos que las columnas del Select no formen parte del índice.
- •Mejor perfomance en el ordenamiento de filas
- •Asegura únicos valores para las filas almacenadas
- •Cuando las columnas que intervienen en un JOIN tienen índices se le da mejor perfomance si el sistema logra recuperar los datos a través de ellas
- •Asegura el cumplimiento de contraints y reglas de negocio.
 - Primary key, foreign keys, unique values

COSTO DE LA UTILIZACIÓN DE INDICES

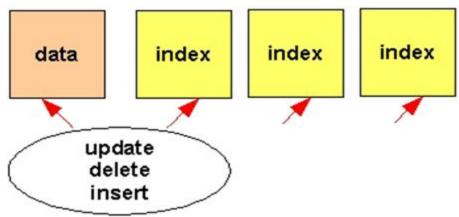
COSTO DE ESPACIO DE DISCO

El primer costo asociado es el espacio que ocupa en disco, que en algunos casos suele ser mayor al que ocupan los datos.



COSTO DE PROCESAMIENTO Y MANTENIMIENTO

El segundo costo es el de procesamiento, hay que tener en cuenta que cada vez que una fila es insertada o modificada o borrada, el índice debe estar bloqueado, con lo cual el sistema deberá recorrer y actualizar los distintos índices.



GUIA RESUMEN CUANDO DEBERIAMOS INDEXAR

- •Indexar columnas que intervienen en Joins
- •Indexar las columnas donde frecuentemente se realizan filtros
- •Indexar columnas que son frecuentemente usadas en orders by
- •Evitar duplicación de índices
 - •Sobre todo en columnas con pocos valores diferentes Ej: Sexo, Estado Civil, Etc.
- •Verificar que el tamaño de índice debería ser pequeño comparado con la fila
 - •Tratar sobre todo en crear índices sobre columnas cuya longitud de atributo sea pequeña
 - •No crear índices sobre tablas con poca cantidad de filas, no olvidar que siempre se recupera de a páginas. De esta manera evitaríamos que el sistema lea el árbol de índices

GUIA RESUMEN CUANDO DEBERIAMOS INDEXAR (Cont.)

- •Limitar la cantidad de índices en tablas que son actualizadas frecuentemente
 - •Porque sobre estas tablas se estarán ejecutando Selects extras
- •Tratar de usar índices compuestos para incrementar los valores únicos
 - •Tener en cuenta que si una o más columnas intervienen en un índice compuesto el optimizador podría decidir acceder a través de ese índice aunque sea solo para la búsqueda de los datos de una columna, esto se denomina "partial key search"
- •Usando cluster index se agiliza la recuperación de filas
 - •Uno de los principales objetivos de la Optimización de bases de datos es reducir la entrada/salida de disco. Reorganizando aquellas tablas que lo necesiten, se obtendría como resultado que las filas serían almacenadas en bloques contiguos, con lo cual facilitaría el acceso y reduciría la cantidad de accesos ya que recuperaría en menos páginas los mismos datos.

CONSTRUCCIÓN DE ÍNDICES EN PARALELO

Los motores de BD usan en general métodos de construcción de índices en paralelo.

El arbol B+ es construido por 2 o más procesos paralelos. Para esto el motor realiza una muestra de la filas a Indexar (aproximadamente 1000) y luego decide como separar en grupos. Luego scanea las filas y las ordena usando el mecanismo de sort en paralelo.

Las claves ordenadas son colocadas en los grupos apropiados para luego ir armarndo en paralelo un subárbol por cada grupo. Al finalizar los subárboles se unen en un único Arbol B+.

Motor SQL SERVER

Sql Server utiliza una estructura de Arbol B+. Máxima cantidad de campos para la clave de un índice compuesto: 16.

Creación de un índice único y simple:

CREATE UNIQUE INDEX ix1_ordenes ON ordenes (n_orden);

Creación de Indice duplicado y compuesto.

CREATE INDEX ix2_ordenes ON ordenes (n_cliente, f_orden);

Creación de Indice clustered

CREATE CLUSTERED INDEX ix3_ordenes ON ordenes (N_orden);

Motor SQL SERVER (cont.)

Manejo del Load Factor

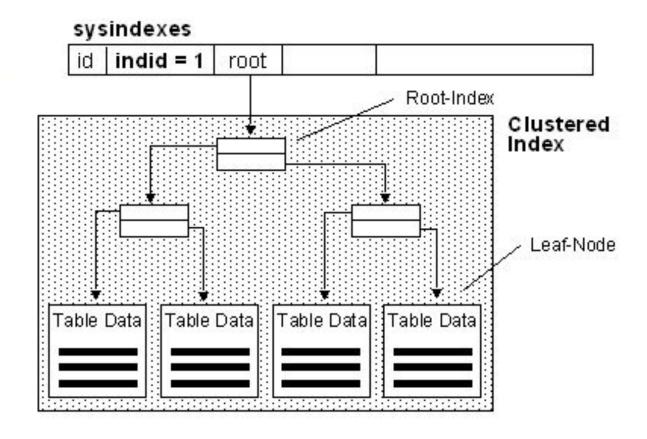
FILLFACTOR- Porcentaje de cada página del índice a ser dejado como espacio libre en su creación.

Por ej. Si el FILLFACTOR=20, en la creación del índice se ocupará hasta el 80% de cada nodo.

CREATE UNIQUE INDEX ix1_ordenes ON ordenes(N_orden) WITH FILLFACTOR = 20;

MOTOR SQL SERVER

Indices Clustered



Tablas Organizadas por Índice (IOT - Index-Organized Tables) (Ora)

Una IOT tiene una organización del almacenamiento que es una variante del Arbol-B. A diferencia de una tabla común, en la que los datos se guardan como un conjunto desordenado, en una IOT se guardan en una estructura de índice de Árbol-B, ordenado a la manera de una clave primaria. Sin embargo, además de almacenar los valores de las columnas de clave primaria de cada tabla, cada entrada en el índice almacena además los valores de las columnas no clave.

De esta forma, en vez de mantener dos estructuras separadas de almacenamiento, una para la tabla y una para el índice de Árbol-B, el sistema mantiene sólo un índice Árbol-B, porque además de almacenar el rowid de las filas, también se guardan las columnas no clave.

Tablas Organizadas por Índice (IOT - Index-Organized Tables) (Ora)

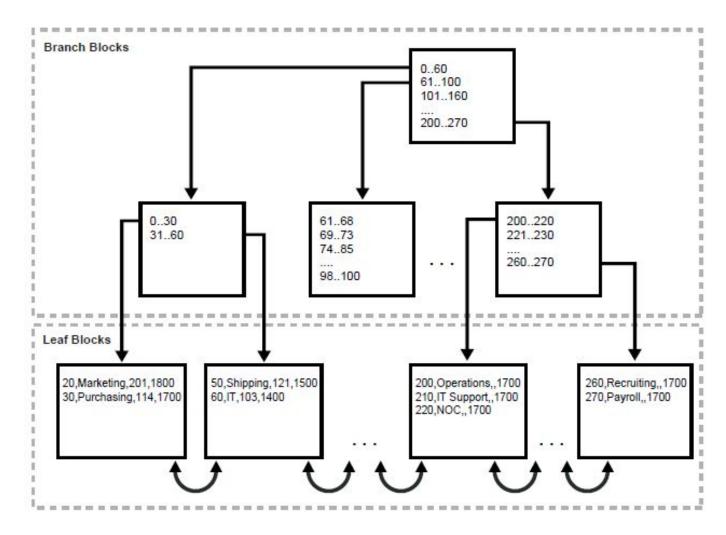
Motor Oracle

La siguiente sentencia crea la tabla ordenes, como una tabla organizada por índice, indicando que la columna N_cliente divide el área del índice, de las columnas no clave.

ORGANIZATION INDEX INCLUDING N_CLIENTE OVERFLOW

Tablas Organizadas por Índice (IOT - Index-Organized Tables) (Ora)

Motor Oracle



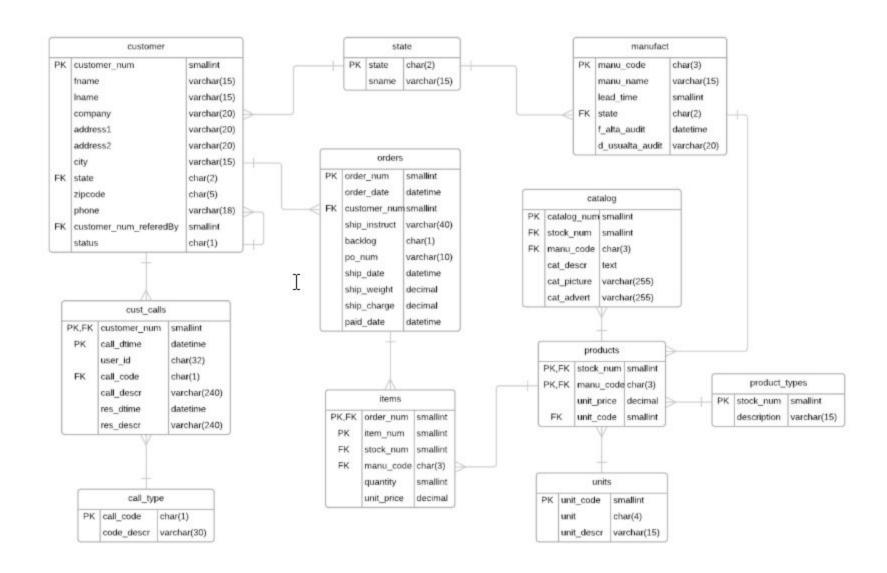
SQL – Hoy Veremos

•DDL – Data Definition Language

-INDICES

- •DML Data Manipulation Language
 - -SUBQUERIES en
 - SELECT
 - DELETE
 - UPDATE
 - INSERT
 - -Operadores Multi Select
 - UNION
 - INTERSECT
 - EXCEPT

Base de Datos de Ejemplo



SQL – Operador INSERT

INSERT INTO closed_orders

SELECT * FROM orders

WHERE paid date IS NOT null

En una tabla creada previamente con la misma estructura que la tabla ordenes, insertamos las filas que devuelve el SELECT.

Es fundamental que el select devuelva las mismas filas y en el mismo orden las columnas que la tabla destino.

En este ejemplo el insert esta acoplado a la definición de la tabla closed_orders y el SELECT está acoplado a la definición de la tabla orders.

DELETE FROM customer

WHERE customer_num NOT IN

(SELECT DISTINCT customer_num FROM cust_calls)



AND customer_num **NOT IN**

(SELECT DISTINCT customer_num FROM orders)



AND customer_num NOT IN

(SELECT DISTINCT customer_num_referedBy FROM customer c2

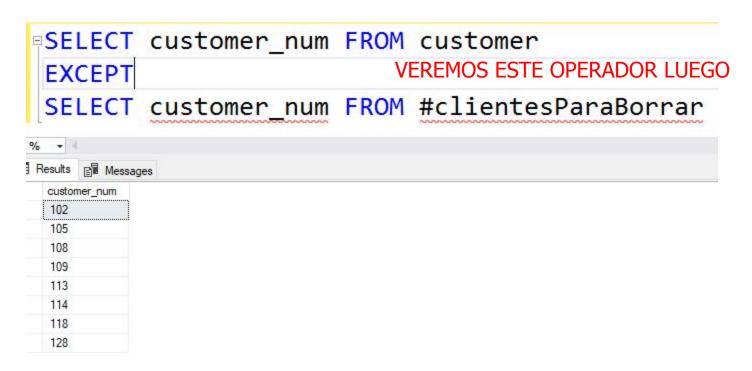


WHERE customer_num_referedBy IS NOT NULL)

Utilizando tres subqueries como condiciones del DELETE, podríamos borrar todos los clientes de la Tabla customer, que no posean ordenes de compra asociadas y que no hayan referenciado a otro cliente y que no tengan llamados telefónicos.

```
SELECT * INTO #clientesParaBorrar FROM customer
DELETE FROM #clientesParaBorrar
     WHERE customer num NOT IN
        (SELECT DISTINCT customer num FROM cust calls)
        AND customer num NOT IN
        (SELECT DISTINCT customer num FROM orders)
        AND customer_num NOT IN
        (SELECT DISTINCT customer_num_referedBy FROM customer c2
         WHERE customer_num_referedBy IS NOT NULL)
0 % + 4
Messages
 (8 row(s) affected)
```

Para hacer la prueba y no borrar datos de nuestra tabla real corrimos el ejemplo en la tabla temporal creada #clientesParaBorrar



En este ejemplo hacemos la comparación entre las tablas para ver que filas se borraron.

```
UPDATE FROM customer

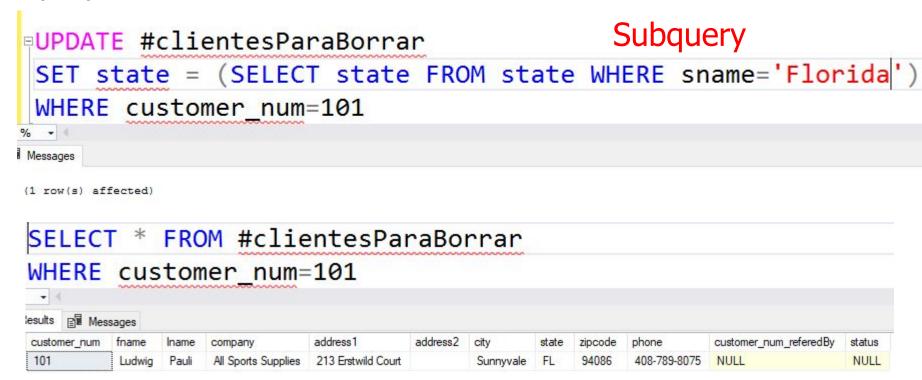
SET columna = Subquery

WHERE columna (=/IN/NOT IN) Subquery

(EXISTS / NOT EXISTS ) Subquery
```

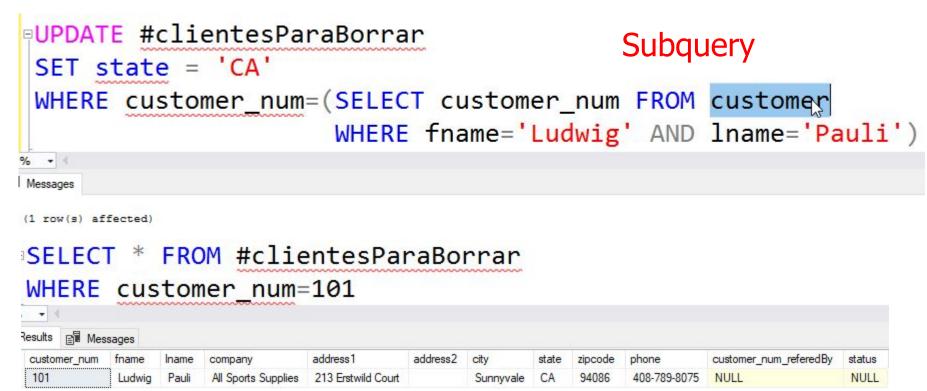
En un update vemos distintos lugares y formas de ejecutar un subquery.

Suquery en CLAUSULA SET



El subquery puede retornar solo una única columna y fila. O sea un valor escalar.

Suquery en CLAUSULA WHERE con Subquery que devuelve valor escalar



El subquery puede retornar solo una única columna y fila. O sea un valor escalar.

Por qué?

Suquery en CLAUSULA WHERE con Subquery que devuelve una columna con multiples filas.

```
Subquery
UPDATE manufact SET lead_time=15
 WHERE manu code IN (SELECT DISTINCT manu code FROM items)
Messages
                     -- COMPROBACION
(9 row(s) affected)
                     SELECT DISTINCT i.manu code, lead time
                       FROM items i INNER JOIN manufact m ON (i.manu code=m.manu code)
                  Results Messages
                    manu_code | lead_time
```

Modificamos el lead_time (tiempo de entrega) a 15 días para todos los proveedores de los cuales ya hayamos entregado productos.

Que sucede si corremos este UPDATE?

```
UPDATE manufact SET lead_time=15
WHERE manu_code = (SELECT DISTINCT manu_code FROM items)
```

Break



SELECT col, col, Subquery

FROM table JOIN Subquery

WHERE columna (=/<=/>
(EXISTS / NOT EXISTS)

Subquery Subquery

GROUP BY ...

HAVING fxAgregada (=/<=/>=/IN/NOT IN)
(EXISTS / NOT EXISTS)

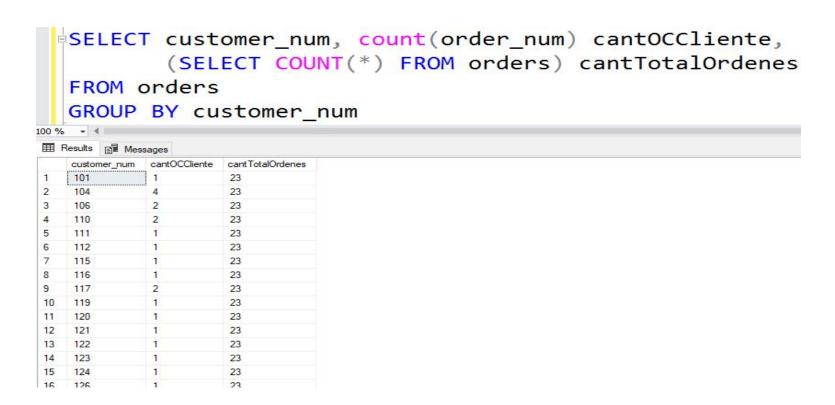
Subquery Subquery

ORDER BY

Subquery

En un SELECT vemos distintos lugares y formas de ejecutar un subquery.

Subquery en Lista de columnas.

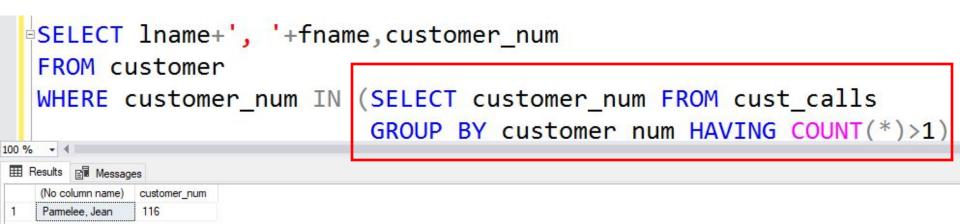


El subquery devuelve la cantidad de ordenes que existen y como si fuese una variable lo muestra en cada fila

Subquery en el FROM.

El subquery devuelve un conjunto de filas se asocian al alias c2 como si fuese el result set de una tabla y luego estos datos son JOINEADOS con la tabla customer c1.

Subquery en el WHERE.



La consulta muestra los clientes que posean mas de una llamada. Muchas veces este tipo de subqueries se pueden resolver mediante joins.

Resuelvanlo en 5 minutos!!

Subquery en el WHERE con múltiples filas.

```
SELECT lname+', '+fname, customer_num

FROM customer
WHERE customer_num IN

SELECT customer_num FROM cust_calls
GROUP BY customer num HAVING COUNT(*)>1)

Results Messages

(No column name) customer_num
1 Parmelee, Jean 116
```

La consulta muestra los clientes que posean mas de una llamada. Muchas veces este tipo de subqueries se pueden resolver mediante joins.

Resuelvanlo en 5 minutos!!

Subquery en el WHERE con un valor escalar.

```
SELECT COUNT(*) FROM customer

WHERE city = (SELECT city FROM customer WHERE lname = 'Higgins')

B Results Messages

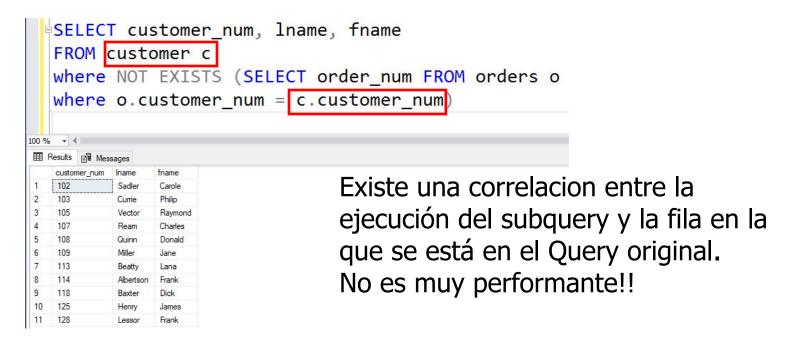
(No column name)

1 5
```

La consulta devuelve la cantidad de clientes que viven en la misma ciudad en la que vive Higgins, incluso el mismo. Se puede resolver con joins.

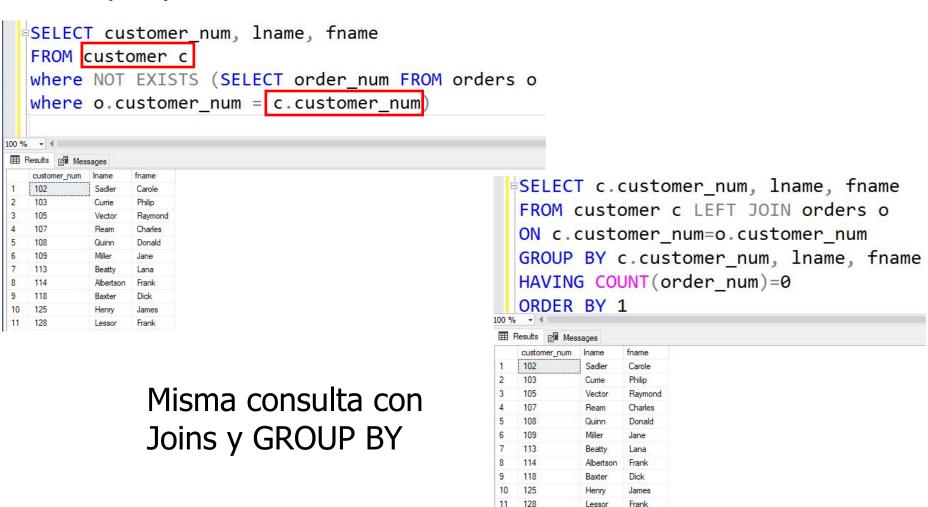
Resuelvanlo en 5 minutos!!

Subquery CORRELACIONADO.



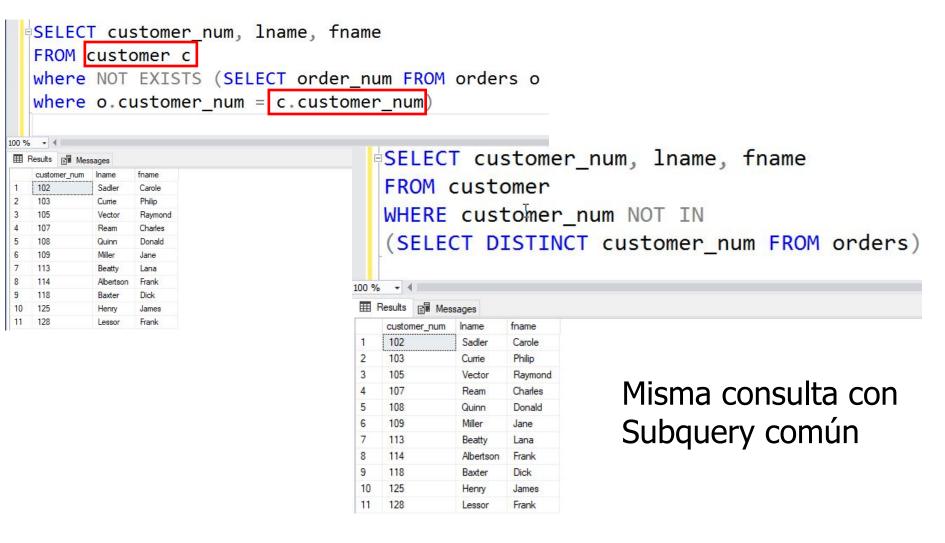
La consulta devuelve los clientes que para los que no existen ordenes de compra.

Subquery CORRELACIONADO.



Frank

Subquery CORRELACIONADO.



listar las ordenes de los clientes con su total comprado pero sólo para los que el total de la orden sea mayor que el promedio comprado de dicho cliente.

```
SQLQuery1.sql - D...HQEMP6\zaffa (53))* - ×
 SELECT o.order_num,o.customer_num, SUM(quantity*unit_price) totalOrden
         promedio
  FROM orders o JOIN items i ON o.order num=i.order num
  JOIN (SELECT customer num, SUM(quantity*unit price)
                                  /COUNT(DISTINCT o2.order num) promedio
         FROM orders o2 JOIN items i2 ON (o2.order_num=i2.order_num)
         GROUP BY customer num) tabSQ
     ON (o.customer_num=tabSQ.customer_num)
  GROUP BY o.order num, o.customer num, tabSQ.promedio
  HAVING SUM(i.quantity*i.unit price) > tabSQ.promedio
Results Resages
  order num
        customer_num
              totalOrden
                   promedio
  1003
               1355.00
                    570.950000
  1008
               1340.00
        110
                    895.000000
                                Resolución con subquery en FROM
```

Ing. Juan Zaffaroni

117

1014 1012 1440.00

2840.00

1428 000000

2268.000000

SQL – Hoy Veremos

•DDL – Data Definition Language

-INDICES

- •DML Data Manipulation Language
 - -SUBQUERIES en
 - SELECT
 - DELETE
 - UPDATE
 - INSERT

–Operadores Multi Select

- UNION
- INTERSECT
- EXCEPT

SELECT stock_num,manu_code FROM products WHERE unit price < 25

UNION

SELECT stock_num,manu_code FROM items WHERE stock_num=5 ORDER BY 1,2 La cantidad de campos en cada consulta debe corresponderse.

Los campos en igual posición deben tener el mismo tipo de datos.

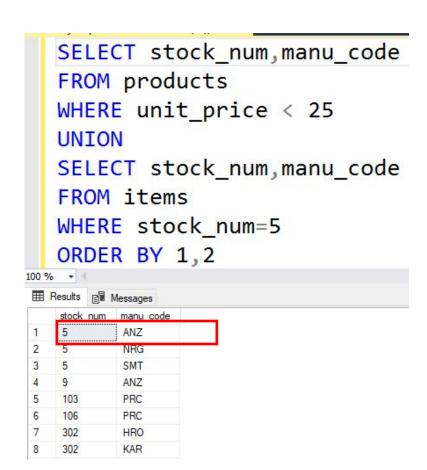
La clausula ORDER BY va al final del ultimo SELECT.

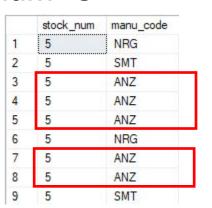
Solo se pude Ordenar referenciando por posición.

En el caso de filas con idénticos valores solo deja una de las dos.

SELECT stock_num,manu_code FROM products WHERE unit_price < 25 SELECT stock_num,manu_code FROM items WHERE stock_num=5

	stock_num	manu_code
1	5	ANZ
2	9	ANZ
3	103	PRC
4	106	PRC
5	302	HRO
6	302	KAR



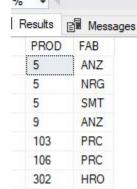


SELECT stock_num PROD,manu_code FAB FROM products WHERE unit_price < 25

UNION

SELECT stock_num,manu_code FROM items WHERE stock_num=5 ORDER BY 1,2

```
SELECT stock_num PROD,manu_code FAB FROM products
WHERE unit_price < 25
UNION
SELECT stock_num,manu_code FROM items WHERE stock_num=5
ORDER BY 1,2
```



KAR

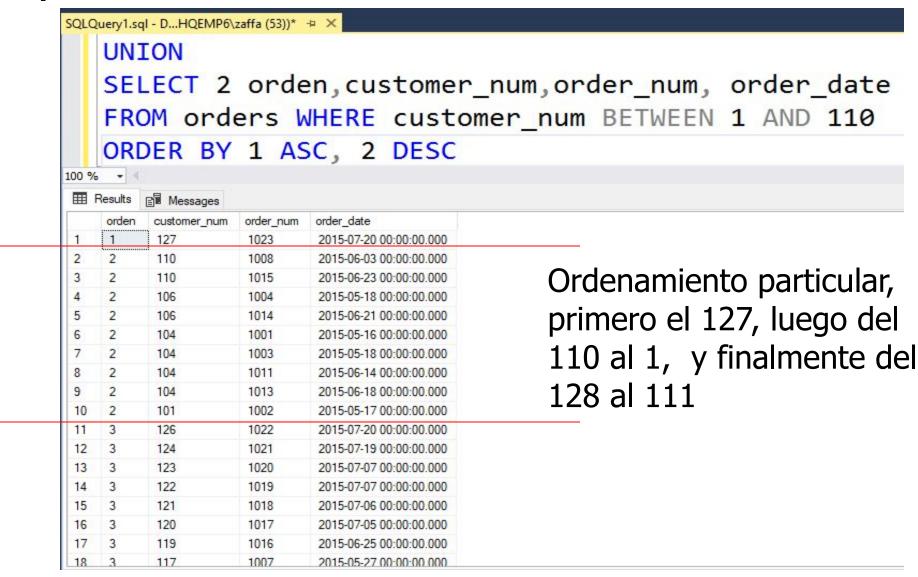
La tabla de salida toma los nombres de columnas del primer SELECT.

Operador Union — Mas de dos tablas

SELECT 1 orden,customer_num,order_num, order_date
FROM orders WHERE customer_num=127
UNION
SELECT 3 orden,customer_num,order_num, order_date
FROM orders WHERE customer_num BETWEEN 111 AND 126
UNION
SELECT 2 orden,customer_num,order_num, order_date
FROM orders WHERE customer_num BETWEEN 1 AND 110
ORDER BY 1 ASC, 2 DESC

En este caso particular vemos que utilizamos un ordenamiento de filas determinado por una constante en cada select que dice que filas irán primero en la salida.

Operador Union – Mas de dos tablas



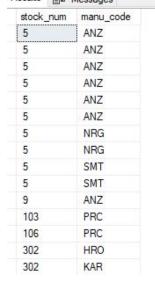
Operador UNION ALL

SELECT stock_num,manu_code FROM products WHERE unit_price < 25 UNION ALL

SELECT stock_num,manu_code FROM items WHERE stock_num=5 ORDER BY 1,2

SELECT stock_num,manu_code FROM products WHERE unit_price < 25
UNION ALL
SELECT stock_num,manu_code FROM items WHERE stock_num=5 ORDER BY 1,2

Results Messages



El Operador UNION con la clausula ALL, repite en la estructura de salida todas las filas de las tablas involucradas, sean o no iguales.

Operador INTERSECT

SELECT stock_num,manu_code FROM products WHERE unit_price < 25 INTERSECT SELECT stock_num,manu_code FROM items WHERE stock_num=5 ORDER BY 1,2

Mismas restricciones del UNION.

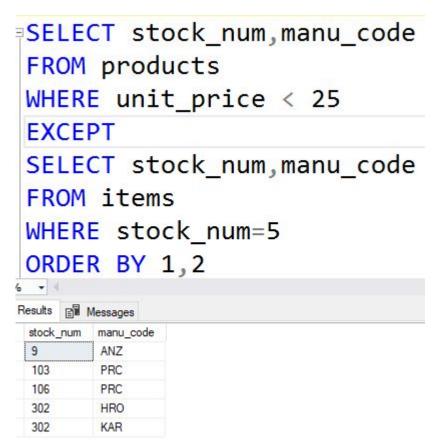
Devuelve las filas que están en ambas consultas.

```
FROM products
WHERE unit_price < 25
INTERSECT
SELECT stock_num,manu_code
FROM items
WHERE stock_num=5
ORDER BY 1,2
```

Operador EXCEPT

SELECT stock_num,manu_code FROM products WHERE unit_price < 25 EXCEPT SELECT stock_num,manu_code FROM items WHERE stock_num=5 ORDER BY 1,2 Mismas restricciones del UNION.

Devuelve las filas del Primer SELECT que no están en la intersección con el Segundo SELECT.



SELECT stock_num,manu_code FROM products WHERE unit_price < 25

	stock_num	manu_code
1	5	ANZ
2	9	ANZ
3	103	PRC
4	106	PRC
5	302	HRO
6	302	KAR

SELECT stock_num,manu_code FROM items WHERE stock_num=5

	stock_num	manu_code
1	5	NRG
2	5	SMT
3	5	ANZ
4	5	ANZ
5	5	ANZ
6	5	NRG
7	5	ANZ
8	5	ANZ
9	5	SMT

UNION

stock_num	manu_code
5	ANZ
5	NRG
5	SMT
9	ANZ
103	PRC
106	PRC
302	HRO
302	KAR

UNION ALL

stock_num	manu_code
5	ANZ
5	NRG
5	NRG
5	SMT
5	SMT
9	ANZ
103	PRC
106	PRC
302	HRO
302	KAR

INTERSECT

stock_num	manu_code
5	ANZ

EXCEPT

stock_num	manu_code
9	ANZ
103	PRC
106	PRC
302	HRO
302	KAR

