

M1B1T1_AI1: Consultas a bases de datos relacionales con SQL

Autor: Leandro Gutierrez

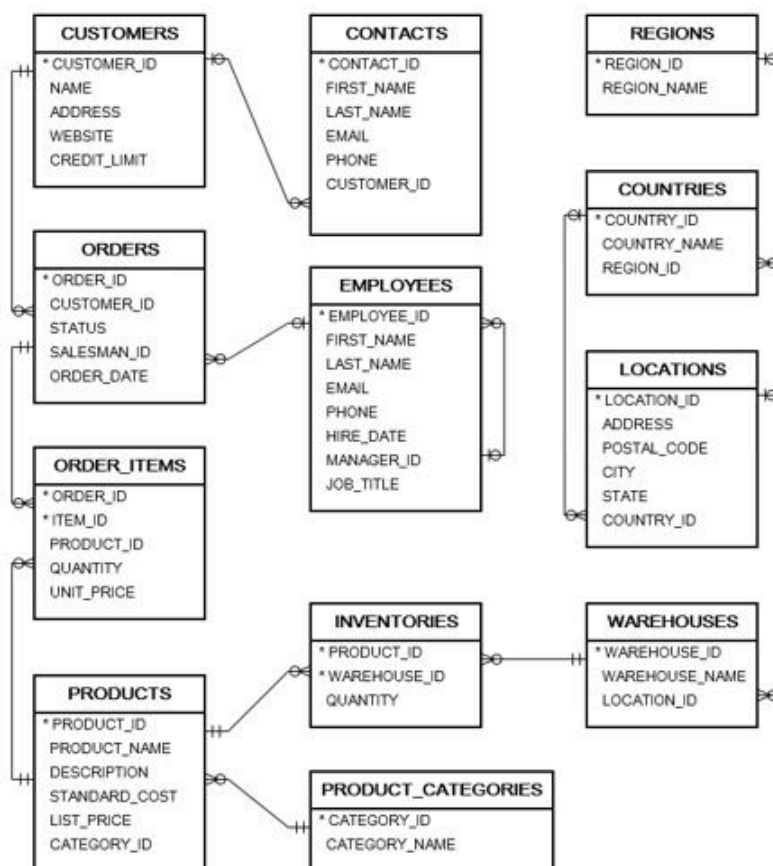
Este documento intenta dar respuesta a las actividades propuestas en el Modulo 1 Bloque 1 del Master en Big Data y Ciencia de Datos. En él se describirán cada uno de los enunciados postulados y los resultados obtenidos a través del uso de la base de datos relacional Oracle y su interfaz SQL.

Abril 05, 2024

Descripción

Has sido contratado por una multinacional como Data Engineer. Se trata de una empresa global que vende hardware informático, incluidos almacenamiento, placas base, RAM, tarjetas de video y CPUs.

Te dan acceso a su base de datos relacional, la cual está montada en un servidor Oracle. El modelo de datos (tablas, campos y relaciones) es el siguiente:



Ejercicio 1

¿Cuáles son los datos de los almacenes que tiene la compañía? Se necesita:

- Identificador del almacén.
- Nombre del almacén.
- Nombre de la ciudad, país y región donde está ubicado.

SQL

```
select  w.warehouse_id,  
        w.warehouse_name,  
        l.city||', '||c.country_name||', '|| r.region_name as LOCATION  
from warehouses w  
inner join locations l on l.location_id = w.location_id  
inner join countries c on c.country_id = l.country_id  
inner join regions r on r.region_id = c.region_id  
order by r.region_id;
```

Resultados

	123 WAREHOUSE_ID	RBC WAREHOUSE_NAME	RBC LOCATION
1	1	Southlake, Texas	Southlake, United States of America, Americas
2	2	San Francisco	South San Francisco, United States of America, Americas
3	3	New Jersey	South Brunswick, United States of America, Americas
4	4	Seattle, Washington	Seattle, United States of America, Americas
5	7	Mexico City	Mexico City, Mexico, Americas
6	5	Toronto	Toronto, Canada, Americas
7	9	Bombay	Bombay, India, Asia
8	8	Beijing	Beijing, China, Asia
9	6	Sydney	Sydney, Australia, Asia

Respuesta

La compañía cuenta con 9 depósitos. Desplegados en:

- 2 regiones: América y Asia.
- 6 países: Estados Unidos, México, Canadá, India, China y Australia
- 9 Ciudades: South Lake, San Francisco, New Jersey, Seattle, Mexico DF, Toronto, Bombay, Beijin y Sydney.

Ejercicio 2

¿Cuál es el nombre del producto que tiene más stock en Asia?

SQL

```

with ranking as (
    select  p.product_id,
            p.product_name,
            sum(i.quantity)
    from products p
    inner join inventories i on i.product_id = p.product_id
    inner join warehouses w on w.warehouse_id = i.warehouse_id
    inner join locations l on l.location_id = w.location_id
    inner join countries c on c.country_id = l.country_id
    inner join regions r on r.region_id = c.region_id
    where r.region_name = 'Asia'
    group by p.product_id, p.product_name, r.region_id
    order by sum(i.quantity) desc
)
select * from ranking where rownum = 1;

```

Resultados

	123 PRODUCT_ID 🔼🔼	ABC PRODUCT_NAME 🔼🔼	123 SUM(I.QUANTITY) 🔼🔼
1	283	G.Skill Trident Z	605

Respuesta

El producto de mayor stock en la región asiática es la **Memoria RAM DDR4 G.Skill Trident Z** con un total de 605 unidades.

Ejercicio 3

¿Cuál es el producto que ha vendido más unidades durante 2016?

SQL

```

with ranking as (
    select  p.product_id,
            p.product_name,
            sum(oi.quantity) venta
    from orders o
    inner join order_items oi on oi.order_id = o.order_id
    inner join products p on p.product_id = oi.product_id
    where extract(year from o.order_date) = '2016' and o.status = 'Shipped'
    group by p.product_id, p.product_name
    order by venta desc
)
select * from ranking where rownum = 1;

```

Resultados

	123 PRODUCT_ID T	ABC PRODUCT_NAME T	123 VENTA T
1	174	MSI GTX 1080 TI AERO 11G OC	364

Respuesta

El producto de más vendido durante el año 2016 fué la **Placa de video MSI GTX 1080 TI AERO 11G OC** con un total de 364 unidades.

Ejercicio 4

¿Cuál es la categoría de productos que ha vendido más unidades durante 2017?

SQL

```
with ranking as (
  select  c.category_id,
          c.category_name,
          sum(oi.quantity) venta
  from orders o
  inner join order_items oi on oi.order_id = o.order_id
  inner join products p on p.product_id = oi.product_id
  inner join product_categories c on c.category_id = p.category_id
  where extract(year from o.order_date) = '2017' and o.status = 'Shipped'
  group by c.category_id, c.category_name
  order by venta desc
)
select * from ranking where rownum = 1;
```

Resultados

	123 CATEGORY_ID T	ABC CATEGORY_NAME T	123 VENTA T
1	5	Storage	6.693

Respuesta

La categoría mas vendida durante 2017 fué **Storage (Almacenamiento)** con un total de 6693 unidades.

Ejercicio 5

¿Cuál es el nombre del cliente cuyo gasto ha sido más alto en 2015?

SQL

```

with ranking as (
  select  c.customer_id,
          c.name,
          sum(oi.quantity * oi.unit_price) venta
  from orders o
  inner join order_items oi on oi.order_id = o.order_id
  inner join products p on p.product_id = oi.product_id
  inner join customers c on c.customer_id = o.customer_id
  where extract(year from o.order_date) = '2015' and o.status = 'Shipped'
  group by c.customer_id, c.name
  order by venta desc
)
select * from ranking where rownum = 1;

```

Resultados

	123 CUSTOMER_ID 🔼🔼	ABC NAME 🔼🔼	123 VENTA 🔼🔼
1	44	Jabil Circuit	1.050.939,97

Respuesta

El cliente que mas gastó durante el año 2015 fué el fabricante de productos electrónicos de alta complejidad a escala mundial **Jabil Circuit** (<https://www.jabil.com/>).

Ejercicio 6

¿Cuánto ha facturado la compañía en cada uno de los años de los que tiene datos?


SQL

```

select  extract(year from o.order_date) as año,
          sum(oi.quantity * oi.unit_price) venta
  from orders o
  inner join order_items oi on oi.order_id = o.order_id
  where o.status = 'Shipped'
  group by extract(year from o.order_date)
  order by año asc;

```

Resultados

	123 AÑO 	123 VENTA 
1	2.013	514.267,91
2	2.015	5.563.747,78
3	2.016	12.921.598
4	2.017	14.408.188,68

Respuesta

Los datos cuentan con ventas desde el año 2013 al 2017, con excepcion del año 2014, en el cual no se registraron ventas. En los resultados se aprecia un crecimiento progresivo en las ventas en cada resultado anual. Logrando una tasa de mejora aproximada del **2690%** Considerando:

$$Tasa\ de\ crecimiento\% = ((Valor\ inicial - Valor\ final) / Valor\ inicial) \times 100$$




Ejercicio 7

¿Cuáles son los nombres de los productos cuyo precio es superior la media?

SQL

```
select  p.product_id,
        p.product_name,
        list_price
from products p
where list_price > (select avg(list_price) from products)
order by list_price asc;
```

Resultados

	123 PRODUCT_ID 	ABC PRODUCT_NAME 	123 LIST_PRICE 
1	176	Intel Xeon E5-2650 V3 (OEM/Tray)	939,49
2	190	Supermicro X10SDV-8C-TLN4F	948,99
3	182	PNY VCQK4200-PB	949,89
4	59	Intel Core i7-5960X (OEM/Tray)	977,99
5	184	AMD 100-506061	999,99
6	181	ATI FirePro R5000	999,99
7	53	Intel Core 2 Extreme QX6800	1.003,98
8	214	Intel Core i7-5960X	1.009,79
9	167	Intel Xeon E5-2650L V3 (OEM/Tray)	1.010,46
10	54	Intel Xeon E5-1660 V3 (OEM/Tray)	1.019,99

Nota: solo se muestran los primeros 10 resultados, en total el listado comprende 85 articulos.

Respuesta

Con una media de precio al rededor de los **903.24**, encontramos unos 85 productos cuyo valor es superior a dicho promedio.

Ejercicio 8

¿Cuáles son los empleados (nombre y apellido) que han facturado más de 50.000 \$ durante 2017?

SQL

```
select  e.employee_id,
        (e.first_name || ' ' || e.last_name) as nombre_completo,
        sum(oi.quantity * oi.unit_price) as venta
from orders o
inner join order_items oi on oi.order_id = o.order_id
inner join employees e on e.employee_id = o.salesman_id
where extract(year from o.order_date) = '2017' and o.status = 'Shipped'
group by e.employee_id, (e.first_name || ' ' || e.last_name)
having sum(oi.quantity * oi.unit_price) > 50000
order by venta asc;
```

Resultados

	123 EMPLOYEE_ID	ABC NOMBRE_COMPLETO	123 VENTA
1	59	Chloe Cruz	277.585,23
2	61	Daisy Ortiz	307.842,27
3	54	Lily Fisher	394.765,27
4	64	Florence Freeman	545.057,39
5	57	Scarlett Gibson	645.379,54
6	55	Grace Ellis	1.048.915,28
7	60	Isabelle Marshall	2.092.044,24
8	62	Freya Gomez	2.499.633,26

Respuesta

Encontramos 8 empleados cuya facturación durante el año 2017 fué superior a los \$50.000. El primer lugar lo ocupa **Freya Gomez, con una facturación total de \$2.499.633,26**

Ejercicio 9

¿Cuánto clientes no tienen persona de contacto?

SQL

```
select count(customer_id)
from customers
where customer_id not in (select customer_id from contacts);
```

Resultados

	123 COUNT(CUSTOMER_ID) 🔼🔼
1	0

Respuesta

No encontramos clientes sin contacto asignado en la bas de datos.

Ejercicio 10

¿Cuál es el Manager (nombre y apellido identificado por el campo manager_id) que menos ha facturado durante 2017?

SQL

```
with ranking as (
  select m.employee_id,
         (m.first_name || ' ' || m.last_name) as nombre_completo,
         sum(oi.quantity * oi.unit_price) as venta
  from orders o
  inner join order_items oi on oi.order_id = o.order_id
  inner join employees e on e.employee_id = o.salesman_id
  inner join employees m on m.employee_id = e.manager_id
  where extract(year from o.order_date) = '2017' and o.status = 'Shipped'
  group by m.employee_id, (m.first_name || ' ' || m.last_name)
  order by venta asc
)
select * from ranking where rownum = 1;
```

Resultados

	123 EMPLOYEE_ID 🔼🔼	ABC NOMBRE_COMPLETO 🔼🔼	123 VENTA 🔼🔼
1	48	Jessica Woods	545.057,39

Respuesta

De los 3 managers que registraron ventas durante el año 2017, fué **Jessica Woods** quién menos facturó, con un total de **\$545057.39** durante el período analizado.