

Trabajo Práctico

Nº 1

Base de Datos II

Alumno: Donnarumma, Cesar Julián

Comisión: 1

- 1) Identificar las tablas y columnas de la base de datos provista que:
 - a. Deberán ser incluidas en el DataWarehouse para poder resolver las necesidades de reporting mencionadas anteriormente.
 - b. Podrían ser incluidas en el DataWarehouse para ofrecer una capacidad de análisis más detallada, sin desviarnos de los requerimientos iniciales (por ejemplo, mayor nivel de detalle, pero no crear nuevos reportes)

Reporte: Clasificación de los productos por categoría.

- Consulta SQL:

```
SELECT NOMBRE, FAMILIA
FROM ARTICULOS;
```

Tabla	Columna	Tipo	Motivo
ARTICULOS	NOMBRE	Requerido	Tener el nombre de los productos
ARTICULOS	FAMILIAR	Requerido	Hacer las categorías por familia

Reporte: Distribución de los clientes por zona (región, ciudad).

- Consulta SQL de distribución de clientes por comuna:

```
SELECT RESIDENCIA, COUNT(RESIDENCIA) AS 'CANTIDAD DE CLIENTES'
FROM CLIENTES GROUP BY RESIDENCIA;
```

Tabla	Columna	Tipo	Motivo
CLIENTES	RESIDENCIA	Requerido	Tener las residencias de los clientes

Reporte: Tipo de cliente y su preferencia sobre los productos (al menos por edad).

- Consulta SQL que podría servir:

```
SELECT cli.EDAD, art.FAMILIA, COUNT(*) AS 'CANTIDAD DE COMPRAS'
FROM CLIENTES cli, VENTAS ven, DETALLE_VENTA det, ARTICULOS art
WHERE cli.ID_CLIENTE = ven.ID_CLIENTE AND ven.DOCUMENTO = det.DOCUMENTO
AND det.CODIGO_PRODUCTO = art.CODIGO_PRODUCTO
GROUP BY cli.EDAD, art.FAMILIA;
```

```
SELECT cli.EDAD, art.NOMBRE, COUNT(*) AS 'CANTIDAD DE COMPRAS'
FROM CLIENTES cli, VENTAS ven, DETALLE_VENTA det, ARTICULOS art
WHERE cli.ID_CLIENTE = ven.ID_CLIENTE AND ven.DOCUMENTO = det.DOCUMENTO
```

```

AND det.CODIGO_PRODUCTO = art.CODIGO_PRODUCTO
GROUP BY cli.EDAD, art.NOMBRE;

```

Tabla	Columna	Tipo	Motivo
CLIENTES	EDAD	Requerido	Para crear tipos de clientes
ARTICULOS	NOMBRE	Requerido	Para identificar preferencia según el nombre
ARTICULOS	FAMILIA	Sugerido	Se podría identificar la preferencia según la familia (categoría)

Reporte: Relación entre las ventas en \$ por vendedor y la cantidad de horas de capacitación que reciben (tipo de capacitación y horas de capacitación).

- Consulta SQL ventas por vendedor y cantidad de hs de capacitación:

```

SELECT vdd.Nombre, vdd.Apellido, SUM(vnt.TOTAL_NETO) as 'TOTAL_VENDIDO',
vdd.HORAS_CAPACITACION
FROM VENEDORES vdd, VENTAS vnt
WHERE vdd.ID_VENDEDOR = vnt.ID_VENDEDOR
GROUP BY vdd.Nombre, vdd.Apellido, vdd.HORAS_CAPACITACION;

```

- Consulta SQL ventas por vendedor y tipo de capacitación:

```

SELECT vdd.Nombre, vdd.Apellido, SUM(vnt.TOTAL_NETO) as 'TOTAL_VENDIDO',
vdd.[TIPO_ CAPACITACION]
FROM VENEDORES vdd, VENTAS vnt
WHERE vdd.ID_VENDEDOR = vnt.ID_VENDEDOR
GROUP BY vdd.Nombre, vdd.Apellido, vdd.[TIPO_ CAPACITACION];

```

Tabla	Columna	Tipo	Motivo
VENEDORES	Nombre	Requerido	Identificar vendedor
VENEDORES	Apellido	Requerido	Identificar vendedor
VENEDORES	TIPO_CAPACITACION	Requerido	Tener el tipo de capacitación
VENEDORES	HORAS_CAPACITACION	Requerido	Tener la cantidad de hs de capacitación
VENTAS	TOTAL_NETO	Requerido	Tener la cantidades que vendió el vendedor

Reporte: Ventas mensuales y anuales (expresadas tanto en \$ como en dólares) por categoría de producto.

- Consulta SQL ventas mensuales por categoría de producto:

```
SELECT art.FAMILIA, MONTH(ven.FECHA) AS 'MES', SUM(ven.TOTAL_NETO)
'VENTAS'
FROM ARTICULOS art, DETALLE_VENTA det, VENTAS ven
WHERE art.CODIGO_PRODUCTO = det.CODIGO_PRODUCTO AND det.DOCUMENTO =
ven.DOCUMENTO
GROUP BY art.FAMILIA, MONTH(ven.FECHA);
```

- Consulta SQL ventas anuales por categoría de producto:

```
SELECT art.FAMILIA, YEAR(ven.FECHA) AS 'AÑO', SUM(ven.TOTAL_NETO) 'VENTAS'
FROM ARTICULOS art, DETALLE_VENTA det, VENTAS ven
WHERE art.CODIGO_PRODUCTO = det.CODIGO_PRODUCTO AND det.DOCUMENTO =
ven.DOCUMENTO
GROUP BY art.FAMILIA, YEAR(ven.FECHA)
```

Tabla	Columna	Tipo	Motivo
ARTICULOS	FAMILIA	Requerido	Para categorizar por familia
VENTAS	FECHA	Requerido	Poder sacar anuales y mensuales
VENTAS	TOTAL_NETO	Requerido	Tener las ganancias

2. Identificar los datos externos que no están disponibles en la base de datos “transaccional” provista pero que deberán ser incluidos en el DataWarehouse para poder resolver las necesidades de reporting ya mencionadas. Sugerir de donde podrían obtenerse.
- Para resolver el reporte de ventas anuales y mensuales en dólares es necesario el dato de cambio de peso a dólar, que se podría sacar de la página del Banco Central de la República Argentina.
 - Para el reporte de distribución de clientes por región es necesario obtener de afuera los datos de a qué región pertenece cada una de las comunas en las que viven los clientes, se podría buscar en internet alguna base de datos donde estén hechas las tablas comunas, regiones, provincias similar a la que hay en la base de datos de TRADEProd pero que contengan todas las comunas que existen con sus respectivas regiones y provincias.

3. Investigar tres motores de base de datos que se podrían utilizar para el Datawarehouse. Mencionar los mismos y elegir uno de ellos para la implementación de nuestro DW considerando lo mencionado en la sección Escenario.
- SQL Server: es el motor de base de datos que ya se encuentra en uso en la empresa. Es un motor de base de datos relacional que ha sido diseñado para manejar grandes cantidades de datos y soporta el procesamiento en paralelo para mejorar el rendimiento. Además, SQL Server tiene integración con otras herramientas de Microsoft, como Excel y Power BI, lo que podría facilitar la generación de reportes.
 - Oracle Database: es otro motor de base de datos relacional que también es muy popular en el mercado. Es escalable y tiene una amplia gama de herramientas y funciones para la gestión de datos. Además, posee una arquitectura distribuida que permite el procesamiento de grandes cantidades de datos en múltiples servidores.
 - Teradata: es un motor de base de datos especializado en Data Warehousing y análisis. Es conocido por su capacidad para manejar grandes volúmenes de datos y su alta escalabilidad. Ofrece herramientas y características como el almacenamiento en columnas y la compresión de datos.

Teniendo en cuenta el escenario mencionado, el motor de base de datos que el elegiría para la implementación del DataWarehouse sería SQL Server. Esto se debe a que la empresa ya posee licencias de SQL Server y cuenta con DBAs con experiencia en dicha plataforma. Además, SQL Server es un motor de base de datos muy popular y tiene una amplia gama de herramientas y funciones para la gestión de datos, lo que podría ser beneficioso para el análisis de las ventas de la empresa.

4. Investigar sobre herramientas de ETL y realizar un cuadro comparativo entre dos de ellas (elegidas por ustedes) considerando al menos 4 características de las mismas justificando la decisión de utilizar una de las dos para la construcción de los ETL.
- Talend: es una herramienta de código abierto que permite a los usuarios extraer, transformar y cargar datos de manera eficiente. Tiene una interfaz gráfica de usuario fácil de usar que permite a los usuarios diseñar y ejecutar flujos de trabajo complejos. También cuenta con una amplia gama de conectores preconstruidos para diferentes fuentes de datos.
 - Informática PowerCenter: es una herramienta comercial que se utiliza para la integración de datos. Es conocida por ser escalable y por tener una amplia gama de características para la extracción, transformación y carga de datos. También posee una amplia gama de conectores preconstruidos para diferentes fuentes de datos.

Características	Talend	Informatica PowerCenter
Tipo de licencia	Código abierto	Comercial

Facilidad de uso	Interfaz gráfica de usuario fácil de utilizar	Curva de aprendizaje más empinada
Conectores preconstruidos	Amplia gama de conectores preconstruidos para diversas fuentes de datos	Amplia gama de conectores preconstruidos para diversas fuentes de datos
Escalabilidad	Escalable para grandes volúmenes de datos	Escalable para grandes volúmenes de datos

Debido a que es una herramienta de código abierto, es decir de acceso gratuito sin costos de licencia, y que tiene una interfaz gráfica fácil de utilizar (más que nada por el poco tiempo que disponemos para aprender a utilizar la herramienta) de las dos herramientas elegiría Talend.