PRUEBA TÉCNICA ARQUITECTO DE DATOS JIKKOSOFT S.A.S

Sección 2: Habilidades prácticas

A. Escritura de consultas SQL:

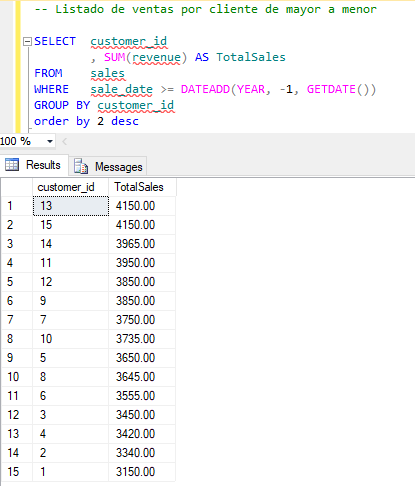
1. Escriba una consulta SQL para encontrar los 5 principales clientes por ingresos del último año en la tabla de ventas.

Usamos scripts:

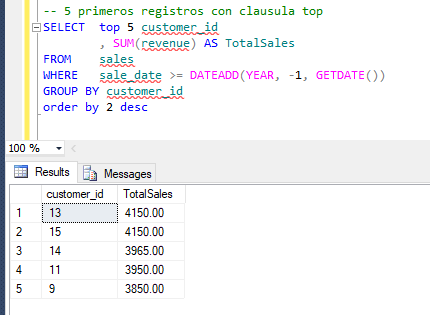
\_Script\_1.sql

Seccion2\_Punto\_A\_Script\_2.sql

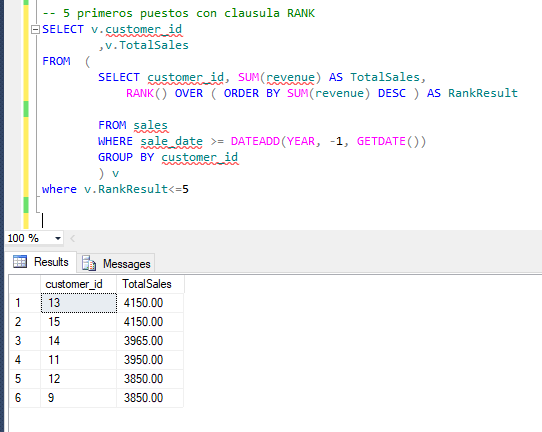
Partimos de la consulta general de ventas:



Dependiendo de los requerimientos del negocio podrían generarse las siguientes opciones:



Pero de acuerdo a los datos existen clientes con los mismos valores de ventas, una posible implementación sería mediante el uso de la función ventana RANK.

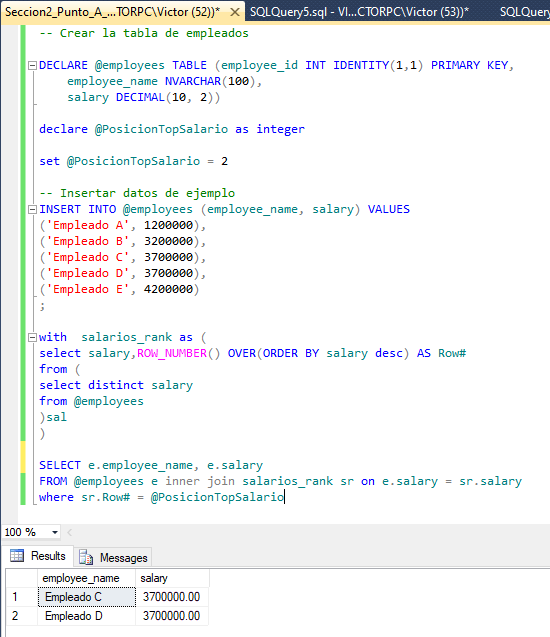


Aquí traemos los registros incluyendo cuando hay empate en la posición 5, por lo que en este caso nos retorna 6 registros.

1. Escriba una consulta SQL para recuperar el segundo salario más alto de la tabla de empleados.

Podríamos usar una solución parecida a la del punto uno, pero podríamos tener inconsistencias cunado varios empleados tienen el mismo salario, por tanto realizamos una implementación diferente.

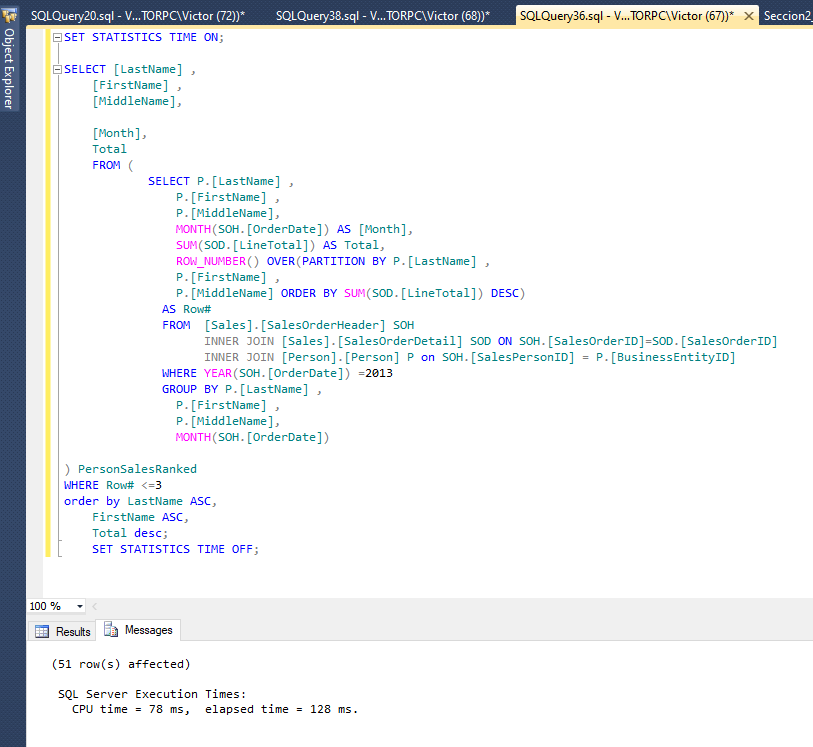
Script: Seccion2\_Punto\_A\_Script\_3.sql



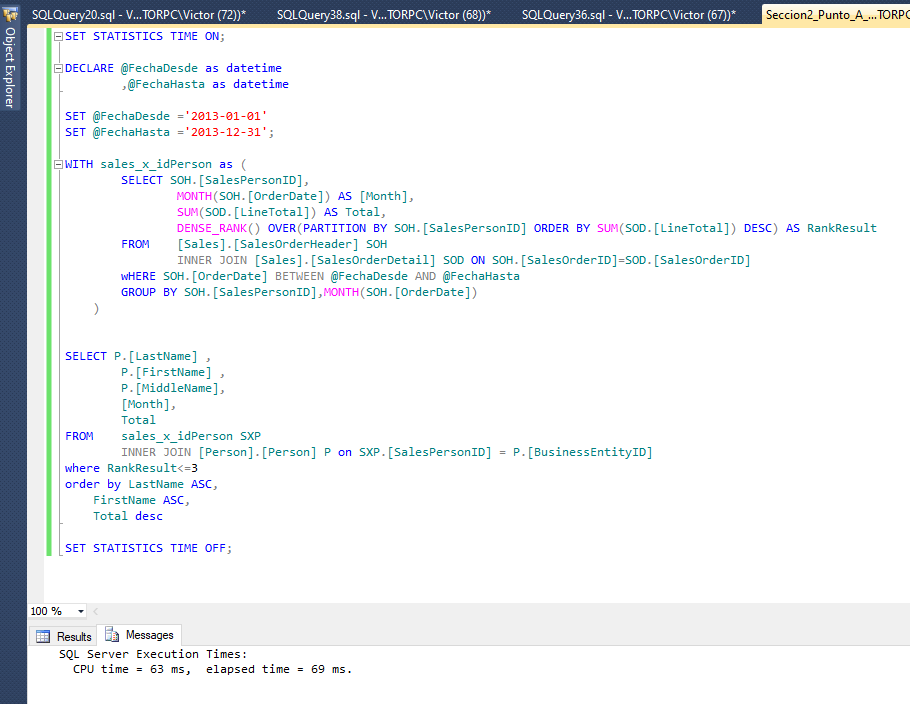
1. Optimice una consulta SQL y explique las mejoras de rendimiento.

Script: Seccion2\_Punto\_A\_Script\_4.sql

Tenemos la siguiente consulta que nos retorna los tres meses de mayor venta por cada vendedor en el año 2013.



Se realiza una reorganización de la consulta y se aplican optimizaciones:



Se reemplaza la forma en que se realiza el filtro de fecha para que se realice en el mismo tipo de datos del campo, de esta manera no se hagan transacciones adicionales de conversión sobre la campo en la consulta, adicionalmente pueda tomar el índice que tiene l tabla sobre el campo [OrderDate].

Se realiza la consulta de agrupación de los datos de ventas en un CTE, para realizar la agregación por vendedor y mes, el ordenamiento por monto de los datos. También es útil para organizar la consulta en secciones que permitan simplificar la complejidad y el mantenimiento futuro.

El proceso de ordenamiento y clasificación por monto y mes se realiza con la Función Ventana DenseRank, que es más apropiada para el objetivo, simplifica la complejidad y evita inconsistencias cuando hay empates de valores en las posiciones buscadas.

Una vez los datos se encuentran Agrupados y clasificados por monto y mes, se cruza el resultado con la tabla de vendedores de esta forma sólo se hace el cruce con los datos ya agrupados, evitando el cruce con el detalle de registros.

Con todos estos ajustes se tiene una consulta menos compleja, con mayor facilidad de mantenimiento y se mejoran los tiempos de respuesta cuando los volúmenes de datos son altos.

B. Diseño de la base de datos:

1. Diseñe un esquema de base de datos normalizado para una librería en línea. Incluya tablas para libros, autores, editores, clientes y pedidos. Nota: se aprecia en gran medida incluir requisitos funcionales adicionales como manejo de stock y seguimiento de envíos.

Se usan los scripts: Seccion2\_Punto\_B\_Diagrama1.bdml (<https://dbdiagram.io/home>)

Imagen: TiendaLibros.png



1. Explique sus decisiones de diseño y cree un diagrama de ER para el esquema diseñado.

Se realiza el diseño de la base de datos, contemplando un sistema para una tienda de libros en línea, que pueda vender copias de obras literarias en distintos formatos (libros, revistas, digitales, video, audio, etc.) en diferentes presentaciones (Tapa dura, Tapa blanda, argollados, digitales, cd, DVD, etc.).

El concepto libro, nos permite definir la metada fundamental de la obra literaria, como son el título, autores, genero, etc.

Cómo tal la tiende vende copias de estos libros a las que llamamos ediciones que corresponden al ejemplar que se está vendiendo, un libro puede tener muchas ediciones, en diferentes formatos, presentaciones, de diferentes editores.

Las copias físicas de cada edición de un libro se pueden controlar mediante un manejo de existencias, las cantidades pueden estar almacenadas en diferentes ubicaciones. El modelo puede extenderse pudiendo realizar identificación geográfica de estas ubicaciones, para efectos del ejercicio, creo que ilustra de manera suficiente la funcionalidad.

Los pedidos se manejan con una estructura sencilla de encabezado y detalla. En donde el detalle corresponde a la identificación de las ediciones que se están facturando, las cantidades, costos, descuentos e impuestos.

A nivel de encabezado del pedido, podemos identificar al cliente del pedido, el estado actual del pedido, nos permite adicionalmente con estructuras relacionadas, que un cliente pueda usar varias formas de pago al tiempo para cancelar su pedido (TC, paypal, cupones, etc.), y generar un envío en una estructura aparte.

Los envíos contendrán los datos de ubicación del destinatario, así como la trasportadora, y numero de guía.

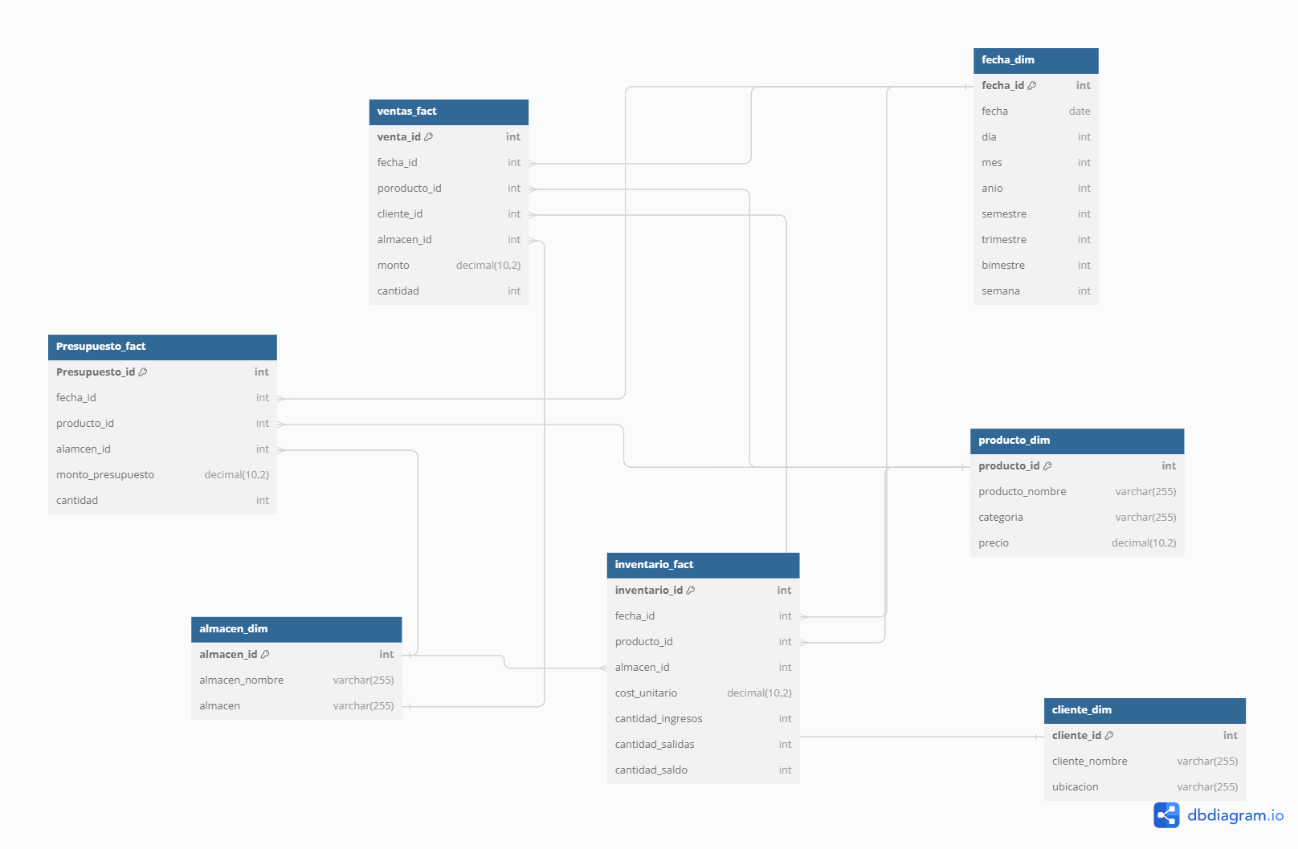
Se tendrá una estructura para almacenar el seguimiento de los envíos con la información que pueda recuperarse de la trasportadora.

C. Modelado de datos:

1. Desarrolle un esquema de estrellas para un almacén de datos de ventas al por menor. Identifique la tabla de hechos y las tablas de dimensiones, incluyendo múltiples tablas de hechos y dimensiones compartidas.

Se usan los scripts: Seccion2\_Punto\_B\_Diagrama2.bdml (<https://dbdiagram.io/home>)

Imagen: DatamartVentas.png



D. Transformación de datos:

1. Proporcione un flujo de trabajo de muestra de ETL para extraer los datos de los clientes de una base de datos transaccional, transformarlos mediante la limpieza y la agregación, y cargarlos en un almacén de datos.

El siguiente es un flujo típico de un proceso de cargue de información a un datawarehouse, en donde se considera un grado de parametrización que flexibilice la mantenibilidad de las ETLs, así como estructuras de control y funcionalidades de notificación.

* Se recibe parámetro de la tabla o consulta que será cargada al almacén de datos, si es carga total o los datos del delta de información, así como la identificación el proceso de trasformación a ser ejecutado, así cómo la información requerida apra notificaciones (emails)
* Se consulta la metadata de la estructura de entrada (consulta) y de la tabla temporal en área staging y de la estructura destino en el datawarehouse.
* Se registra en la tabla de control los datos de seguimiento, fecha inicio, estructura, etc.
* Se realiza el proceso de extracción de la información de acuerdo a la metadata y el delta, almacenando la información en el área staging.
* Se registra en la tabla de control la cantidad de registros extraidos.
* Se ejecuta el proceso de trasformación identificado al inicio de la ETL el cual contiene las reglas de trasdomación y limpieza de los datos. Dejando la información preparada para la carga en una estructura de datos y en una estructura de errores, se registran la información correspondiente para los registros descartados o con incosistencias.
* Se realiza la cargada de la información que fue preparada para el cargue a la estructura destino.
* Se registra en la tabla de control las cantidades cargadas.
* Se registra en la tabla de control el éxito del proceso o la falla de acuerdo a lo definido por los requerimientos.