

Trabajo Práctico Especial

Base De Datos I (72.37)

Integrantes:

- Kim, Azul Candelaria (Legajo: 60264)
- Lombardi, Matias Federico (Legajo: 60527)
- Rosati, Santos Matias (Legajo: 60052)

Roles:

- Encargado del informe: Kim, Lombardi, Rosati
- Encargado de las funciones: Kim, Lombardi
- Encargado del trigger: Lombardi, Rosati
- Encargado del funcionamiento global del proyecto: Kim, Lombardi, Rosati
- Encargado de investigación: Kim, Rosati

Cabe destacar que la realización del trabajo fue en simultáneo, por lo cual estos roles no fueron completamente estrictos, sino que se fueron intercambiando de acuerdo al contexto.

Investigación:

Para la correcta realización del Trabajo Práctico, mientras surgían distintas dudas y dificultades al momento de implementarlo (explicadas en la siguiente sección) fue necesario investigar sobre ciertas funcionalidades de PostgreSQL (ya que hay varias reglas en las que se diferencia ampliamente con el ANSI) o SQL en general. A continuación se explican brevemente las búsquedas realizadas (no se limitan únicamente a las mismas, se mencionan las que tienen una mayor relevancia en el *scope* del trabajo).

Al momento de obtener el mes para insertar en la tabla *definitiva*, se utilizó una función que con los parámetros recibidos, retorna el número de mes correspondiente de acuerdo a las 3 letras que lo definen (por ejemplo para el mes de enero, su correspondiente sería Jan).

Respecto a la obtención de datos para las tablas, se buscó en la documentación de PostgreSQL la manera de importar la información del archivo CSV provisto por la cátedra. Para eso, se utilizó el comando COPY con las opciones WITH DELIMITER ',' CSV HEADER.

Dentro de la función *MedianaMargenMovil*, se buscó la manera de al restarle meses a una fecha, obtener una nueva. La forma de hacerlo es creando un intervalo de 1 mes y multiplicarlo por la cantidad de meses deseada.

Para la validación de los datos ingresados a la tabla *intermedia*, se debió usar CHECKS junto con una expresión regular para validar el formato del input que se estaba recibiendo.

Los links correspondientes se encuentran en la Bibliografía.

Dificultades encontradas:

Se generaron dudas respecto a la consigna del cálculo de la mediana donde en el caso en el que el segundo parámetro sea menor o igual a 0, "no se debe mostrar nada". Además, se presentaron dificultades en el intervalo de fechas que se utiliza en el primer parámetro. Al invocar la función con los mismos datos plasmados en los ejemplos de uso, no se obtenían los mismo resultados. Estos dos convenientes fueron consultados con las profesoras y, consecuentemente, se resolvieron.

A su vez, al realizar la importación de datos desde el CSV fuente, al utilizar el comando *COPY*, no se podía utilizar un path relativo por lo que la consulta fallaba. Buscando los posibles errores, se encontró en la documentación de PostgreSQL que cuando se utiliza este comando, lo mejor es utilizar un path absoluto al archivo, por lo que este ítem se encuentra clarificado en la sección correspondiente.

También surgieron dudas al momento de consultar los campos *Quarter*, *Week* y *Month* de la tabla *intermedia*, ya que al hacer un análisis de los datos recibidos, el formato de los mismos siempre se mantenía. Para garantizar el correcto funcionamiento del sistema (tanto de los insert, como las funciones auxiliares) se decidió agregar un chequeo para estos campos mediante la cláusula *CHECK* para que se reciba el formato esperado.

Otra duda que surgió fue respecto de la generalización. En la consigna se pide que las consultas sean genéricas. Sin embargo, fue necesario ser más específicos en algunos sectores (especialmente al utilizar cursores). Para la generalización de la *query* que define las tuplas recuperadas en el cursor, se consideró pasarle como parámetro un String que refiere a la columna y utilizando una sentencia CASE en el código. Sin embargo, esto hace más engorroso el código y no hay una mejora sustancial en la calidad del código, por lo que se decidió conservar la implementación original de tres cursores, haciendo más legible el código.

En base al análisis de los datos provistos por la cátedra, se encontró que el campo *Quarter* está definido por los campos *Week* y *Month*, ya que los mismos contienen también al año al que pertenecen, y entre estos dos datos se puede descifrar el valor de *Quarter* (la vuelta no se cumple, ya que en un mismo trimestre hay 3 meses, valga la redundancia). Debido a esto, se decidió que este campo no formen parte de la clave de la tabla *intermedia*. Además, los campos *Revenue* y *Cost* no identifican a ninguna tupla en particular (ya que puede haber tuplas con el mismo valor de estos campos en momentos o categorías diferentes), por lo cual estos tampoco pertenecen a la clave.

Importación de Datos:

Para la importación de los datos se creó una base de datos local, en donde se crearon las tablas *intermedia* y *definitiva*. Ambas comparten los mismos atributos, a excepción del formato de las fechas: en la tabla *intermedia* se utilizan 3 atributos que especifican el trimestre, la semana y el mes, mientras que en *definitiva* se guarda la fecha utilizando un mapeo para la semana.

Luego, ejecutando el comando *COPY*, como se encuentra en el archivo funciones.sql (cambiando el path por el path absoluto al archivo donde se encuentran las ventas) se importan los datos desde el csv hacia la tabla *intermedia*. Para insertar los mismos en la tabla *definitiva*, se creó un *TRIGGER* que se dispara luego de insertar una tupla en *intermedia* y tiene en cuenta el mapeo mencionado para el día de la fecha.

Bibliografía:

- [Convertir mes en número](#)
- [Importar CSV a tabla](#)
- [Crear un intervalo dinámicamente](#)
- [Cómo usar una regex en SQL.](#)