



Título:

TRABAJO PRÁCTICO ESPECIAL

Nivel y Área:

72.37 – Bases de Datos I
1º Cuatrimestre 2018

Comisión: S - **Carrera:** Informática - **Grupo:** 1

Fecha: 14 de junio de 2018

Alumnos Expositores:

- AQUILI, Alejo Ezequiel 57432
- BASSANI, Santiago 57435
- IZAGUIRRE, Agustín Emilio 57774

Trabajo Práctico Especial

Introducción:

El trabajo práctico especial, presenta como objetivo administrar una migración de una base de datos conformada por el registro de las bicicletas publicas del gobierno de la ciudad “ecoBici” del año 2016 exportada a un archivo CSV hacia una base de datos PostgreSQL, construyendo tablas con restricciones que nos permite el uso de SQL avanzado (PSM y Triggers) lo cual nos provee funcionalidades que nos permiten representar restricciones más “fuertes” que las que podemos representar con Primary Keys, Foreign Key, etc.

Dificultades:

La primera dificultad que se presentó durante la implementación del trabajo se produjo en la Importación con tipo interval el campo tiempo_uso, pues existían tuplas en el archivo CSV que contenían espacios en los números de horas minutos o segundos. Para solucionarlo antes de realizar el parseo para la conversión a interval se realiza una validación del formato exportado en la función `remove_invalid_null_fields_and_time_used_invalid_format()`.

Luego se presento un problema con el tiempo de ejecución, al realizar la migración sobre el CSV con todos los recorridos, transcurrían 30 minutos y no terminaba la ejecución. Se consultó a los docentes y Valeria nos sugirió incorporar una Primary Key en la tabla auxiliar recorrido_temp como una mejora en la performance, lo cual genera índices y se obtiene una ejecución con tiempo promedio de 2 minutos para la migración del archivo CSV con todos los recorridos.

Finalmente, en lo que respecta a la construcción del Trigger, el cual no verificaba que coincidiera el usuario en el recorrido para validar el solapamiento de la tupla a insertar con respecto a las tuplas existentes, de esta forma en el archivo test1.csv funcionaba correctamente por no haber solapamiento entre los distintos usuarios, pero si en el archivo con todos los recorridos del año 2016. Para detectar esta falla en una etapa de debugging, Leticia nos sugirió utilizar “RAISE NOTICE” para poder tener registro de los valores con los que fallaba la inserción por solapamiento dentro del Trigger, lo que nos permitió poder corregir este problema.

Comentarios y contenidos por fuera de la materia:

Se optó por guardar el primer periodo cuando se detectaba solapamiento de los intervalos, ya que la consigna no aclaraba nada al respecto.

Se usaron dos tablas auxiliares las cuales se eliminan al finalizar la ejecución de la migración y se decidió garantizar que la tabla original migrada del CSV no se altere durante el proceso de migración.

Se decidió utilizar funciones de manejo de strings y cualquier otro tipo de función que respeten el estándar ANSI como lo son SUBSTRING, POSITION, CAST, etc. Las únicas prestaciones propias del DBMS PostgreSQL fueron: los tipos de datos TEXT (solicitado en la tabla recorrido_final requerida en la consigna) el cual es fácilmente reemplazable por tipos VARCHAR que si respetan el estándar. Por otra parte, en lo que respecta al manejo de los Triggers, este manejo es específico para cada DBMS.

Siendo estas dos cuestiones mencionadas las únicas para tener en cuenta a la hora de migrar a otro motor que no sea PostgreSQL y que respete el estándar ANSI.

En cuanto a la tarea de investigación y como consecuencia de la decisión de adherirse al estándar ANSI, se investigo si las funciones utilizadas respetaban el estándar y se investigo del uso y funcionamiento de la función POSITION.

Roles:

El alumno Alejo Aquili fue encargado de informe y de funcionamiento global. El alumno Santiago Bassani fue el encargado del trigger. Finalmente, el alumno Agustin Izaguirre fue el encargado de funciones y testeo integral de las mismas.

Instalación:

Se debe extraer los archivos del trabajo práctico especial, tener instalado make. Ejecutar make.

Se debe especificar el archivo CSV a manipular, para ello se debe cambiar el string de la línea 15 del archivo funciones.sql.

Conclusión:

Concluimos que existe una fuerte dicotomía entre los distintos DBMS, y experimentamos lo que implica un proceso de migración a una base de datos.

Por otra parte, se concluyó que es de vital importancia para implementar restricciones más complejas sobre la base de datos que tengamos mediante el uso de Triggers y PSM.