



## **PRACTICO Nº 5**

### **DATA WAREHOUSING Y DATA MINNING**

- 1) ¿Que es un almacén de datos? ¿En que se diferencia de una base de datos?
- 2) Defina los términos: OLAP, ROLAP, MOLAP, DSS
- 3) Describa los pasos a la hora de crear un almacén
- 4) ¿Cual es la relación entre el concepto de vista relacional y el concepto de almacén de datos y data marts?
- 5) Mencione y describa los dos esquemas mas conocidos para la implementación de los Almacenes de Datos
- 6) Sobre la BD Pedidos provista por la cátedra esbozar las sentencias SQL para resolver las siguientes consultas (utilizando GROUPING SETS, ROLLUP o CUBE):
  - a. Total productos pedidos discriminado por producto empleado, por producto cliente y por producto cliente/empleado
  - b. Cantidad de cada producto pedido por año, año-mes
  - c. Obtener el ranking de los productos mas pedidos a los menos pedidos
- 7) Muéstrese el modo de expresar group by cube(a, b, c, d) utilizando rollup; la respuesta sólo debe tener una cláusula group by.
- 8) Describir las diferentes técnicas para la Minería de Datos



## **LABORATORIO Nº 5**

### **DATA WAREHOUSING Y DATA MINNING**

- 1) Implementar la bd operacional de la cooperativa PatSur con sus dos sistemas de facturación
- 2) Hacer inserciones masivas a todas las tablas mediante scripts similares al implementado en Lab 1
- 3) Implementar un DW (en otra BD u otro esquema de la misma BD en postgres) para la cooperativa PatSur, basado en la BD operacional creada y utilizando el esquema de Copo de Nieve presentado en la teoría (Tabla de hecho Ventas, medidas unidad y precio, dimensiones: sucursal, región, provincia, ciudad, cliente, tipo cliente, producto, categoría, subcategoría, y la dimensión Tiempo existente en todo DW). Además implementar los scripts para hacer la ETL correspondiente desde la BD operacional.
- 4) Implementar consultas para las vistas que los analistas requirieron
  - Venta vista por mes o por año, por sucursal, por región, por cliente y demás combinaciones entre las perspectivas.
  - El mínimo nivel de detalle que se quiere tener disponible para el análisis de las ventas (\$ vendidos y unidades vendidas) es el de la facturas.
  - Es necesario conocer también de que manera influye, en las ventas de productos, la zona geográfica en la que están ubicados los locales.
  - De cada cliente se desea conocer cuales son los que generan mayores ingresos a la cooperativa.
  - Se necesitará hacer análisis diarios, mensuales, trimestrales y anuales.

Para las mismas se deben hacer los SELECT y comentarlos con las funciones GROUPING SETS, ROLLUP y CUBE brindadas por las extensiones del lenguaje SQL:1999 para facilitar la agrupación de los datos.

5) Se deberá hacer una implementación en una plataforma Open Source de Business Intelligence, tomando los datos del Data Warehouse generado y armando un cubo con las medidas y dimensiones señaladas, luego utilizar la herramienta OLAP para generar las vistas indicadas en el punto 4. Basándonos en el siguiente artículo que menciona diferentes suites open source para reporting o business intelligence:

<https://opensource.com/business/16/6/top-business-intelligence-reporting-tools>

Cada grupo va a tener que investigar/implementar una de ellas para explotar el Dw según el siguiente detalle:

- Grupo 1: BIRT
- Grupo 2: JasperReport
- Grupo 3: Pentaho
- Grupo 4: SpagoBI
- Grupo 5: KNIME
- Grupo 6: ReportServer

En el mismo artículo se dan los links a los sitios principales y de descarga de cada plataforma.



El presente laboratorio tiene las siguientes pautas adicionales:

- El Lab 5 Oficia también de Trabajo Práctico final
- Puede hacerse individualmente o de a dos como máximo
- En el informe debe constar
  - o Como hicieron los scripts de carga de la base de datos operacional y su llenado en general
  - o Como hicieron el script de carga del Data Warehouse para cada sistema de facturación
  - o Las ejecuciones que hacen del script para llenar el DW
  - o Respuesta a las consultas del punto 4 (si no pueden hacer todas, algunas) con SQL con las extensiones de agrupamiento, etc
  - o El desarrollo de una pequeña reseña de la herramienta asignada de reporting o business inteligente y de como la pudieron implementar (o si no pudieron los problemas con los que se encontraron) y el o los reportes del punto 4 que pudieron implementar con dicha herramienta.
- Además de la entrega del informe y los scripts y archivos de trabajo, deberá exponerse la solución ante la cátedra y compañeros de cursada en fecha a designar