



**Universidad Católica de Santiago del Estero**

Departamento Académico Rafaela

B. Hip. Irigoyen 1502 – Rafaela – Santa Fe – República Argentina

Te: 03492-432832-433408-433550 - [ucsedar@ucse.edu.ar](mailto:ucsedar@ucse.edu.ar)

---

## **Electiva III: Business Intelligence y Big Data**

### **Trabajo Práctico N° 1: Data Warehouse y Data Mining**

#### **Fecha de Entrega:**

- 24/10/2016 – 23:59 hs.

#### **Fecha de Exposición:**

- 26/10/2016.

### **Equipo 2: Chiabotto, Faraudello, Pieruccioni, Toledo.**

#### **Objetivo**

Este trabajo práctico tiene como objetivo la implementación de un Data Warehouse que, a partir de una o varias Bases de Datos Operativas previamente existentes correspondiente a sistemas transaccionales que se encuentran operando y con datos cargados, permitirá responder a una serie de requerimientos planteados por la dirección

#### **Consignas**

Se deberá implementar un Data Warehouse que ofrezca al usuario las herramientas, en este caso algunos cubos OLAP, para responder con facilidad a los requerimientos detallados en el Trabajo Práctico.

- La presentación del TP (una vez terminado) se realizará subiendo el archivo comprimido (o link desde donde poder bajarlo) al entorno virtual de la materia, en la tarea abierta a tal fin. La fecha tope para la entrega del TP es el 24/10/2016 a las 23:59 hs.

Los archivos que deberán entregarse:

- Documento (preferentemente pdf) donde detalle:
  - Integrantes del grupo. Hasta 4 personas.
  - Tiempo efectivo utilizado en la resolución del TP.
  - Comentario acerca de la dificultad que le ofreció cada una de las partes y cualquier otra observación que le parezca relevante mencionar.
  - Esquemas Dimensionales que documenten la solución implementada.
  - Una captura por cada reporte solicitado en el enunciado, con los datos pedidos desplegados.
- Agregar además un archivo zip o rar que contenga:



**Universidad Católica de Santiago del Estero**

Departamento Académico Rafaela

B. Hip. Irigoyen 1502 – Rafaela – Santa Fe – República Argentina

Te: 03492-432832-433408-433550 - [ucsedar@ucse.edu.ar](mailto:ucsedar@ucse.edu.ar)

---

- Backup de la DB del DW (en formato plano sql).
- Exportación de las transformaciones generadas.
- Archivo/s con el diseño de los cubos.
- El nombre de los 2 archivos (un documento de texto y un zip o rar) debe ser: <apellidos de los integrantes del grupo en minúsculas, separados por "-" en orden alfabético>.<extensión>
- Por favor, realizar una sola entrega por cada grupo del TP.

### **Caso de Estudio: Distribuidora Mayorista**

La empresa distribuidora de productos alimenticios Distribuidora Mayorista desea instalar un sistema de DW para hacer un seguimiento más eficiente de sus productos.

Se trata de una empresa nacional, que cuenta con diversos centros de fabricación y/o elaboración de productos alimenticios y trabaja también en cooperación con productores agrícolas de la región. La empresa se encarga también de la distribución de los productos en todo el territorio nacional.

Se comenzó con la distribución de productos envasados y bebidas, incorporándose luego los lácteos y panificados. Recientemente, gracias a los acuerdos con cooperativas agrarias se incluyó la distribución de productos agrícolas. Muchos de los productos que se distribuyen son muy perecederos (la mayor parte de los lácteos, panificados y vegetales), por lo que se debe ajustar muy bien las cantidades en stock de estos productos.

La empresa trabaja con empresas mayoristas y supermercados, pero también con almacenes y restaurantes. Algunos de estos clientes tienen casas en varias ciudades del país por lo que debe resolverse el traslado de mercaderías al interior. Actualmente se está apuntando a incrementar las ventas en las ciudades del interior y ganar mercado incorporando comercios locales.

La empresa desea resolver los siguientes requerimientos:

#### Evolución de las ventas:

- Se desea hacer un seguimiento de las ventas comparando los distintos meses del año, y del año anterior, estudiando la evolución por familia de productos, y pudiéndola refinar hasta un producto en concreto.
- Se desea también observar las variaciones en las ventas para las distintas ciudades del país.

#### Análisis de mercado:



## Universidad Católica de Santiago del Estero

Departamento Académico Rafaela

B. Hip. Irigoyen 1502 – Rafaela – Santa Fe – República Argentina

Te: 03492-432832-433408-433550 - [ucsedar@ucse.edu.ar](mailto:ucsedar@ucse.edu.ar)

- Las diferentes promociones están orientadas a un determinado perfil de clientes, por lo que es necesario medir los volúmenes de venta para los diferentes rubros (mayoristas, supermercados, almacenes y restaurantes) estudiando los efectos positivos y/o negativos de la promoción en cada sector. No interesa comparar cliente por cliente, alcanza con un fraccionamiento vertical por rubros.

### Distribución geográfica.

- Interesa comparar las ventas por departamentos y ciudades. Esto nos indica las regiones que están en riesgo y necesitan de mayor atención.

### Desempeño de vendedores.

- Se necesita comparar el desempeño de los distintos vendedores, y la evolución de dicho desempeño a lo largo del tiempo.
- Es importante saber, cuanto influye en las ventas las horas de capacitación de los vendedores, como así también la antigüedad en la empresa.
- Un estudio de ventas por producto ayuda a planificar qué productos se asignarán a qué vendedores, y un estudio de ventas por ciudad ayuda a planificar las giras a las que se asignarán los mismos.

Luego de realizar un relevamiento inicial, se llega a las siguientes conclusiones:

1. Los orígenes de datos que alimentarán el data warehouse serán:
  - a. El sistema de facturación y ventas.
  - b. El sistema de administración de RRHH y liquidación de haberes.

## Bases de Datos Operativas

### Sistema de Facturación y ventas:

- Departamentos (Id\_depto, Nom\_depto, Zona)
- Ciudades (Id\_ciudad, Id\_depto, Nom\_ciudad, Población, Clasificación)
- Rubros (Id\_rubro, Nom\_rubro)
- Clientes (Id\_cliente, Nombre, Dirección, Teléfono, Ciudad, Departamento, Rubro, Categoría, Fecha alta)
- Facturas (Factura, Fecha, Cliente, Vendedor)
- Registros-Facturas (Factura, Artículo, Importe, Unidades)
- Artículos (Id\_artículo, Id\_producto, Id\_tamaño)
- Productos (Id\_producto, Id\_familia, Id\_duracion)
- Códigos (Tipo, Código, Descripción)



**Universidad Católica de Santiago del Estero**

Departamento Académico Rafaela

B. Hip. Irigoyen 1502 – Rafaela – Santa Fe – República Argentina

Te: 03492-432832-433408-433550 - [ucsedar@ucse.edu.ar](mailto:ucsedar@ucse.edu.ar)

---

- Vendedores (Id\_vendedor, Nombre, Dirección, Teléfono, Especialidad)

**Nota:** los productos son una agrupación de los artículos (por ejemplo: un producto puede ser "Salsa portuguesa" y uno de sus artículos "Salsa portuguesa, lata de 1/2 kg"). Los códigos contienen descripciones de códigos utilizadas por el sistema. El campo tipo indica a qué se refiere el código (se encuentran códigos de artículos, productos, tamaños, duraciones y familias).

#### **Sistema de Recursos Humanos:**

- Empleado(legajo, nombre, apellido, dirección, sueldo, horas\_capacitacion, fecha\_ingreso, id\_local)
- Telefono\_empleado(legajo, teléfono\_empleado)

## **OBSERVACIONES**

- Cada grupo puede usar para la resolución del TP el conjunto de herramientas que considere conveniente o con la que se sienta más cómodo.
- Desde la materia se provee un entorno de trabajo completo e integrado con todas las herramientas necesarias instaladas en una máquina virtual que puede ser utilizada con el producto (de uso gratuito) "Oracle Virtual Box".
- En dicha máquina está instalado el motor PostgreSQL donde pueden generarse tanto las BD operativas como el DW.
- En la maquina virtual entregada, además, se encuentra un conjunto de herramientas que pueden ser utilizadas para la resolución. Tiene los siguientes elementos:
  - SO: Linux Debian Stable (squeeze)
  - Entorno de Escritorio: KDE
  - Herramienta para Diagrama de Tablas: Druid Data Modeler
  - Motor de DB: PostgreSQL
  - Herramienta para ETL: Pentaho Data Integration
  - Herramienta para generación de Cubos: Schema Workbench
  - Motor OLAP: Mondrian
  - Herramienta para la Visualización de los cubos: JPivot
  - Herramienta para Despliegue de los cubos: Pentaho