



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

Electiva III: Business Intelligence y Big Data

Trabajo Práctico N° 1: Data Warehouse y Data Mining

Fecha de Entrega:

- 23/10/2017 – 23:55 hs.

Fecha de Exposición:

- 25/10/2017.

Equipo 1: Barsotti, Pellegrinet, Sclerandi.

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 23/10/2017 a las 23:55 hs.

Los archivos que deberán entregarse:

- Documento (preferentemente pdf) donde detalle:
 - Integrantes del grupo.
 - Tiempo efectivo (real) 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:
 - Backup de las DB's originales (en formato sql plano).



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

- 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: Zumba Gym

Luego de cinco años de estar operando, la empresa que administra diferentes gimnasios en la ciudad de Santa Fe ya cuenta con 10 (diez) gimnasios distribuidos por en la ciudad, y ciudades cercanas (Santo Tomé, Recreo, Paraná).

Aprovechando el éxito del merchandising relacionado con la marca registrada Zumba, la empresa logró desarrollar fuertemente su unidad de negocios de venta de productos en cada uno de los gimnasios. Entre los productos se incluyen indumentaria utilizada para la actividad deportiva, calzados deportivos y complejos vitamínicos. Además, en todos los gimnasios, hay heladeras con bebidas y algunos snack que se ofrecen a las personas que realizan actividades físicas.

Los problemas surgen en cada reunión gerencial: los datos sobre las ventas de abonos y productos no parecen confiables ni están listos en tiempo y forma para que el equipo de gerentes pueda tomar decisiones acertadas. Esto ocurre porque los distintos gerentes se manejan con diferentes fuentes de información (en algunos casos usando planillas de cálculo, hechas con ingreso manual de datos, minutos antes de las reuniones) y además porque cada vez que se requiere un informe de mediana complejidad, hay que solicitarlo al encargado de sistemas que demora por lo menos una semana en confeccionarlo ya que está demasiado ocupado con la atención a otros clientes corporativos.

En resumen: cada reunión es un caos por las diferentes versiones de la verdad que cada gerente tiene y en algunos casos porque ni siquiera se llega a tener la información necesaria o está demasiado desactualizada.

Debido a esta situación, la empresa decide contratar a su equipo, para que le dé una solución a este problema. El directorio ha escuchado hablar de palabras como sistemas de ayuda a la decisión, business intelligence, data warehouse y otras pero lo que realmente necesita es una solución de negocio para afianzar el crecimiento de la empresa.



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

En principio se quiere trabajar con la información de ventas, tanto de abonos a las actividades del gimnasio como de productos, para luego ir extendiendo el alcance del proyecto a otras áreas.

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 emisor y facturador de actividades físicas, que está instalado en todos los gimnasios.
 - b. Dos (2) sistemas distintos de facturación de productos y comidas para los gimnasios (ya que los más nuevos tiene un sistema que se contrató hace un año a una consultora, pero los gimnasios más antiguos tienen todavía un sistema desarrollado por el sobrino de uno de accionistas de la empresa).
 - c. El sistema de administración de RRHH y liquidación de haberes.
2. Se detecta que los analistas de negocio de la empresa necesitan analizar la información de ventas de un producto o de los abonos desde diferentes perspectivas. Por ejemplo, la venta de abonos vista por mes, por estación o por año, por gimnasio, por actividad física, por profesor que la dicta y demás combinaciones entre las perspectivas.
3. Realizar un análisis de los clientes, teniendo en cuenta sus horarios de asistencia al gimnasio, de forma de detectar horarios de mayor concurrencia, diferentes características de asistencia, etc.
4. El mínimo nivel de detalle que se quiere tener disponible para el análisis de las ventas de productos (\$ vendidos y unidades vendidas) es el de la línea de los ticket.
5. De cada empleado/instructor es necesario poder hacer análisis de sus actividades considerando también el impacto de las horas de capacitación recibidas.
6. Es necesario conocer también de que manera influye, en las ventas de productos, la zona geográfica en la que están ubicados los gimnasios.
7. De cada cliente se sabe si es mujer u hombre, y su nivel de estudio. Se desea conocer cuáles son las clientes que generan mayores ingresos a la empresa.
8. Se necesitará hacer análisis diarios, mensuales, trimestrales y anuales.
9. Realizar análisis acerca de la influencia de las diferentes ofertas en la venta de abonos y productos.

Bases de Datos Operativas

Sistema de Emisión y Facturación de abonos:

- Venta(nro_ticket, fecha, código_empleado, código_cliente, monto)
- Empleado(código_empleado, nombre, apellido, fecha_nacimiento)
- Cliente(código_cliente, nombre, apellido, dirección, email, cod_nivel_estudio, sexo)



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

- Telefono_cliente(código_cliente, ítem, teléfono_cliente)
- Item_venta(nro_ticket, código_producto, cantidad, monto_item)
- Producto(código_producto, descripción_producto, costo, cod_tipo_producto, cod_familia)
- Tipo_producto(cod_tipo_producto, descripción)
- Familia(cod_familia, descripción_familia)
- Asistencia(código_cliente, fecha)

Sistema de Facturación I de productos:

- Familia(id_categoria, descripción)
- Subcategoria(id_subcategoria, id_categoria, descripción)
- Producto(id_producto, id_subcategoria, descripción, precio_actual)
- Venta(numero_ticket, fecha_venta, id_empleado, id_cliente)
- Item_Venta(numero_ticket, id_producto, cantidad, precio)
- Empleado(id_empleado, nombre, apellido)
- Cliente(id_cliente, nombre, domicilio)
- Telefono_cliente(id_cliente, teléfono_cliente)

Sistema de Facturación II de productos:

- Cat(cod_cat, descripción)
- Subcat(cod_subcat, cod_cat, desc)
- Prod(cod_prod, cod_subcat, desc, precio_actual)
- Venta(nro_ticket, fecha_venta, cod_empleado, cod_cliente)
- Item_Venta(nro_ticket, cod_prod, cantidad, precio)
- Empleado(cod_empleado, nombre, apellido)
- Cliente(cod_cliente, nombre, domicilio)
- Telefono_cliente(cod_cliente, tel_cliente)

Sistema de Recursos Humanos:

- Empleado(legajo, nombre, apellido, dirección, sueldo, fecha_ingreso)
- Especialidades_empleado(cod_especialidad, legajo, Nivel_conocimiento, horas_capacitación)
- Especialidades(cod_especialidad, descripción)
- Actividad(id_actividad, descripción, fecha_inicio)
- Actividad_instructor(id_actividad, legajo, hora_inicio, hora_fin)
- Telefono_empleado(legajo, teléfono_empleado)



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

Nota: tenga en cuenta que la ubicación geográfica no se encuentra registrada, ya que en cada gimnasio existen diferentes instalaciones del sistema, que indican según los tickets a que gimnasio corresponden, pero esto debe ser tenido en cuenta en el diseño del DW. Además, cada sistema de facturación cuenta con su propia codificación de empleados y clientes (esto significa, por ejemplo, que un mismo código de empleado puede ser usado para dos empleados distintos de un gimnasio a otro, lo mismo con los códigos de clientes).

Además, indique que modificaciones se deberían realizar para responder a estos nuevos requerimientos:

1. Si Zumba Gym decidiera incorporar información acerca del stocks disponible de sus productos en el data warehouse. El stock deberá estar disponible a nivel de producto, día y local.
2. Si se lanzara una tarjeta de fidelización de clientes y se pudiera identificar a qué cliente se hizo cada venta (tanto de abonos como de otros tipos de productos).
3. Poder evaluar el impacto que tiene en las ventas, la antigüedad, capacitación y el salario de los empleados.
4. De cada local es necesario poder analizar las ventas (importe y cantidad) con respecto a la cantidad de actividades definidas y los metros cuadrados que disponen.

OBSERVACIONES

- Toda suposición que realice debe ser aclarada al iniciar el trabajo.
- Para este trabajo, deben utilizar las herramientas dadas en la materia, que se listan a continuación.
- 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 la máquina 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