Sistema de Inteligencia de negocio para el análisis publicitario

Juan Manuel Heras Gomez

Master en Ingeniería Informática

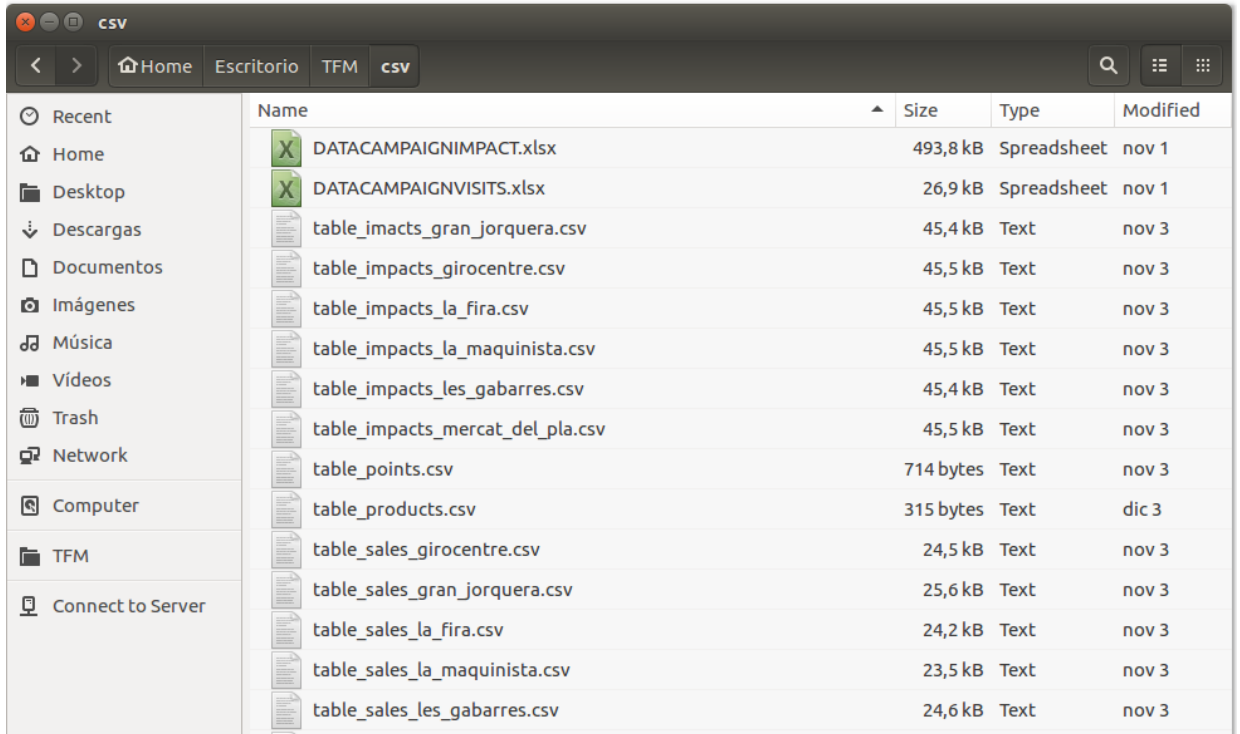
M1.221. TFM – Business Intelligence

Introducción

- Contexto: Crear un entorno BI que posibilite el análisis de la información generada durante una campaña publicitaria.
- Objetivo: A partir de la información generada, determinar la eficacia de la campaña publicitaria.
- Objetivos del TFM:
  - Diseñar un almacenamiento de datos (Data Warehouse) que permita almacenar la información adquirida de los distintos orígenes.
  - Implementar este almacen de datos y programar los procesos ETL (extracción, transformación y carga) que permita alimentar el Data Warehouse a partir de los ficheros base facilitados.
  - Analizar las diferentes plataformas BI Open Source disponibles en el mercado que nos permitan explotar la información almacenada.
  - Elegir e implantar una de estas soluciones Open Source de tal forma que se disponga de una capa de software para el análisis de la información.

Cuestiones a responder

- *En general, ¿la publicidad basada en impactos visuales, es efectiva?*
- *¿Hay determinadas zonas geográficas donde los impactos son más efectivos?*
- *Según la familia de producto, ¿existe alguna relación entre los artículos impactados y el resultado obtenido?*
- *¿Son más efectivos los impactos según el momento temporal en el que se producen?*
- *¿Existe algún artículo o familia que, según la zona, su publicidad tiene mayor impacto? ¿Y según el momento cuando se realiza?*



Conocimientos Previos

Tiempo: Se quiere agrupar toda la información con una granularidad mínima de día. Además se desea realizar agrupaciones semanales, mensuales, trimestrales y anuales.

Centros Comerciales: La campaña se llevará a cabo en distintos centros comerciales. Estos centros estarán clasificados como urbanos o extrarradio según se encuentren en el área metropolitana de Barcelona u otras comarcas. Se dispondrá de información sobre la ciudad y la zona donde se encuentra el centro comercial. También se dispondrá de información sobre el punto final donde se hacen los impactos. Aquí tendremos opciones de grandes pantallas del centro comercial, pantallas en las tiendas o monitores situados en los puntos de información.

Artículos: La campaña se realizará para distintas familias de productos. No todos los productos de una misma familia tendrán el mismo número de impactos visuales.

Indicadores: Para un determinado artículo, un punto publicitario tendremos los impactos que se han producido a lo largo del tiempo.
Además también dispondremos de las ventas realizadas de aquel artículo durante aquel día.
Paralelamente también tendremos el número de visitantes del centro comercial.

Metodología y Planificación

- Metodología Scrum: Metodología Ágil centrado en producir entregables que funcionen y en ofrecer resultados visibles
- Planificación (12 fases principales):
  1. Planificación del Proyecto
  2. Análisis de las herramientas del mercado
  3. Propuesta de la herramienta BI
  4. Creación del entorno de trabajo
  5. Diseño del DWH
  6. Implementación del DWH
  7. Diseño de los procesos ETL
  8. Implementación de los procesos ETL
  9. Creación del Entorno BI para el análisis
  10. Análisis de los datos
  11. Conclusiones
  12. Entrega

Análisis de las herramientas del Mercado

- BIRT
- SpagoBI
- Pentaho Community
- Jaspersoft Community
- Qlikview

Herramienta Seleccionada: Pentaho

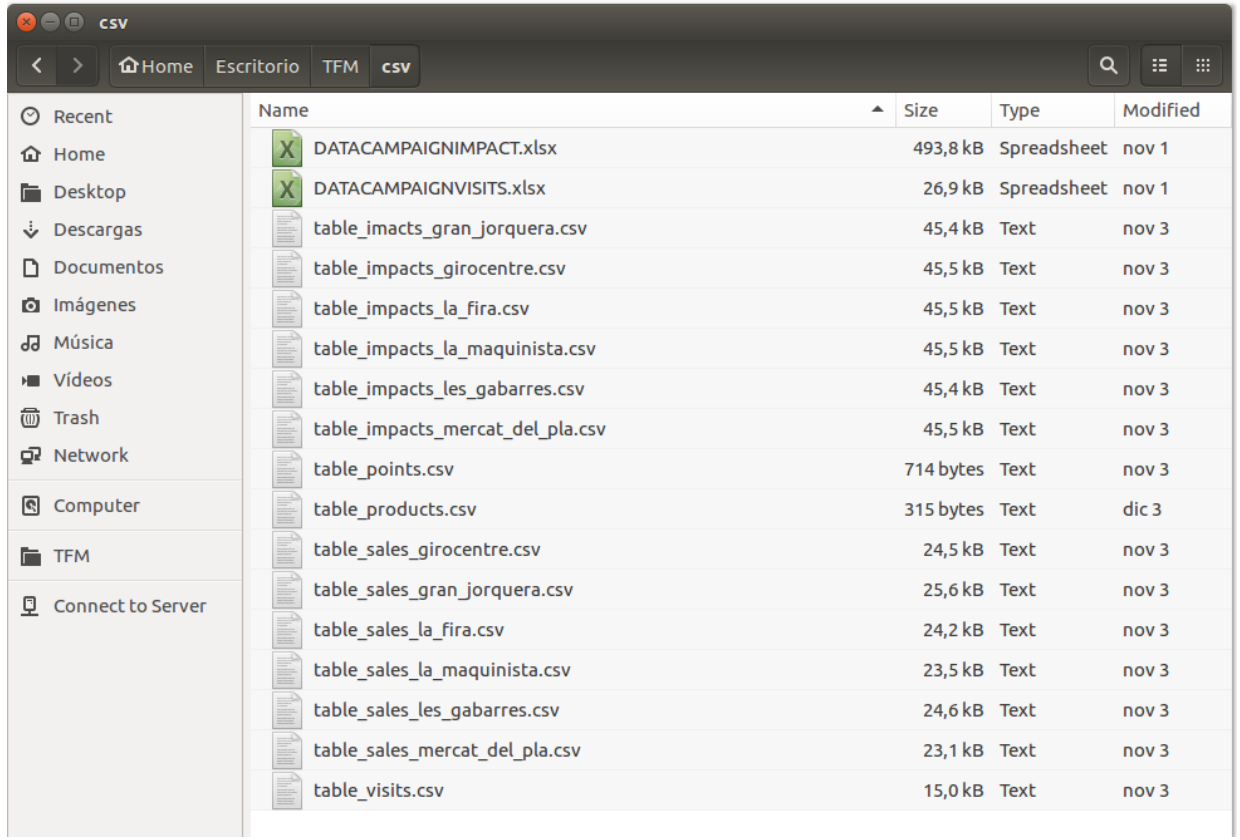
Entorno de Trabajo

Virtual Machine con Ubuntu Linux 16,04,3 LTS (Xenial Xerus) y las siguientes características:

- RAM: 6GB
- vCPUs: 1
- Almacenamiento: 20 GB
- Interfaces de Red: 1 con configuración NAT
- Sistema Operativo: Ubuntu Linux 16.04
- Base de datos: MySQL
- DWH: Schema de MySQL
- Herramienta BI: Pentaho versión 6.1 con los siguientes módulos:
  - Pentaho Business Analytics Server o BI server
  - Pentaho Schema Workbench
  - Pentaho Report Designer
  - Pentaho Data Integration (antiguo kettle)
  - Plugin Saiku analyzer

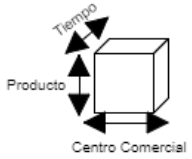
Fuentes de Datos

- Base de datos Mysql y Archivos CSV

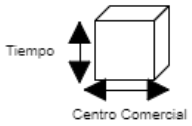


Datawarehouse - Diseño

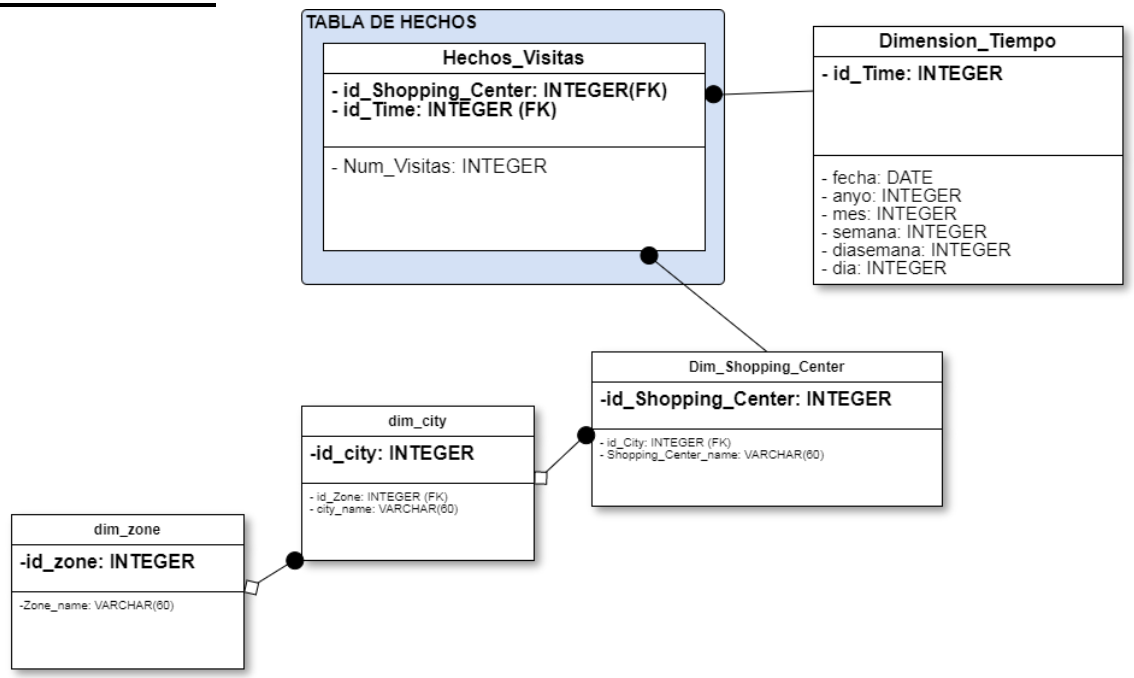
Cubo Ventas - Impactos



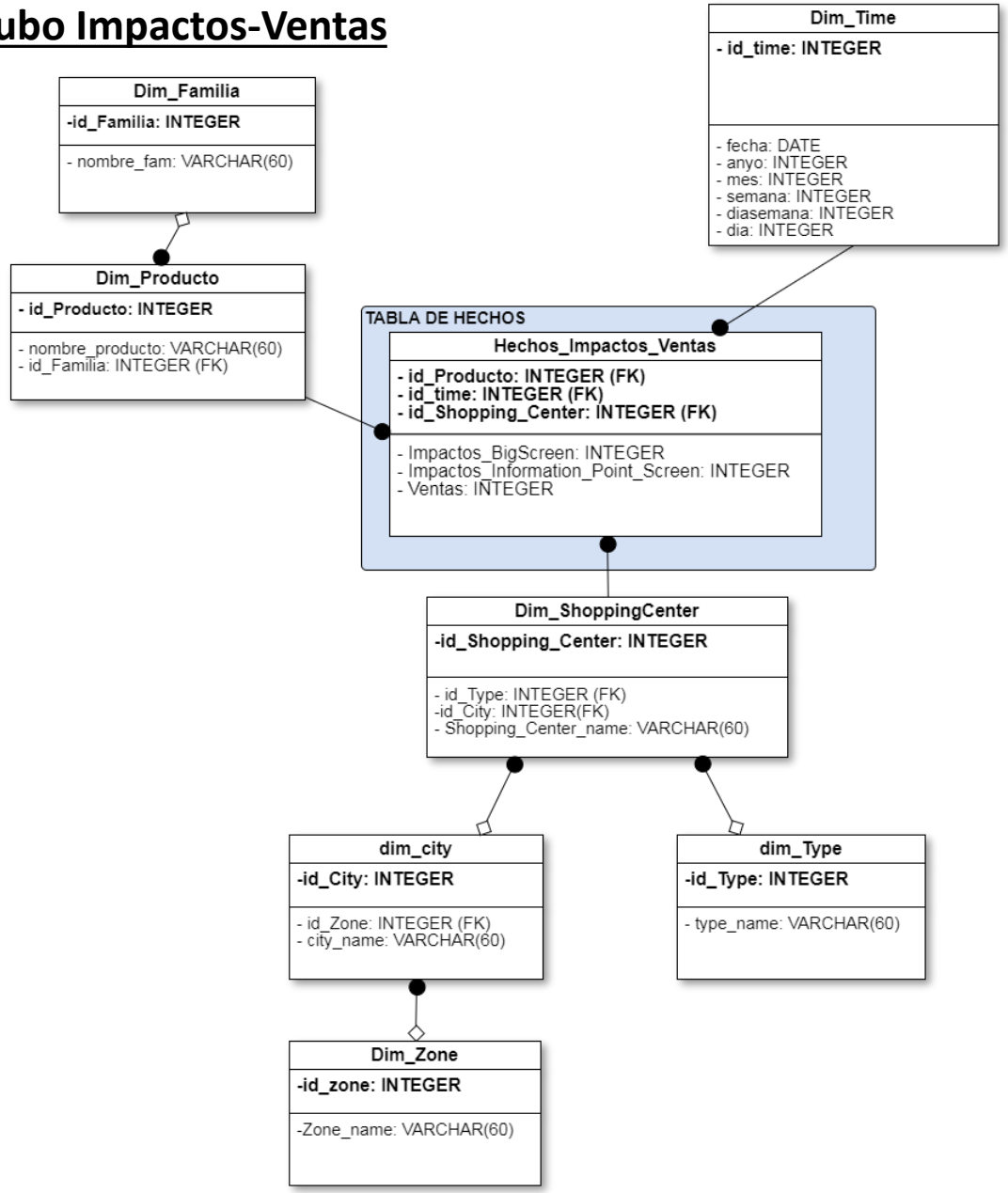
Cubo Visitas



Cubo Visitas

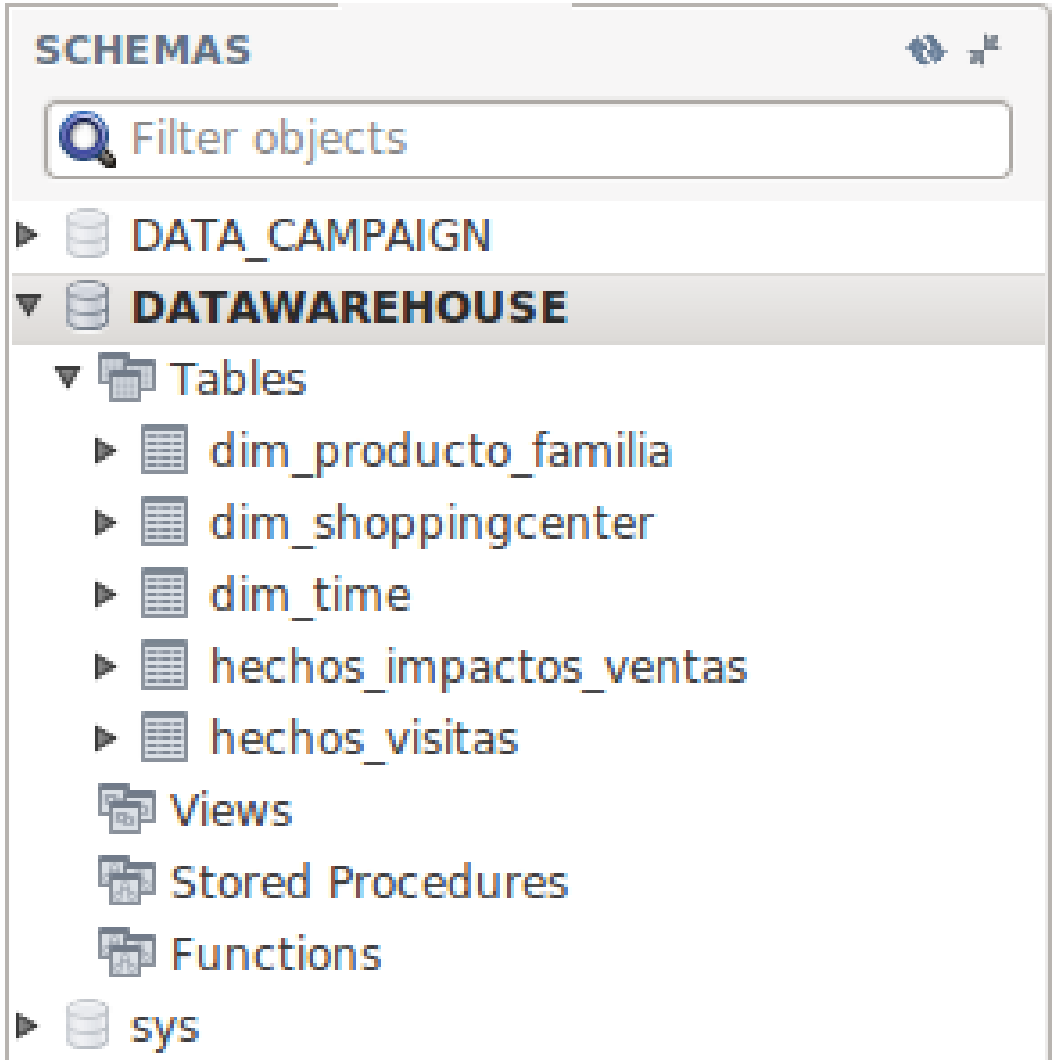


Cubo Impactos-Ventas

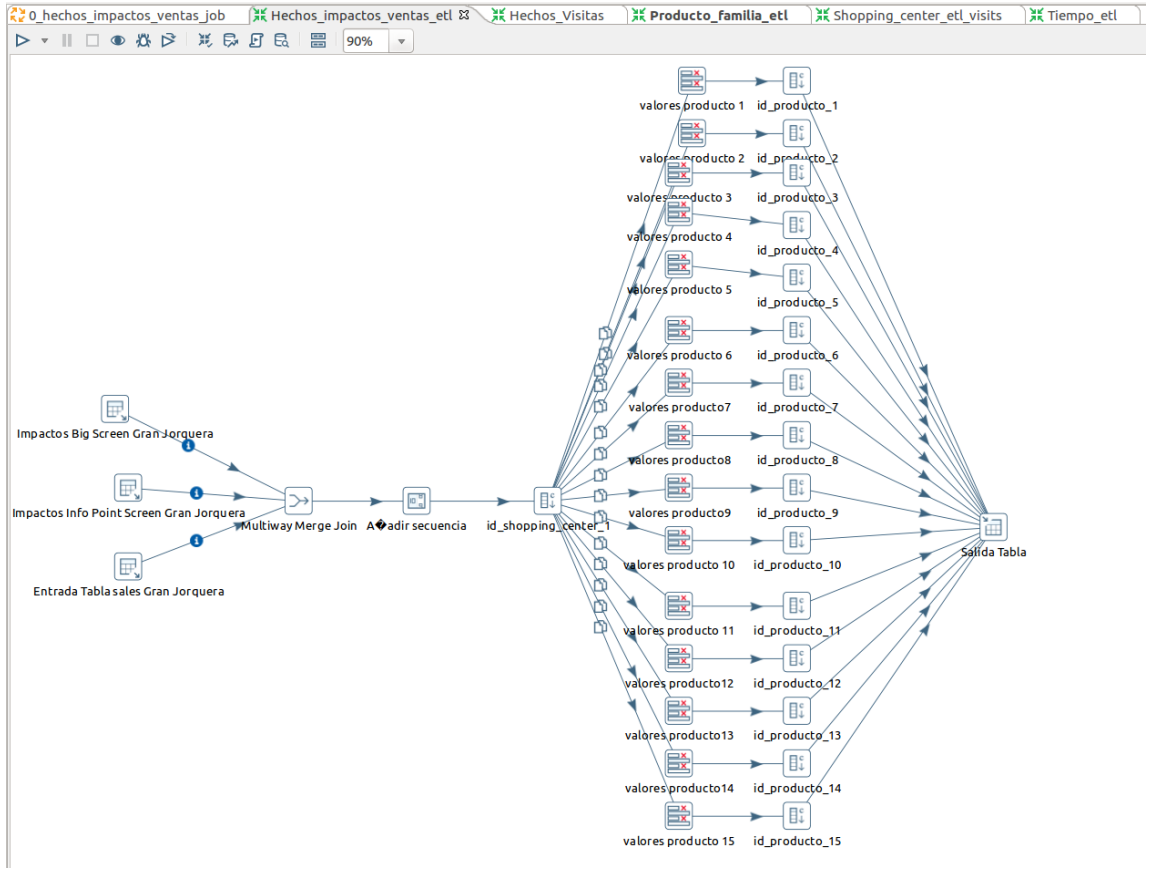


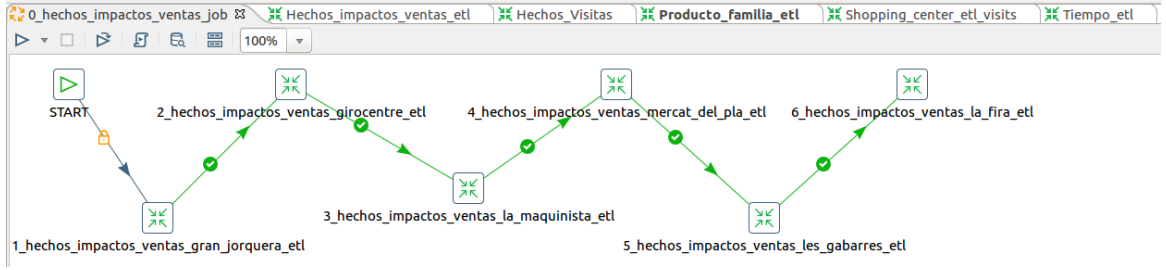
Datawarehouse - Implementación

- BBDD MySQL



ETL – Extract-Transform-Load





Cubos

Schema - Ventas-Impactos y Visitas (cubo.xml)

- xSchema
  - xVentas - Impactos
    - Table: hechos_impactos_ventas
    - Centro Comercial
    - Producto
    - Tiempo
    - Tipo
    - Total Ventas
    - Total Impactos Big Screen
    - Total Information Point Screen
  - xVisitas
    - Table: hechos_visitas
    - Tiempo
    - Centro Comercial
    - Total Visitas
  - xdim_shopping_center
    - xdim_shopping_center
      - Zona
      - Ciudad
      - Centro Comercial
      - Table: dim_shoppingcenter
  - xdim_producto
    - xdim_producto
      - Familia
      - Producto
      - Table: dim_producto_familia
  - xdim_time
    - xdim_tiempo
      - Año
      - Mes
      - Día
      - Table: dim_time
  - xdim_shopping_center_type
    - xdim_shopping_center_type
      - Tipo
      - Centro Comercial
      - Table: dim_shoppingcenter

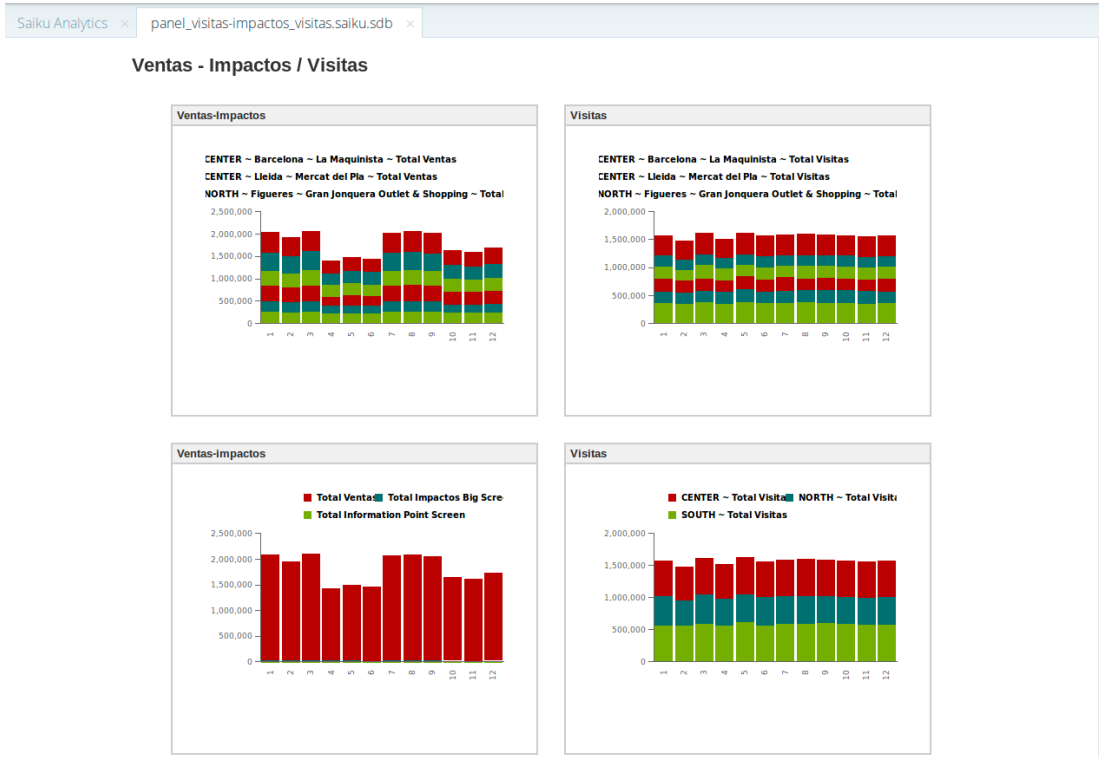
Cube

Attribute	Value
name	Ventas - Impactos
description	
caption	
cache	☑
enabled	☑
visible	☑

Entorno BI

- La herramienta: Pentaho + Saiku
- Queries y Paneles de Mando

		CENTER			NORTH			SOUTH		
Familia	Producto	Total Ventas	Total Impactos Big Screen	Total Information Point Screen	Total Ventas	Total Impactos Big Screen	Total Information Point Screen	Total Ventas	Total Impactos Big Screen	Total Information Point Screen
Food	PIZZA TDRL	33.987	3.007	3.041	72.269	3.005	3.026	107.845	3.017	3.004
	PIZZA DROE	34.165	3.600	3.664	82.105	3.694	3.648	131.276	3.599	3.691
	PIZZA CUSN	34.229	4.605	4.587	97.916	4.554	4.569	158.939	4.558	4.590
Electronics	TV SMRT42SNG	2.121.120	3.002	2.970	1.521.169	3.046	3.035	906.989	2.988	2.948
	TV SMRT42PHI	2.533.950	3.682	3.636	1.682.160	3.698	3.702	915.932	3.631	3.583
	TV SMRT42SNY	3.108.840	4.631	4.589	1.895.015	4.565	4.558	921.284	4.577	4.588
Clothing	TROUSERS LVS	152.406	3.037	3.016	136.790	3.044	3.040	55.062	3.032	2.982
	TROUSERS MNG	174.151	3.672	3.738	157.385	3.644	3.643	55.402	3.664	3.675
	TROUSERS LEE	201.577	4.544	4.560	184.280	4.583	4.545	55.986	4.588	4.582
Sports	SHOES NKE	66.114	3.053	2.992	308.962	3.017	3.022	516.330	3.046	3.023
	SHOES ADS	67.342	3.644	3.675	356.466	3.708	3.689	614.538	3.602	3.677
	SHOES ACS	66.218	4.607	4.529	437.937	4.569	4.592	748.890	4.610	4.556
Home	FRAME 18X10 CK	68.189	3.078	3.005	88.717	3.007	3.033	83.262	3.007	3.038
	FRAME 18X10 CH	78.152	3.673	3.664	91.477	3.669	3.675	84.028	3.719	3.654
	FRAME 18X10 MD	73.156	2.294	2.291	95.117	4.632	4.599	86.411	4.580	4.581



Cuestiones a responder

- *¿Hay determinadas zonas geográficas donde los impactos son más efectivos?*
- *Según la familia de producto, ¿existe alguna relación entre los artículos impactados y el resultado obtenido?*
- *¿Son más efectivos los impactos según el momento temporal en el que se producen?*
- *¿Existe algún artículo o familia que, según la zona, su publicidad tiene mayor impacto? ¿Y según el momento cuando se realiza?*

- *En general, ¿la publicidad basada en impactos visuales, es efectiva?*

Resultados del Análisis

- **Los números reflejan que a mayor número de impactos, mayor número de Ventas.**
- **Con un número similar de impactos, la Zona Centro muestra una mayor efectividad obteniendo un mayor número de ventas con respecto a la zona Norte o Sur.**
- **Los artículos de la familia de productos “Electronics” son los que mayor eficacia presentan en líneas generales.**
- **El primer trimestre del año y los meses de verano presentan una mayor eficacia, siendo los meses de más impactos y más ventas.**
- **Existe una eficacia en la campaña dependiendo de la zona y el producto.**
- **En líneas generales la eficacia de la campaña es la misma en todas las zonas.**
- **Los números reflejan una afluencia de visitantes constante durante todo el año. Esto no influye en el número de ventas o impactos.**

GRACIAS