

Entorno de inteligencia de negocios para el análisis publicitario basado en anuncios de impacto visual de pantallas ubicadas en centros comerciales.

Lina Marcela Quitian Gómez
Máster en Ingeniería Informática
Business Intelligence

María Isabel Guitart Hormigo
David Amorós Alcaraz

08 de enero de 2018



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

Licencias alternativas (elegir alguna de las siguientes y sustituir la de la página anterior)

A) Creative Commons:



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-CompartirIgual [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-sa/3.0/es/)



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc/3.0/es/)



Esta obra está sujeta a una licencia de Reconocimiento-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nd/3.0/es/)



Esta obra está sujeta a una licencia de Reconocimiento-CompartirIgual [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-sa/3.0/es/)



Esta obra está sujeta a una licencia de Reconocimiento [3.0 España de Creative Commons](https://creativecommons.org/licenses/by/3.0/es/)

B) GNU Free Documentation License (GNU FDL)

Copyright © 2018 Lina Marcela Quitian Gómez.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3

or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

C) Copyright

© (Lina Marcela Quitian Gómez)

Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Crear un entorno de inteligencia de negocio para el análisis publicitario basado en anuncios de impacto visual en pantallas ubicadas en varios centros comerciales.</i>
Nombre del autor:	<i>Lina Marcela Quitian Gómez</i>
Nombre del consultor/a:	<i>María Isabel Guitart Hormigo</i>
Nombre del PRA:	<i>Entrega Final</i>
Fecha de entrega (mm/aaaa):	01/2018
Titulación:::	<i>Máster en Ingeniería Informática</i>
Área del Trabajo Final:	<i>Business Intelligence (BI)</i>
Idioma del trabajo:	<i>Español</i>
Palabras clave	<i>ETL, Business Intelligence, Data Warehouse, OLAP</i>

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.*

El presente trabajo final de máster consiste en el desarrollo de una herramienta de análisis publicitario basado en anuncios de impacto visual en pantallas ubicadas en centros comerciales. La idea central gira sobre el diseño e implementación de un sistema de Business Intelligence que facilita la adquisición, el almacenamiento y la explotación de datos. Los datos son obtenidos mediante una campaña publicitaria llevada a cabo en ubicaciones diferentes y se procesan mediante una herramienta Open Source. Adicionalmente, dicho almacén de datos es adaptado al modelo multidimensional que dotará al producto de las capacidades de análisis requeridas. Finalmente, se diseña e implementa el Data WereHouse y los procesos ETL, para crear cubos OLAP necesarios para hacer un análisis valorativo de la herramienta, en términos del impacto visual generado sobre las ventas.

Abstract (in English, 250 words or less):

The present final master's project consists of the development of an advertising analysis tool based on visual impact of screens located in shopping centers. The central idea revolves around the design and implementation of a Business Intelligence system that facilitates the acquisition, storage and exploitation of data. The data is obtained by an advertising campaign was carried out in different places and processed through an Open Source tool. Additionally, said data warehouse is adapted to the multidimensional model that will provide the product with the required analysis capabilities. Finally, the Data WereHouse and the ETL processes are designed and implemented, to create OLAP cubes

necessary to make an evaluative analysis of the tool, in terms of the visual impact generated on sales.

Índice

1.	Introducción	1
1.1	Contexto y justificación del Trabajo	1
1.2	Objetivos del Trabajo.....	1
1.3	Enfoque y método seguido	2
1.4	Planificación del Trabajo.....	2
1.5	Breve resumen de productos obtenidos	5
1.6	Breve descripción de los otros capítulos de la memoria.....	5
2.	Análisis y propuesta de la herramienta BI a utilizar	7
2.1	¿Qué es Business Intelligence?	7
2.2	Los principales componentes de Business Intelligence.....	7
2.3	Beneficios que aporta Business Intelligence	9
2.4	Herramientas BI a utilizar	10
3.	Diseño e implementación del DWH del proyecto	13
3.1	Estudio de los datos proporcionados.....	13
3.2	Modelo conceptual de datos	14
3.3	Diseño del modelo lógico de los datos	15
3.4	Modelo físico de los datos	16
4.	Diseño e implementación de los procesos ETL	18
4.1	Transformación y carga de producto.	20
4.2	Transformación y carga de tiempo.	21
4.3	Transformación y carga de centro comercial.	23
4.4	Transformación y carga de visita.	25
4.5	Transformación y carga de venta.	27
4.6	Transformación y carga de impacto.....	29
5.	Creación del entorno BI para el análisis	32
1.	¿En general, la publicidad basada en impactos visuales, es positiva?	32
2.	¿Hay determinadas zonas geográficas donde los impactos son más efectivos?	33
3.	¿Según la familia de producto, existe alguna relación entre los artículos impactados y el resultado obtenido?	34
4.	¿Son más efectivos los impactos, según el momento temporal en que se producen?.....	35
5.	¿Existe algún artículo o familia que según la zona, su publicidad resulta más efectiva? ¿Y según el momento en que se realiza?	36
6.	Análisis de los datos y Conclusiones	37
4.	Glosario	38
5.	Bibliografía	39
6.	Anexos	40

Lista de figuras

Figura 1: Hitos del proyecto y su duración	4
Figura 2: Planificación Inicial representado en Diagrama de Gantt.....	4
Figura 3: Modelo conceptual de datos	15
Figura 4: modelo lógico de datos	16
Figura 5: modelo físico de datos	17
Figura 6: Implementación del proceso ETL.....	18
Figura 7: Herramienta Spoon para el proceso ETL.....	18
Figura 8: creación de la transformación	19
Figura 9: conexión con MySql.....	20
Figura 10: Transformación y carga de d_producto.....	20
Figura 11: Obtiene valores de secuencia para d_producto.....	21
Figura 12: Productos cargados a d_producto en MySql.....	21
Figura 13: Transformación y carga de d_tiempo.....	22
Figura 14: transformación del campo trimestre.....	22
Figura 15: preparación de los datos para d_tiempo.....	22
Figura 16: Productos cargados a d_tiempo en MySql.....	23
Figura 17: Transformación y carga de d_ccomercial.....	23
Figura 18: Merge Rows (Diff)	24
Figura 19: Obtiene valores de secuencia para d_ccomercial.....	24
Figura 20: Productos cargados a d_ccomercial en MySql.....	25
Figura 21: Transformación de la tabla de hecho visita.....	25
Figura 22: Selección y renombrado de los valores.....	26
Figura 23: Adición de valores constante.....	26
Figura 24: Verificación de los datos cargados en MySql.....	27
Figura 25: transformación hecho_venta.....	28
Figura 26: Verificación de la carga de h_venta en SQL.....	29
Figura 27: transformación de h_impacto	30
Figura 28: Verificación de la carga de h_impacto en SQL.....	31
Figura 29: promedio de la publicidad visual.....	32
Figura 30: estadística básica de la publicidad en general	33
Figura 31: Estadísticas básicas de las zonas impactadas.....	33
Figura 32: impacto de la publicidad en los puntos de información.....	33
Figura 33: impacto de la publicidad en los puntos de información.....	34
Figura 34: relación de los productos impactados.....	34
Figura 35: estadísticas de los productos impactados.....	35
Figura 36: estadísticas básicas de impactos en los trimestres.....	35
Figura 37: impacto de la publicidad en los trimestres.....	35
Figura 38: efectividad de productos por zonas.....	36
Figura 39: efectividad de productos en tiempo.....	36
Figura 40: efectividad de productos publicitados.....	36

1. Introducción

1.1 Contexto y justificación del Trabajo

En la actualidad las empresas buscan poseer su marca o producto en el mercado, por esta razón recurren a diversos métodos para dar a conocer, a la mayor cantidad de personas como sea posible, su producto. Un método que está dando gran resultado y que catapultará las ventas de las empresas es el merchandising visual, que utiliza las pantallas, la tecnología para captar la atención de los consumidores y establecer ciertas expectativas en el cliente para orientar su compra. El merchandising visual ha demostrado ser un incrementador en clientes y en el efecto sobre las ventas. Sus radicales funciones distintivas ayudan al diseño de tiendas de todo tipo de comercio a nivel global. Los resultados demuestran, que el diseño de un establecimiento tanto interno como externo también surge de la parte seductora hacia los clientes, sus elementos cobran acción en las decisiones de los compradores puesto que de manera inconsciente se interesan por la facilidad de comprar.

Las nuevas tecnologías han hecho evolucionar el concepto de merchandising visual incorporando nuevas formas de captar la atención desde la reproducción de anuncios en todo tipos de pantallas interconectadas, publicidad dentro de las apps o en las redes sociales más comunes.

Para determinar el impacto que tiene el merchandising visual con respecto a una forma tradicional de anunciar un producto se creará un entorno Business Intelligence (BI) que posibilite el análisis de la información generada durante una campaña publicitaria basada en la reproducción de pequeños anuncios en forma de impactos visuales en pantallas o monitores situados dentro de varios centros comerciales.

Con este análisis se podrá valorar si la reproducción de estos pequeños anuncios de forma controlada hace que el artículo anunciado incremente el número de ventas respecto de los otros artículos publicitados de forma tradicional. Además, la información suministrada podrá ser analizada en forma de cubos de datos el que permitirá un análisis mucho más cuidadoso y la extracción de conclusiones mucho más elaboradas.

1.2 Objetivos del Trabajo

Diseñar e implementar un sistema de Business Intelligence que facilite la adquisición, el almacenamiento y la explotación de datos obtenidos de la campaña publicitaria llevada a cabo en ubicaciones diferentes.

Los objetivos específicos del trabajo son:

- 1.2.1 Diseñar un almacén de datos (Data Warehouse) que permita almacenar la información adquirida en los diferentes orígenes.

- 1.2.2 Implementar este almacén de datos y programar los procesos ETL (extracción, transformación y carga) que permitan alimentar el DWH a partir de los ficheros base facilitados.
- 1.2.3 Analizar las diferentes plataformas BI OS disponibles en el mercado que permitan explorar la información almacenada.
- 1.2.4 Seleccionar una herramienta de software BI OS para disponer de una capa de software dedicada al análisis de información.

1.3 Enfoque y método seguido

La metodología que se usará para llevar a cabo este proyecto es la metodología SCRUM, ya que esta se adapta mejor al tipo de proyecto que se pretende realizar, ya que está conformado por un equipo pequeño y el tiempo que se tiene para terminar el TFM es corto.

Debido que SCRUM ofrece trabajar por partes, se divide el triángulo o trabajo verticalmente para ir, desde abajo hasta arriba, dibujando una espiral vertical sobre el triángulo. Lo cual permitirá obtener resultados de manera continua, de acuerdo a las entregas de las PEC. Una vez desarrollada la tarea, se pasa a corregir y se sigue con la siguiente tarea.

Con SCRUM se trabaja con ciclos cortos para cada tarea, donde cada ciclo tiene una duración de 2 a 4 semanas, cada ciclo lleva acabo los siguientes pasos:

- Definir
- Planificar
- Construir
- Presentar
- Revisar

Cada requerimiento del Sprint o ciclo se puede permutar y priorizar[1]. Los ciclos corresponderán a cada entrega de las PEC. En los Sprint además de las tablas de dimensiones (ETL) se pueden ir cargando de comunas en otros Sprint futuros.

1.4 Planificación del Trabajo

En la siguiente tabla se muestran los hitos del proyecto, estos hitos corresponden a los ciclos de la metodología SCRUM a utilizar.

Hitos	Tiempo (días)	Inicio	Final
Pec1	12	20 septiembre 2017	02 octubre 2017
Análisis y propuesta de la herramienta BI a utilizar			
Pec2	34	03 octubre 2017	06 noviembre 2017
Creación del entorno de trabajo			
Diseño e implementación del DWH del proyecto			
Pec3	28	07 noviembre 2017	04 diciembre 2017
Diseño e implementación de los procesos ETL			
Pec4	34	05 diciembre 2017	08 enero 2018
Creación del entorno BI para el análisis			
Análisis de los datos y conclusiones.			
TEF	3	15 enero 2018	18 enero 2018

Tabla 1: Hitos a cumplir

Para detallar mejor los hitos a cumplir durante el proyecto se utilizará la herramienta Open Project para la elaboración del diagrama de Gantt. Con este diagrama se pretende mostrar el tiempo previsto para cada tarea y así lograr los objetivos previstos.

Open Project es un software gratuito que ofrece todas las opciones de una herramienta de proyecto, como asignar recursos a tareas, crear dependencias de tareas, ver los recursos disponibles, filtrar las tareas por estado y ordenarlas según prioridad, coste y otros factores.

A continuación se muestra el diagrama de Gantt que representa las tareas y actividades que se pretenden llevar a cabo en el proyecto y el tiempo de cada una de ellas.

	Nombre	Duración	Inicio	Terminado
1	TFM	87 days?	20/09/17 08:00 AM	18/01/18 05:00 PM
2	Inicio	2 days?	20/09/17 08:00 AM	21/09/17 05:00 PM
3	PEC1_Análisis y propuesta de la herramienta BI a utilizar	7 days	22/09/17 08:00 AM	2/10/17 05:00 PM
4	Justificar Trabajo	1 day	22/09/17 08:00 AM	22/09/17 05:00 PM
5	Redactar objetivos	1 day	25/09/17 08:00 AM	25/09/17 05:00 PM
6	Redactar enfoque y método seguido	2 days	27/09/17 08:00 AM	28/09/17 05:00 PM
7	Redactar planificación del Trabajo	1 day	28/09/17 08:00 AM	28/09/17 05:00 PM
8	Redactar los productos a obtener	1 day	30/09/17 08:00 AM	2/10/17 05:00 PM
9	Redactar la descripción de los otros capítulos de la memoria	1 day	2/10/17 08:00 AM	2/10/17 05:00 PM
10	PEC2	25 days?	3/10/17 08:00 AM	6/11/17 05:00 PM
11	Creación del entorno de trabajo	13 days?	3/10/17 08:00 AM	19/10/17 05:00 PM
12	Redactar teoría de BI	4 days?	3/10/17 08:00 AM	6/10/17 05:00 PM
13	Redactar componentes de BI	3 days?	7/10/17 08:00 AM	11/10/17 05:00 PM
14	Redactar el impacto de usar BI	3 days?	12/10/17 08:00 AM	16/10/17 05:00 PM
15	Redactar beneficios que aporta BI	3 days?	17/10/17 08:00 AM	19/10/17 05:00 PM
16	Diseño e implementación del DWH del proyecto	12 days?	20/10/17 08:00 AM	6/11/17 05:00 PM
17	Teoría del DWH	3 days?	20/10/17 08:00 AM	24/10/17 05:00 PM
18	Diseño del DWH	5 days?	25/10/17 08:00 AM	31/10/17 05:00 PM
19	Implementación del DWH	4 days?	1/11/17 08:00 AM	6/11/17 05:00 PM
20	PEC3	20 days?	7/11/17 08:00 AM	4/12/17 05:00 PM
21	Diseño e implementación de los procesos ETL	20 days?	7/11/17 08:00 AM	4/12/17 05:00 PM
22	Teoría de los procesos ETL	5 days?	7/11/17 08:00 AM	13/11/17 05:00 PM
23	Diseño de los procesos ETL	11 days?	14/11/17 08:00 AM	28/11/17 05:00 PM
24	Implementación de los procesos ETL	4 days?	29/11/17 08:00 AM	4/12/17 05:00 PM
25	PEC4	25 days?	5/12/17 08:00 AM	8/01/18 05:00 PM
26	Creación del entorno BI para el análisis	12 days?	5/12/17 08:00 AM	20/12/17 05:00 PM
27	Escoger la herramienta para el análisis	5 days?	5/12/17 08:00 AM	11/12/17 05:00 PM
28	Construcción de la BD para el análisis	7 days?	12/12/17 08:00 AM	20/12/17 05:00 PM
29	Análisis de los datos y conclusiones	13 days?	21/12/17 08:00 AM	8/01/18 05:00 PM
30	Análisis de la información	5 days?	21/12/17 08:00 AM	27/12/17 05:00 PM
31	construcción de Informes	8 days?	28/12/17 08:00 AM	8/01/18 05:00 PM
32	TEF	4 days?	15/01/18 08:00 AM	18/01/18 05:00 PM
33	Responder preguntas	4 days?	15/01/18 08:00 AM	18/01/18 05:00 PM

Figura 1: Hitos del proyecto y su duración

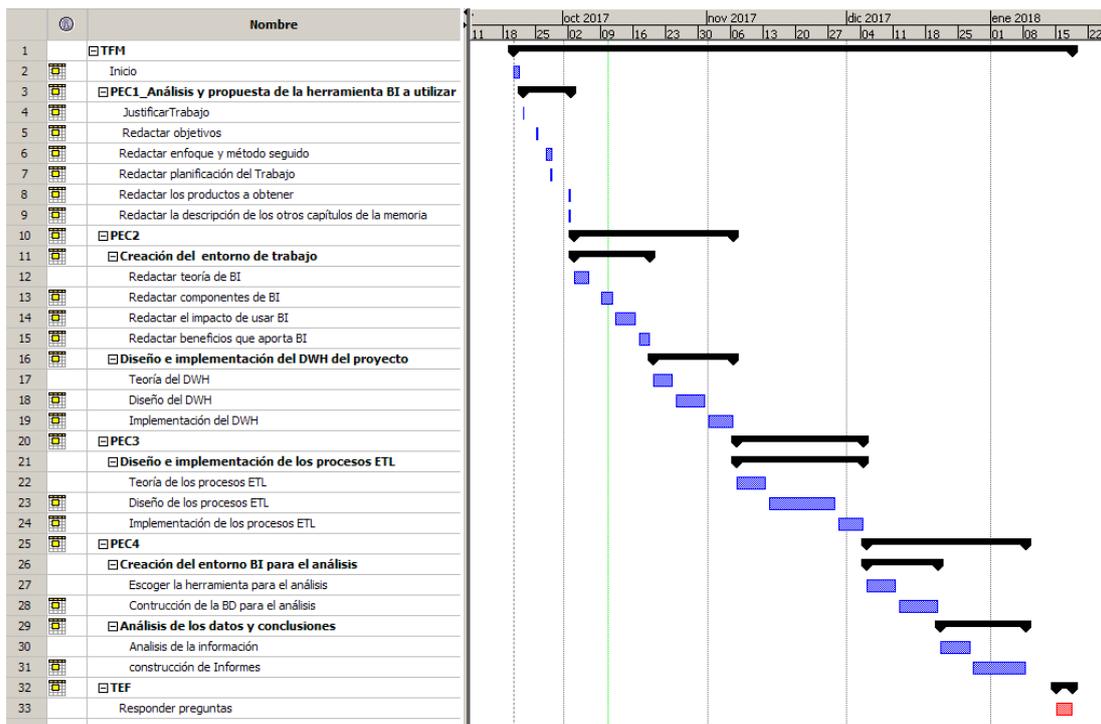


Figura 2: Planificación Inicial representado en Diagrama de Gantt.

1.5 Breve resumen de productos obtenidos

Cuando se verifique el impacto de implantar las cuatro fases del proyecto sobre el sistema Business Intelligence, luego se ayudará a la obtención de datos representativos de una campaña publicitaria para luego ser enseñada de manera fácil.

Las herramientas de Business Intelligence son las más apropiadas para el aporte de conocimientos las empresas e instituciones; estas además contienen KPIs, que permiten mostrar la información de manera segura y fiable en tiempo real.

Con el diseño e implantación del DWH se obtendrá la base de datos y los scripts, para la base de datos, donde se cargarán los datos generados.

En el diseño e implementación de los procesos ETL se crean un conjunto de transformaciones que cargan la información de ventas del producto en un modelo intermedio, luego se convierten para cargarlos en el DWH creado.

En la creación del entorno BI para el análisis se obtiene un proceso para organizar los datos sobre la publicidad del producto, así se recuperan y se manipulan los datos, para luego hacer consultas e informes.

Por último, se generan unos informes donde se resume toda la información recaudada en el análisis publicitario solicitado, para que con estos informes se puedan tomar las decisiones pertinentes.

1.6 Breve descripción de los otros capítulos de la memoria

Capítulo 1: Como primera parte se explica el trabajo a realizar, se justifica la selección de la materia sobre la que se realizará el Trabajo. Asimismo, se hace referencia a los objetivos generales y a la metodología empleada para materializarlos. Además, se muestra el diagrama de Gantt, en el que se precisa la planificación temporal del proyecto. Y por último se expone los productos obtenidos y un breve resumen dónde se explica cada capítulo.

Capítulo 2: Luego en la creación del entorno de trabajo -esta sección es teórica- se explica el concepto de Business Intelligence, en dónde es mejor aplicarlo. La teoría de Business Intelligence se ajusta al tipo de proyecto al cual se pretende realizar. También se expone los principales componentes, el impacto y los beneficios que aporta Business Intelligence aun estudio.

Capítulo 3: Después, en el apartado de diseño e implementación del DWH del proyecto, se hace un análisis detallado de los ficheros de datos proporcionados para detectar las relaciones existentes entre ellos, que determinarán el diseño final. Para así, generar el modelo conceptual del DWH que sirve de entrada para obtener el modelo lógico y el modelo físico final.

Capítulo 4: En el apartado diseño e implementación de los procesos ETL se describen las transformaciones necesarias para cargar los ficheros proporcionados en un staging area, que permite eliminar valores nulos o huérfanos y que facilitan su manipulación posterior.

Capítulo 5: En la creación del entorno BI para el análisis se detalla la tecnología escogida para el análisis multidimensional a través de tablas matriciales o pivotantes. Como esta tecnología organiza los datos en el sistema para recuperar y manipular los datos a través de consultas e informes.

Capítulo 6: Por último, en el análisis de los datos y conclusiones se muestran algunos ejemplos de las pruebas llevadas a cabo, estas pruebas son

mostradas en forma de informes que incorporan métricas para poder evaluar la publicidad de impactos visuales en los centros comerciales de un determinado producto.

2. Análisis y propuesta de la herramienta BI a utilizar

2.1 ¿Qué es Business Intelligence?

Es un proceso en la cual una empresa usa una metodología, aplicaciones y tecnología en unos datos existentes para tomar estos datos, estudiarlos y convertirlos en información; para luego tomar unas decisiones o responder a unos problemas encontrados en el análisis. La Business Intelligence nos facilita el proceso de la toma de decisiones.

Para explicar mejor el concepto de Business Intelligence, es una comprensión total del estado en que se encuentra la empresa, los futuros eventos que se puedan presentar, para exponer una solución a la empresa con el soporte adquirido en el proceso de análisis, para luego mejorar su rendimiento desde un punto de vista estratégico y operativo.

Los datos que utiliza Business Intelligence son extraídos de distintas fuentes mediante herramientas como ETL (extraer, transformar y cargar), que después son cargados en un almacén de datos DWH.

“Cuando una empresa desea mejorar su productividad es obligatorio que considere factores como la vigilancia de sus procesos de producción, la optimización de costes, la fidelización de clientes y la optimización de sus tiempos de respuesta ante los cambios a los que nos somete un mercado tan competitivo como el actual. En este proceso de vigía, cambio y reacción es cuando la Inteligencia de Negocio juega un factor fundamental ya que las herramientas que la componen nos ayudan a verificar la situación actual y el futuro comportamiento según las condiciones del mercado”[2].

2.2 Los principales componentes de Business Intelligence.

2.2.1 Datawarehouse

Es un lugar físico o lógico, donde la información de una empresa puede ser guardada, ya sea de uno o varios orígenes. Además de estar integrada, estructurada, organizada y puede ser de distintos tipos.

Esta corresponde a toda la información relacionada con la empresa reunida, la cual es generada en forma de informes, para que pueda ser consultada, utilizada y analizada, así darle un tratamiento especial para tomar futuras decisiones y poder dar soporte a las necesidades de la empresa en el menor tiempo posible de manera eficiente y a un costo muy bajo.

2.2.1.1 **Elementos del Data WereHouse:** La idea principal es que la información sea presentada desnormalizada para optimizar las consultas. De esta manera se habla de:

- **Tabla de hecho:** Es la representación en el data warehouse de los procesos de negocio de la organización.

- **Dimensión:** es la representación en el data warehouse de una vista para un cierto proceso de negocio.
- **Métrica:** son los indicadores de negocio de un proceso de negocio. Aquellos conceptos cuantificables que permiten medir nuestro proceso de negocio. [3]

2.2.2 Proceso ETL

Sus siglas traducen Extracción, Transformación y Carga. Hace parte del componente de integración y, de su implementación adecuada dependen la integridad, uniformidad, consistencia y disponibilidad de los datos utilizados en el componente de análisis de una solución de BI. Su función es extraer, limpiar, transformar, resumir, y formatear los datos que se almacenarán en la bodega de datos de la solución de BI. [4]

Este proceso es conformado por 6 tareas:

- ✓ Seleccionar los datos para la extracción: se definen los datos de las fuentes (generalmente provienen de diversas fuentes heterogéneas).
- ✓ Transformar las fuentes: una vez que los datos se hayan extraído de las fuentes de datos pueden ser transformados o esos nuevos datos pueden ser derivados. Algunas de las tareas más comunes de este paso son: filtración de datos, conversión de códigos, cálculos de valores derivados, transformación entre diversos formatos de datos, generación automática de números secuenciales (llaves derivadas), etc.
- ✓ Unir las fuentes: las diversas fuentes pueden unirse para ser cargadas al almacén como una sola fuente.
- ✓ Seleccionar el destino para la carga: el destino o los destinos son seleccionados para cargar los datos posteriormente.
- ✓ Unir los atributos de las fuentes de datos con los atributos del destino: los atributos (campos) que se obtuvieron de las fuentes de datos pueden ser mapeados con los correspondientes destinos.
- ✓ Cargar los datos: el almacén es poblado con los datos transformados. [5]

2.2.3 Motor OLAP

Traduce proceso analítico en línea. Permite a los usuarios organizar los datos de manera óptima para luego ser analizados, de manera que se pueda leer una gran cantidad de datos para extraer información útil de manera rápida y consistente, como el estado de las áreas de la empresa, comportamiento de los compradores o las ventas realizadas.

El acceso a los datos es de solo lectura y las consultas se hacen a una gran velocidad. OLAP se compone de hechos numéricos llamados medidas las cuales se clasifican por dimensiones. El cubo de metadatos se crea a partir de un esquema de estrella o copo de nieve, el esquema de las tablas en una base de datos relacional. Las medidas se obtienen de los registros de una tabla de hechos y las dimensiones se derivan de la dimensión de los cuadros.

La funcionalidad de los sistemas OLAP se caracteriza por ser un análisis multidimensional de datos corporativos, que soportan los análisis del usuario y unas posibilidades de navegación, seleccionando la información a obtener. Normalmente este tipo de selecciones se ve reflejado en la visualización de la estructura multidimensional, en unos campos de selección que nos permitan elegir el nivel de agregación (jerarquía) de la dimensión, y/o la elección de un dato en concreto, la visualización de los atributos del sujeto, frente a una(s) dimensiones en modo tabla, pudiendo con ello realizar, entre otras las siguientes acciones:

- ✓ **Rotar (Swap):** alterar las filas por columnas (permutar dos dimensiones de análisis)
- ✓ **Bajar (Down):** bajar el nivel de visualización en las filas a una jerarquía inferior
- ✓ **Detallar (Drilldown):** informar para una fila en concreto, de datos a un nivel inferior
- ✓ **Expandir (Expand):** id. Anterior sin perder la información a nivel superior para éste y el resto de los valores
- ✓ **Colapsar (Collapse):** operación inversa de la anterior. [6]

2.3 Beneficios que aporta Business Intelligence

- Genera una ventaja competitiva: gracias al análisis hecho en la empresa se obtiene un control financiero, se logra optimizar los costos, se calcula la realización de productos de la empresa, se analizan los clientes y logra saber que producto genera más rentabilidad.
- Se logra tomar una decisión de manera rápida: Gracias a la implementación de BI en una empresa se logra tener una respuesta a una pregunta de gran importancia en muy poco tiempo, ya que gracias a los informes generados por BI se logra obtener estadísticas de ventas, los costos de la empresa, que tan eficiente es el marketing, etc.
- Se obtiene el estado evolutivo de la empresa: BI brinda la posibilidad de saber el estado de la empresa en un determinado periodo de tiempo, como es el caso de las ganancias del negocio en su pasado, presente o futuro.
- Se logra estudiar al cliente: Con BI se analiza los hábitos de consumidores a la hora de comprar y se utiliza esa información a favor; para luego generar estrategias para poder fidelizar estos consumidores como clientes en la empresa.
- Se tiene un mayor control sobre las áreas de la empresa: Cada área de la empresa maneja datos que son utilizados por la BI, que son almacenados en un solo lugar, con esto se tiene la posibilidad de analizar la empresa en poco tiempo.

2.4 Herramientas BI a utilizar

De las distintas plataformas BI *open source* que existen hoy por hoy, se escogió unas herramientas que a pesar de tener versión comercial es de código abierto, que ofrece economía y grandes soluciones. Estas ofrecen la posibilidad de acceder al código fuente, para así estudiarlo, modificarlo y mejorarlo. A continuación se detallan las herramientas escogida y cada una de sus características.

2.4.1 MySQL: Es un gestor relacional de bases de datos de fácil instalación, administración y a un costo muy bajo. Los datos son guardados en diferentes tablas y no coloca los datos en una gran tabla, esto brinda flexibilidad y agilidad. Está desarrollado en C y C++. Está desarrollado por MySQL AB, Sun Microsystems y Oracle Corporation. Sus características más importantes son:

- MySQL es multiplataforma, lo cual permite utilizarlo en varios sistemas operativos.
- Soporta distintos motores de almacenamiento y solo diez tienen sus soportes correspondientes.
- Es libre con la licencia GNU y algunas versiones son comerciales pero aun costo bajo.
- Tiene un alto rendimiento.
- Agrupa múltiples transacciones de varias conexiones y así se incrementa el número de transacciones por segundo.
- Es uno de los sistemas más populares para desarrollar bases de datos.
- permite ir a bases de datos multiusuario por medio de la web y con distintos lenguajes de programación que se adaptan a diferentes necesidades y requerimientos.
- Conectividad segura

Dentro del paquete MySQL Community se escoge workbench como herramienta en la creación y administración de bases de datos.

2.4.2 Workbench: Fue creada por la empresa Sun Microsystems y es usada con el sistema de gestión de bases de datos MySQL. Es una herramienta de entorno visual la cual permite diseñar, crear y mantenimiento para el sistema de bases de datos, además que brinda la posibilidad de hacer consultas SQL. A continuación se mencionan las características más importantes:

- Permite conexiones a bases de datos & Instance Management.
- Soporte para plugins personalizados.
- Soporte Unicode
- Sincronización de esquema
- Brinda libertad a los usuarios

2.4.3 Pentaho: Es una herramienta de BI orientada a dar soluciones y centrada en procesos. Las soluciones que Pentaho pretende ofrecer se componen fundamentalmente de una infraestructura de herramientas de análisis e informes integrados con un motor de workflow de procesos de negocio. La plataforma será capaz de ejecutar las reglas de negocio necesarias, expresadas en forma de procesos, actividades, además es capaz de presentar y entregar la información adecuada en el momento adecuado.

Su modelo de ingresos parece estar orientado a los servicios (soporte, formación, consultoría y soporte a ISVs y distribuciones OEM. [7] Las características de Pentaho son:

- Es una herramienta flexible que cubre una amplia gama de necesidades empresariales
- La versión Community Edition es software libre.
- Se puede ajustar según las necesidades y escoger las funcionalidades que más le convengan a usuario.

Los módulos de la plataforma Pentaho son:

- 2.4.3.1 **Reporting:** un módulo de los informes ofrece la solución adecuada a las necesidades de los usuarios. Pentaho Reporting es una solución basada en el proyecto JFreeReport y permite generar informes ágil y de gran capacidad.
- 2.4.3.2 **Pentaho Reporting:** permite la distribución de los resultados del análisis en múltiples formatos – todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto. Los reportes Pentaho permiten también programación de tareas y ejecución automática de informes con una determinada periodicidad.
- 2.4.3.3 **Análisis:** suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (pivot tables, crosstabs), generadas por Mondrian y JPivot, el usuario puede navegar por los datos, ajustando la visión de los datos, los filtros de visualización, añadiendo o quitando los campos de agregación. Los datos pueden ser representados en una forma de SVG o Flash, los dashboards widgets, o también integrados con los sistemas de minería de datos y los portales web (portlets). Además, con el Microsoft Excel Analysis Services, se puede analizar los datos dinámicos en Microsoft Excel (usando la conexión a OLAP server Mondrian).
- 2.4.3.4 **Dashboards:** todos los componentes del módulo Pentaho Reporting y Pentaho Análisis pueden formar parte de un Dashboard. En Pentaho Dashboards es muy fácil incorporar una gran variedad en tipos de gráficos, tablas y velocímetros (dashboard widgets) e integrarlos con los Portlets JSP, en donde podrá visualizar informes, gráficos y análisis OLAP.
- 2.4.3.5 **Data Mining** – análisis en Pentaho se realiza con una herramienta WeKa. [8]

2.4.4 Spoon: Es una herramienta de ETL de Pentaho Data Integration (PDI). Ayuda a la extracción, transformación y carga de los datos. Debido a que es una interfaz gráfica de usuario (GUI), la cual permite diseñar transformaciones y tareas que se pueden ejecutar con la herramienta kettle.

Kettle posee dos herramientas, la primera **Pan** es un motor de transformación de datos y que puede hacer distintas funciones como lectura, manipulación y escritura de datos hacia y desde distintas fuentes de datos. La segunda es Kitchen, es un programa que ejecuta los Trabajos diseñados por Spoon en XML o en un catálogo de base de datos.

Los Trabajos normalmente se planifican en modo batch (por lotes) para ejecutarlos automáticamente en intervalos regulares.

Las Transformaciones y Trabajos se pueden describir usando un archivo XML o se pueden colocar en un catálogo de base de datos de Kettle.

Luego Pan o Kitchen pueden leer los datos para ejecutar los pasos que se describen en la Transformación o ejecutar el Trabajo[9].

2.4.5 Mondrian Schema Workbench: Es una interfaz de diseñador que le permite crear y probar esquemas de cubo OLAP Mondrian visualmente. El motor Mondrian procesa solicitudes MDX con los esquemas ROLAP (OLAP Relacional). Estos archivos de esquema son modelos de metadatos XML que se crean en una estructura específica utilizada por el motor Mondrian. Estos modelos XML se pueden considerar estructuras tipo cubo que utilizan tablas FACT y DIMENSION existentes que se encuentran en su RDBMS. No requiere que se construya o mantenga un cubo físico real; solo que se crea el modelo de metadatos[10].

2.4.6 Saiku: es un visor OLAP que proporciona al usuario final una herramienta para realizar análisis de forma fácil e intuitiva, por medio de tabla pivotante o en modo gráfico. Es un plugin de Pentaho, además es una interfaz fácil limpia e intuitiva que permite a los usuarios construir sus propias vistas arrastrando y soltando campos. En Saiku se pueden visualizar los datos de origen.

3. Diseño e implementación del DWH del proyecto

Debido a los anuncios que fueron reproducidos en diferentes sitios de Cataluña y en dos diferentes tipos de pantallas, se requiere analizar dicha campaña publicitaria, para generar unos informes que le permitan al cliente tomar decisiones para la el bien su negocio, para esto se debe diseñar un almacén de datos (Data WereHouse) para guardar la información de dicha campaña.

En este capítulo se detallarán las etapas que lleva implementar el data Werehouse adecuadamente. Se describirán lo datos suministrado del análisis publicitario que se pretende realizar; se expondrá el diseño del modelo conceptual de datos, con el cual se identifican vistas necesarias para dar respuesta a las preguntas; el diseño del modelo lógico de los datos, por medio de este se identificarán las métricas necesarias y, por último, el modelo físico de los datos, que materializa el diseño en un gestor de base de datos.

3.1 Estudio de los datos proporcionados

Se ha proporcionado una base de datos en Excel con unas hojas de cálculos que contiene información sobre los puntos donde se instalaron las pantallas que reproducen un contenido para la publicidad de unos productos. Cada hoja de cálculo contiene la siguiente información:

Points: Aquí está toda la información sobre la zona en que se implantó la publicidad (norte, sur o centro), la ciudad, que establecimiento o lugar, el tipo de lugar (periferia o urbano) y el impacto del punto (si fue emitido en pantalla grande o en la pantalla del punto de información).

Products: Se describen los productos utilizados para llevar a cabo la publicidad y la familia a la cual pertenece; ya que si es comida, electrónicos, ropa, deporte y hogar.

Visits: Se detallan las visitas de personas a los diferentes lugares comerciales y la fecha en que hicieron la visita.

Impacts Gran Jonquera Sc: Aquí se muestran el impacto de cada producto en el Gran Jonquera, con la fecha y en cada tipo de pantalla donde fue emitida la publicidad del producto.

Sales Gran Jonquera Sc: Se describen las ventas en el Gran Jonquera de cada producto y fecha de la venta.

Impacts Girocentre: Aquí se muestran el impacto de cada producto en el Girocentre, con la fecha y en cada tipo de pantalla donde fue emitida la publicidad del producto.

Sales Girocentre: Se describen las ventas en el Girocentre de cada producto y fecha de la venta.

Impacts la maquinista: Aquí se muestran el impacto de cada producto en la maquinista, con la fecha y en cada tipo de pantalla donde fue emitida la publicidad del producto.

Sales la maquinista: Se describen las ventas en la maquinista de cada producto y fecha de la venta.

Impacts mercat del pla: Aquí se muestran el impacto de cada producto en el mercat del pla, con la fecha y en cada tipo de pantalla donde fue emitida la publicidad del producto.

Sales mercat del pla: Se describen las ventas en el mercat del pla de cada producto y fecha de la venta.

Impacts les gabarres: Aquí se muestran el impacto de cada producto en les gabarres, con la fecha y en cada tipo de pantalla donde fue emitida la publicidad del producto.

Sales les gabarres: Se describen las ventas en les gabarres de cada producto y fecha de la venta.

Impacts la fira: Aquí se muestran el impacto de cada producto en la fira, con la fecha y en cada tipo de pantalla donde fue emitida la publicidad del producto.

Sales la fira: Se describen las ventas en la fira de cada producto y fecha de la venta.

3.2 Modelo conceptual de datos

El modelo conceptual es el primer paso para crear el almacén de datos ya que nos permitirá identificar los procesos y vistas para darle respuesta al cliente.

Para llevar a cabo este modelo se identifican las tablas de hechos y las dimensiones, para analizar los datos desde distintos puntos de vista.

Para llevar a cabo el siguiente modelo se basa en el esquema de constelación que está compuesto por tres tablas de hechos *hecho_venta*, *hecho_impacto* y *hecho_Visita*.

Luego se identifican las dimensiones, las cuales son reutilizadas por las tablas de hechos. Donde las dimensiones (*d_tiempo*, *d_producto* y *d_ccomercial*) están vinculadas a las tablas de hechos *hecho_impacto* y *hecho_venta*; y las dimensiones (*d_tiempo* y *d_ccomercial*) están vinculadas a la tabla de hecho (*hecho_Visita*)

Las visitas, los impactos y las ventas pueden analizarse desde distintos puntos de vista en función de verificar que la publicidad basada en impactos visuales es favorable para los vendedores de los productos:

Dimensión tiempo (*d_tiempo*): Momento en el que se ha realizado la compra

Dimensión producto (*d_producto*): información del producto que se vende

Dimensión centro comercial (*d_ccomercial*): Desde que centro comercial se ha realizado la venta.

Dimensión tiempo (*d_tiempo*): Momento en el que se ha hecho la visita

Dimensión centro comercial (*d_ccomercial*): a que centro comercial se ha realizado la visita.

A partir de estas tablas de hechos y dimensiones se obtiene el siguiente modelo conceptual:

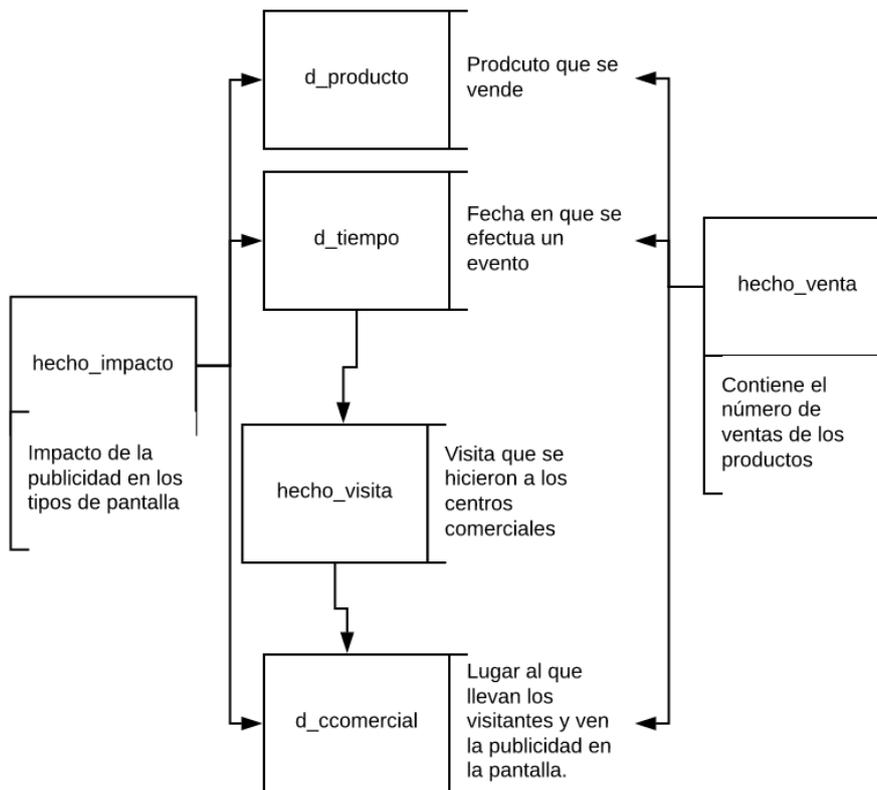


Figura 3: Modelo conceptual de datos

3.3 Diseño del modelo lógico de los datos

Luego de realizar el modelo conceptual se pasa a desarrollar el modelo lógico de datos, donde se establecen las métricas de las tablas de hechos identificadas y en las dimensiones los atributos.

Las tablas de hechos tienen su clave subrogada para identificar cada registro de manera única, además las diferentes claves foráneas que relaciona cada tabla de hecho con las dimensiones, contiene por último las métricas para medir el impacto de las pantallas grandes y las pantallas en los puntos de información.

Así se definen las tablas de hechos identificadas:

Tabla de hecho	Claves foráneas	Métricas
hecho_visita	Id_visita, Id_tiempo, id_ccomercial	num_visita
hecho_impacto	Id_impacto, Id_tiempo, id_ccomercial, id_producto	impacto_gpantalla, impacto_pinfo
Hecho_venta	Id_venta, Id_tiempo, id_ccomercial, id_producto	num_venta

Tabla 2: tablas de hechos.

Los atributos de cada una de las dimensiones son:

Dimensión	Clave primaria	Atributos
d_tiempo	id_tiempo	dia, semana, mes, trimestre, año
d_producto	id_producto	nombre, categoría
d_ccomercial	id_ccomercial	nombre, area, ciudad, zona

Tabla 3: dimensiones.

A partir de las tablas de hechos anteriores y las dimensiones, se obtiene el siguiente esquema:

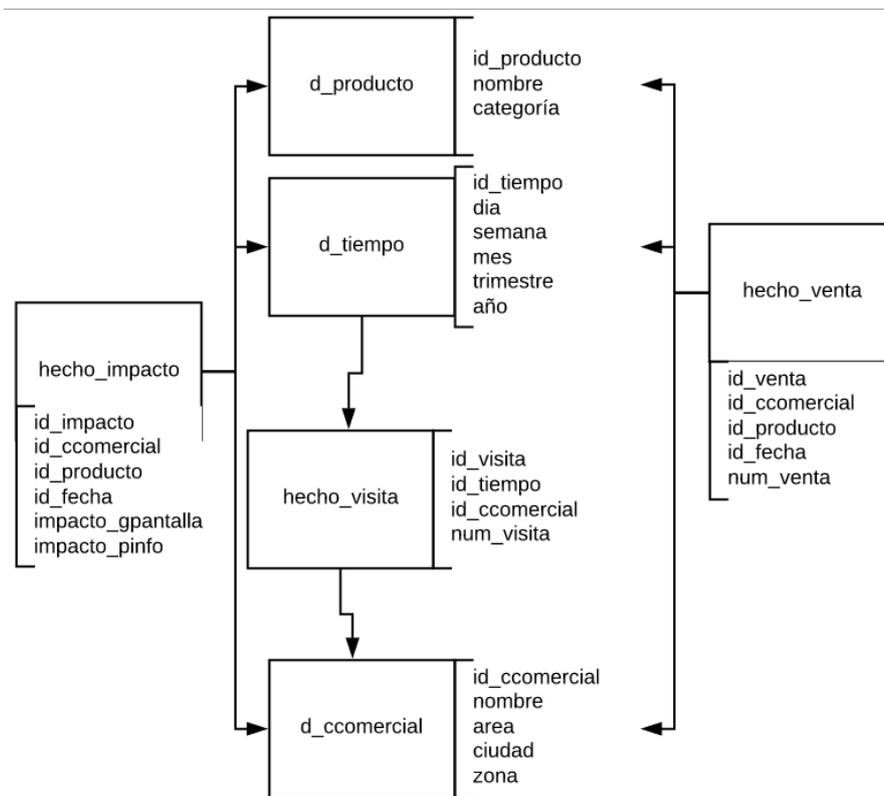


Figura 4: modelo lógico de datos

3.4 Modelo físico de los datos

Una vez terminado la creación del modelo lógico de datos, el siguiente y último paso del data WareHouse es la creación del modelo físico de datos. Para llevar a cabo este paso se trabajará con MySql y su herramienta de modelización visual como lo es Workbench.

La creación del modelo físico estará basado en el modelo lógico obtenido anteriormente, con ello se define cada formato de las claves y atributos de las tablas.

A continuación se muestra el diseño del modelo físico obtenido:

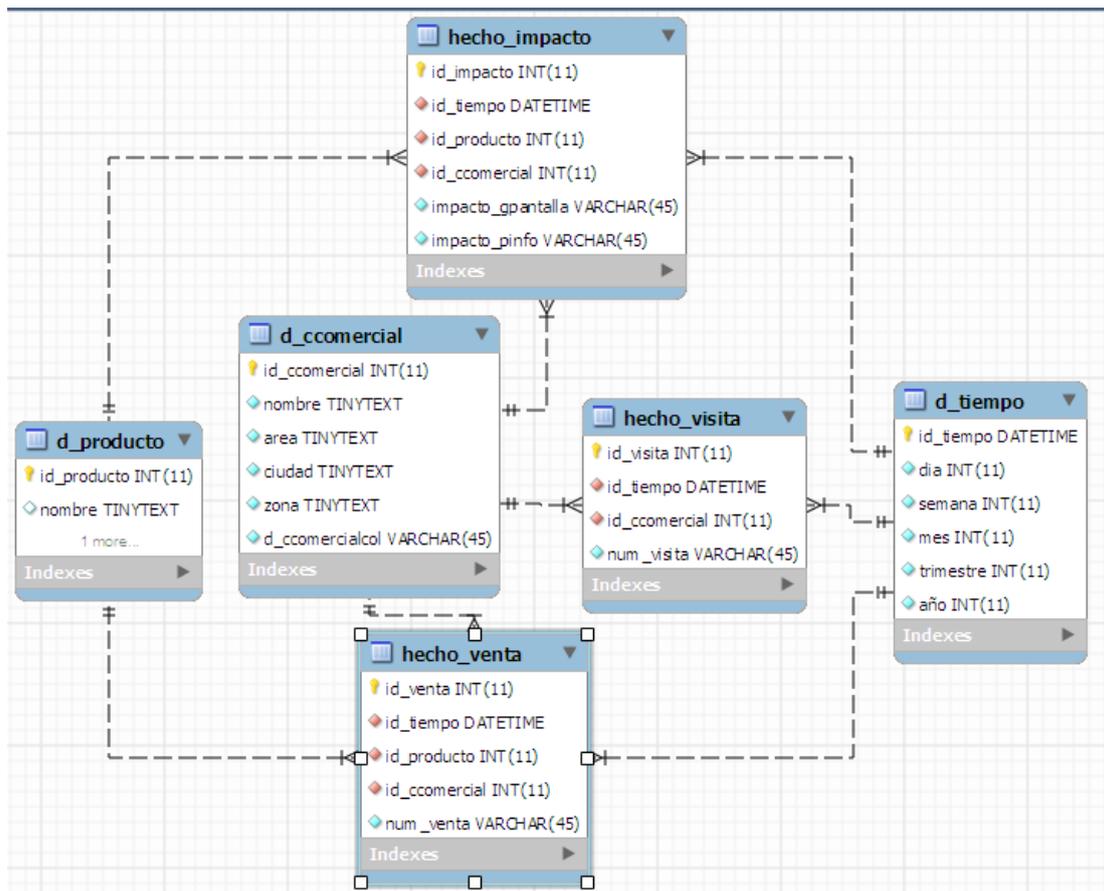


Figura 5: modelo físico de datos

4. Diseño e implementación de los procesos ETL

Una vez que se diseña e implementa el data warehouse, el siguiente paso es el diseño e implementación del proceso ETL (extracción, transformación y carga). El proceso ETL toma los datos de su origen, luego se transforman y se cargan en el sistema destino, en este caso el DWH que ya se diseñó. Al emplear la herramienta ETL se realiza una integración de los datos impecables de distintas fuentes, en este caso los datos solo se tomaran de una fuente, como la base de datos suministrada llamada DATACAMPAIGNIMPACT en formato Excel. Esto va a permitir generar los informes necesarios.

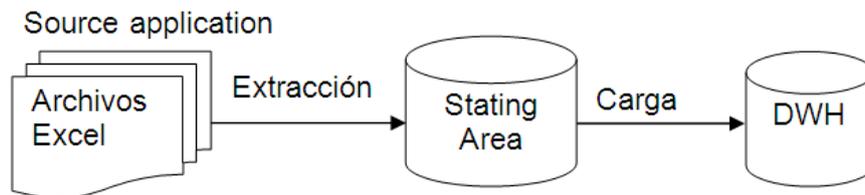


Figura 6: Implementación del proceso ETL.

La herramienta que se usará para la integración de datos, es Spoon, el entorno grafico que permite trabajar con Kettle. Con la herramienta Spoon se extraerán los datos desde Excel donde está la base de datos, que contiene la información de la compañía publicitaria hecha en varias centros comerciales de España y luego serán cargados en MySQL estos datos serán de objeto de análisis. A continuación se muestra el entorno inicial de Spoon.

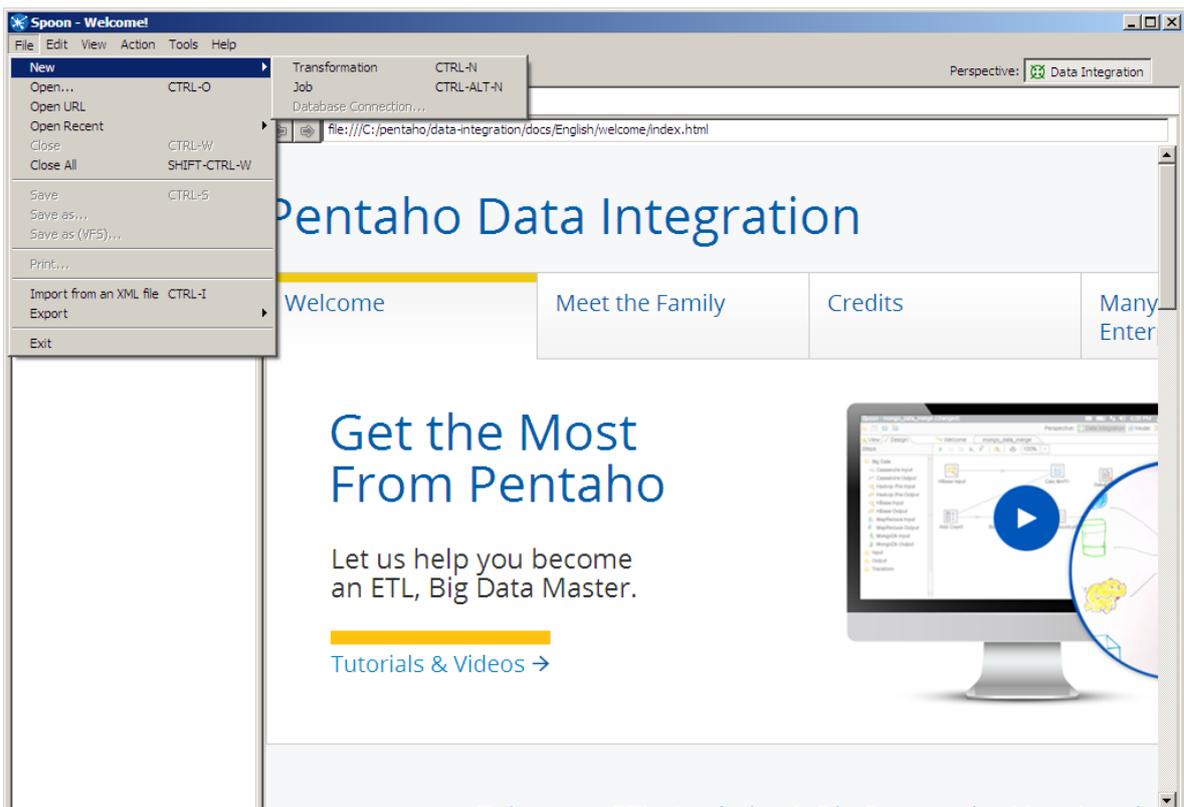


Figura 7: Herramienta Spoon para el proceso ETL.

Como primer paso en Spoon se creará la transformación con su nombre y la dirección donde se guardará el archivo, a continuación se mostrará la figura donde se muestra la creación de la transformación.

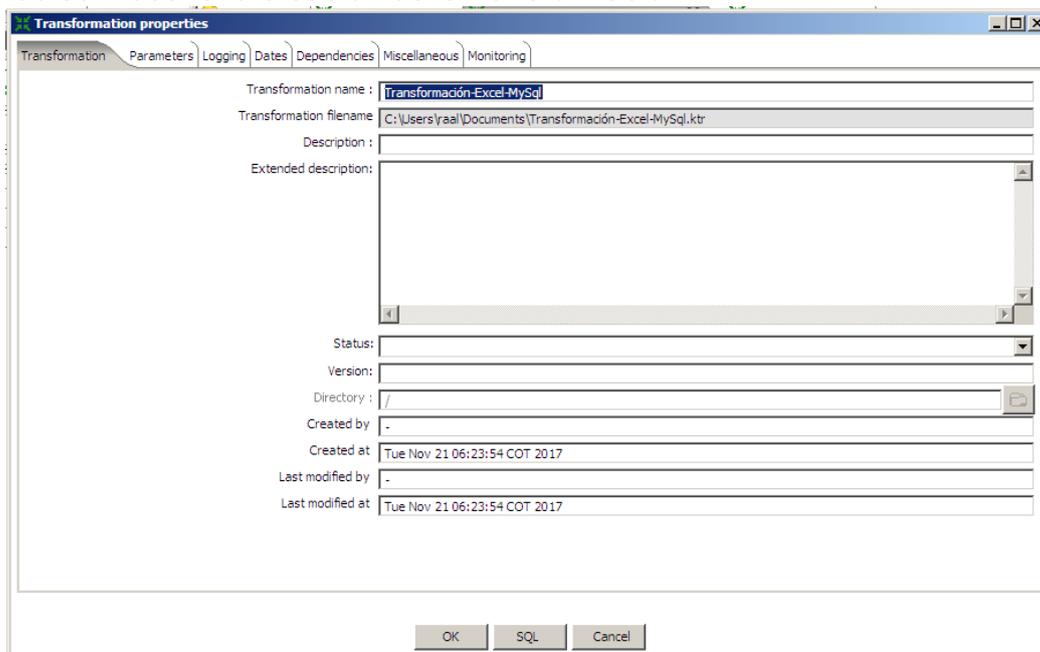


Figura 8: creación de la transformación

Luego se establece una conexión con MySQL donde se guardarán los datos del DWH provenientes de la base de datos en formato Excel suministrada. A continuación se observa la conexión hecha con el nombre del host, el nombre de la base de datos, el puerto, el usuario y la contraseña.

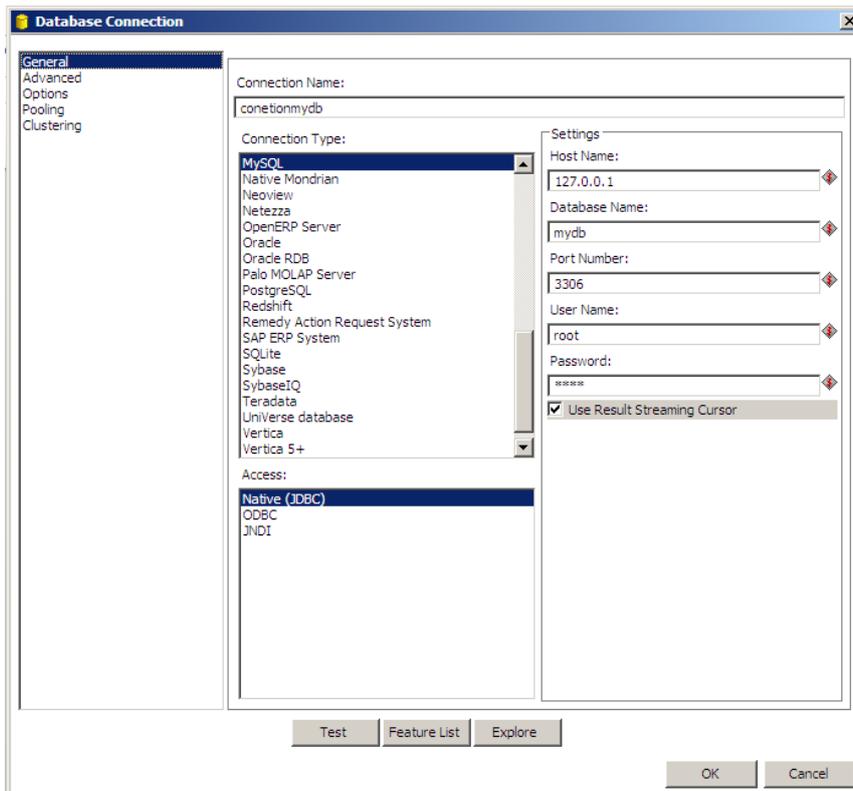


Figura 9: conexión con MySql.

Para cada carga de datos se crea su correspondiente tabla de salida. Para facilitar el proceso de transformación se utiliza un área intermedia de secuencia, que además de las columnas propias de cada tabla, se ha ampliado con la columna id, que se emplea para generar las correspondientes autoincrementales que se utilizarán en las tablas de dimensiones y hechos del DWH.

4.1 Transformación y carga de producto.



Figura 10: Transformación y carga de d_producto.

Entrada_productos: Paso en el que se lee el archivo Excel que contiene los datos de los productos.

Autonumerico: añade un valor entero que inicia en uno e incrementa de uno en uno, agrega los datos dependiendo de los que haya en el archivo excel al momento de cargarlos.

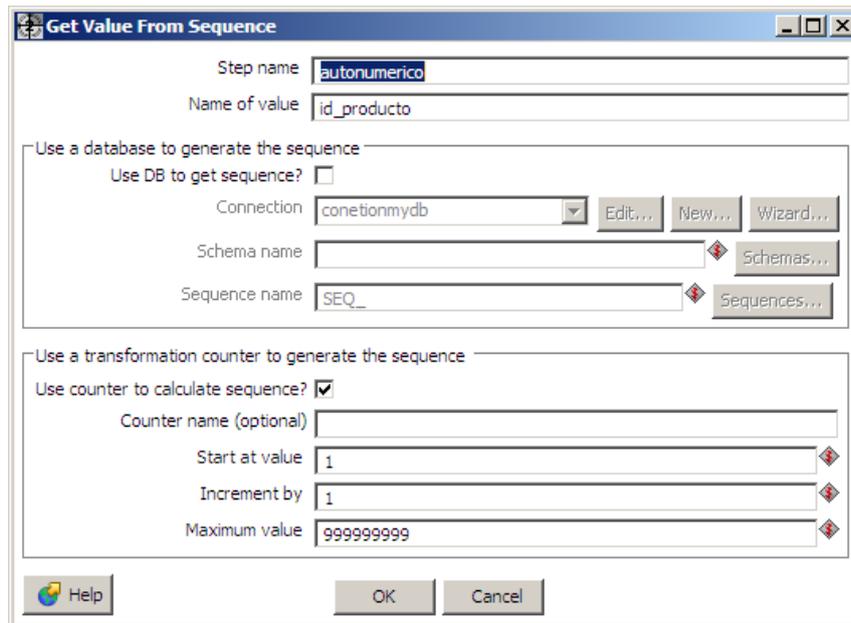


Figura 11: Obtiene valores de secuencia para d_producto.

Carga de productos: Carga los datos de los productos del flujo de entrada en la tabla d_producto.

A continuación, se puede ver en MySQL los productos cargados a la tabla d_producto del Data Warehouse.

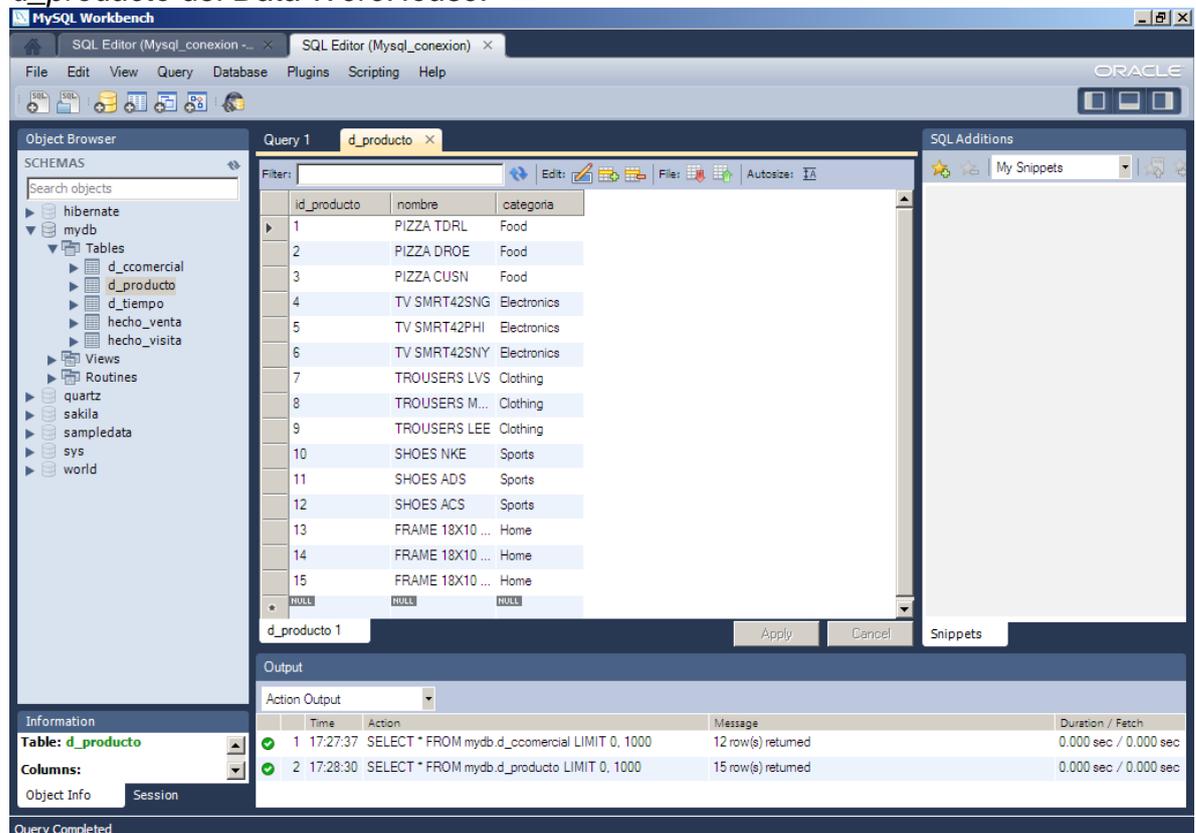


Figura 12: Productos cargados a d_producto en MySQL.

4.2 Transformación y carga de tiempo.

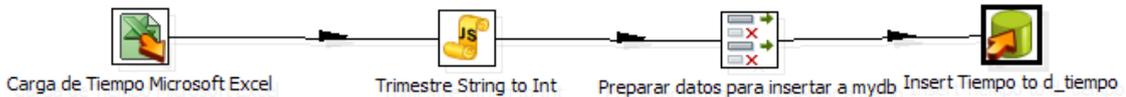


Figura 13: Transformación y carga de d_tiempo.

Carga de tiempo Microsoft Excel: Paso en el que se lee el archivo Excel que contiene los datos de tiempo.

Trimestre String to Int: Paso javascript que transforma el campo de trimestre.

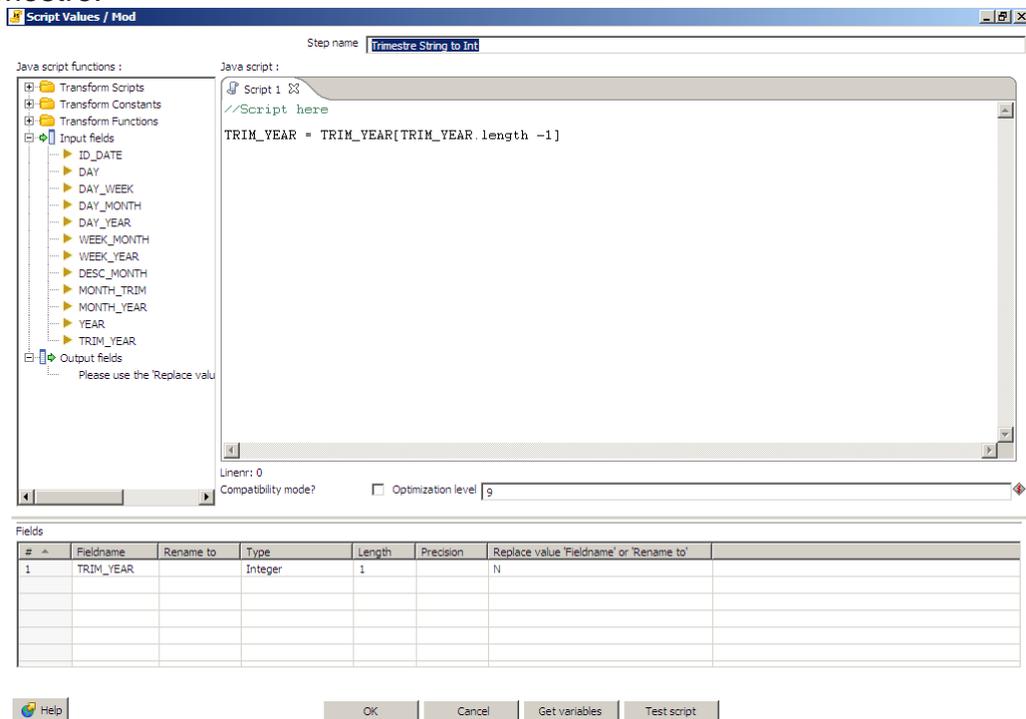


Figura 14: transformación del campo trimestre.

Preparar datos para insertar a mydb: selecciona y renombrar los valores, y así prepararlos para d_tiempo.

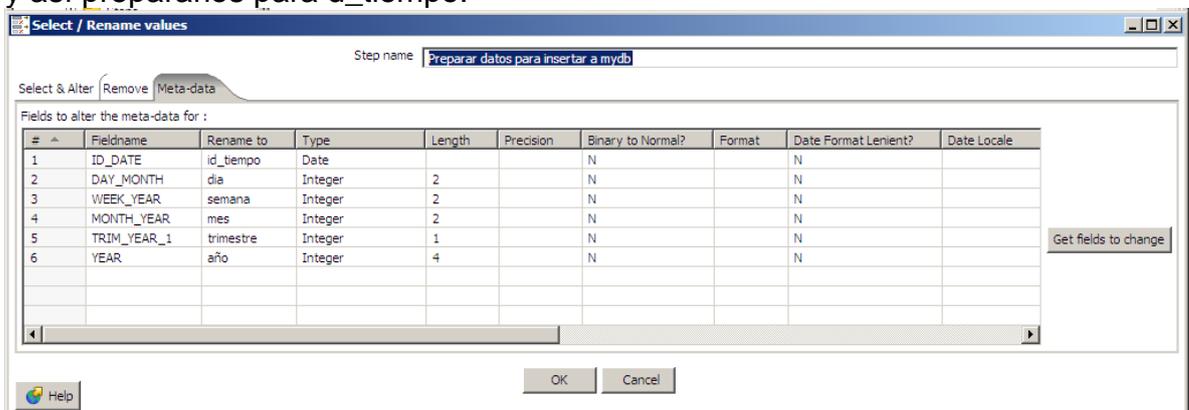


Figura 15: preparación de los datos para d_tiempo.

Insert Tiempo to d_tiempo: Carga o inserta los datos de fecha del flujo de entrada en la tabla d_tiempo del DWH.

A continuación, se puede ver en MySql las fechas cargadas a la tabla *d_tiempo* del *Data WereHouse*.

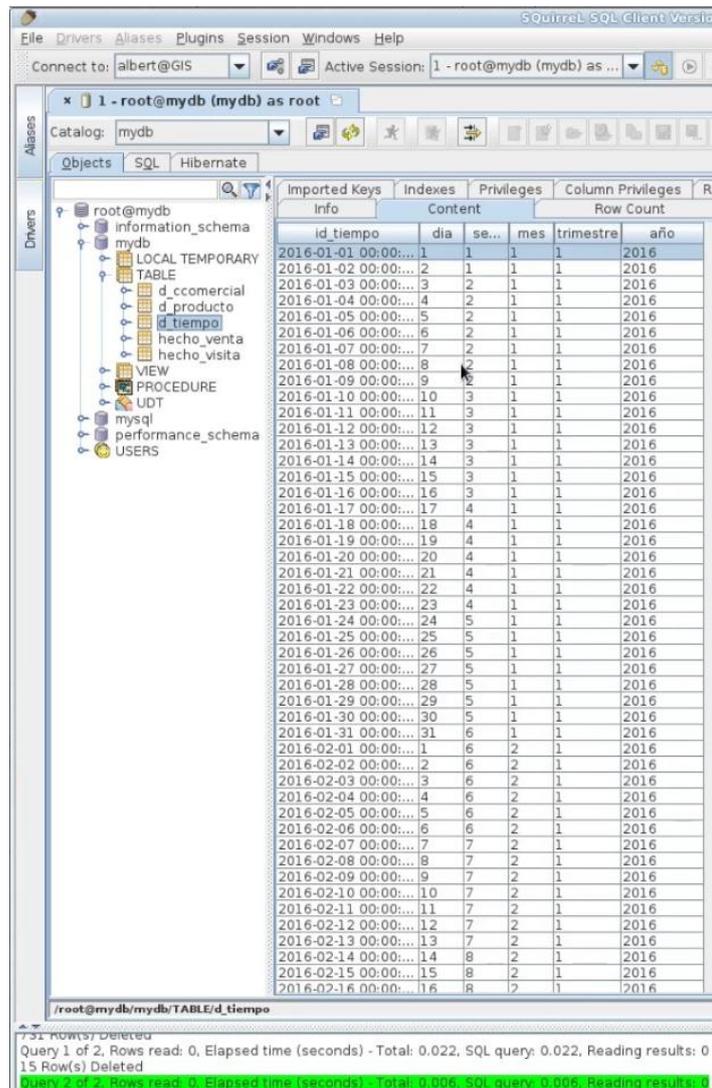


Figura 16: Productos cargados a *d_tiempo* en MySql.

4.3 Transformación y carga de centro comercial.

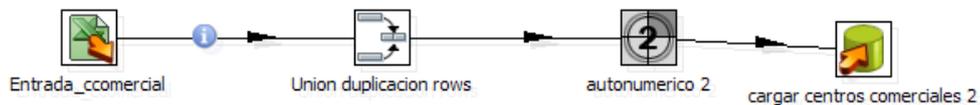


Figura 17: Transformación y carga de *d_ccomercial*.

Entrada_ccomercial: Paso en el que se lee el archivo Excel que contiene los datos de los puntos o centros comerciales.

Union duplicacion rows: este paso permite comparar y hacer coincidir las filas ZONE, CITY Y SHOPPING CENTER que vienen de *entrada_ccomercial*.

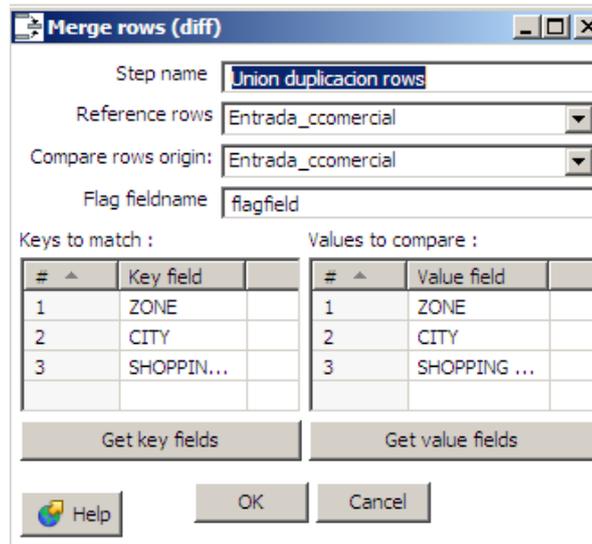


Figura 18: Merge Rows (Diff)

Autonumerico 2: añade un valor entero que inicia en uno, e incrementa de uno en uno, agrega los datos dependiendo de los que haya en el archivo excel al momento de cargarlos.

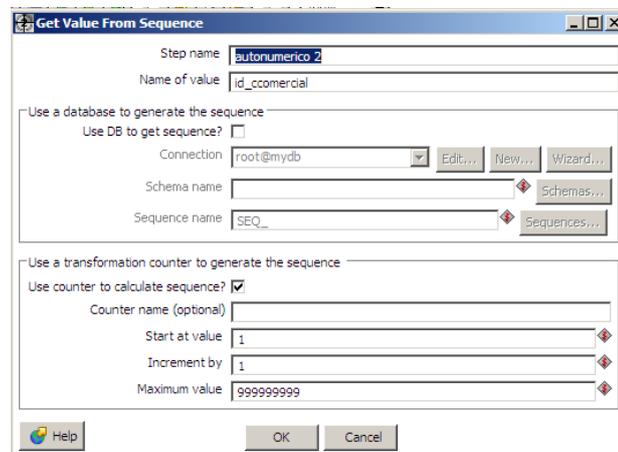


Figura 19: Obtiene valores de secuencia para d_ccomercial.

Cargar centros comerciales: Carga los datos de los centros comerciales del flujo de entrada en la tabla *d_ccomercial*.

A continuación, se puede ver en MySQL los centros comerciales cargados a la tabla *d_ccomercial* del *Data Warehouse*.

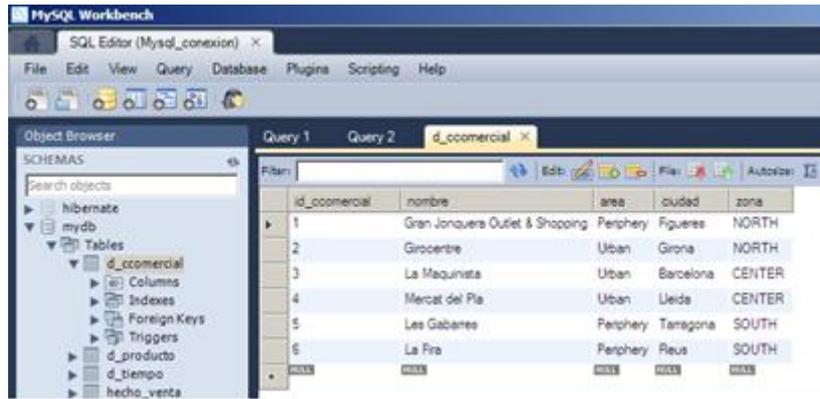


Figura 20: Productos cargados a d_ccomercial en MySql.

4.4 Transformación y carga de visita.

La figura 20 muestra la transformación de visita ejecutada de manera exitosa.

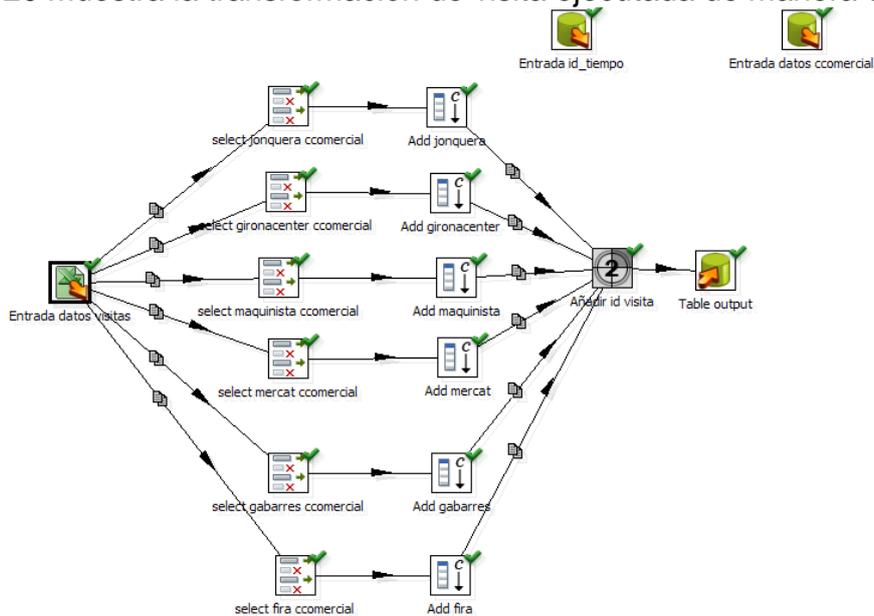


Figura 21: Transformación de la tabla de hecho visita.

Lo primero que se hizo fue seleccionar una tabla de entrada input tanto para id_tiempo, como para los datos del centro comercial, de esta manera se extraen estos datos que van a ser claves para la carga de los datos. Luego selecciona una entrada de microsoft Excel, para poder traer la hoja de visita de la base de datos en excel con las columnas correspondientes.

A continuación, se muestra el siguiente paso que fue insertar la herramienta Select/ Rename values, como su nombre lo indica selecciona y renombra valores de la hoja de calculo visits, en este caso primero selecciona el DATE y lo renombra en id_date con formato date; luego se extraen los valores de la columna de cada centro comercial y se convierten en num_visita en formato entero con máximo ocho valores.

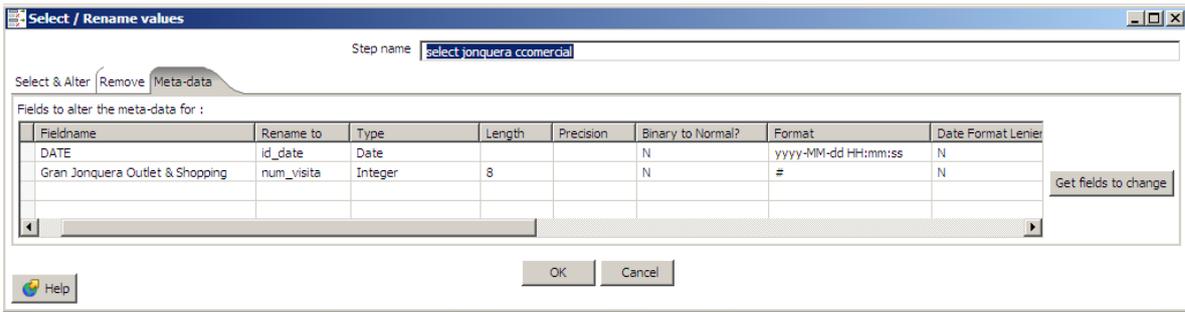


Figura 22: Selección y renombre de los valores.

En la figura 22, se muestra la adición de la herramienta add constant values, que permite seleccionar el campo id_ccomercial de tipo entero y adicionarle el número que corresponda al centro comercial.

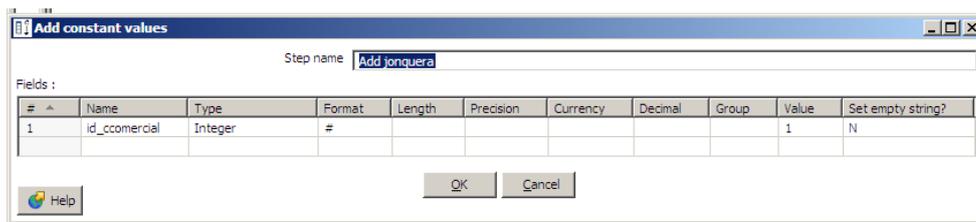


Figura 23: Adición de valores constante.

Adicionalmente se le añade el autonumerico para que inserte los valores de la columna id_visita.

Por último, se puede apreciar en la figura 23 que en MySQL los centros comerciales fueron cargados a la tabla h_visita del Data WereHouse.

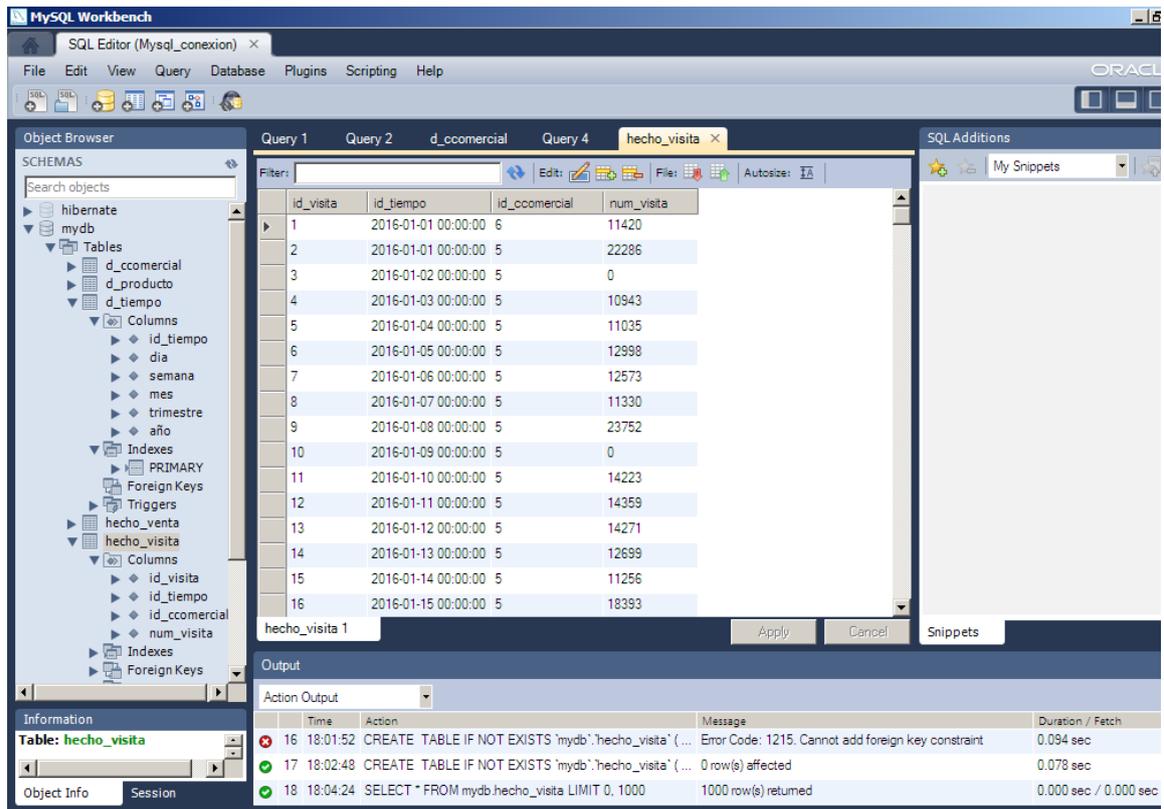


Figura 24: Verificación de los datos cargados en MySql.

4.5 Transformación y carga de venta.

En la figura 24, se muestra toda la transformación de venta que va permitir tomar los datos de la base de datos en Excel, pasarla por una serie de pasos hasta cargarlos en la base de datos de MySql en la tabla llamada hecho_venta. Con las entradas de los datos en Excel se recopilan las hojas de cálculo Sales Gran Jonquera SC, Sales Giron Center, Sales Maquinista, Sales Mercat, Sales Gabarres y Sales La Fira. Luego se introduce la opción de add constants, para que se adicione el id que corresponde al nombre del centro comercial. Al adicionar como tercer paso la opción de select rename values con el nombre Get all information se muestran todos los productos vendidos y que se encontraban en las hojas de cálculos Sales, se renombran y se le especifica el tipo de dato Integer. El cuarto paso nuevamente con la opción select rename values llamada remove columns más el nombre del producto, se retiran las columnas de los nombre de los productos. En el quinto paso se adiciona un Script para que cuente cada producto de las ventas, a manera de ejemplo se selecciona el producto pi_tdlr y se le agrega el 1, lo que quiere decir que el id del producto pi_tdlr es 1; y así sucesivamente para cada producto. El sexto paso añade el autonumerico de id_ventas, para que comience en 1 e incremente de 1 en 1. Por último se especifican los campos que se van añadir y con qué nombre.

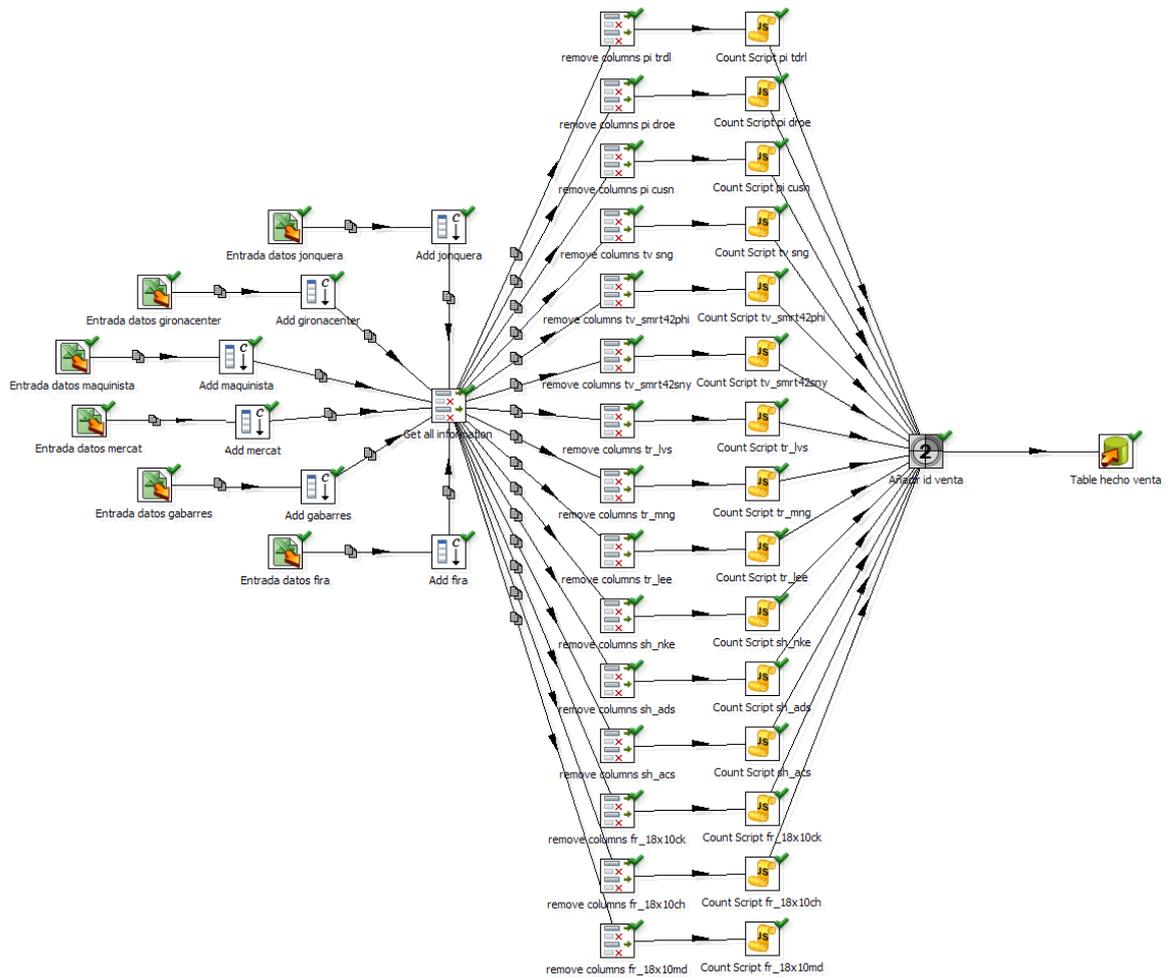


Figura 25: transformación hecho_venta.

Una vez hechos todos los pasos anteriores de la transformación se ejecuta dicha transformación y se comprueban que se hayan cargado correctamente los datos en MySQL Workbench, en la figura 25 se puede comprobar que los datos han sido cargados correctamente en la tabla hecho_venta.

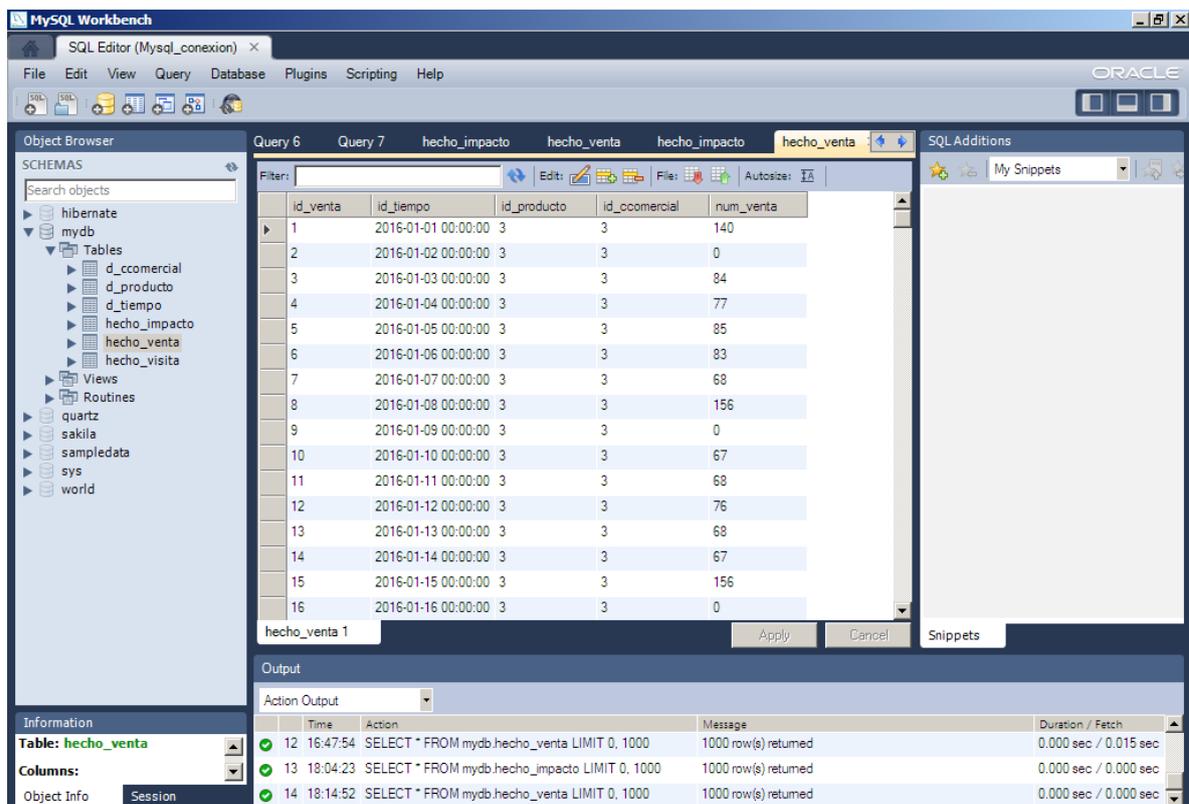


Figura 26: Verificación de la carga de h_venta en SQL.

4.6 Transformación y carga de impacto.

En la figura 26, se muestra toda la transformación de impacto que va permitir tomar los datos de la base de datos en Excel, pasarla por una serie de pasos hasta cargarlos en la base de datos de MySQL en la tabla llamada hecho_impacts.

Con las entradas de los datos en Excel se recopilan las hojas de cálculo Impacts Gran Jonquera SC, Impacts Giron Center, Impacts Maquinista, Impacts Mercat, Impacts Gabarres y Impacts La Fira. Luego se introduce la opción de add constants, para que se adicione el id que corresponde al nombre del centro comercial.

El tercer paso es adicionar la opción de select rename values con el nombre Get all information con esta opción se obtienen la fecha, se renombra y se especifica que es de tipo date; luego se obtiene el impacto de los puntos, se renombra a impacto_points y se le especifica que es de tipo string; después se obtienen todos los productos vendidos y que se encontraban en las hojas de cálculos Impacts, se renombran y se le especifica el tipo de dato Integer.

En el cuarto paso, se adiciona la opción Dummy para comprobar que los productos no sean falsos y luego se añade la opción filter rows con el nombre filter más el nombre del producto; esta opción reconoce que si los datos son falsos no es buen dato para cargar en la base de datos de salida.

El quinto paso nuevamente con la opción select rename values llamada remove columns más el nombre del producto, se retiran las columnas de los nombre de los productos. En el sexto paso se adiciona un Script para que cuente cada producto de impacto, a manera de ejemplo se selecciona el producto pi_tdrl y se le agrega el 1, lo que quiere decir que el id del producto pi_tdrl es 1, además

se le especifica al producto que el punto de impacto puede ser Big Screen o sino Information Point Screen, dependiendo de la información que contenga el producto en excel; y así sucesivamente para cada producto. El séptimo paso añade el autonumerico de id_impacto, para que comience en 1 e incremente de 1 en 1. Por último se especifican los campos que se van añadir y con qué nombre.

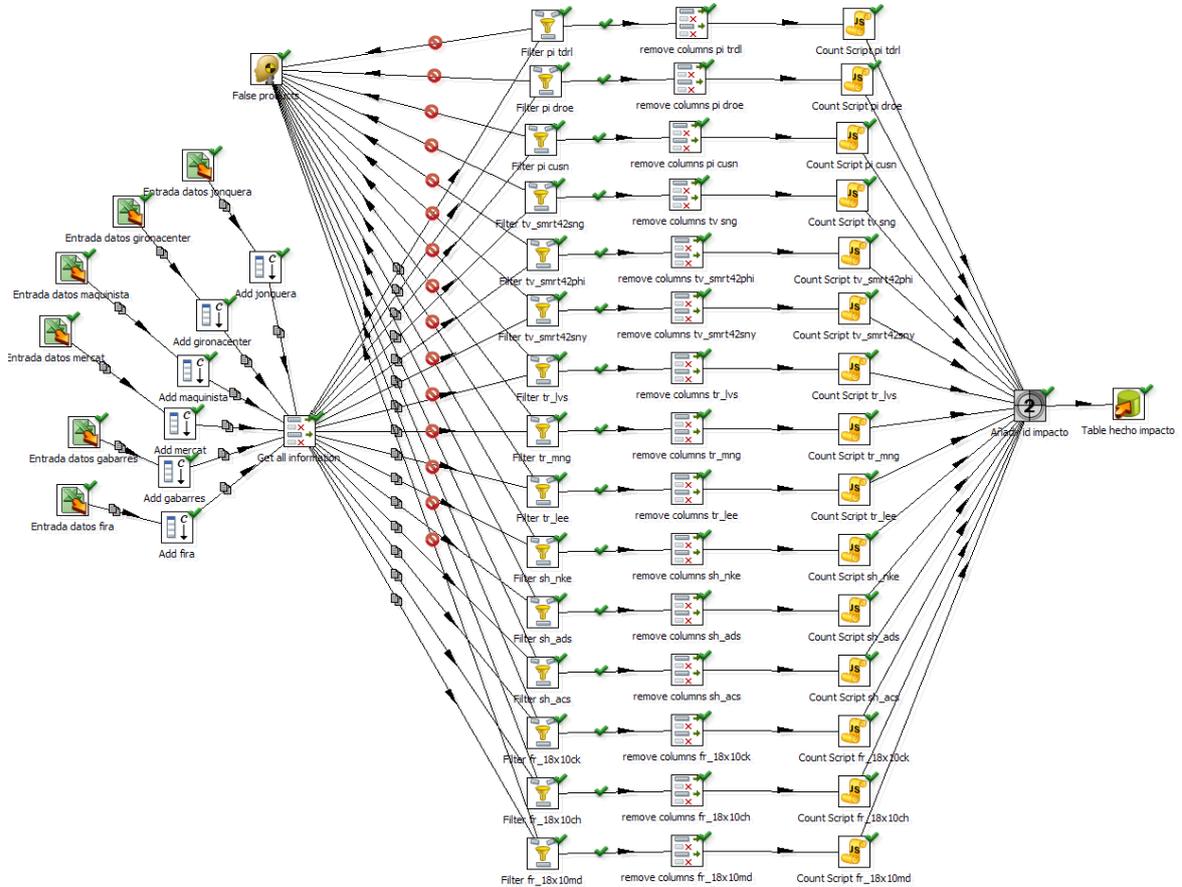


Figura 27: transformación de h_impacto

Una vez terminados los pasos anteriores de la transformación, se ejecuta dicha transformación y se comprueban que se hayan cargado correctamente los datos en MySQL Workbench, en la figura 27 se puede comprobar que los datos han sido cargados correctamente en la tabla hecho_impacts.

The screenshot shows a database management interface with the following components:

- Object Browser:** Shows a tree view of schemas including hibernate, mydb, quartz, sakila, sampledata, sys, and world. The 'mydb' schema is expanded to show tables like d_ccomercial, d_producto, d_tiempo, hecho_impacto, hecho_venta, and hecho_visita.
- Table View:** Displays the 'hecho_impacto' table with columns: id_impacto, id_tiempo, id_producto, id_ccomercial, impacto_goantalla, and impacto_pinfo. The table contains 16 rows of data.
- Output Panel:** Shows the execution results of three SQL queries. Each query is a 'SELECT * FROM mydb.hecho_impacto LIMIT 0, 1000' statement, and each returned 1000 rows.

id_impacto	id_tiempo	id_producto	id_ccomercial	impacto_goantalla	impacto_pinfo
1	2016-01-01 00:00:00	9	5	9	Big Screen
2	2016-01-02 00:00:00	9	5	10	Big Screen
3	2016-01-03 00:00:00	9	5	10	Big Screen
4	2016-01-04 00:00:00	9	5	11	Big Screen
5	2016-01-05 00:00:00	9	5	12	Big Screen
6	2016-01-06 00:00:00	9	5	12	Big Screen
7	2016-01-07 00:00:00	9	5	12	Big Screen
8	2016-01-08 00:00:00	9	5	12	Big Screen
9	2016-01-09 00:00:00	9	5	12	Big Screen
10	2016-01-10 00:00:00	9	5	10	Big Screen
11	2016-01-11 00:00:00	9	5	10	Big Screen
12	2016-01-12 00:00:00	9	5	9	Big Screen
13	2016-01-13 00:00:00	9	5	12	Big Screen
14	2016-01-14 00:00:00	9	5	12	Big Screen
15	2016-01-15 00:00:00	9	5	9	Big Screen
16	2016-01-16 00:00:00	9	5	12	Big Screen

Time	Action	Message	Duration / Fetch
11 16:47:44	SELECT * FROM mydb.hecho_impacto LIMIT 0, 1000	1000 row(s) returned	0.000 sec / 0.016 sec
12 16:47:54	SELECT * FROM mydb.hecho_venta LIMIT 0, 1000	1000 row(s) returned	0.000 sec / 0.015 sec
13 18:04:23	SELECT * FROM mydb.hecho_impacto LIMIT 0, 1000	1000 row(s) returned	0.000 sec / 0.000 sec

Figura 28: Verificación de la carga de h_impacto en SQL.

5. Creación del entorno BI para el análisis

Ya terminado el proceso de diseño e implementación del proceso ETL, donde se cargan los datos de la campaña publicitaria al DWH, se procede a crear los cubos por cada tabla de hecho y luego a generar los informes, que permitirá saber si la campaña publicitaria basada en la reproducción de pequeños anuncios en forma de impactos visuales en pantallas o monitores situados dentro de varios centros comerciales, incrementa el número de ventas de los artículos anunciados.

El análisis se hará en forma de cubos de datos, en donde se hace más cuidadoso y la extracción de las conclusiones se hacen mejor elaboradas. Los cubos se realizan con la herramienta Mondrian Schema Workbench y los informes se generarán utilizando el plugin de pentaho Saiku. De esta forma, se podrá dar respuesta a las siguientes preguntas:

- ¿En general, la publicidad basada en impactos visuales, es positiva?
- ¿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 que se producen?
- ¿Existe algún artículo o familia que según la zona, su publicidad resulta más efectiva? ¿Y según el momento en que se realiza?

1. ¿En general, la publicidad basada en impactos visuales, es positiva?

Sí, ya que hay un aumento significativo en las ventas de los productos en los centros comerciales publicitados en las pantallas de forma controlada, obteniendo un resultado del 31,2% de las ventas con respecto al impacto.



Figura 29: promedio de la publicidad visual.

Año	C_PInfo	DC_PInfo	SUM_PInfo	MIN_PInfo	MAX_PInfo	AVG_PInfo	C_GPantalla	DC_GPantalla	SUM_GPantalla	MIN_GPantalla	MAX_GPantalla	AVG_GPantalla
2016	54900	13	140585	0,00	12	3	54900	13	140866	0,00	12	3

Figura 30: estadística básica de la publicidad en general

2. ¿Hay determinadas zonas geográficas donde los impactos son más efectivos?

En la zona de Figueres la publicidad presentada en los puntos de información es más efectiva, causando así un aumento en las ventas. Luego le sigue las zonas de Barcelona y Reus, en Lleida hay un aumento pero no es tan significativo como en las otras zonas. En la zona de Tarragona la publicidad presentada en los puntos de información no causan gran impacto en los visitantes, ya que las ventas no dependen de la publicidad en los puntos de información.

Por otra parte la publicidad presentada en las grandes pantallas causa el mayor impacto en la zona de Lleida, en Figueres muestra un impacto positivo en las ventas, dónde decrece el impacto son en las zonas de Barcelona y Reus. Dejando claro que en Tarragona la publicidad presentada en la pantalla grande no causa casi impacto, por lo que las ventas no crecen a causa de este tipo de publicidad.

Ciudad	Zona	C_PInfo	DC_PInfo	SUM_PInfo	MIN_PInfo	MAX_PInfo	AVG_PInfo	C_GPantalla	DC_GPantalla	SUM_GPantalla	MIN_GPantalla	MAX_GPantalla	AVG_GPantalla
Figueres	NORTH	10980	13	28178	0,00	12	3	10980	13	28227	0,00	12	3
Reus	SOUTH	10980	13	28137	0,00	12	3	10980	13	28117	0,00	12	3
Barcelona	CENTER	10980	13	28140	0,00	12	3	10980	13	28154	0,00	12	3
Tarragona	SOUTH	10980	13	28035	0,00	12	3	10980	13	28101	0,00	12	3
Lleida	CENTER	10980	13	28095	0,00	12	3	10980	13	28267	0,00	12	3

Figura 31: Estadísticas básicas de las zonas impactadas.



Figura 32: impacto de la publicidad en los puntos de información.



Figura 33: impacto de la publicidad en los puntos de información.

3. ¿Según la familia de producto, existe alguna relación entre los artículos impactados y el resultado obtenido?

La familia de producto Electronics tiene un impacto de 34.5 % para cada pantalla y disminuye el promedio del resultado obtenido en un 15.5%, con respecto a las otras familias de productos (home, food, Sports, Clothing), que en promedio el impacto en la pantalla del punto de información y la gran pantalla es del 34.4 % y el resultado obtenido en realidad fue en promedio del 15.6% para cada pantalla.

En Electronics hay una disminución mínima por lo que no hay una mayor variabilidad con respecto a las otras familias de productos.

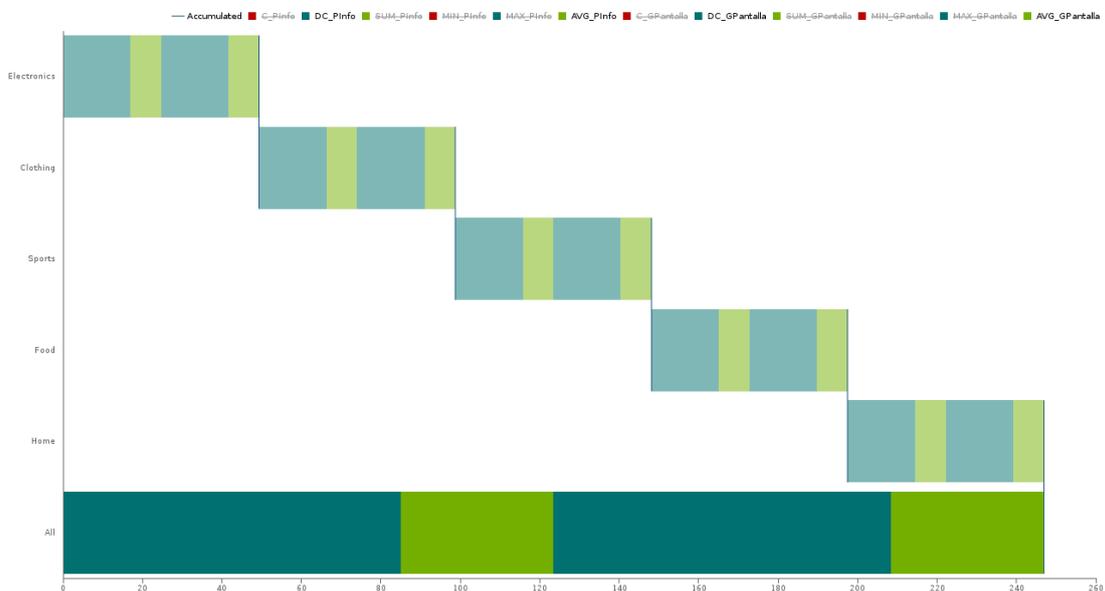


Figura 34: relación de los productos impactados.

Categoría	C_PInfo	DC_PInfo	SUM_PInfo	MIN_PInfo	MAX_PInfo	AVG_PInfo	C_GPantalla	DC_GPantalla	SUM_GPantalla	MIN_GPantalla	MAX_GPantalla	AVG_GPantalla
Home	3660	9	9173	0,00	8	3	3660	9	9213	0,00	8	3
Home	3660	8	7549	0,00	12	2	3660	8	7581	0,00	12	2
Home	3660	8	11437	0,00	12	3	3660	8	11482	0,00	12	3
Food	3660	8	11427	0,00	12	3	3660	8	11430	0,00	12	3
Food	3660	9	9186	0,00	8	3	3660	9	9065	0,00	8	2
Food	3660	8	7566	0,00	12	2	3660	8	7535	0,00	12	2
Sports	3660	8	11390	0,00	12	3	3660	8	11481	0,00	12	3
Sports	3660	9	9176	0,00	8	3	3660	9	9145	0,00	8	2
Sports	3660	8	7529	0,00	12	2	3660	8	7606	0,00	12	2
Clothing	3660	8	11419	0,00	12	3	3660	8	11414	0,00	12	3
Clothing	3660	8	7530	0,00	12	2	3660	8	7585	0,00	12	2
Clothing	3660	9	9263	0,00	8	3	3660	9	9141	0,00	8	2
Electronics	3660	9	9059	0,00	8	2	3660	9	9177	0,00	8	3
Electronics	3660	8	7450	0,00	12	2	3660	8	7519	0,00	12	2
Electronics	3660	8	11431	0,00	12	3	3660	8	11492	0,00	12	3

Figura 35: estadísticas de los productos impactados.

4. ¿Son más efectivos los impactos, según el momento temporal en que se producen?

El impacto de la publicidad mostrada en las pantallas de información es más efectivos durante el primer y tercer trimestre, durante el segundo y cuarto trimestre el impactos no es tan positivo.

En cambio la publicidad en pantallas grandes es menos efectiva en el segundo trimestre, los demás trimestres son más efectivos.

Año	Trimestre	C_PInfo	DC_PInfo	SUM_PInfo	MIN_PInfo	MAX_PInfo	AVG_PInfo	C_GPantalla	DC_GPantalla	SUM_GPantalla	MIN_GPantalla	MAX_GPantalla	AVG_GPantalla
2016	1	13650	13	42150	0,00	12	3	13650	13	42115	0,00	12	3
	2	13650	13	42072	0,00	12	3	13650	4	13659	0,00	3	1
	3	13800	13	42559	0,00	12	3	13800	13	42541	0,00	12	3
	4	13800	4	13804	0,00	3	1	13800	13	42551	0,00	12	3

Figura 36: estadísticas básicas de impactos en los trimestres.

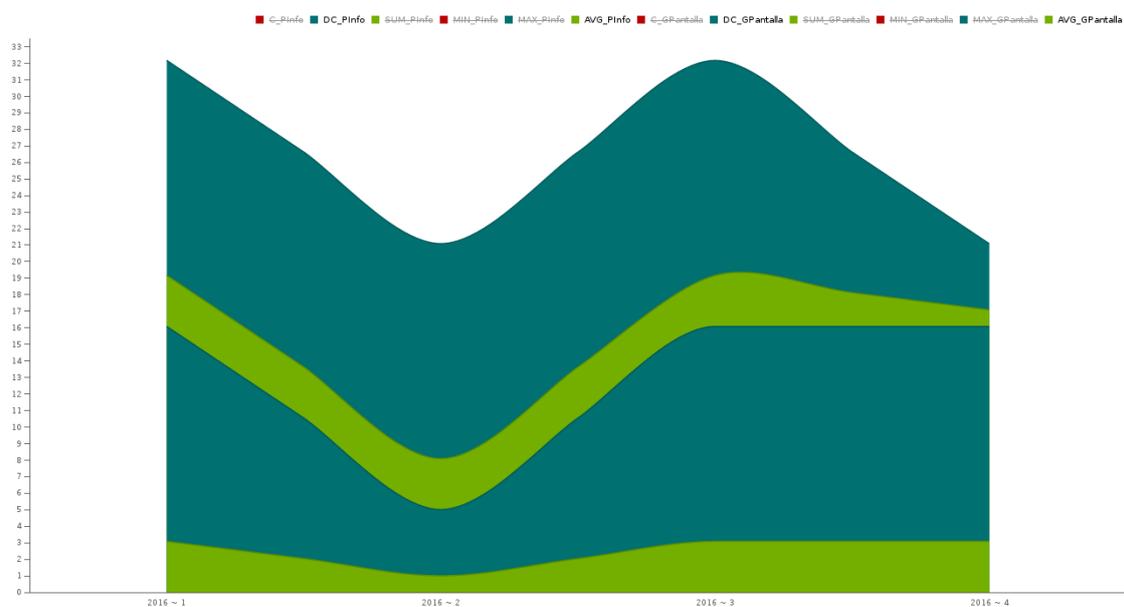


Figura 37: impacto de la publicidad en los trimestres.

5. ¿Existe algún artículo o familia que según la zona, su publicidad resulta más efectiva? ¿Y según el momento en que se realiza?

Todos los productos publicitados en las pantallas tienen la misma efectividad en cualquier zona o en cualquier momento, lo que quiere decir que sin importar el tipo de producto que se ofrezca la efectividad es la misma para todos.

(All)	Ciudad	Zona	C_Pinfo	DC_Pinfo	SUM_Pinfo	MIN_Pinfo	MAX_Pinfo	AVG_Pinfo	C_GPantalla	DC_GPantalla	SUM_GPantalla	MIN_GPantalla	MAX_GPantalla	AVG_GPantalla
Productos	Figueras	NORTH	10980	13	28178	0,00	12	3	10980	13	28227	0,00	12	3
	Reus	SOUTH	10980	13	28137	0,00	12	3	10980	13	28117	0,00	12	3
	Barcelona	CENTER	10980	13	28140	0,00	12	3	10980	13	28154	0,00	12	3
	Tarragona	SOUTH	10980	13	28035	0,00	12	3	10980	13	28101	0,00	12	3
	Lleida	CENTER	10980	13	28095	0,00	12	3	10980	13	28267	0,00	12	3

Figura 38: efectividad de productos por zonas.

(All)	Ciudad	Zona	Año	Trimestre	C_Pinfo	DC_Pinfo	SUM_Pinfo	MIN_Pinfo	MAX_Pinfo	AVG_Pinfo	C_GPantalla	DC_GPantalla	SUM_GPantalla	MIN_GPantalla	MAX_GPantalla	AVG_GPantalla
Productos	Figueras	NORTH	2016	1	2730	13	8418	0,00	12	3	2730	13	8506	0,00	12	3
				2	2730	13	8448	0,00	12	3	2730	4	2706	0,00	3	1
				3	2760	13	8549	0,00	12	3	2760	13	8542	0,00	12	3
				4	2760	4	2763	0,00	3	1	2760	13	8473	0,00	12	3
	Reus	SOUTH	2016	1	2730	13	8465	0,00	12	3	2730	13	8360	0,00	12	3
				2	2730	13	8415	0,00	12	3	2730	4	2738	0,00	3	1
				3	2760	13	8506	0,00	12	3	2760	13	8510	0,00	12	3
				4	2760	4	2751	0,00	3	1	2760	13	8509	0,00	12	3
	Barcelona	CENTER	2016	1	2730	13	8461	0,00	12	3	2730	13	8410	0,00	12	3
				2	2730	13	8430	0,00	12	3	2730	4	2777	0,00	3	1
				3	2760	13	8475	0,00	12	3	2760	13	8522	0,00	12	3
				4	2760	4	2774	0,00	3	1	2760	13	8445	0,00	12	3
	Tarragona	SOUTH	2016	1	2730	13	8378	0,00	12	3	2730	13	8408	0,00	12	3
				2	2730	13	8386	0,00	12	3	2730	4	2695	0,00	3	1
				3	2760	13	8518	0,00	12	3	2760	13	8435	0,00	12	3
				4	2760	4	2753	0,00	3	1	2760	13	8563	0,00	12	3
	Lleida	CENTER	2016	1	2730	13	8428	0,00	12	3	2730	13	8431	0,00	12	3
				2	2730	13	8393	0,00	12	3	2730	4	2743	0,00	3	1
				3	2760	13	8511	0,00	12	3	2760	13	8532	0,00	12	3
				4	2760	4	2763	0,00	3	1	2760	13	8561	0,00	12	3

Figura 39: efectividad de productos en tiempo.

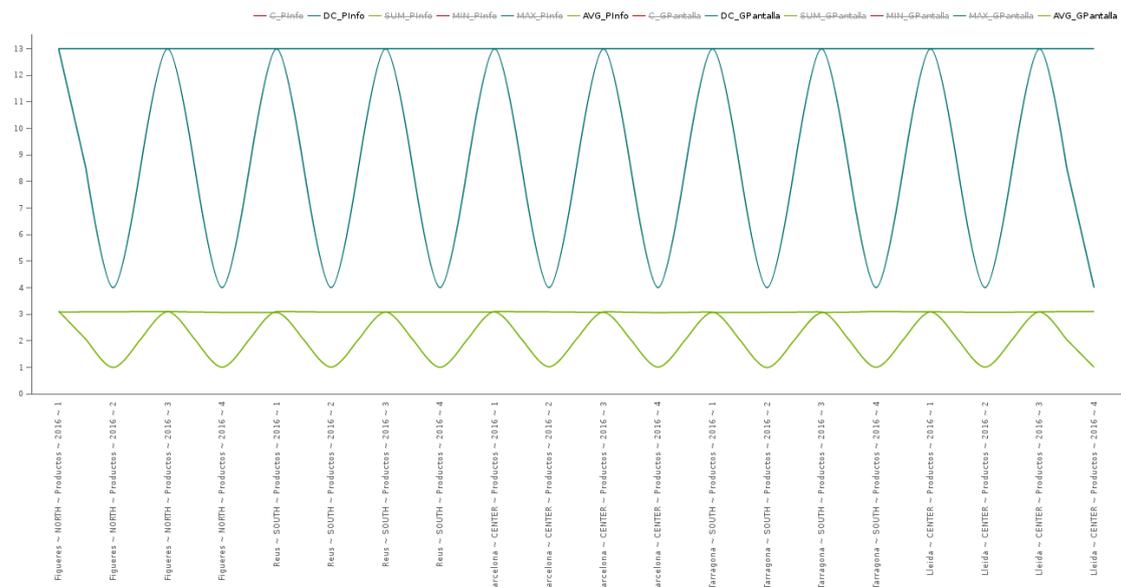


Figura 40: efectividad de productos publicitados.

6. Análisis de los datos y Conclusiones

- El general, se pudo conseguir alcanzar los objetivos planteados desde el principio del semestre, en el trabajo final de Máster; ya se pudo verificar la efectividad de campaña publicitaria basada en reproducción de pequeños anuncios de forma controlada.
- El enfoque seguido para abordar el proyecto ha sido el adecuado, ya que se ha comprobado si la publicidad digital aumenta el número de ventas.
- La planificación ha sido la correcta y no ha habido grandes desviaciones con respecto la planteada inicialmente.
- Es importante señalar que la herramienta Pentaho ha requerido una dedicación extra en cuanto a la instalación, ya que presentaba error con la compatibilidad de la máquina que se estaba trabajando, para desarrollar los procesos ETL.
- Al incorporar una plataforma BI en una empresa, da la posibilidad de recopilar y distribuir de manera eficiente la información de varias fuentes de datos, ya sea de un negocio o empresa. Dada la gran cantidad de datos que se maneja en una empresa esta herramienta resulta muy útil para tomar decisiones y planificar la estrategia del negocio.
- La parte más importante del proyecto consiste en la elaboración de un modelo de datos capaz de cubrir las necesidades de información solicitadas por el usuario. Esto es así, con independencia de lo compleja que se convierta la ETL encargada de poblarlo, ya que el tiempo y los recursos que se invierten en esta parte, son recompensados al disponer de un modelo fácil de explotar y rápido en cuanto a tiempos de respuesta se refiere. El modelo creado cumple estas características.
- Se dispone de un conjunto de informes predefinidos que son un buen punto de partida para mostrar las capacidades que ofrece el sistema al usuario. Esto le permitirá comprender la utilidad y el valor que la publicidad basada en impactos visuales proporciona, para en una siguiente fase poder evolucionarlo según sus necesidades y teniendo una idea más clara de lo que le puede aportar.
- El trabajo final de Máster propuesto en la asignatura ha sido muy interesante y de gran utilidad, ya que ha permitido adquirir conocimientos y experiencia en proyectos relacionados con la inteligencia de negocio, los cuales pueden generar mucho valor a las empresas mediante la exploración y el análisis de sus propios datos.
- Las herramientas Open Source ofrecen una gran variedad de posibilidades como es el caso de Pentaho, que ponen al alcance un sistema BI útil para conocer el impacto de la publicidad en pantallas, demostrando así que es una herramienta muy interesante para un proyecto como este con poco presupuesto.

4. Glosario

Datos: Es una representación simbólica o un atributo de una entidad.

Información: conjunto organizado de datos procesados, que constituyen un mensaje que cambia el estado de conocimiento del sujeto o sistema que recibe dicho mensaje.[11]

BI: Business Intelligence. Conjunto de técnicas y herramientas para transformar los datos e información de una organización en conocimiento con el cual tomar decisiones estratégicas.

ETL: Extraer, transformar y cargar. Proceso por el cual se extraen datos de una fuente de información que corresponde a una empresa, estos datos se transforman y luego se cargan para luego ser analizados.

OLAP: Online Analytical Processing. Permite extraer fácilmente y de manera selectiva datos y verlos desde distintos puntos de vista, estos datos son almacenados en una bases de datos multidimensional.

DWH: Data Warehouse. Almacén de datos, donde una empresa guarda su información.

Visual merchandising: conjunto de técnicas que se aplican en una tienda para poner los productos a disposición del consumidor obteniendo una rentabilidad, seduciéndolo a la compra.[12]

5. Bibliografía

- [1] S. Beca, “Agregando Valor: Desarrollo Ágil y Business Intelligence,” *Blog*, 2009. .
- [2] J. Diaz, “La Inteligencia de Negocio (Business Intelligence),” *Blog*, 2010. .
- [3] J. C. Díaz, *Introducción al Business Intelligence*. 2012.
- [4] A. Bustamante Martínez, E. A. Galvis Lista, and L. C. Gómez Flórez, “Técnicas de modelado de procesos de ETL: una revisión de alternativas y su aplicación en un proyecto de desarrollo de una solución de BI. (Spanish),” *ETL Process. Model. Tech. an Altern. Rev. its Appl. a BI Solut. Dev. Proj.*, vol. 18, pp. 185–191, 2013.
- [5] L. Muñoz, J. N. Mazón, J. Trujillo, L. Munoz, and J.-N. Mazon, “ETL Process Modeling Conceptual for Data Warehouses: A Systematic Mapping Study,” *IEEE Lat. Am. Trans.*, vol. 9, pp. 360–365, 2011.
- [6] B. C. P. IBARRA, “TAREAS 1-4.” slideshare, México, p. 15, 2010.
- [7] Wikipedia, “Pentaho,” *Pentaho*, 2016. [Online]. Available: <https://es.wikipedia.org/wiki/Pentaho>. [Accessed: 18-Oct-2017].
- [8] Rodryfuyo, “Qué es Pentaho y para que sirve?,” *Blog*, 2010. [Online]. Available: <https://rodryfuyol.wordpress.com/que-es-pentaho-y-para-que-sirve/>. [Accessed: 18-Oct-2017].
- [9] B. R. Dario, “.01 Introducción a Spoon,” *wikipedia*, 2010. [Online]. Available: <https://wiki.pentaho.com/pages/viewpage.action?pagelId=14844841>. [Accessed: 20-Nov-2017].
- [10] A. H. G. Company, “Mondrian Documentation,” *Mondrian Schema Workbench*, 2007. [Online]. Available: <https://mondrian.pentaho.com/documentation/workbench.php>.
- [11] WIKIPEDIA, “Información,” *WIKI*, 2017. .
- [12] G. Quijano, “¿Qué es el Visual Merchandising o Vitrinismo?,” *Artículo*, 2015. .

6. Anexos

Cubos OLAP realizados para cada tabla de hechos.

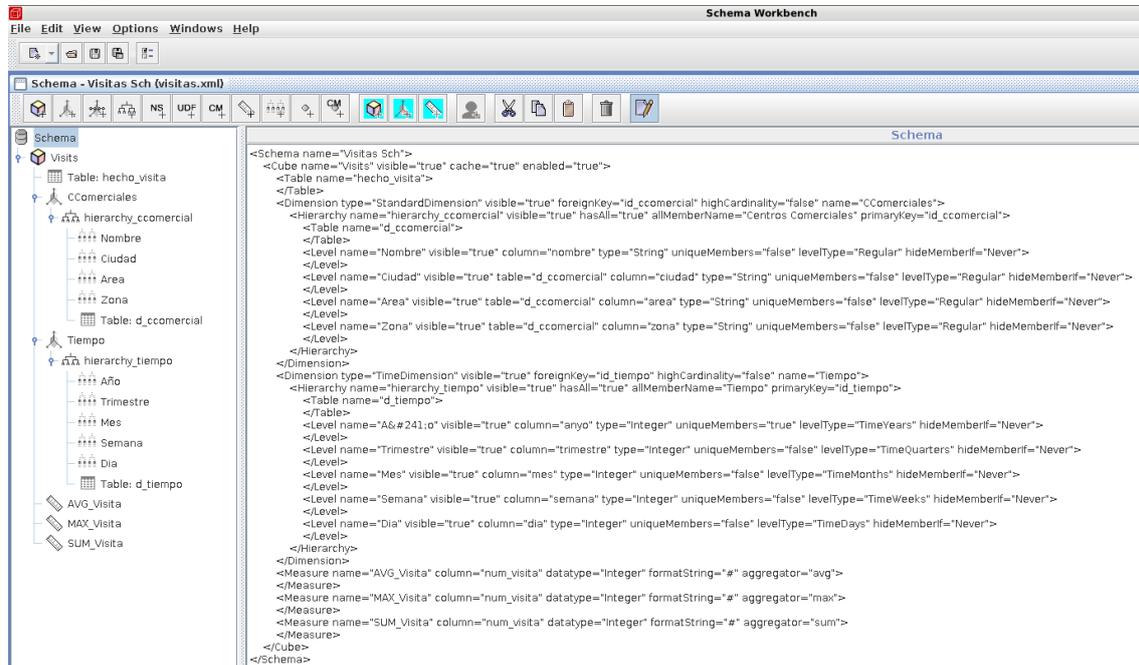


Ilustración 1: cubo de visitas

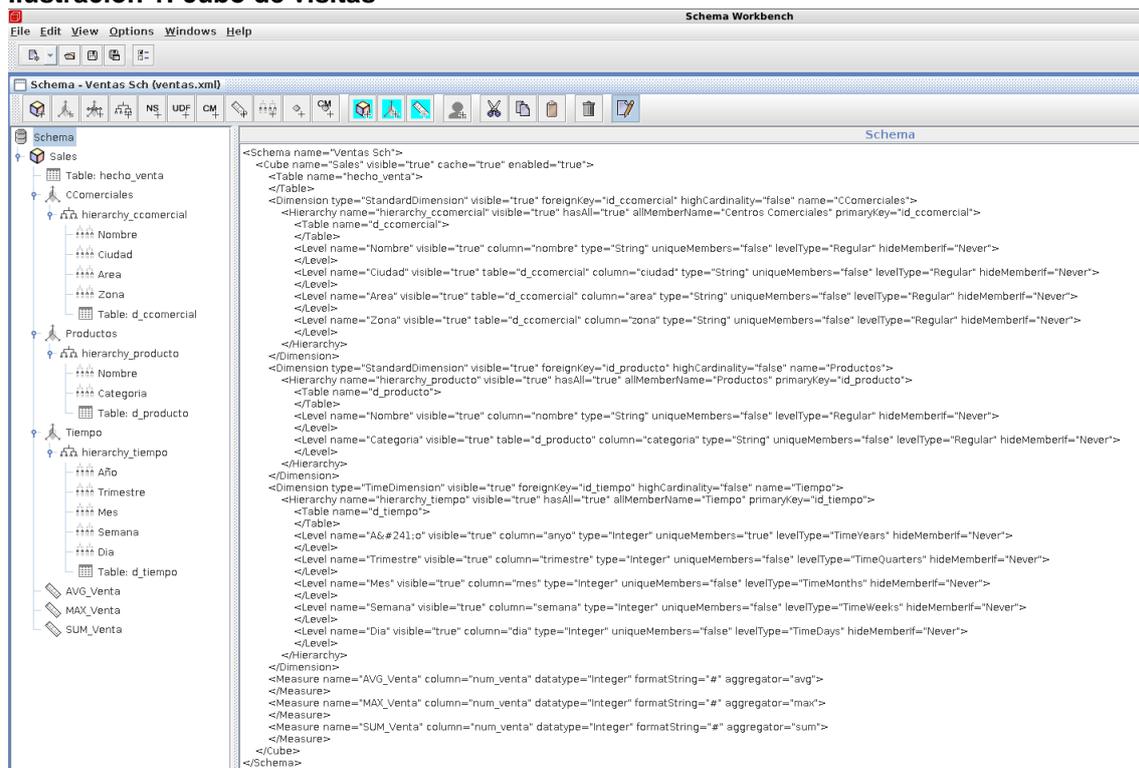


Ilustración 2: cubo de ventas

The screenshot displays the Schema Workbench interface for a cube named 'Impactos'. The left pane shows a hierarchical tree of the schema, including tables like 'hecho_impacto', 'd_ccomercial', 'd_producto', and 'd_tiempo', along with various hierarchies and measures. The right pane shows the corresponding XML schema definition.

```

<Schema name="Impactos Sch">
  <Cube name="Impactos" visible="true" cache="true" enabled="true">
    <Table name="hecho_impacto">
    </Table>
    <Dimension type="StandardDimension" visible="true" foreignKey="id_ccomercial" highCardinality="false" name="CComerciales">
      <Hierarchy name="hierarchy_ccomercial" visible="true" hasAll="true" allMemberName="Centros Comerciales" primaryKey="id_ccomercial">
        <Table name="d_ccomercial">
        </Table>
        <Level name="Nombre" visible="true" column="nombre" type="String" uniqueMembers="false" levelType="Regular" hideMemberif="Never">
        </Level>
        <Level name="Ciudad" visible="true" table="d_ccomercial" column="ciudad" type="String" uniqueMembers="false" levelType="Regular" hideMemberif="Never">
        </Level>
        <Level name="Area" visible="true" table="d_ccomercial" column="area" type="String" uniqueMembers="false" levelType="Regular" hideMemberif="Never">
        </Level>
        <Level name="Zona" visible="true" table="d_ccomercial" column="zona" type="String" uniqueMembers="false" levelType="Regular" hideMemberif="Never">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" foreignKey="id_producto" highCardinality="false" name="Productos">
      <Hierarchy name="hierarchy_producto" visible="true" hasAll="true" allMemberName="Productos" primaryKey="id_producto">
        <Table name="d_producto">
        </Table>
        <Level name="Nombre" visible="true" column="nombre" type="String" uniqueMembers="false" levelType="Regular" hideMemberif="Never">
        </Level>
        <Level name="Categoria" visible="true" table="d_producto" column="categoria" type="String" uniqueMembers="false" levelType="Regular" hideMemberif="Never">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="TimeDimension" visible="true" foreignKey="id_tiempo" highCardinality="false" name="Tiempo">
      <Hierarchy name="hierarchy_tiempo" visible="true" hasAll="true" allMemberName="Tiempo" primaryKey="id_tiempo">
        <Table name="d_tiempo">
        </Table>
        <Level name="A&#241;o" visible="true" column="anyo" type="Integer" uniqueMembers="true" levelType="TimeYears" hideMemberif="Never">
        </Level>
        <Level name="Trimestre" visible="true" column="trimestre" type="Integer" uniqueMembers="false" levelType="TimeQuarters" hideMemberif="Never">
        </Level>
        <Level name="Mes" visible="true" column="mes" type="Integer" uniqueMembers="false" levelType="TimeMonths" hideMemberif="Never">
        </Level>
        <Level name="Semana" visible="true" column="semana" type="Integer" uniqueMembers="false" levelType="TimeWeeks" hideMemberif="Never">
        </Level>
        <Level name="Dia" visible="true" column="dia" type="Integer" uniqueMembers="false" levelType="TimeDays" hideMemberif="Never">
        </Level>
      </Hierarchy>
    </Dimension>
    <Measure name="AVG_GPantalla" column="impacto_gpantalla" datatype="Integer" formatString="#" aggregator="avg">
    </Measure>
    <Measure name="AVG_Pinfo" column="impacto_pinfo" datatype="Integer" formatString="#" aggregator="avg">
    </Measure>
    <Measure name="MAX_GPantalla" column="impacto_gpantalla" datatype="Integer" formatString="#" aggregator="max">
    </Measure>
    <Measure name="MAX_Pinfo" column="impacto_pinfo" datatype="Integer" formatString="#" aggregator="max">
    </Measure>
    <Measure name="MIN_GPantalla" column="impacto_gpantalla" datatype="Integer" formatString="#" aggregator="min">
    </Measure>
    <Measure name="MIN_Pinfo" column="impacto_pinfo" datatype="Integer" formatString="#" aggregator="min">
    </Measure>
    <Measure name="SUM_GPantalla" column="impacto_gpantalla" datatype="Integer" formatString="#" aggregator="sum">
    </Measure>
  </Cube>
</Schema>

```

Ilustración 3: cubo de impactos