



Almacén de datos: Análisis de las ventas de una compañía

David Rodríguez Jurado
Máster U. Ing. Informática

Víctor Ruiz Marques

16/06/2015



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/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	Almacén de datos: Análisis de las ventas de una compañía
Nombre del autor:	David Rodríguez Jurado
Nombre del consultor:	Víctor Ruiz Marques
Fecha de entrega (mm/aaaa):	06/2015
Área del Trabajo Final:	Business Intelligence
Titulación:	Máster U. Ing. Informática
Resumen del Trabajo (máximo 250 palabras):	
<p>Creación de un entorno de <i>business intelligence</i> que permita analizar las ventas de la empresa ficticia TOTSALES. La empresa solicita que se realice un estudio de las ventas de sus cinco delegaciones, con el objetivo de entender y analizar su negocio.</p> <p>El proyecto se inicia realizando un estudio de las plataformas <i>open source</i> más utilizadas, para elegir la que mejor cubra las necesidades de TOTSALES. Una vez seleccionada, se continúa con las fases propias del desarrollo de este tipo de plataformas: diseño e implementación de un <i>Data Warehouse</i> y creación de los procesos ETL para cargar los datos proporcionados por la empresa.</p> <p>Finalmente, se diseñan y generan los informes solicitados, con el objetivo de obtener unos indicadores clave que proporcionen a TOTSALES una visión más analítica de su negocio.</p>	
Abstract (in English, 250 words or less):	
<p>Create a business intelligence environment in order to analyze the sales of TOTSALES company. A sales study of the five company's delegations is requested to achieve a better knowledge of its business.</p> <p>These project begins with a study of the most used <i>open source</i> platforms so as to identify which one meets TOTSALES's needs. Once the software has been chosen, it is time to define the required steps to design a business intelligence platform: the design and implementation of a <i>Data Warehouse</i> and the creation of the ETL processes to load all data given by the company.</p> <p>Finally, the requested reports are designed and generated in order to get the key indicators that will provide TOTSALES with a more analytic perspective of its business.</p>	
Palabras clave (entre 4 y 8):	
Business Intelligence, <i>Pentaho</i> , ETL, PostgreSQL, <i>Data Warehouse</i>	

Índice

1. INTRODUCCIÓN	1
1.1. CONTEXTO Y JUSTIFICACIÓN DEL TRABAJO	1
1.2. OBJETIVOS DEL TRABAJO	1
1.3. ENFOQUE Y MÉTODO SEGUIDO	1
1.4. PLANIFICACIÓN DEL TRABAJO	2
1.4.1. <i>Tabla de hitos</i>	2
1.5. BREVE SUMARIO DE PRODUCTOS OBTENIDOS	3
1.6. BREVE DESCRIPCIÓN DE LOS OTROS CAPÍTULOS DE LA MEMORIA	4
2. RESTO DE CAPÍTULOS	5
2.1. COMPARATIVA PLATAFORMAS BI	5
2.1.1. <i>Licenciamiento</i>	5
2.1.2. <i>Arquitectura</i>	5
2.1.3. <i>Funcionalidades</i>	7
2.1.4. <i>Usabilidad</i>	8
2.1.5. <i>Conclusiones</i>	8
2.2. DISEÑO E IMPLEMENTACIÓN DE UN <i>DATA WAREHOUSE</i>	9
2.2.1. <i>Análisis de los datos proporcionados</i>	9
2.2.2. <i>Modelo conceptual de datos</i>	10
2.2.3. <i>Modelo lógico de datos</i>	11
2.2.4. <i>Modelo físico de datos</i>	14
2.3. DISEÑO DE PROCESOS ETL	15
2.3.1. <i>Staging Area</i>	16
2.3.1.1. Transformación almacenes	16
2.3.1.2. Transformación artículo	17
2.3.1.3. Transformación clientes	17
2.3.1.4. Transformación población clientes	18
2.3.1.5. Transformación comerciales	19
2.3.1.6. Transformación conceptos	19
2.3.1.7. Transformación empresas	20
2.3.1.8. Transformación familias	20
2.3.1.9. Transformación referencia entre artículo y familia	21
2.3.1.10. Transformación formas de pago	22
2.3.1.11. Transformación tipo de dato	23
2.3.1.12. Transformación ventas	23
2.3.1.13. Tarea para ejecutar las transformaciones del Staging Area	24
2.3.2. <i>Carga del Data Warehouse</i>	25
2.3.2.1. Carga de la dimensión d_almacenes	25
2.3.2.2. Carga de la dimensión d_articulo	26
2.3.2.3. Carga de la dimensión d_cliente	27
2.3.2.4. Carga de la dimensión d_tipo_cliente	28
2.3.2.5. Carga de la dimensión d_cliente_ubicacion	29
2.3.2.6. Carga de la dimensión d_comercial	30
2.3.2.7. Carga de la dimensión d_concepto	31
2.3.2.8. Carga de la dimensión d_delegacion	32
2.3.2.9. Carga de la dimensión d_familia	33
2.3.2.10. Carga de la dimensión d_fecha	34

2.3.2.11. Carga de la dimensión d_forma_pago	35
2.3.2.12. Carga de la dimensión d_tipo_dato	36
2.3.2.13. Carga de la tabla de hechos h_ventas	37
2.3.2.14. Tarea para ejecutar las transformaciones esquema del <i>Data Warehouse</i>	39
2.4. EXPLOTACIÓN DE LOS DATOS	40
2.4.1. <i>Evolución de las ventas por delegación y año</i>	41
2.4.1.1. Estructura del informe	42
2.4.2. <i>Productos y familias de productos más vendidos</i>	43
2.4.2.1. Estructura del informe	56
2.4.3. <i>Top clientes</i>	57
2.4.3.1. Estructura del informe	58
2.4.4. <i>Distribución ventas según zona cliente</i>	59
2.4.4.1. Estructura del informe	61
2.4.5. <i>Evolución de las ventas en función del tiempo</i>	62
2.4.5.1. Estructura del informe	64
2.4.6. <i>Comisiones a liquidar a los comerciales por año y trimestre</i>	65
2.4.6.1. Estructura del informe	67
2.4.7. <i>Margen anual obtenido por familia de productos</i>	69
2.4.7.1. Estructura del informe	71
3. CONCLUSIONES	73
4. GLOSARIO	74
5. BIBLIOGRAFÍA	75
6. ANEXOS	76
6.1. MODELO FÍSICO DEL ESQUEMA INTERMEDIO	76

Lista de figuras

Ilustración 1 Arquitectura SpagoBI	6
Ilustración 2 Arquitectura <i>Pentaho</i>	6
Ilustración 3 Arquitectura Jaspersoft	7
Ilustración 4 Modelo conceptual <i>Data Warehouse</i>	11
Ilustración 5 Modelo lógico <i>Data Warehouse</i>	14
Ilustración 6 Modelo físico del <i>Data Warehouse</i>	15
Ilustración 7 Flujo de trabajo en la implementación del <i>Data Warehouse</i>	16
Ilustración 8 Transformación almacenes - intermedio	16
Ilustración 9 Transformación artículo - intermedio	17
Ilustración 10 Transformación clientes - intermedio	17
Ilustración 11 Transformación ubicación - intermedio	18
Ilustración 12 Transformación comerciales - intermedio	19
Ilustración 13 Transformación conceptos - intermedio	19
Ilustración 14 Transformación empresas - intermedio	20
Ilustración 15 Transformación familias - intermedio	20
Ilustración 16 Transformación familia por producto	22
Ilustración 17 Transformación formas de pago - intermedio	22
Ilustración 18 Transformación tipo de dato - intermedio	23
Ilustración 19 Transformación ventas - intermedio	23
Ilustración 20 Transformación Carga d_almacenes	26
Ilustración 21 Transformación Carga d_articulo	26
Ilustración 22 Transformación Carga d_cliente	27
Ilustración 23 Transformación Carga d_cliente_tipo	28
Ilustración 24 Transformación Carga d_cliente_ubicacion	29
Ilustración 25 Transformación Carga d_comercial	30
Ilustración 26 Transformación Carga d_concepto	31
Ilustración 27 Transformación Carga d_delegacion	32
Ilustración 28 Transformación Carga d_familia	33
Ilustración 29 Transformación Carga d_fecha	34
Ilustración 30 Transformación Carga d_forma_pago	36
Ilustración 31 Transformación Carga d_tipo_dato	36
Ilustración 32 Transformación Carga h_ventas	37
Ilustración 33 Data set evolución de las ventas por delegación	43
Ilustración 34 Estructura del informe evolución de las ventas por delegación y año	43
Ilustración 35 Data set evolución de las ventas por familia de productos	56
Ilustración 36 Data set artículos vendidos por familia y año	56
Ilustración 37 Estructura informe familias y artículos más vendidos	57
Ilustración 38 Data set top 10 clientes	58
Ilustración 39 Estructura informe top 10 clientes	59
Ilustración 40 Data set distribución de las ventas por provincia	61
Ilustración 41 Estructura del informe ventas distribuidas por provincia	62
Ilustración 42 Data set evolución de las ventas en el tiempo	64
Ilustración 43 Data set evolución ventas mensuales	64
Ilustración 44 Data set evolución ventas trimestral	64
Ilustración 45 Estructura informe evolución ventas temporal	65
Ilustración 46 Data set comisiones comercial y año	68
Ilustración 47 Data set porcentaje comisiones por año	68

Ilustración 48 Data set porcentaje trimestral por año	68
Ilustración 49 Estructura informe comisiones a liquidar a los comerciales por año y trimestre	69
Ilustración 50 Data set margen anual por familias de productos	71
Ilustración 51 Estructura informe margen anual por familia de productos	72

1. Introducción

1.1. Contexto y justificación del Trabajo

El trabajo final de máster consiste en el diseño de un sistema de BI que permita analizar la información de las ventas de la empresa TOTSALLES. La motivación de la empresa para implantar un sistema de este tipo se debe principalmente a:

- La percepción de los responsables de que las ventas están decreciendo año tras año.
- Que el coste de venta de sus productos ha subido durante los últimos años y afecta negativamente a sus márgenes de beneficio.

En los últimos años, los datos que generan las compañías han incrementado en complejidad y, principalmente, en tamaño. Esto se debe en gran medida al aumento de los productos que permiten generar datos, a la digitalización de activos y a los propios mercados de datos. Por lo tanto, es necesario disponer de herramientas que permitan analizar y tomar decisiones en base a ellos, siendo un sistema BI la mejor opción para realizarlo con éxito.

1.2. Objetivos del Trabajo

La competitividad de una empresa viene determinada por la calidad y la cantidad de información que sea capaz de utilizar en la toma de decisiones. Con la implementación de un sistema BI se proporcionan las herramientas necesarias para aprovechar los datos almacenados en ERP o bases de datos transaccionales, reduciendo el efecto negativo que puede tener una incorrecta determinación. Por ello, con la implantación de un sistema BI la empresa busca:

- Analizar su negocio actual para detectar posibles pérdidas en la cuota de mercado.
- Hacer un seguimiento en la evolución de los costes de venta y, en consecuencia, del margen de beneficio.

1.3. Enfoque y método seguido

Teniendo en cuenta los objetivos que tiene la empresa TOTSALLES y los indicadores que se quieren obtener, se analizan las 3 plataformas BI *open source* más populares para encontrar la que mejor se ajusta a sus necesidades.

Una vez decidida la plataforma que se utilizará, se continúa con las diferentes fases propias de la implantación de un sistema de BI:

- Diseño e implantación de un *Data Warehouse*. Una de las tareas más importantes cuando se implanta una solución BI es la de diseñar un *Data Warehouse* que permita manejar y controlar toda la información de una organización. Esto garantiza disponer en una vista única de todos los datos aunque provengan de diferentes fuentes. El *Data Warehouse* facilitará la comprensión de los datos y, en consecuencia, permitirá

disponer de información útil que servirá como apoyo en la toma de decisiones de la organización. Con la implementación del *Data Warehouse*, TOTSALLES no sólo comprenderá qué está pasando, sino que podrá predecir qué va a suceder.

- Creación de los procesos ETL. Partiendo de los datos proporcionados por la empresa en ficheros CSV se realizan las transformaciones necesarias para cargarlos en el *Data Warehouse* diseñado en la etapa anterior. Con estos procesos se pretende eliminar duplicidades innecesarias y evitar que queden conjuntos de datos que no tengan conexión entre ellos.
- Diseño y ejecución de informes. Con el objetivo de que la empresa pueda entender cómo está evolucionando su negocio, se generan una serie de informes con los que ésta puede analizar desde diferentes puntos de vista la evolución de las ventas y emprender las acciones correctivas oportunas.

1.4. Planificación del Trabajo

1.4.1. Tabla de hitos

En la siguiente tabla se presentan los hitos principales del proyecto, indicando los entregables de las diferentes PEC planificadas.

Nombre	Durada	Inicio	Final
PEC 1	12 días	26/02/2015	10/03/2015
Plan de trabajo			
Análisis de las 3 plataformas de BI			
PEC 2	41 días	11/03/2015	21/04/2015
Diseño e implementación de un DW			
PEC 3	23 días	22/04/2015	15/05/2015
Carga de datos mediante ETL			
PEC 4	31 días	16/05/2015	16/06/2015
Generación de informes e indicadores			
Entrega memoria del TFM			
Entrega Presentación			
Debate virtual	3 días	29/06/2015	01/07/2015

PEC 1 - Plan de trabajo

Elaboración del plan de trabajo que define los hitos a seguir durante la realización del TFM. En esta PEC también se entrega el análisis de 3 plataformas BI *open source* de las que se elige una para realizar el TFM.

PEC 2 - Diseño e implementación de un DW

Creación de un repositorio de datos que proporcione una visión global de los datos de TOTSALLES. Se utilizará una base de datos relacional pero desnormalizada para poder optimizar las consultas. Para implementar el *Data Warehouse* se utilizará MySQL.

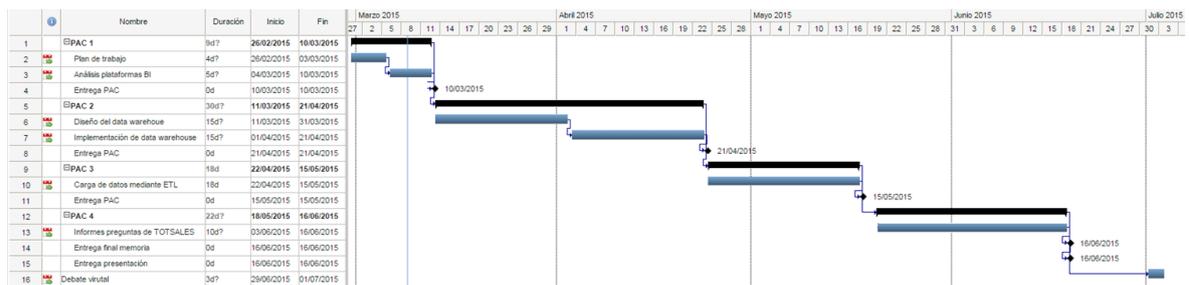
PEC 3 - Carga de datos mediante ETL

Definición de los procesos de extracción, transformación y carga que alimentarán al *Data Warehouse*. Partiendo de los datos semiestructurados (archivos csv) proporcionados por TOTSALLES éstos se transformarán según las necesidades definidas para poderlos cargar en el entorno destino. Dado que se ha seleccionado *Pentaho* como plataforma BI, esta tarea se realizará mediante el módulo *Pentaho Data Integration* (anteriormente Kettle).

PEC 4 - Generación de informes e indicadores

Generación de los informes solicitados para responder a las preguntas planteadas por TOTSALLES, utilizando el módulo Report Designer de *Pentaho*. Dado que es la última entrega también se incluye la memoria y la presentación del TFM.

Diagrama de Gantt



1.5. Breve resumen de productos obtenidos

Cada una de las tres fases en las que se estructura el proyecto genera una serie de productos que sirven de entrada a la siguiente fase:

- En el diseño e implementación del *Data Warehouse* se obtiene el diseño físico de la base de datos junto con los scripts para generar la base de datos en la que posteriormente se cargarán los datos proporcionados.
- En los procesos ETL se generan un conjunto de transformaciones que inicialmente cargan los datos de ventas de la empresa en el modelo intermedio y, a continuación, los transforman para cargarlos en el *Data Warehouse* diseñado.
- En la última fase se generan los seis informes solicitados, que muestran de forma resumida y mediante diagramas la información solicitada, de forma que la empresa pueda tomar las decisiones oportunas.

1.6. Breve descripción de los otros capítulos de la memoria

La elección de la plataforma de BI es una de las decisiones más importantes, ya que determina no sólo si quedan cubiertas las necesidades actuales, sino las que puedan surgir en el futuro, ya sea por un aumento en el volumen y la complejidad de los datos como por un cambio en las necesidades analíticas de la empresa. El apartado **Comparativa plataformas BI** pretende realizar un análisis de las plataformas más utilizadas desde diferentes puntos de vista. Se han comparado aspectos como los modelos de licenciamiento, la arquitectura o las funcionalidades incorporadas, analizando los pros y contras de cada una de ellas.

En la siguiente etapa se parte de un análisis detallado de los ficheros de datos proporcionados para así detectar las relaciones existentes entre ellos, que determinarán el diseño final. El apartado **Diseño e implementación de un Data Warehouse** parte de este análisis para generar el modelo conceptual del *Data Warehouse* que sirve de entrada para obtener el modelo lógico y el modelo físico final.

Las transformaciones necesarias para cargar los ficheros proporcionados se describen en el apartado **Diseño de procesos ETL**. Inicialmente se detallan las transformaciones utilizadas para cargar los datos en un *staging area*, que permite eliminar valores nulos o huérfanos y que facilitan su manipulación posterior.

El principal objetivo de disponer de una plataforma BI es la de poder explotar la información del *Data Warehouse* mediante una serie de informes que incorporan las métricas para poder evaluar el negocio. El apartado **Explotación de los datos** incluye la generación de los reports utilizando los datos de empresa, así como un detalle del formato y la estructura de los informes que se han generado utilizando *Pentaho Report Designer*.

2. Resto de capítulos

2.1. Comparativa plataformas BI

De entre las diferentes plataformas BI *open source* se han considerado JasperSoft, SpagoBI y *Pentaho* para realizar el análisis. El principal objetivo es conocer cuáles son sus limitaciones y funcionalidades para determinar cuál se adapta mejor a los requerimientos de TOTSALLES.

2.1.1. Licenciamiento

Hay que tener en cuenta los diferentes modelos de licenciamiento que tiene cada una de las plataformas, ya que puede implicar ciertas limitaciones u obligaciones cuando se le da un uso comercial. En este sentido, existen dos modelos de licenciamiento diferenciado:

- Versión comercial. Además de una versión gratuita, incorporan una *enterprise* que permite obtener una serie de servicios adicionales como soporte o formaciones:
 - *Jaspersoft* (<https://www.jaspersoft.com/es/node/77688>). La versión gratuita tiene licencia GNU Affero General Public License y diferentes opciones de pago en función de los servicios ofrecidos.
 - *Pentaho*. La versión gratuita tiene licencia GPL v2.
- Versión gratuita. Este modelo es el que utiliza *SpagoBI*, que incorpora todas las funcionalidades disponibles en la versión gratuita bajo licencia Mozilla Public License v. 2.0 (<http://www.spagobi.org/homepage/opensource/license/>)

2.1.2. Arquitectura

Las tres plataformas están desarrolladas en Java y se despliegan en un servidor JBoss, por lo que el acceso a los módulos de administración y de usuarios se realiza mediante un navegador web. A nivel de arquitectura, las tres plataformas utilizan el mismo modelo: el núcleo es una aplicación web que hace de *middleware* entre los motores BI, que implementan las diferentes funcionalidades disponibles en cada plataforma y el motor de base de datos que guarda toda la información. El usuario utiliza un navegador web que o bien se conecta directamente al servidor web o bien realiza el acceso mediante un portal.

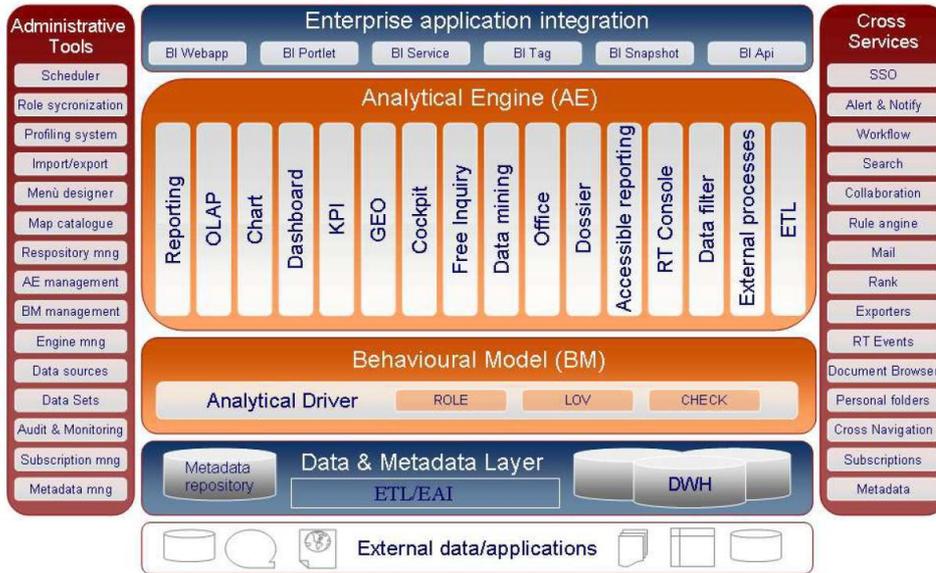


Ilustración 1 Arquitectura SpagoBI

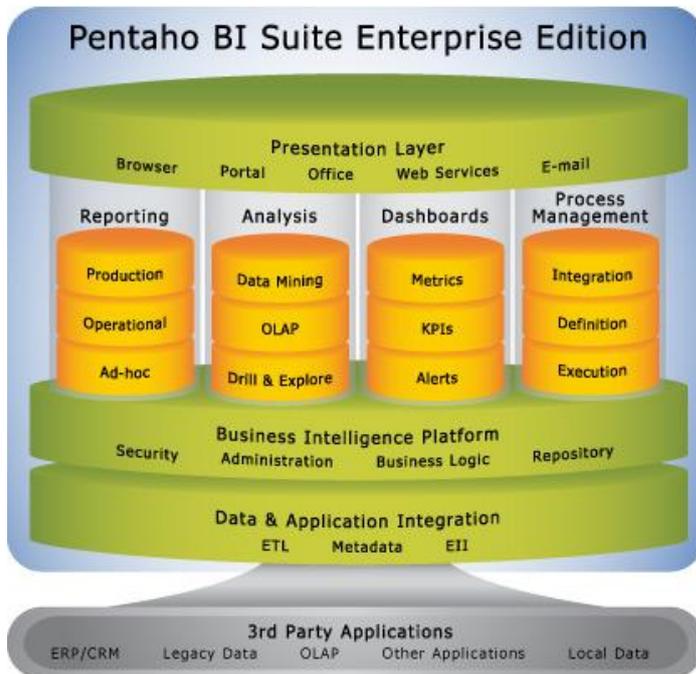


Ilustración 2 Arquitectura Pentaho

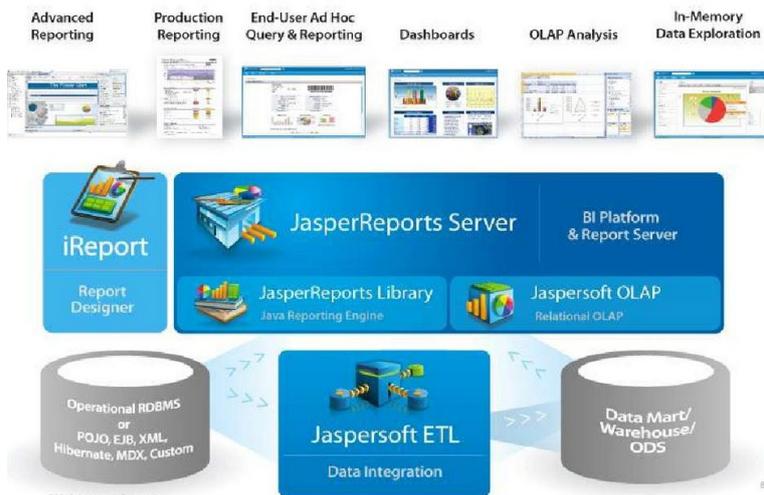


Ilustración 3 Arquitectura JasperSoft

2.1.3. Funcionalidades

Considerando las opciones gratuitas, *SpagoBI* supera a las otras dos plataformas ya que incorpora de serie funcionalidades avanzadas que son de pago en *Jaspersoft* y *Pentaho*. La siguiente tabla resume la disponibilidad de las funcionalidades más destacadas en las plataformas analizadas:

	Jaspersoft	Pentaho	SpagoBI
Informes	SÍ	SÍ	SÍ
Gráficos	SÍ	SÍ	SÍ
Dashboard	SÍ	SÍ	SÍ
OLAP	SÍ	SÍ	SÍ
ETL	SÍ	SÍ	SÍ
Data Mining	NO	SÍ	SÍ
KPI	NO	SÍ	SÍ
Data export	SÍ	SÍ	SÍ
GEO/GIS	SÍ	SÍ	SÍ

En muchos de los casos las plataformas comparten el mismo módulo para implementar una cierta funcionalidad o han evolucionado a partir de proyectos *open source*, como es el caso del ETL de JasperSoft, que surgió de Talend Open Studio. Otros se han desarrollado internamente, como el caso de JasperReport, y han pasado a utilizarse como herramienta para el diseño de informes en las tres plataformas analizadas.

A nivel de seguridad, todas las plataforma permiten acceso HTTPS y autenticación de usuarios, aunque la funcionalidad de gestión de los mismos y uso de perfiles es bastante diferente en cada una de ellas. Por ejemplo, SpagoBI incorpora una completa gestión de perfiles de usuario (http://wiki.spagobi.org/xwiki/bin/view/spagobi_server/Profile+Management), pero para el resto sólo están disponibles en la versión enterprise. *Pentaho* únicamente incorpora autenticación de usuario y JasperSoft una gestión muy básica de perfiles.

2.1.4. Usabilidad

En general, la usabilidad de las tres plataformas a nivel de usuario es buena, ya que se trata de aplicaciones web con un interfaz con el que resulta fácil familiarizarse. Aunque sus diseños no tienen el mismo nivel de detalle que soluciones comerciales, como por ejemplo Tableau (<https://www.tableau.com/products/online>), integran correctamente perfiles de usuario y administración en la misma interfaz gráfica, mostrando u ocultando ciertas opciones. A nivel de administración se pueden destacar los siguientes aspectos:

1. Facilidad en la instalación y en el proceso de configuración (*Jaspersoft* y *Pentaho* se instalan mediante un *wizard*, pero en SpagoBI hay que realizar ciertas parametrizaciones manualmente).
2. La creación de nuevos informes o análisis se realiza cómodamente mediante formularios. En el caso de *Pentaho* es un poco más complejo ya que se basa en el concepto de secuencia de acciones y hay menos información que permita solucionar eventuales problemas.
3. A nivel de documentación o formación, hay una comunidad activa en foros, wikis... que permiten solucionar posibles problemas.

2.1.5. Conclusiones

Después de analizar las diferentes opciones queda patente el elevado grado de madurez de los sistemas BI *open source*, que los convierte en una opción que las organizaciones deben tener en cuenta. En este sentido creo que cubren ampliamente su necesidades funcionales en el ámbito del BI y aportan una serie de características comunes que favorecen su adopción:

- Disponer de una versión completamente gratuita.
- Una sólida comunidad que da soporte y está implicada en el desarrollo del producto.
- Las opciones *open source* proporcionan unos servicios que tradicionalmente sólo estaban disponibles en opciones de pago.

En el caso de versión community de JasperSoft, ésta no incorpora la funcionalidad de Data Mining, lo que la convierte en una herramienta de reporting más que una solución BI completa. En este sentido, el diseñador de informes es bastante completo al permitir incorporar imágenes, gráficos, crosstab... Veo que hay una gran diferencia a nivel de funcionalidades en las diferentes versiones de pago, que lo convierten en una muy buena opción de plataforma BI. Una prueba de ello es la extensa lista de clientes (<https://www.jaspersoft.com/customers>) que lo utilizan como solución BI.

Sin lugar a dudas SpagoBI es la solución *open source* que incorpora más funcionalidades, lo que la convierte en una plataforma business intelligence completa. Incorpora funcionalidades que van más allá de lo tradicional en un sistema de este tipo (Data Mining o Business Process Model) y que ha hecho que grandes empresas como Fiat lo adopten como solución BI.

Como opción para implementar la solución BI de la empresa TOTSALES, se opta por *Pentaho*. Además de los informes solicitados, dispone de las funcionalidades de data mining necesarias

para realizar análisis predictivos. El hecho de que haya una extensa comunidad que se mantiene activa en los foros genera confianza ante posibles problemas que puedan surgir durante el desarrollo.

2.2. Diseño e implementación de un *Data Warehouse*

En los siguientes apartados se detallan las etapas necesarias para definir correctamente el *Data Warehouse*: un análisis de los datos proporcionados; el diseño del modelo conceptual de datos, que identifica vistas de negocio necesarias para dar respuesta a las preguntas de los usuarios finales; el diseño del modelo lógico de datos, que identifica las métricas necesarias y, finalmente, el modelo físico de datos, que materializa el diseño en un gestor de base de datos.

2.2.1. Análisis de los datos proporcionados

La empresa TOTSALLES ha proporcionado una serie de ficheros en formato csv que contienen información relativa a las ventas realizadas durante el periodo comprendido entre el 2009 y el 2013. En cada uno de los ficheros proporcionados se almacena la siguiente información:

- *Ventas.csv*. Contiene todas las transacciones realizadas por la empresa TOTSALLES. Cada registro tiene la información necesaria para relacionarlo con el resto de ficheros proporcionados (conceptos, delegación, comercial, artículo...). A destacar las últimas cuatro columnas que cuantifican el descuento, el importe, la comisión del comercial y el coste de cada una de las ventas.
- *Empresas.csv*. Identifica las 5 delegaciones de la empresa TOTSALLES. Se relaciona con el fichero *Ventas.csv* por el código de la empresa.
- *Clientes.csv*. Contiene los clientes (persona física) y cuál es su tipo de negocio (Pequeño comercio, Grandes superficies, Hoteles, Industria, Servicios, Distribuidores o Mayoristas). Se relaciona con el fichero *Ventas.csv* por el NIF del cliente.
- *ClientesZona.csv*. Lugar de residencia y provincia de los clientes. Se relaciona con el fichero *Ventas.csv* por el NIF del cliente.
- *Comerciales.csv*. Contiene los comerciales que efectúan las ventas. Se relaciona con el fichero *Ventas.csv* por el NIF del comercial.
- *Almacenes.csv*. Identifica los 6 almacenes de la empresa TOTSALLES. Se relaciona con el fichero *Ventas.csv* por el código del almacén.
- *Articulos.csv*. Contiene todos los artículos que vende la compañía. Se relaciona con el fichero *Ventas.csv* por el código del artículo.
- *Conceptos.csv*. Identifica a qué concepto se refiere la venta realizada (Portes, Embalajes...). Se relaciona con el fichero *Ventas.csv* por el código del concepto.
- *FormasPago.csv*. Identifica las 10 formas en las que se puede realizar el pago. Se relaciona con el fichero *Ventas.csv* por el código de la forma de pago.
- *FamiliasArticulo.csv*. Contiene todas las familias de productos correspondientes a los artículos vendidos. Se relaciona con el fichero *Ventas.csv* por el código del artículo.

- *TipoDato.csv*. Identifica si la venta es real, se trata de un presupuesto o de una previsión de venta. Se relaciona con el fichero *Ventas.csv* por el código del tipo de dato.
- *TipoLinea.csv*. Identifica si la venta hace referencia a un artículo o a un concepto, como puede ser un coste por embalaje o portes. Se relaciona con el fichero *Ventas.csv* por el código del tipo de línea.

2.2.2. Modelo conceptual de datos

El análisis de los datos proporcionados es el punto de partida para diseñar el modelo conceptual, que incluye los principales indicadores identificados. Permite tener una visión de alto nivel y, adicionalmente, determinar el alcance del sistema BI que se va a implantar. El principal objetivo en esta etapa es identificar la tabla de hechos y las dimensiones, que permitirán analizar los datos desde los diferentes puntos de vista solicitados por la empresa.

Como tabla de hechos podemos identificar las ventas (*h_ventas*), que se utiliza como indicador para medir el principal evento de negocio de la empresa TOTSALLES. Cada una de las ventas pueden analizarse desde diferentes puntos de vista, en función de las necesidades de la empresa, que determinan las siguientes dimensiones del proceso de negocio:

- Dimensión delegación (*d_delegacion*). Desde cuál de las 5 delegaciones de la empresa se ha realizado la venta.
- Dimensión artículo (*d_articulo*). Información del artículo que se ha vendido.
- Dimensión familia (*d_familia*). Familia de productos a la que pertenece el artículo vendido.
- Dimensión fecha (*d_fecha*). Momento en el que se ha realizado la venta.
- Dimensión cliente (*d_cliente*). Información del cliente al que se le ha vendido un artículo.
- Dimensión ubicación cliente (*d_cliente_ubicacion*). Población y provincia del cliente al que se le ha vendido un artículo.
- Dimensión tipo cliente (*d_cliente_tipo*). Clasificación de los clientes en función de si es un pequeño comercio, grandes superficies, hoteles, industria, servicios, distribuidores o mayoristas.
- Dimensión comercial (*d_comercial*). Información del comercial que ha efectuado la venta.
- Dimensión tipo de dato (*d_tipo_dato*). Información de si la venta es real, es un presupuesto o una previsión de venta.
- Dimensión concepto (*d_concepto*). Para las ventas que no son de artículos, identifica de qué tipo de concepto se trata.
- Dimensión forma de pago (*d_forma_pago*). Identificación de la forma de pago utilizada en la venta.
- Dimensión almacén (*d_almacen*). Desde cuál de los almacenes de la empresa se ha realizado la venta.

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

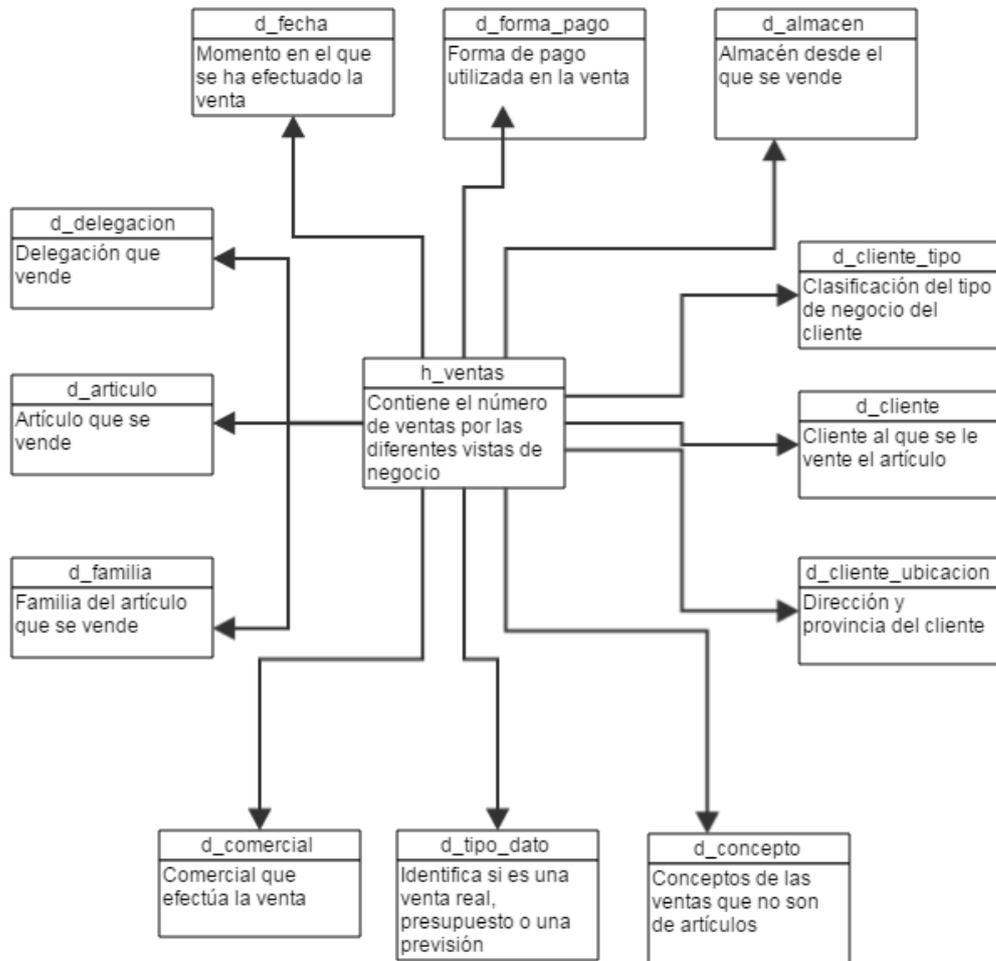


Ilustración 4 Modelo conceptual Data Warehouse

2.2.3. Modelo lógico de datos

Una vez realizado el modelo conceptual se detallan cada una de las dimensiones, enumerando cada uno de los atributos que forman estas dimensiones y construyendo el modelo lógico de datos.

La tabla de hechos h_ventas contiene la clave subrogada para identificar cada registro de forma unívoca, las diferentes claves foráneas para relacionar la venta con cada una de las dimensiones y métricas:

Columna	Descripción
id_venta	Clave primara que identifica cada una de las ventas.
id_delegacion	Clave foránea a la dimensión d_delegacion.

id_articulo	Clave foránea a la dimensión d_articulo.
id_familia	Clave foránea a la dimensión d_familia.
id_comercial	Clave foránea a la dimensión d_comercial.
id_fecha	Clave foránea a la dimensión d_fecha.
id_cliente	Clave foránea a la dimensión d_cliente.
id_ubicacion	Clave foránea a la dimensión d_ubicacion_cliente.
id_tipo	Clave foránea a la dimensión d_tipo_cliente.
id_tipo_dato	Clave foránea a la dimensión d_tipo_dato.
id_concepto	Clave foránea a la dimensión d_concepto.
id_pago	Clave foránea a la dimensión d_forma_pago.
id_almacen	Clave foránea a la dimensión d_almacen.
descuento	Métrica del descuento aplicado a una venta.
importe	Métrica del importe de una venta.
comision	Métrica de la comisión del vendedor.
coste	Métrica del coste de producción del artículo.

Los atributos de cada una de las dimensiones son:

Dimensión	Clave primaria	Atributos
d_delegacion	id_delegacion	desc_delegacion
d_articulo	id_articulo	desc_articulo
d_familia	id_familia	desc_familia
d_comercial	id_comercial	nombre_comercial
d_fecha	id_fecha	dia_semana desc_dia_semana semana_anyo desc_semana_anyo mes

		desc_mes trimestre desc_trimestre anyo
d_cliente	id_cliente	nombre_cliente
d_cliente_ubicacion	id_ubicacion	desc_pobl_cliente desc_prov_cliente
d_cliente_tipo	id_tipo	desc_cliente_tipo
d_tipo_dato	id_tipo_dato	desc_tipo_dato
d_concepto	id_concepto	desc_concepto
d_forma_pago	id_pago	desc_forma_pago
d_almacen	id_almacen	desc_almacen

Considerando la tabla de hechos y las dimensiones detalladas, se obtiene el siguiente esquema:

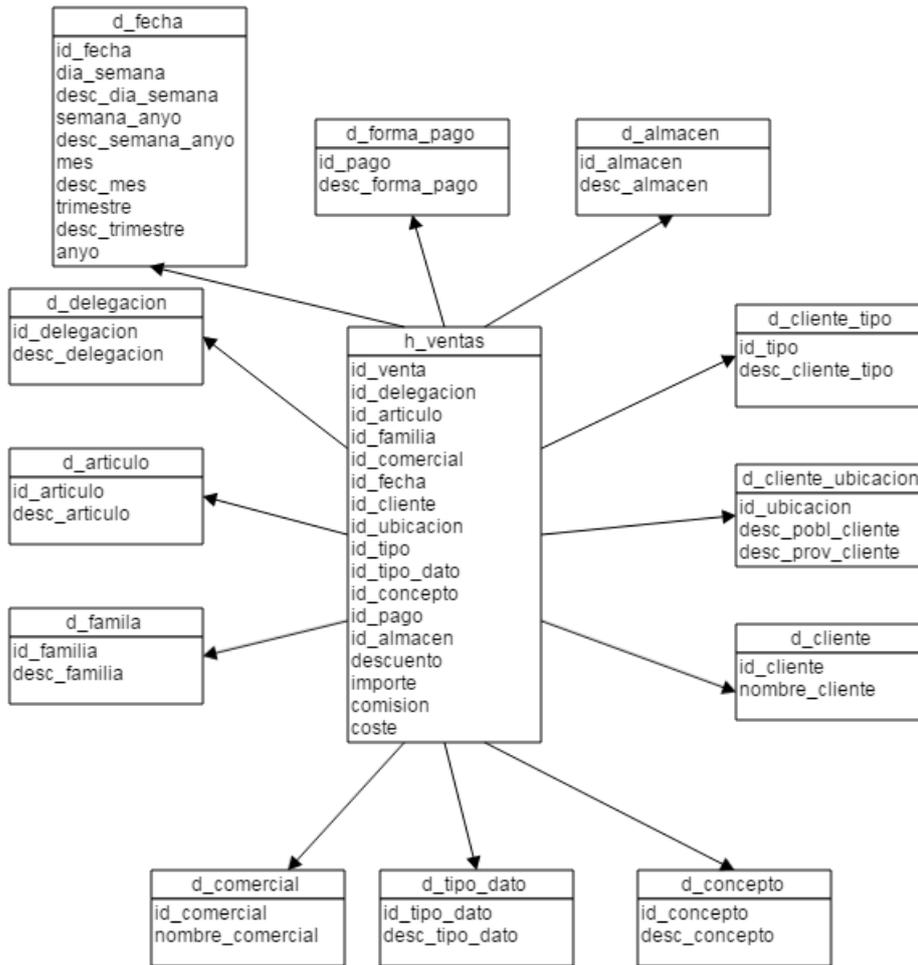


Ilustración 5 Modelo lógico *Data Warehouse*

2.2.4. Modelo físico de datos

Para el diseño físico se va a utilizar como base de datos MySQL Community Edition, que está bajo licencia GPL. Adicionalmente se utiliza MySQL Workbench 6.2 CE como herramienta visual de modelización de la base de datos MySQL.

Para diseñar el modelo físico de datos, se parte del modelo lógico obtenido y se define qué formato van a tener cada una de las claves y atributos:

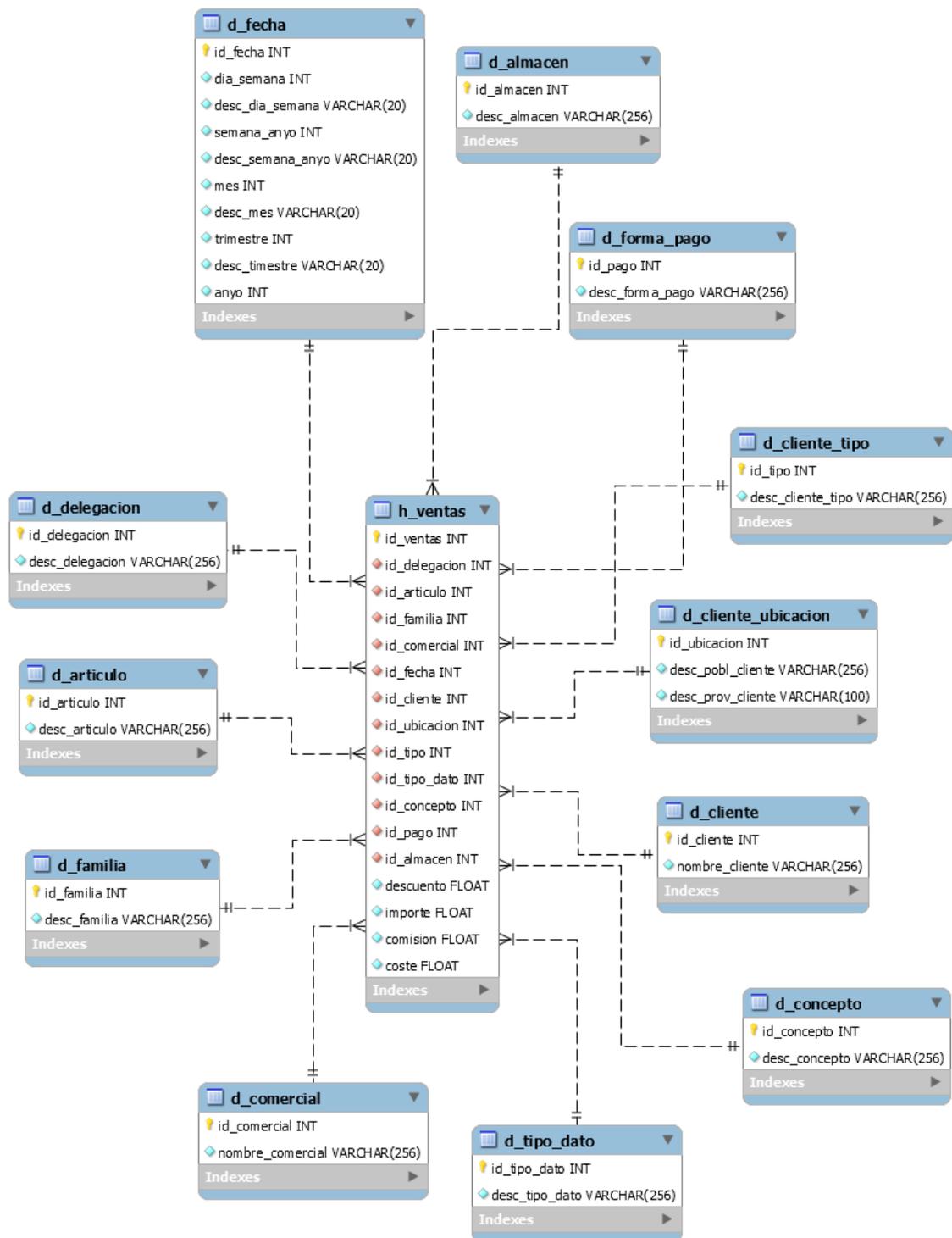


Ilustración 6 Modelo físico del *Data Warehouse*

2.3. Diseño de procesos ETL

El siguiente paso, una vez diseñado el *Data Warehouse*, es el de cargar en él los ficheros de datos de TOTSALLES y así poder generar los informes solicitados. En la mayoría de los casos, ni

el formato de la información ni el de la estructura de datos se adaptan al diseño del *Data Warehouse*, por lo que hay que realizar una serie de transformaciones para poderlos cargar.

Para facilitar estas transformaciones se utiliza un sistema *Staging Area*, que permite manipular cómodamente los datos de entrada de los ficheros csv en un sistema intermedio y así hacer las transformaciones necesarias para la posterior carga en el *Data Warehouse*:

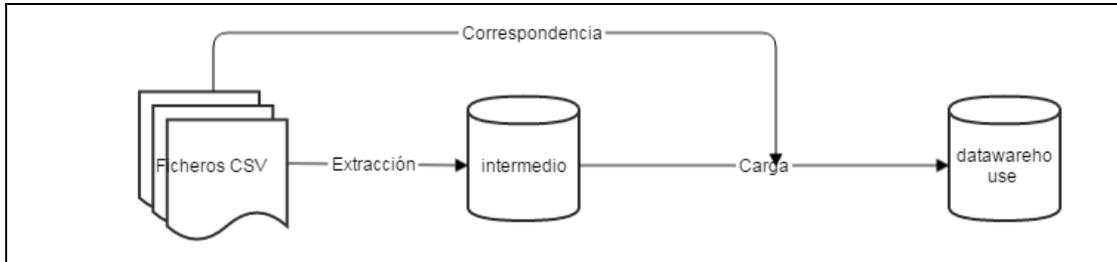


Ilustración 7 Flujo de trabajo en la implementación del *Data Warehouse*

Para desarrollar los procesos ETL se ha utilizado el módulo *Pentaho Data Integration*.

2.3.1. Staging Area

Para la manipulación de los datos antes de cargarlos, se ha creado el esquema *intermedio* en la base de datos MySQL. Para cada uno de los ficheros proporcionados se ha creado su correspondiente tabla, que además de las columnas propias de cada uno de los ficheros, se ha ampliado con la columna *idNuevo*, que se emplea para generar las correspondientes pk autoincrementales que se utilizarán en las tablas de dimensiones y hechos.

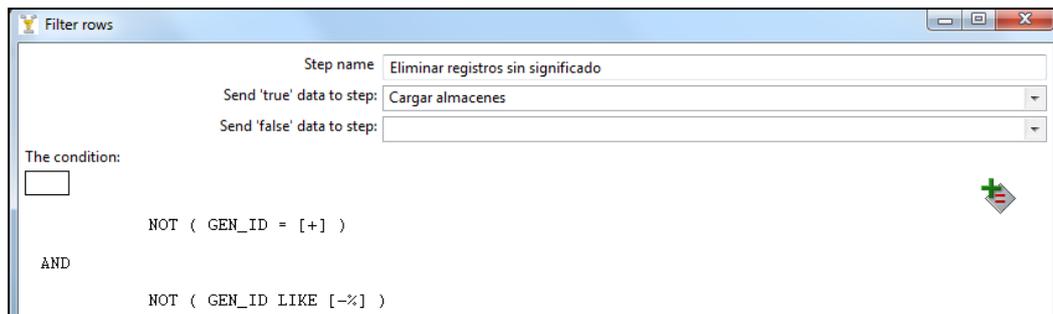
El planteamiento general ha sido la creación de una *transformation* para cada uno de los ficheros proporcionados y un *job* que las agrupa y ejecuta secuencialmente, y que sirve como proceso de carga del esquema intermedio. En los siguientes apartados se detallan cada una de las transformaciones realizadas.

2.3.1.1. Transformación almacenes



Ilustración 8 Transformación almacenes - intermedio

- **Leer CSV almacenes.** Paso para leer el fichero *Almacenes.csv*.
- **Eliminar registros sin significado.** Paso para filtrar del flujo de entrada los registros que corresponden a agrupadores y que no se corresponden a ningún almacén.



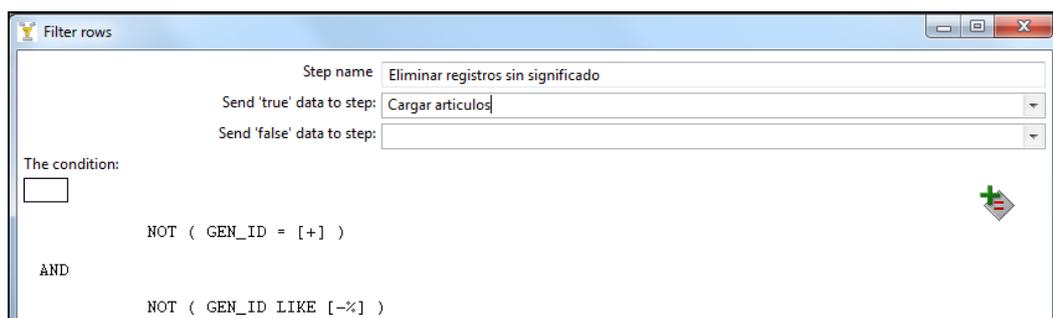
- **Cargar almacenes.** Paso para guardar en la tabla *almacen* del esquema *intermedio* los registros que se corresponden a almacenes reales.

2.3.1.2. Transformación artículo



Ilustración 9 Transformación artículo - intermedio

- **Leer CSV articulo.** Paso para leer el fichero *Articulos.csv*.
- **Eliminar registros sin significado.** Paso para filtrar del flujo de entrada los registros que corresponden a agrupadores y que no se corresponden a ningún artículo.



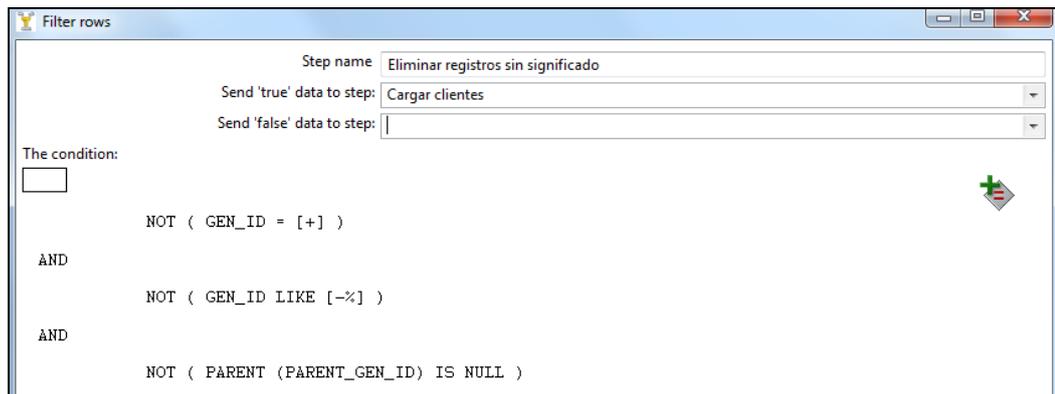
- **Cargar articulos.** Paso para guardar en la tabla *articulos* del esquema *intermedio* los registros que se corresponden a artículos reales.

2.3.1.3. Transformación clientes



Ilustración 10 Transformación clientes - intermedio

- **Leer CSV clientes.** Paso para leer el fichero *Cientes.csv*.
- **Eliminar registros sin significado.** Paso para filtrar del flujo de entrada los registros que corresponden a agrupadores y que no se corresponden a ningún cliente.



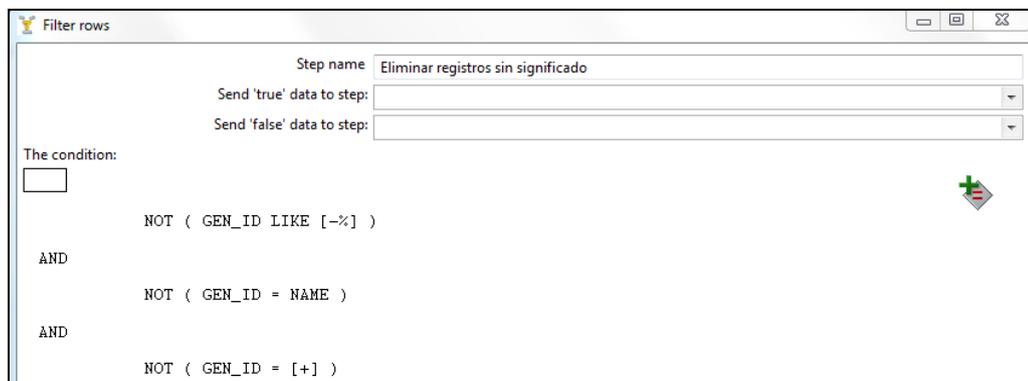
- **Cargar clientes.** Paso para guardar en la tabla *clientes* del esquema *intermedio* los registros que se corresponden a clientes reales.

2.3.1.4. Transformación población clientes



Ilustración 11 Transformación ubicación - intermedio

- **Leer CSV clientes zona.** Paso para leer el fichero *CientesZona.csv*.
- **Eliminar registros sin significado.** Paso para filtrar del flujo de entrada los registros que corresponden a agrupadores y que no se corresponden a ninguna población de un cliente.



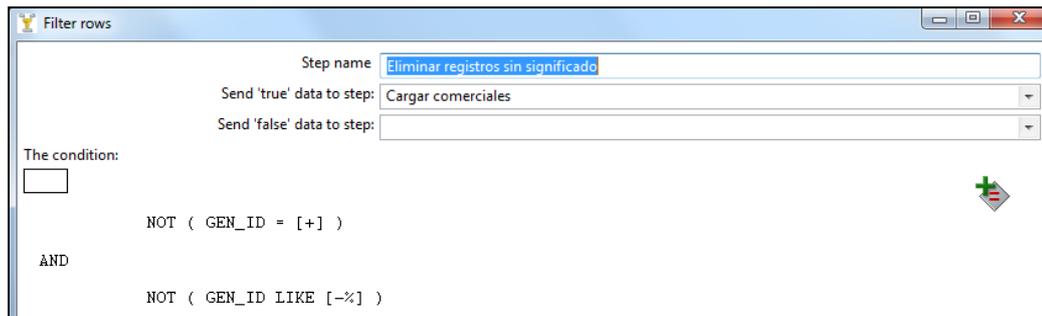
- **Cargar clientes zona población.** Paso para guardar en la tabla *clientes_zona_poblacion* del esquema *intermedio* los registros que se corresponden a poblaciones de clientes.

2.3.1.5. Transformación comerciales



Ilustración 12 Transformación comerciales - intermedio

- **Leer CSV comerciales.** Paso para leer el fichero *Comerciales.csv*.
- **Eliminar registros sin significado.** Paso para filtrar del flujo de entrada los registros que corresponden a agrupadores y que no se corresponden a ningún comercial.



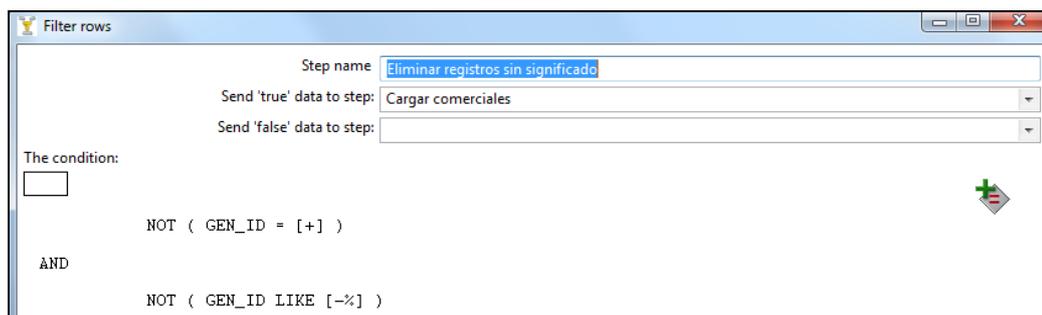
- **Cargar comerciales.** Paso para guardar en la tabla *comerciales* del esquema *intermedio* los registros que se corresponden a comerciales.

2.3.1.6. Transformación conceptos



Ilustración 13 Transformación conceptos - intermedio

- **Leer CSV conceptos.** Paso para leer el fichero *Conceptos.csv*.
- **Eliminar registros sin significado.** Paso para filtrar del flujo de entrada los registros que corresponden a agrupadores y que no se corresponden a ningún concepto.



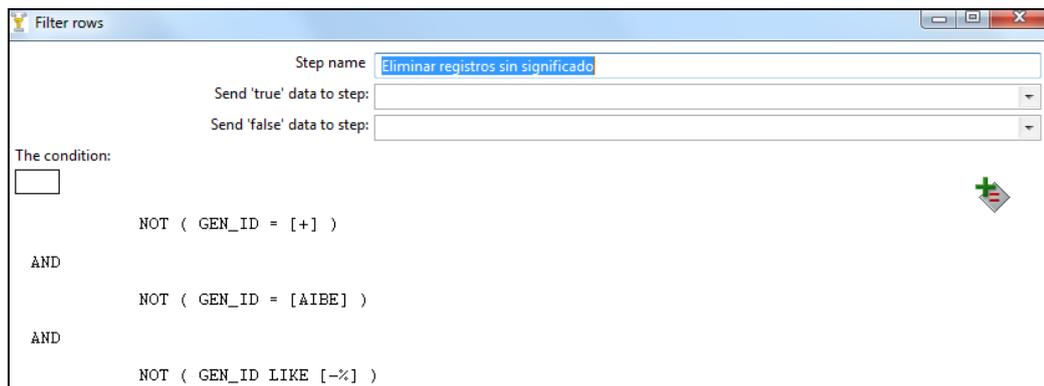
- **Cargar concepto.** Paso para guardar en la tabla *conceptos* del esquema *intermedio* los registros que se corresponden a conceptos facturados.

2.3.1.7. Transformación empresas



Ilustración 14 Transformación empresas - intermedio

- **Leer CSV empresas.** Paso para leer el fichero *Empresas.csv*.
- **Eliminar registros sin significado.** Paso para filtrar del flujo de entrada los registros que corresponden a agrupadores y que no se corresponden a ninguna delegación.



- **Cargar empresas.** Paso para guardar en la tabla *empresas* del esquema *intermedio* los registros que se corresponden a delegaciones de la empresa.

2.3.1.8. Transformación familias

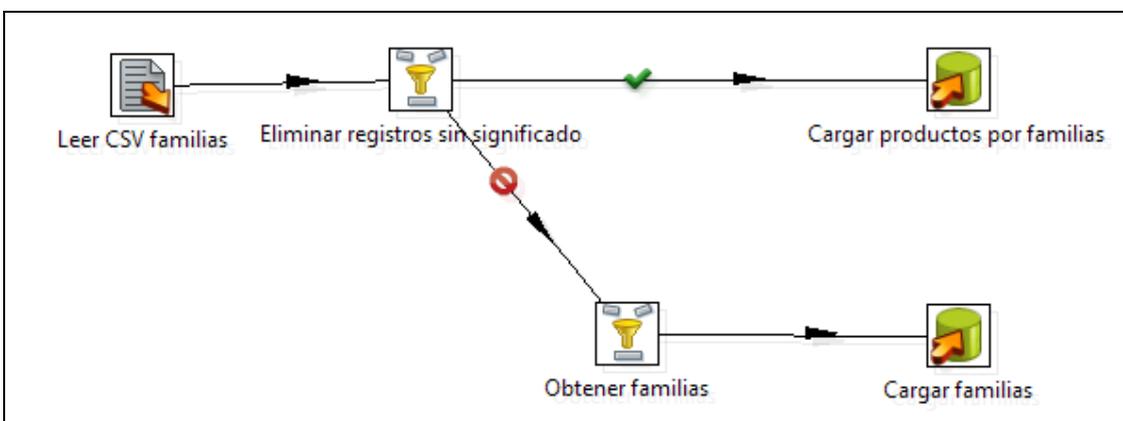
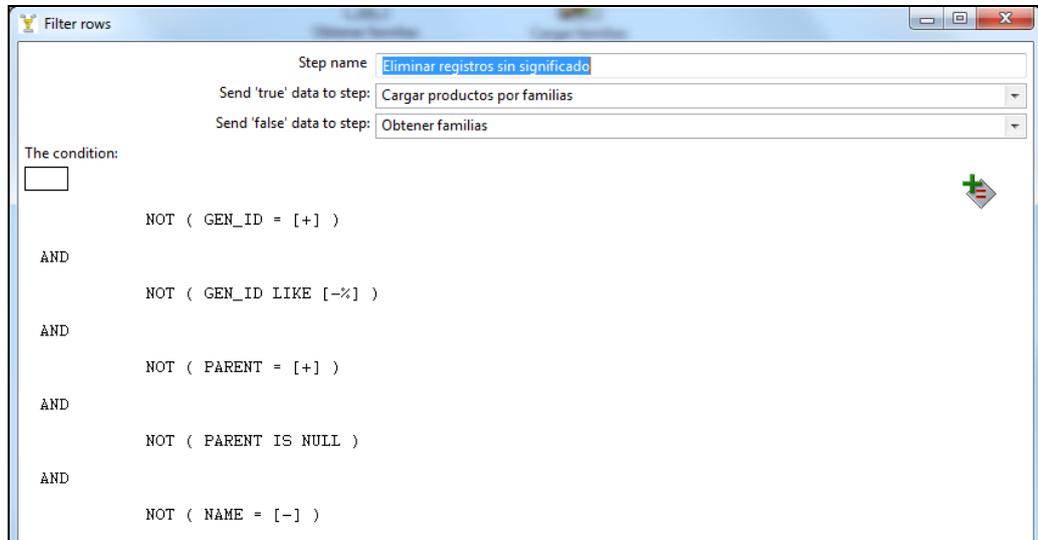


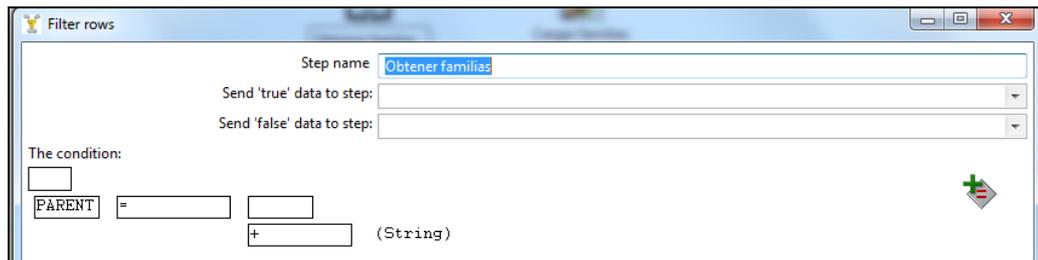
Ilustración 15 Transformación familias - intermedio

- **Leer CSV familias.** Paso para leer el fichero *FamiliasArticulos.csv*.

- **Eliminar registros sin significado.** El objetivo de este filtro es eliminar del flujo de entrada los registros correspondientes a agrupadores y los que no son familias de productos. Los registros que pasan el filtro se cargan directamente como artículos junto a la familia a la que pertenecen. Los que no lo pasan se procesan en un nuevo filtro que obtiene familias finales.



- **Obtener familias.** Filtro para obtener del fichero de entrada los registros que corresponden realmente a familias de artículos.



- **Cargar productos por familias.** Paso para guardar en la tabla *familia_producto* del esquema *intermedio* los registros que se corresponden a artículos junto a la familia a la que pertenecen.
- **Cargar familias.** Paso para guardar en la tabla *familia* del esquema *intermedio* los registros que se corresponden a las familias de artículos.

2.3.1.9. Transformación referencia entre artículo y familia

Una vez cargados los artículos y las familias hay que establecer la referencia entre la clave primaria de las familias a cada uno de los artículos:

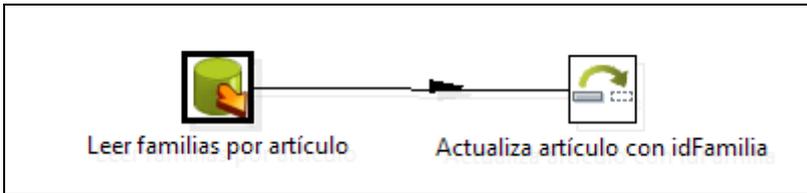
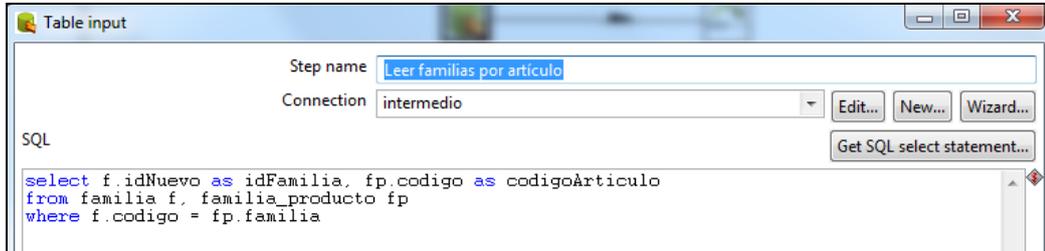


Ilustración 16 Transformación familia por producto

- **Leer familias por artículo.** Paso para leer del esquema intermedio los datos necesarios para vincular a cada artículo su identificador de familia.



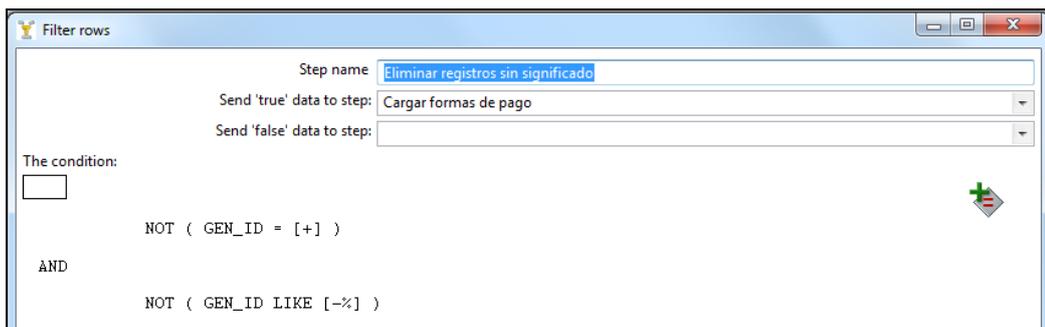
- **Actualiza artículo con idFamilia.** Paso para actualizar la referencia de cada artículo a la familia correspondiente del esquema *intermedio*.

2.3.1.10. Transformación formas de pago



Ilustración 17 Transformación formas de pago - intermedio

- **Leer CSV formas de pago.** Paso para leer el fichero *FormasPago.csv*.
- **Eliminar registros sin significado.** Paso para filtrar del flujo de entrada los registros que corresponden a agrupadores y que no se corresponden a ninguna forma de pago.



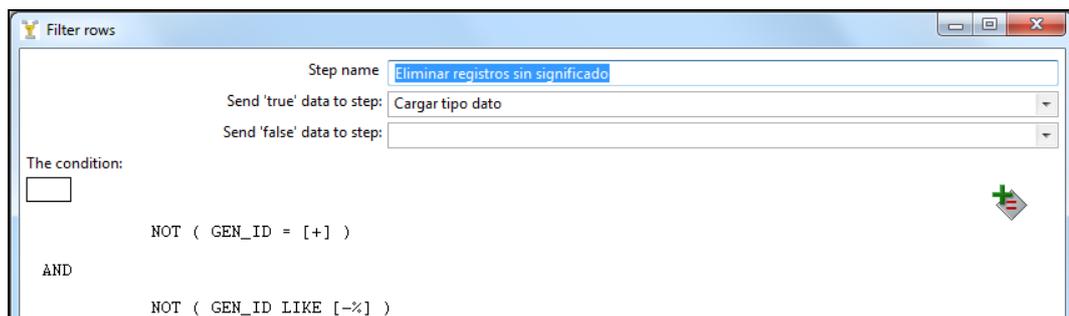
- **Cargar formas de pago.** Paso para guardar en la tabla *formas_pago* del esquema *intermedio* los registros que se corresponden a formas de pago reales.

2.3.1.11. Transformación tipo de dato



Ilustración 18 Transformación tipo de dato - intermedio

- **Leer CSV tipo dato.** Paso para leer el fichero *TipoDato.csv*.
- **Eliminar registros sin significado.** Paso para filtrar del flujo de entrada los registros que corresponden a agrupadores y que no se corresponden a ningún tipo de dato.



- **Cargar tipo dato.** Paso para guardar en la tabla *tipo_dato* del esquema *intermedio* los registros que se corresponden a tipos de dato reales.

2.3.1.12. Transformación ventas

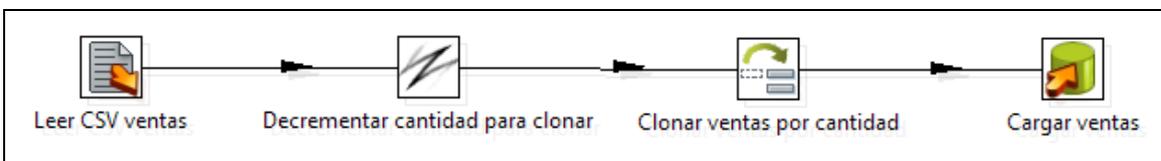
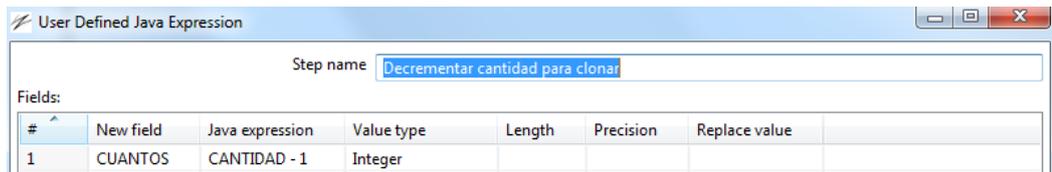
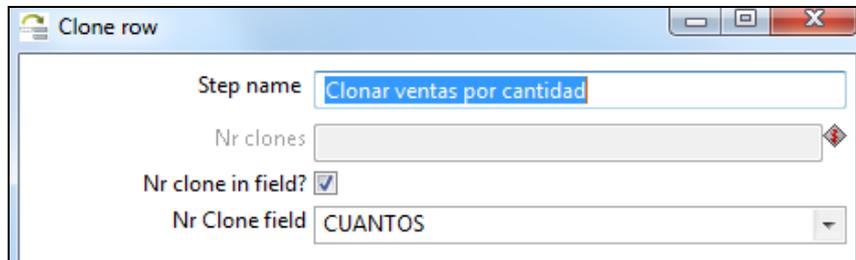


Ilustración 19 Transformación ventas - intermedio

- **Leer CSV ventas.** Paso para leer el fichero *Ventas.csv*.
- **Decrementar cantidad para clonar.** En el fichero de ventas, cada uno de los registros corresponde a la venta de una cantidad de artículos. Para tener un registro para cada uno de los artículos vendidos será necesario clonarlos según indica en el campo CANTIDAD del fichero *Ventas.csv*. Primero se crea un paso con una expresión java para decrementar el valor de clonación, dado que si se clonan según la cantidad indicada en el campo CANTIDAD se crearían registros de más. Por ejemplo, si el valor fuese 1, realmente se crearían 2.



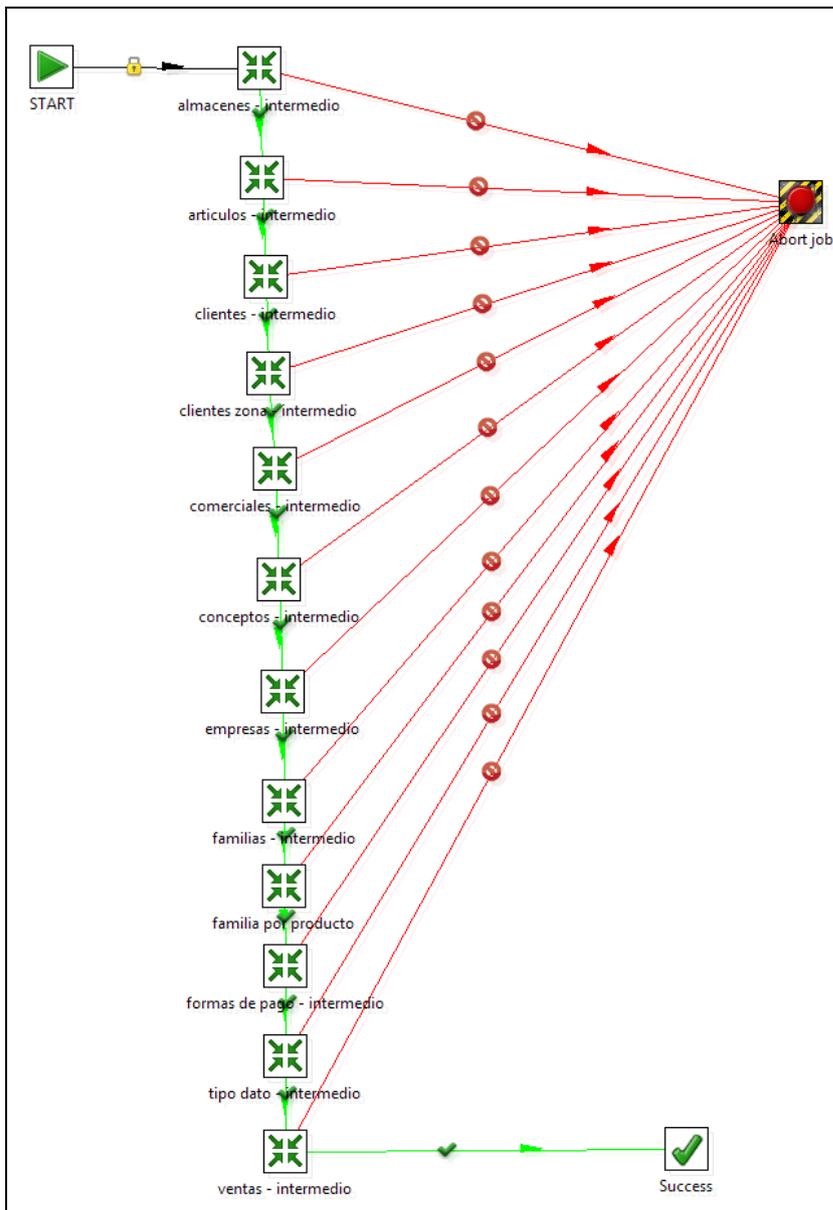
- **Clonar ventas por cantidad.** Clona los registros de ventas según se ha indicado en el paso anterior.



- **Cargar ventas.** Paso para guardar en la tabla *ventas* del esquema *intermedio* los registros que se corresponden a cada una de las ventas realizadas.

2.3.1.13. Tarea para ejecutar las transformaciones del Staging Area

Para ejecutar todas las transformaciones del esquema *intermedio*, se crea un *job* que las ejecuta secuencialmente.



2.3.2. Carga del *Data Warehouse*

Una vez hechas las transformaciones necesarias, en el esquema *intermedio* se encuentran los datos de TOTSALAS para cargar las diferentes dimensiones y tablas de hechos del *Data Warehouse*. El planteamiento ha sido realizar *transformations* que leen los registros de las tablas del esquema *intermedio* y los cargan en cada una de las dimensiones del *Data Warehouse*.

2.3.2.1. Carga de la dimensión *d_almacenes*

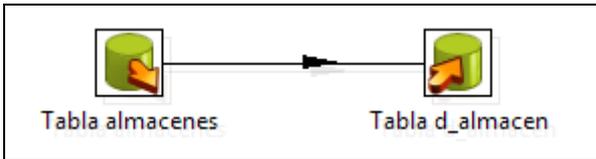
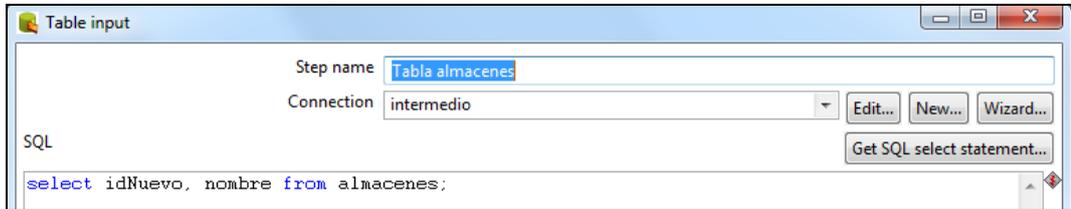
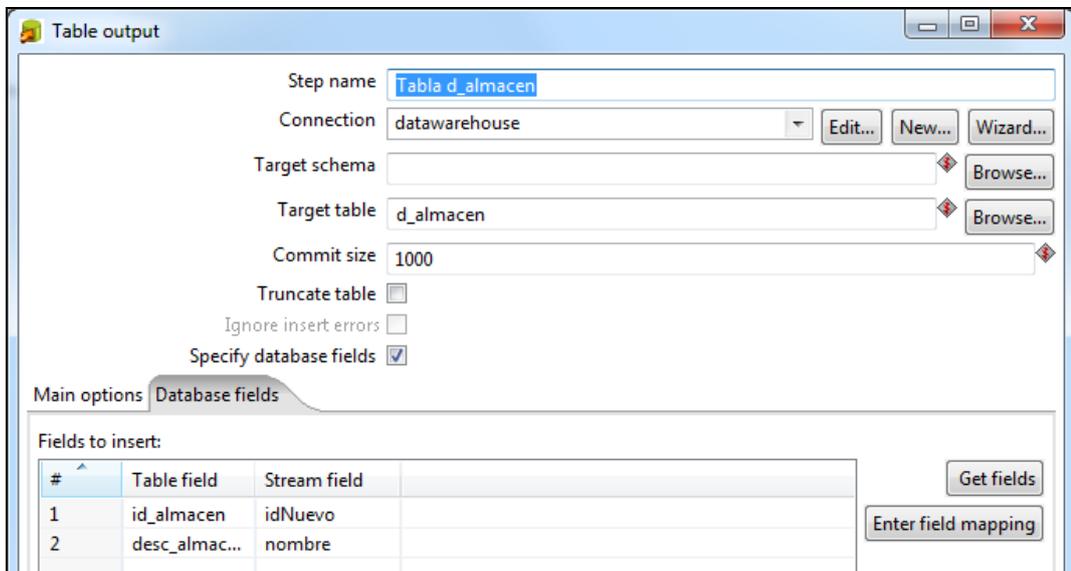


Ilustración 20 Transformación Carga d_almacenes

- **Tabla almacenes.** Ejecuta la SQL para leer los datos de almacenes del esquema *intermedio*.



- **Tabla d_almacen.** Carga los almacenes del flujo de entrada en la tabla *d_almacen*.



2.3.2.2.Carga de la dimensión d_articulo

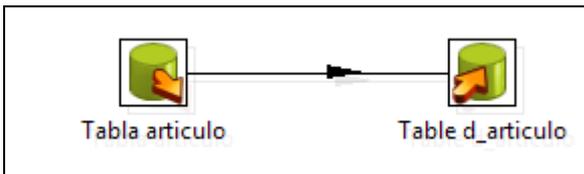
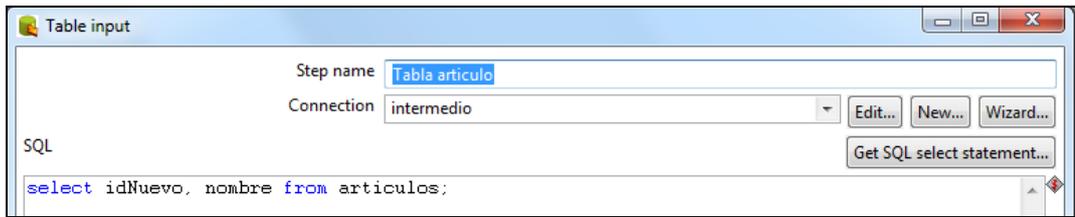
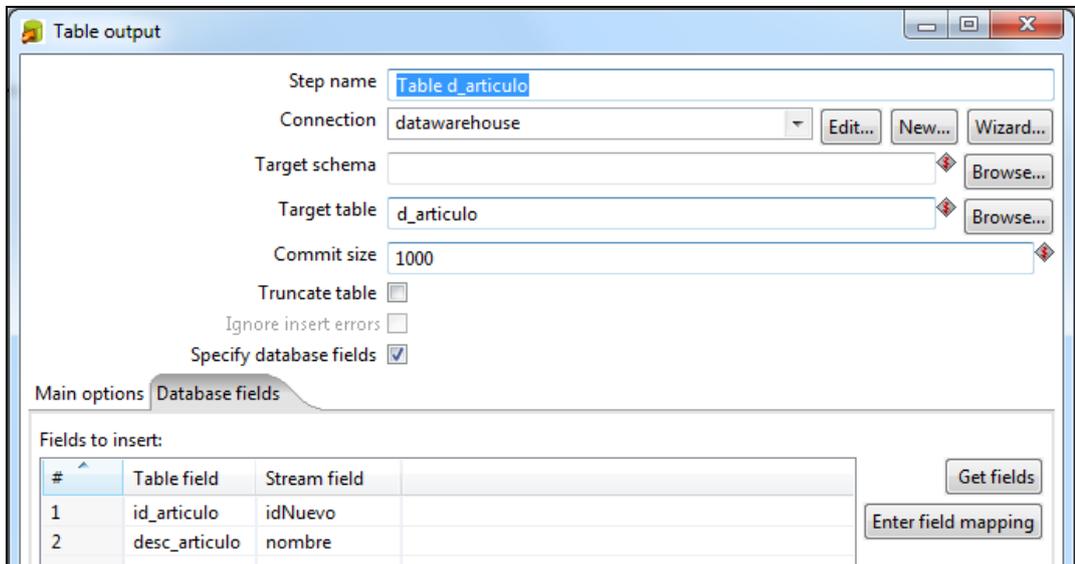


Ilustración 21 Transformación Carga d_articulo

- **Tabla articulo.** Ejecuta la SQL para leer los datos de artículos del esquema *intermedio*.



- **Tabla d_articulo.** Carga los artículos del flujo de entrada en la tabla d_articulo.



2.3.2.3.Carga de la dimensión d_cliente

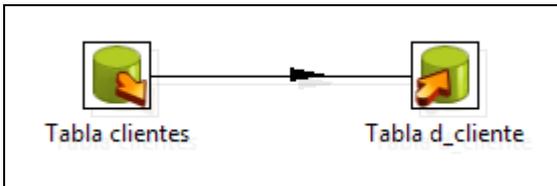
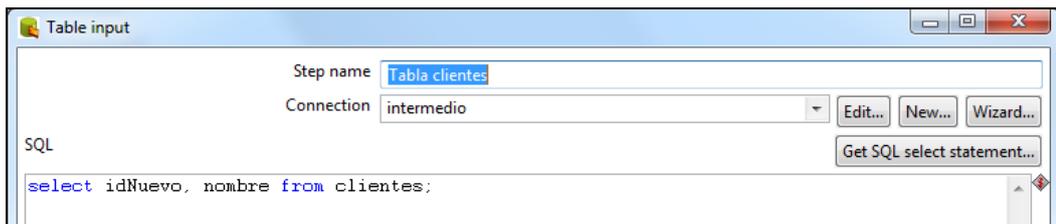
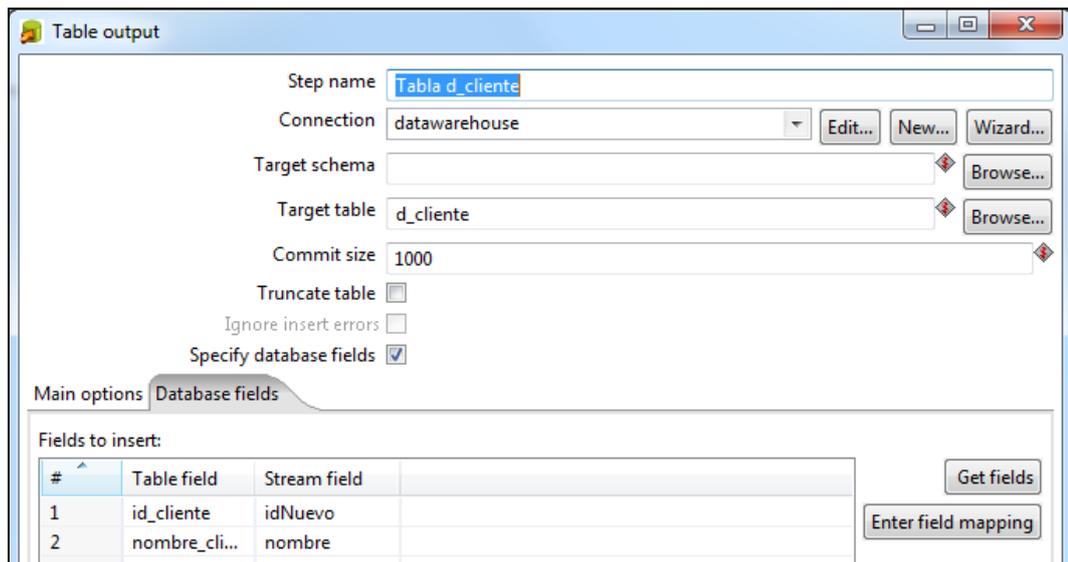


Ilustración 22 Transformación Carga d_cliente

- **Tabla clientes.** Ejecuta la SQL para leer los datos de clientes del esquema *intermedio*.



- **Tabla d_cliente.** Carga los clientes del flujo de entrada en la tabla d_cliente.



2.3.2.4. Carga de la dimensión d_tipo_cliente

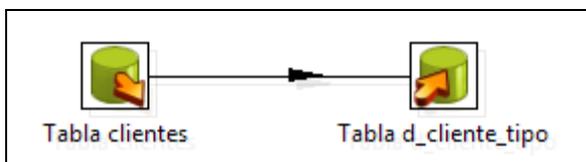
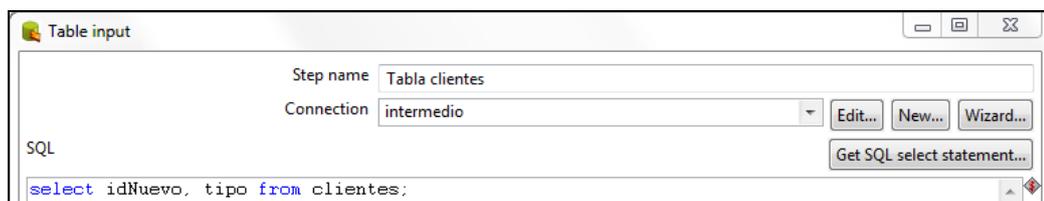
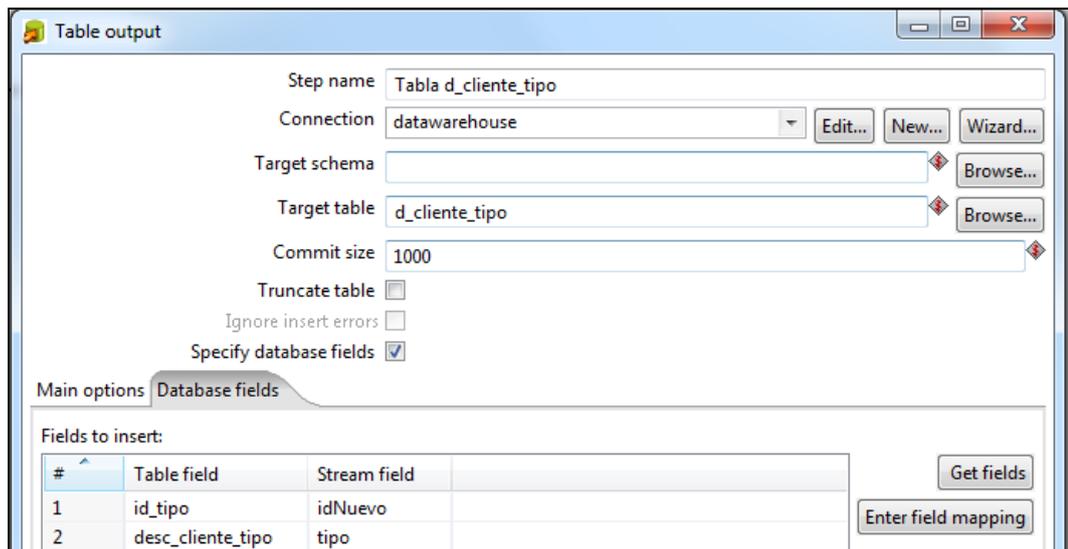


Ilustración 23 Transformación Carga d_cliente_tipo

- **Tabla clientes.** Ejecuta la SQL para leer los datos de tipos de cliente del esquema *intermedio*.



- **Tabla d_cliente_tipo.** Carga los tipos de cliente del flujo de entrada en la tabla *d_cliente_tipo*.



2.3.2.5.Carga de la dimensión d_cliente_ubicacion

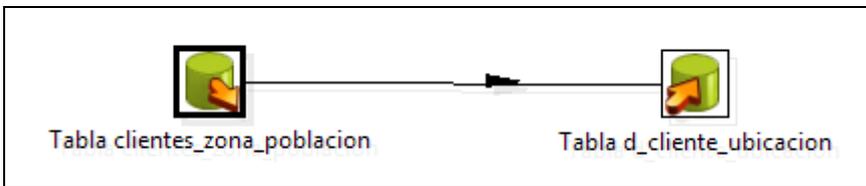
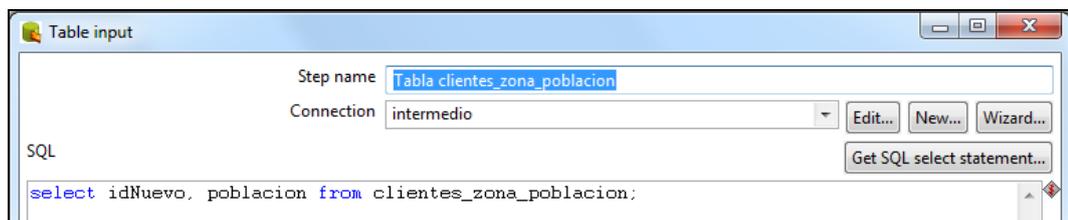
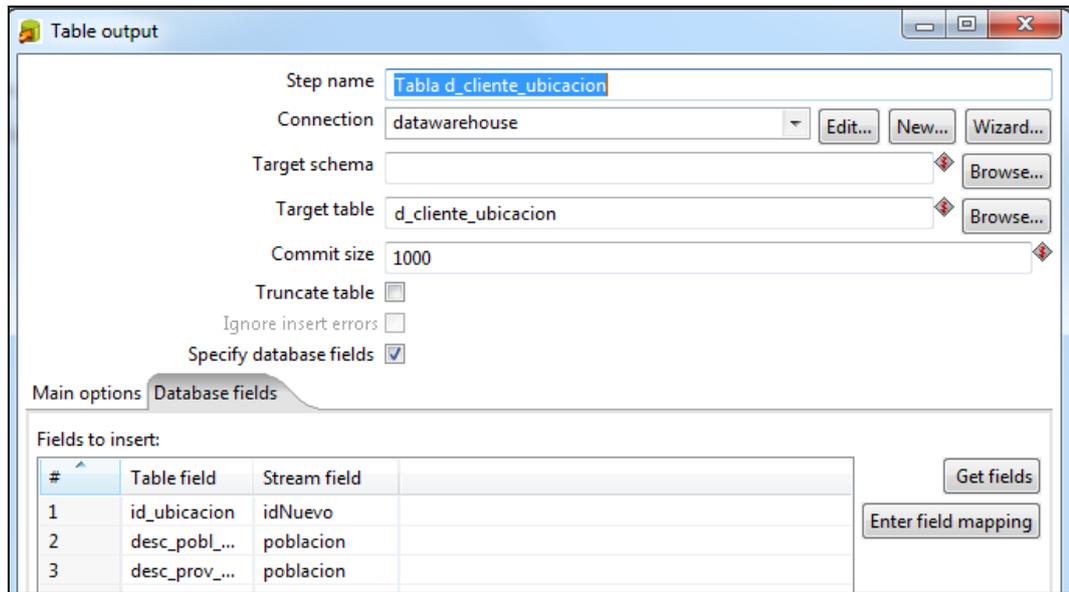


Ilustración 24 Transformación Carga d_cliente_ubicacion

- **Tabla clientes_zona_poblacion.** Ejecuta la SQL para leer la población de los clientes del esquema *intermedio*.



- **Tabla d_cliente_ubicacion.** Carga la población de los clientes del flujo de entrada en la tabla *d_cliente_ubicacion*.



2.3.2.6.Carga de la dimensión d_comercial

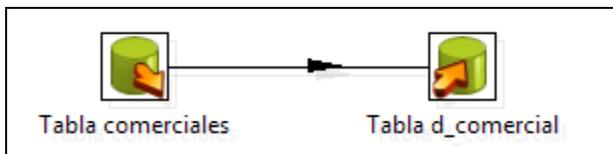
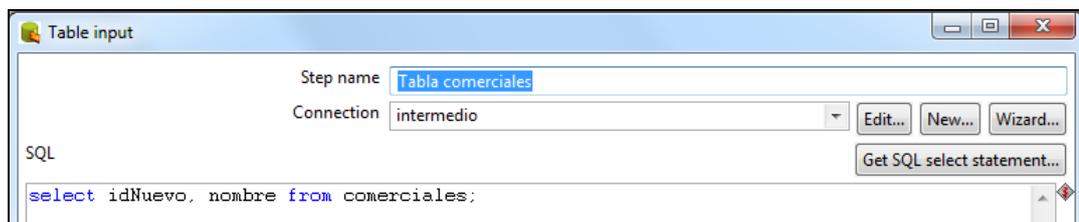
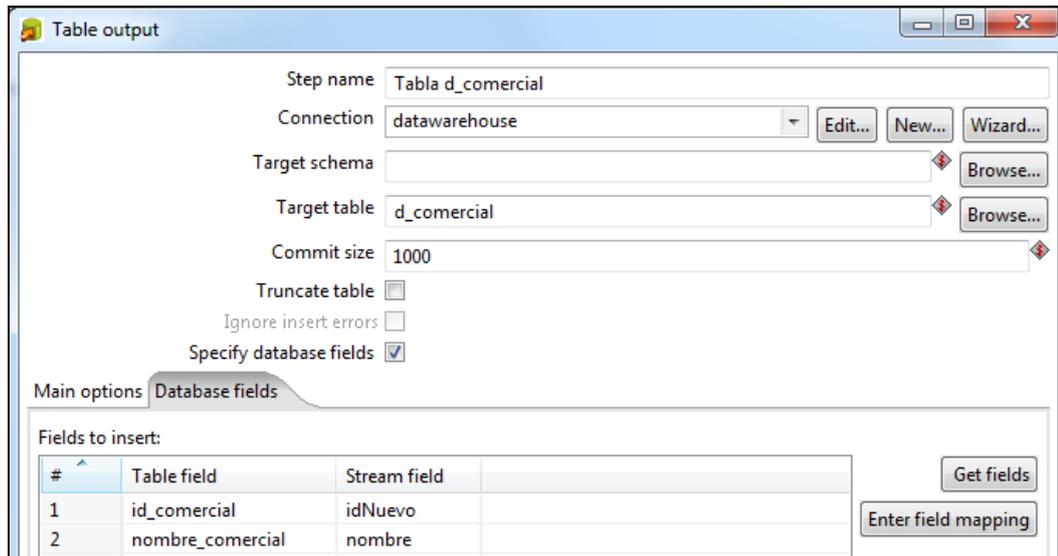


Ilustración 25 Transformación Carga d_comercial

- **Tabla comerciales.** Ejecuta la SQL para leer los comerciales del esquema *intermedio*.



- **Tabla d_comercial.** Carga los comerciales del flujo de entrada en la tabla *d_comercial*.



2.3.2.7. Carga de la dimensión d_concepto

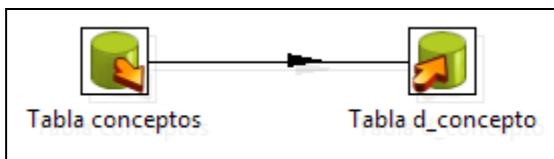
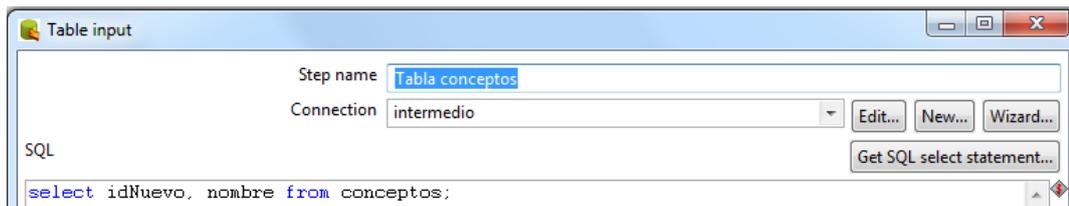
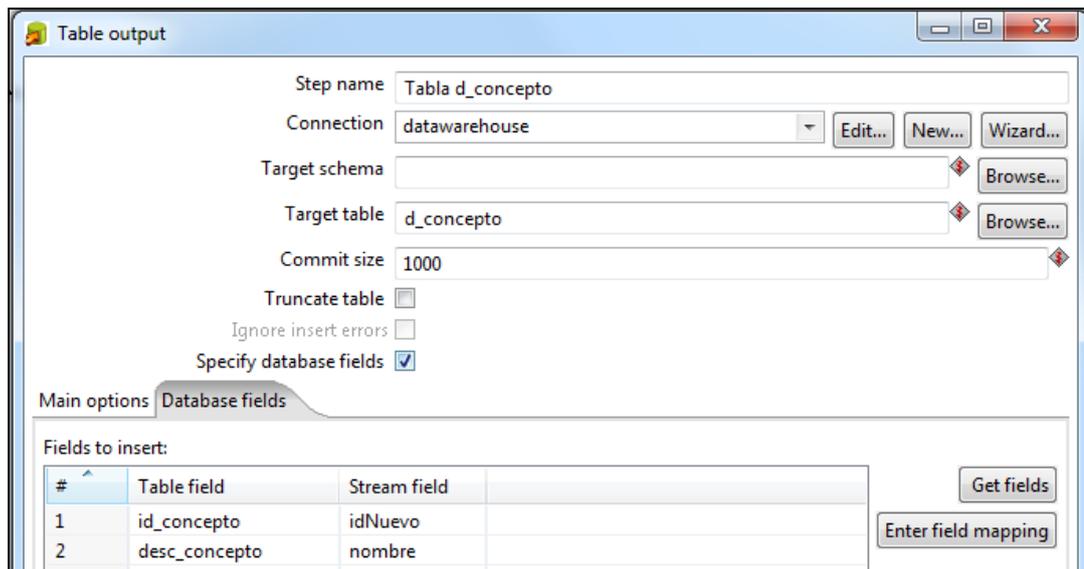


Ilustración 26 Transformación Carga d_concepto

- **Tabla conceptos.** Ejecuta la SQL para leer los conceptos del esquema *intermedio*.



- **Tabla d_concepto.** Carga los conceptos del flujo de entrada en la tabla *d_concepto*.



2.3.2.8.Carga de la dimensión d_delegacion

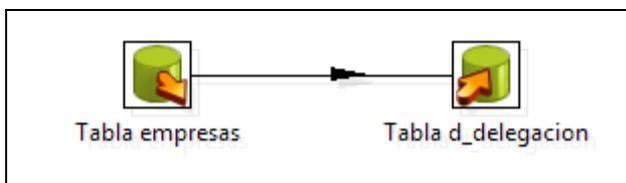
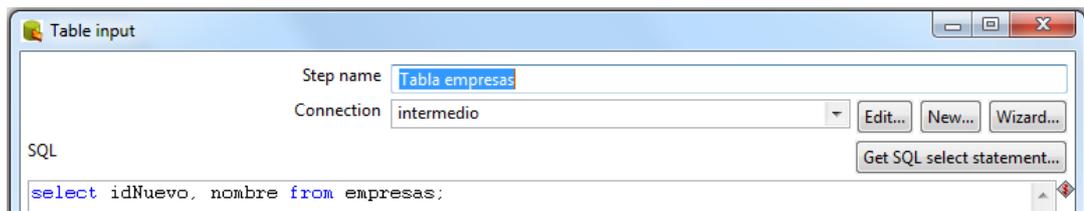
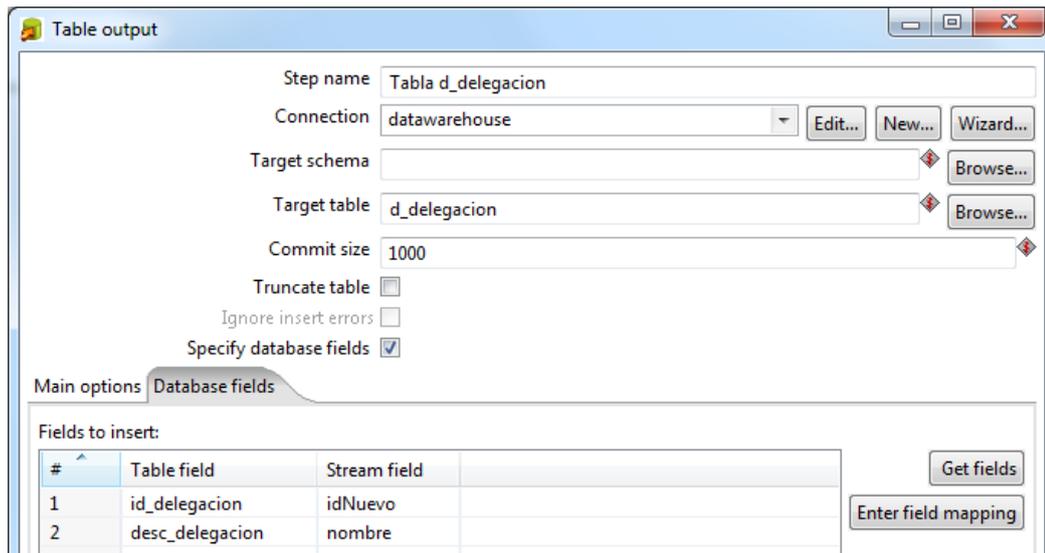


Ilustración 27 Transformación Carga d_delegacion

- **Tabla empresas.** Ejecuta la SQL para leer las empresas del esquema *intermedio*.



- **Tabla d_delegacion.** Carga las delegaciones del flujo de entrada en la tabla *d_delegacion*.



2.3.2.9. Carga de la dimensión d_familia

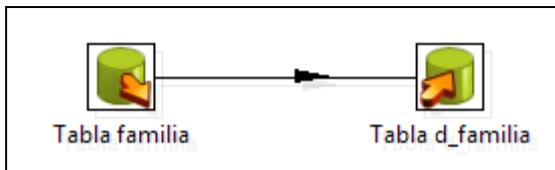
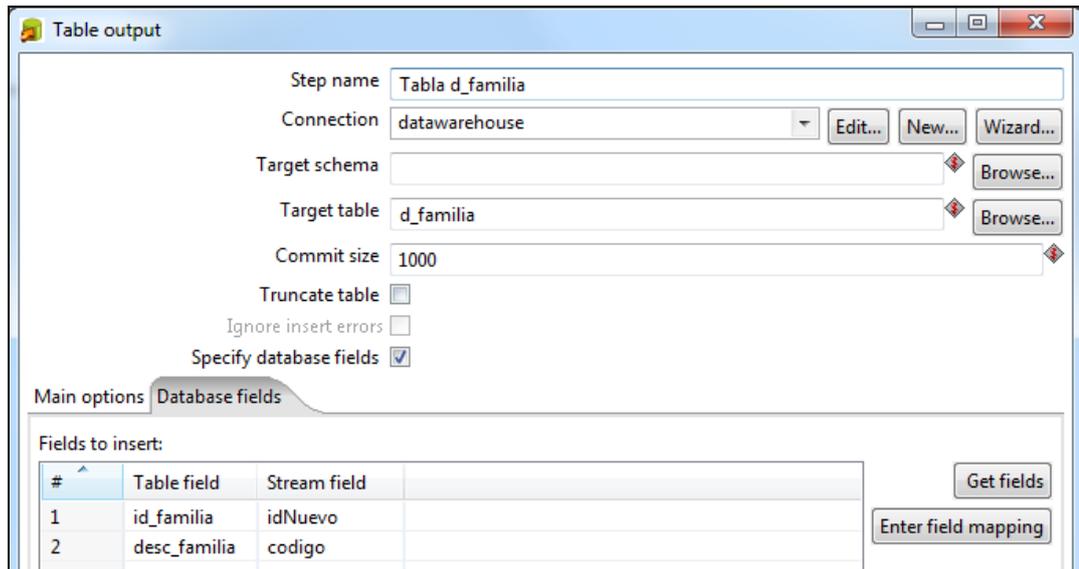


Ilustración 28 Transformación Carga d_familia

- **Tabla familia.** Ejecuta la SQL para leer las familias de productos del esquema *intermedio*.



- **Tabla d_familia.** Carga las familias del flujo de entrada en la tabla *d_familia*.

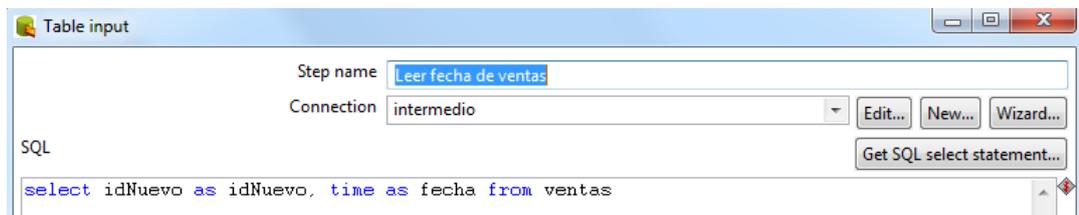


2.3.2.10. Carga de la dimensión d_fecha



Ilustración 29 Transformación Carga d_fecha

- **Tabla ventas (fecha).** Ejecuta la SQL para leer la fecha de cada venta del esquema *intermedio*.



- **Obtener formatos de fecha.** Paso javascript que procesa la fecha de entrada para obtener los diferentes formatos por los que se va a poder consultar la fecha de una venta.

```

Step name: Obtener formatos de fecha

Java script:
Script1
Date.prototype.getWeek = function() {
    var determinedate = new Date();
    determinedate.setFullYear(this.getFullYear(), this.getMonth(), this.getDate());
    var D = determinedate.getDay();
    if(D == 0) D = 7;
    determinedate.setDate(determinedate.getDate() + (4 - D));
    var YN = determinedate.getFullYear();
    var ZBDoCY = Math.floor((determinedate.getTime() - new Date(YN, 0, 1, -6)) / 86400000);
    var WN = 1 + Math.floor(ZBDoCY / 7);
    return WN;
}

var mes = 0;
var trimestre = 0;
var desc_mes = "N";
var desc_trimestre = "N";

var dia_flujo = fecha.substr(0,2);
var mes_flujo = fecha.substr(3,5);
var anyo = fecha.substr(6,10);

var dia = new Array("Domingo", "Lunes", "Martes", "Miércoles", "Jueves", "Viernes", "Sábado", "Domingo");
var fechaJS = new Date(fecha.substr(0,10));

var desc_dia_semana = dia[fechaJS.getDay()];
var dia_semana = fechaJS.getDay()+1;

var semana_anyo = fechaJS.getWeek();
var desc_semana_anyo = "Semana "+semana_anyo;

switch (mes_flujo) {
    case "01": mes = 1; desc_mes = "Enero"; trimestre = 1; desc_trimestre = "Primer trimestre"; break;
    case "02": mes = 2; desc_mes = "Febrero"; trimestre = 1; desc_trimestre = "Primer trimestre"; break;
    case "03": mes = 3; desc_mes = "Marzo"; trimestre = 1; desc_trimestre = "Primer trimestre"; break;
    case "04": mes = 4; desc_mes = "Abril"; trimestre = 2; desc_trimestre = "Segundo trimestre"; break;
    case "05": mes = 5; desc_mes = "Mayo"; trimestre = 2; desc_trimestre = "Segundo trimestre"; break;
    case "06": mes = 6; desc_mes = "Junio"; trimestre = 2; desc_trimestre = "Segundo trimestre"; break;
    case "07": mes = 7; desc_mes = "Julio"; trimestre = 3; desc_trimestre = "Tercer trimestre"; break;
    case "08": mes = 8; desc_mes = "Agosto"; trimestre = 3; desc_trimestre = "Tercer trimestre"; break;
    case "09": mes = 9; desc_mes = "Septiembre"; trimestre = 3; desc_trimestre = "Tercer trimestre"; break;
    case "10": mes = 10; desc_mes = "Octubre"; trimestre = 4; desc_trimestre = "Cuarto trimestre"; break;
    case "11": mes = 11; desc_mes = "Noviembre"; trimestre = 4; desc_trimestre = "Cuarto trimestre"; break;
    case "12": mes = 12; desc_mes = "Diciembre"; trimestre = 4; desc_trimestre = "Cuarto trimestre"; break;
}

```

- **Tabla d_fecha.** Carga los diferentes criterios de fechas de ventas del flujo de entrada en la tabla *d_fecha*.

Step name: Tabla d_fecha

Connection: datawarehouse

Target schema:

Target table: d_fecha

Commit size: 1000

Truncate table:

Ignore insert errors:

Specify database fields:

Main options Database fields

Fields to insert:

#	Table field	Stream field
1	dia_semana	dia_semana
2	desc_dia_semana	desc_dia_semana
3	mes	mes
4	desc_mes	desc_mes
5	trimestre	trimestre
6	anyo	anyo
7	semana_anyo	semana_anyo
8	desc_semana_an...	desc_semana_anyo
9	desc_trimestre	desc_trimestre
10	id_fecha	idNuevo

2.3.2.11. Carga de la dimensión d_forma_pago

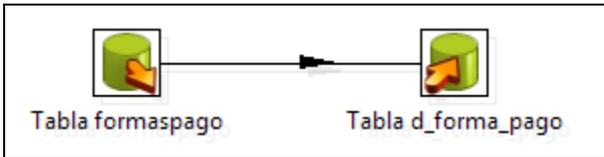
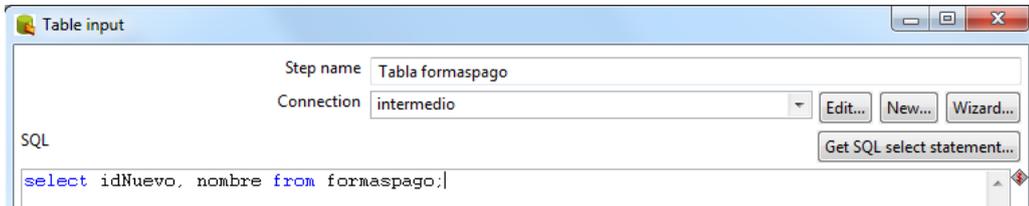
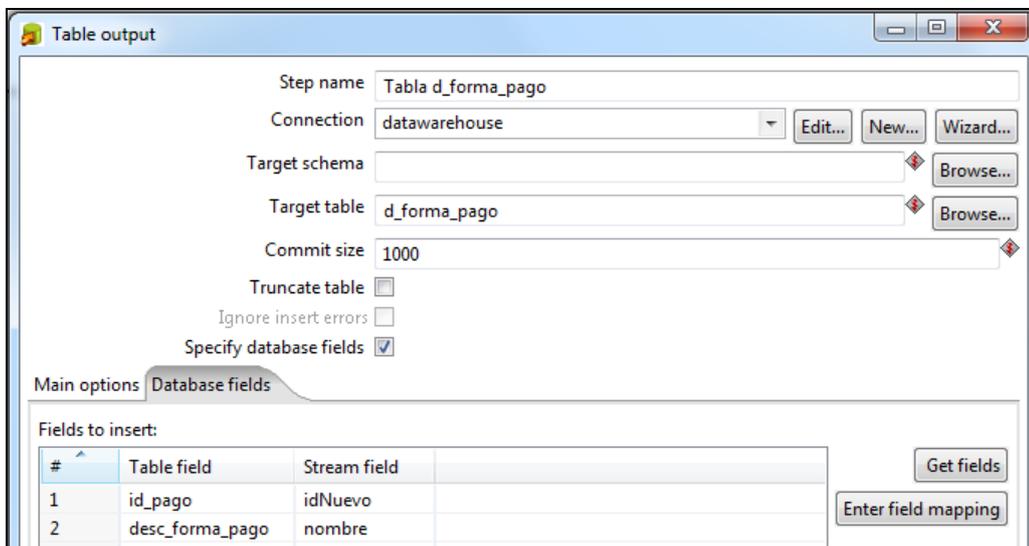


Ilustración 30 Transformación Carga d_forma_pago

- **Tabla formas_pago.** Ejecuta la SQL para leer las formas de pago del esquema *intermedio*.



- **Tabla d_forma_pago.** Carga las formas de pago del flujo de entrada en la tabla *d_forma_pago*.



2.3.2.12.Carga de la dimensión d_tipo_dato

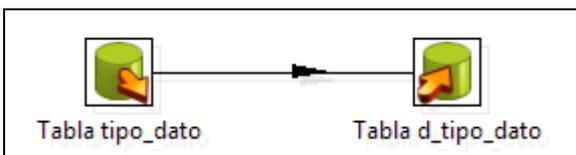
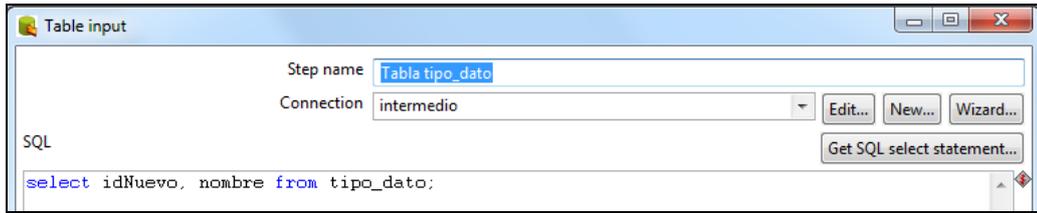
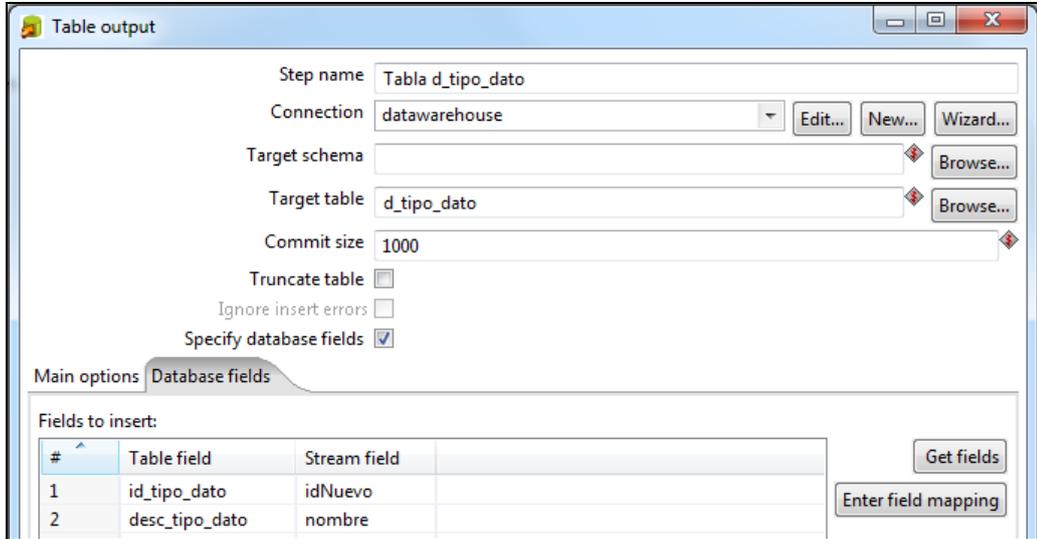


Ilustración 31 Transformación Carga d_tipo_dato

- **Tabla tipo_dato.** Ejecuta la SQL para leer los tipos de dato del esquema *intermedio*.



- **Tabla d_tipo_datos.** Carga el tipo de dato del flujo de entrada en la tabla *d_tipo_datos*.



2.3.2.13. Carga de la tabla de hechos h_ventas



Ilustración 32 Transformación Carga h_ventas

- **Obtener datos intermedios.** Ejecuta la SQL para obtener del esquema *intermedio* todas las ventas e información asociada.

Table input

Step name:

Connection:

SQL

```

select a.idNuevo as idAlmacen,
ar.idNuevo as idArticulo,
c.idNuevo as idCliente,
cz.idNuevo as idClienteZona,
co.idNuevo as idComercial,
cn.idNuevo as idConcepto,
e.idNuevo as idEmpresa,
f.idFamilia as idFamilia,
fp.idNuevo as idFormaPago,
td.idNuevo as idTipoDato,
v.idNuevo as idFecha,
abs(v.descuento) as descuento,
abs(v.importe) as importe,
abs(v.comision) as comision,
abs(v.coste) as coste
from ventas v
left outer join almacenes a on v.almacen = a.codigo
left outer join articulos ar on v.articulo = ar.codigo
left outer join clientes c on v.cliente = c.nif
left outer join clientes_zona_poblacion cz on v.clientezona = cz.nif
left outer join comerciales co on v.comercial = co.nif
left outer join conceptos cn on v.concepto = cn.codigo
left outer join empresas e on v.empresa = e.nif
left outer join familia_producto f on v.familia = f.codigo
left outer join formasPago fp on v.empresa = fp.codigo
left outer join tipo_dato td on v.tipodato = td.codigo

```

- **Completar valor de fk nulas.** Las ventas que no tengan asociada una determinada dimensión se actualizan para que tengan una referencia al valor por defecto de la dimensión correspondiente.

Replace null value

Step name:

Replace Null for all fields

Replace by value:

Set empty string?

Mask (Date):

Select fields:

Select value type:

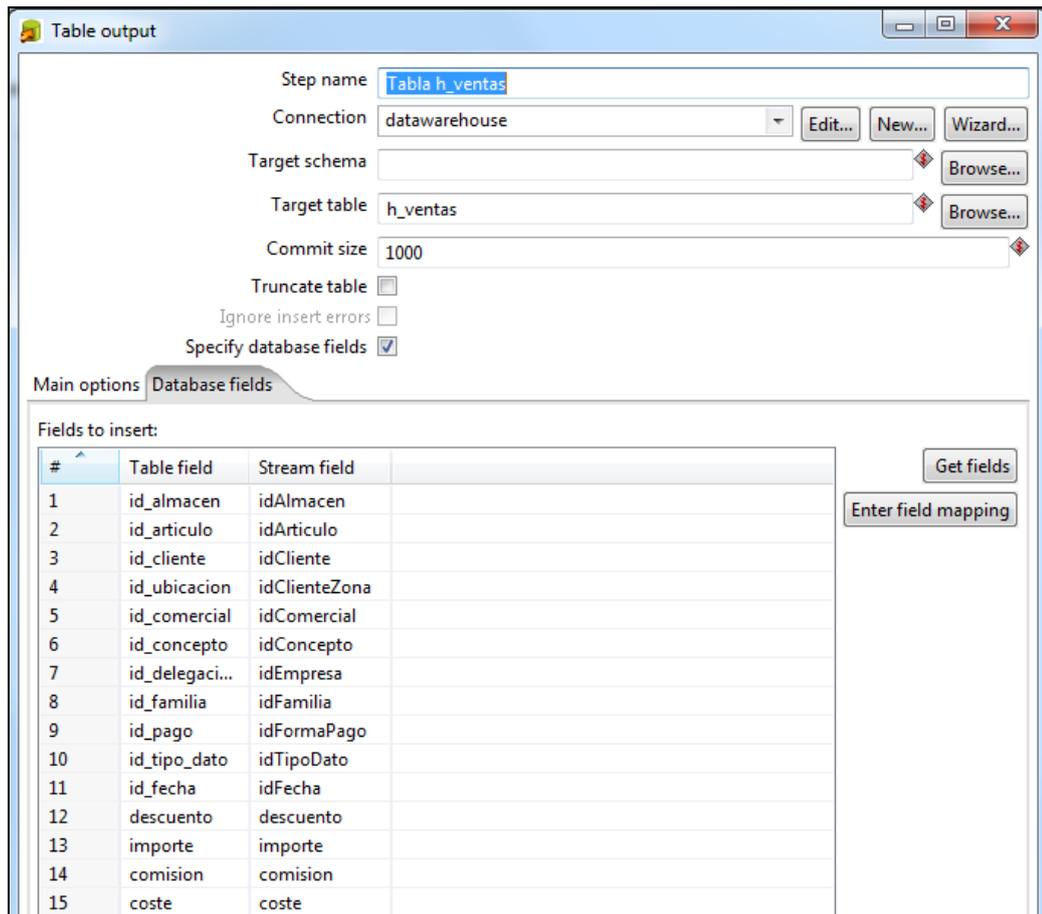
Value types

#	Type	Replace by value	Conversion mask (Date)	Set empty string?

Fields

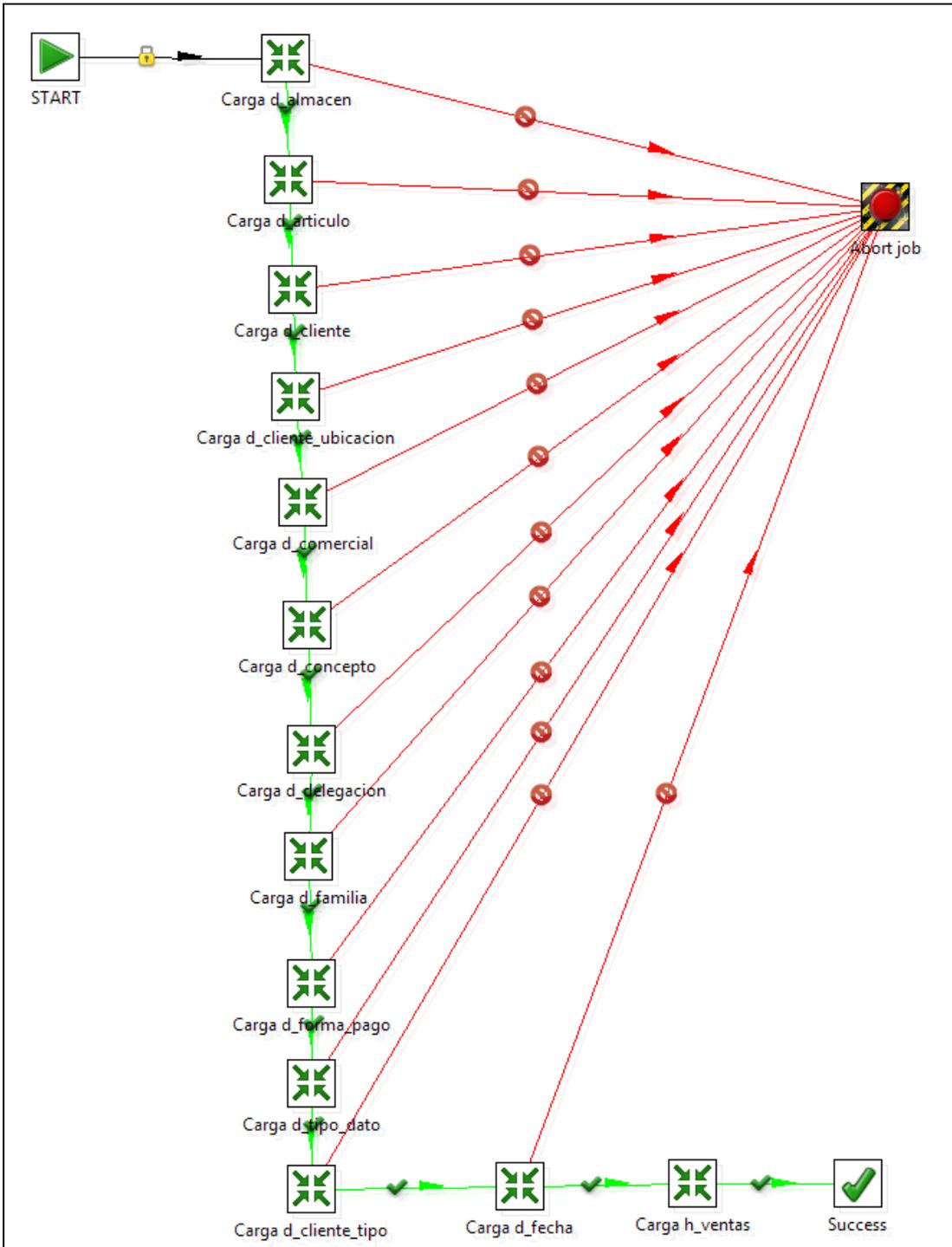
#	Field	Replace by value	Conversion mask (Date)	Set empty string?
1	idAlmacen	-1		N
2	idArticulo	-1		N
3	idCliente	-1		N
4	idClienteZona	-1		N
5	idComercial	-1		N
6	idConcepto	-1		N
7	idEmpresa	-1		N
8	idFamilia	-1		N
9	idFormaPago	-1		N
10	idTipoDato	-1		N

- **Tabla h_ventas.** Carga las ventas del flujo de entrada en la tabla *h_ventas*.



2.3.2.14. Tarea para ejecutar las transformaciones esquema del *Data Warehouse*

Para ejecutar todas las transformaciones del esquema final, se crea un trabajo que las ejecuta secuencialmente.



2.4.Explotación de los datos

Una vez finalizada la etapa ETL, en la que se ha cargado el *Data Warehouse* con los datos proporcionados por TOTSALLES, se generan los informes, utilizando *Pentaho Report Designer*, que permitirán a la compañía saber cómo ha evolucionado el negocio durante el periodo comprendido entre el 2009 y 2013. De esta forma, podrá tomar las acciones necesarias para

aprovechar oportunidades de negocio o emprender acciones correctivas que aumenten su beneficio.

A continuación se muestran los informes solicitados, que incluyen la generación del informe con los datos almacenados en el *Data Warehouse*, una explicación de la estructura del informe y las SQL utilizadas para generar las tablas y los diagramas.

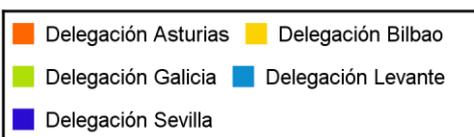
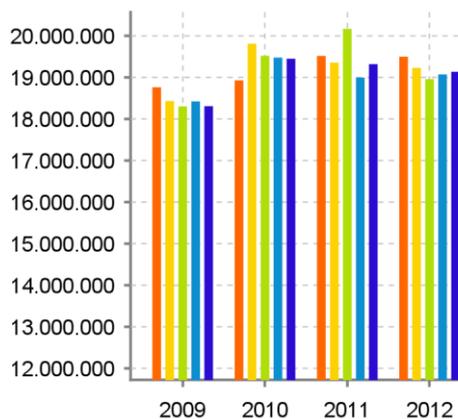
2.4.1. Evolución de las ventas por delegación y año

A partir del año 2010 se produce un incremento de las ventas en todas las delegaciones que se mantiene durante los años 2011 y 2012. Aunque este aumento es más discreto en unas delegaciones que otras, se mantiene hasta el 2013 en el que se produce una importante caída de facturación en todas las delegaciones.

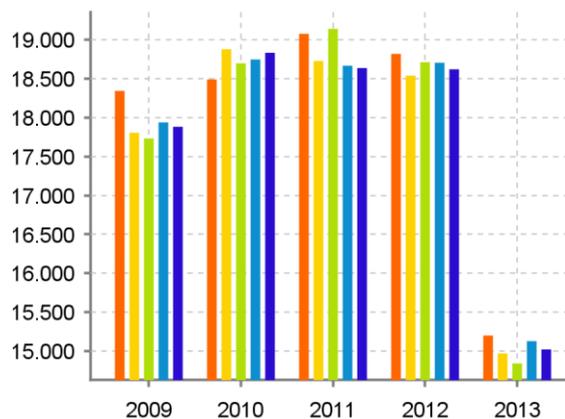
Delegación	Año	Unidades	Importe
Delegación Asturias	2009	18,340	18,757,821
Delegación Asturias	2010	18,487	18,924,414
Delegación Asturias	2011	19,072	19,511,826
Delegación Asturias	2012	18,815	19,493,206
Delegación Asturias	2013	15,203	12,453,444
Delegación Bilbao	2009	17,803	18,425,569
Delegación Bilbao	2010	18,874	19,805,344
Delegación Bilbao	2011	18,724	19,348,925
Delegación Bilbao	2012	18,536	19,223,823
Delegación Bilbao	2013	14,973	12,328,896
Delegación Galicia	2009	17,730	18,297,879
Delegación Galicia	2010	18,694	19,515,484
Delegación Galicia	2011	19,136	20,163,034
Delegación Galicia	2012	18,708	18,952,806
Delegación Galicia	2013	14,845	12,115,540
Delegación Levante	2009	17,936	18,417,285
Delegación Levante	2010	18,743	19,471,419

Delegación Levante	2011	18,663	18,991,230
Delegación Levante	2012	18,701	19,066,918
Delegación Levante	2013	15,131	12,409,472
Delegación Sevilla	2009	17,880	18,303,397
Delegación Sevilla	2010	18,829	19,443,876
Delegación Sevilla	2011	18,633	19,314,043
Delegación Sevilla	2012	18,617	19,130,501
Delegación Sevilla	2013	15,024	12,270,407

Importe ventas



Unidades vendidas



2.4.1.1. Estructura del informe

El informe muestra inicialmente una tabla con el importe de las ventas y el número de artículos vendidos agrupados por delegación y año. A continuación se muestran dos diagramas de barras que visualizan la evolución de las ventas y las unidades vendidas, agrupadas por año y delegación.

Para obtener estos datos se ha utilizado el siguiente *data set* a partir de la consulta:

```

1 select f.anyo as anyo, d.desc_delegacion as delegacion, SUM(h.importe) as importe, count (1) as unidades
2 from public.h_ventas h,
3 public.d_delegacion d,
4 public.d_fecha f
5 where h.id_delegacion = d.id_delegacion
6 and h.id_fecha = f.id_fecha
7 group by anyo, delegacion
8 order by delegacion, anyo

```

Ilustración 33 Data set evolución de las ventas por delegación

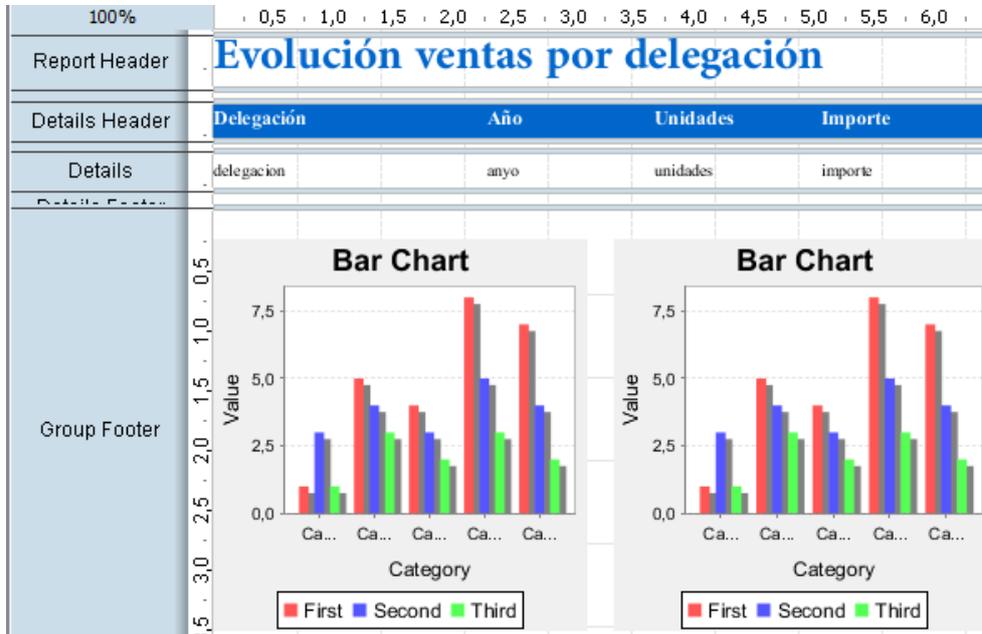


Ilustración 34 Estructura del informe evolución de las ventas por delegación y año

2.4.2. Productos y familias de productos más vendidos

Las familias de productos más vendidas en las 5 delegaciones se mantiene durante todos los años y sigue un patrón muy similar en cada una de ellas. La familia más importante es la de los electrodomésticos, que en todas las delegaciones se sitúa en torno a las 5000 unidades vendidas.

Cabe destacar el gran número de ventas que no tienen una familia de productos asociada, lo que dificulta hacer un análisis más exhaustivo de las familias en las que es necesario potenciar la venta de artículos.

Delegación Asturias		
Año	Familia de producto	Unidades vendidas
2009	Sin Familia	5,692
2009	ELECTRODOMESTICOS	5,002
2009	INFORMATICA	2,393
2009	ELECTRONICA CONSUMO	1,819
2009	CASA / JARDIN	1,403

2009	FOTOGRAFIA / VIDEO / ÓPTICA	1,249
2009	CONSOLAS Y VIDEOJUEGOS	782
2010	Sin Familia	5,925
2010	ELECTRODOMESTICOS	4,928
2010	INFORMATICA	2,358
2010	ELECTRONICA CONSUMO	1,791
2010	CASA / JARDIN	1,433
2010	FOTOGRAFIA / VIDEO / ÓPTICA	1,277
2010	CONSOLAS Y VIDEOJUEGOS	775
2011	Sin Familia	6,160
2011	ELECTRODOMESTICOS	4,990
2011	INFORMATICA	2,567
2011	ELECTRONICA CONSUMO	1,886
2011	CASA / JARDIN	1,410
2011	FOTOGRAFIA / VIDEO / ÓPTICA	1,266
2011	CONSOLAS Y VIDEOJUEGOS	793
2012	Sin Familia	5,971
2012	ELECTRODOMESTICOS	5,074
2012	INFORMATICA	2,404
2012	ELECTRONICA CONSUMO	1,858
2012	CASA / JARDIN	1,540
2012	FOTOGRAFIA / VIDEO / ÓPTICA	1,195
2012	CONSOLAS Y VIDEOJUEGOS	773
2,013	Sin Familia	4,858
2,013	ELECTRODOMESTICOS	4,139
2,013	INFORMATICA	1,893
2,013	ELECTRONICA CONSUMO	1,515
2,013	CASA / JARDIN	1,182
2,013	FOTOGRAFIA / VIDEO / ÓPTICA	981
2,013	CONSOLAS Y VIDEOJUEGOS	635

Delegación Bilbao

Año	Familia de producto	Unidades vendidas
2009	Sin Familia	5,537
2009	ELECTRODOMESTICOS	4,742
2009	INFORMATICA	2,367
2009	ELECTRONICA CONSUMO	1,798
2009	CASA / JARDIN	1,429
2009	FOTOGRAFIA / VIDEO / ÓPTICA	1,178
2009	CONSOLAS Y VIDEOJUEGOS	752
2010	Sin Familia	5,808
2010	ELECTRODOMESTICOS	5,049
2010	INFORMATICA	2,484

2010	ELECTRONICA CONSUMO	1,936
2010	CASA / JARDIN	1,540
2010	FOTOGRAFIA / VIDEO / ÓPTICA	1,202
2010	CONSOLAS Y VIDEOJUEGOS	855
2011	Sin Familia	5,951
2011	ELECTRODOMESTICOS	5,034
2011	INFORMATICA	2,354
2011	ELECTRONICA CONSUMO	1,871
2011	CASA / JARDIN	1,556
2011	FOTOGRAFIA / VIDEO / ÓPTICA	1,273
2011	CONSOLAS Y VIDEOJUEGOS	685
2012	Sin Familia	5,844
2012	ELECTRODOMESTICOS	5,058
2012	INFORMATICA	2,379
2012	ELECTRONICA CONSUMO	1,804
2012	CASA / JARDIN	1,481
2012	FOTOGRAFIA / VIDEO / ÓPTICA	1,293
2012	CONSOLAS Y VIDEOJUEGOS	677
2,013	Sin Familia	4,781
2,013	ELECTRODOMESTICOS	4,100
2,013	INFORMATICA	1,917
2,013	ELECTRONICA CONSUMO	1,394
2,013	CASA / JARDIN	1,162
2,013	FOTOGRAFIA / VIDEO / ÓPTICA	1,052
2,013	CONSOLAS Y VIDEOJUEGOS	567

Delegación Galicia

Año	Familia de producto	Unidades vendidas
2009	Sin Familia	5,629
2009	ELECTRODOMESTICOS	4,893
2009	INFORMATICA	2,198
2009	ELECTRONICA CONSUMO	1,709
2009	CASA / JARDIN	1,374
2009	FOTOGRAFIA / VIDEO / ÓPTICA	1,212
2009	CONSOLAS Y VIDEOJUEGOS	715
2010	Sin Familia	6,028
2010	ELECTRODOMESTICOS	5,025
2010	INFORMATICA	2,408
2010	ELECTRONICA CONSUMO	1,825
2010	CASA / JARDIN	1,388
2010	FOTOGRAFIA / VIDEO / ÓPTICA	1,301
2010	CONSOLAS Y VIDEOJUEGOS	719
2011	Sin Familia	5,953

2011 ELECTRODOMESTICOS	5,184
2011 INFORMATICA	2,470
2011 ELECTRONICA CONSUMO	1,861
2011 CASA / JARDIN	1,554
2011 FOTOGRAFIA / VIDEO / ÓPTICA	1,323
2011 CONSOLAS Y VIDEOJUEGOS	791
2012 Sin Familia	5,877
2012 ELECTRODOMESTICOS	5,106
2012 INFORMATICA	2,392
2012 ELECTRONICA CONSUMO	1,814
2012 CASA / JARDIN	1,528
2012 FOTOGRAFIA / VIDEO / ÓPTICA	1,241
2012 CONSOLAS Y VIDEOJUEGOS	750
2,013 Sin Familia	4,693
2,013 ELECTRODOMESTICOS	3,875
2,013 INFORMATICA	1,966
2,013 ELECTRONICA CONSUMO	1,449
2,013 CASA / JARDIN	1,240
2,013 FOTOGRAFIA / VIDEO / ÓPTICA	992
2,013 CONSOLAS Y VIDEOJUEGOS	630

Delegación Levante

Año	Familia de producto	Unidades vendidas
2009 Sin Familia		5,908
2009 ELECTRODOMESTICOS		4,746
2009 INFORMATICA		2,273
2009 ELECTRONICA CONSUMO		1,620
2009 CASA / JARDIN		1,467
2009 FOTOGRAFIA / VIDEO / ÓPTICA		1,240
2009 CONSOLAS Y VIDEOJUEGOS		682
2010 Sin Familia		6,045
2010 ELECTRODOMESTICOS		4,856
2010 INFORMATICA		2,424
2010 ELECTRONICA CONSUMO		1,897
2010 CASA / JARDIN		1,534
2010 FOTOGRAFIA / VIDEO / ÓPTICA		1,332
2010 CONSOLAS Y VIDEOJUEGOS		655
2011 Sin Familia		5,879
2011 ELECTRODOMESTICOS		5,110
2011 INFORMATICA		2,295
2011 ELECTRONICA CONSUMO		1,875
2011 CASA / JARDIN		1,462
2011 FOTOGRAFIA / VIDEO / ÓPTICA		1,348

2011	CONSOLAS Y VIDEOJUEGOS	694
2012	Sin Familia	5,841
2012	ELECTRODOMESTICOS	4,934
2012	INFORMATICA	2,446
2012	ELECTRONICA CONSUMO	1,885
2012	CASA / JARDIN	1,505
2012	FOTOGRAFIA / VIDEO / ÓPTICA	1,343
2012	CONSOLAS Y VIDEOJUEGOS	747
2,013	Sin Familia	4,643
2,013	ELECTRODOMESTICOS	4,154
2,013	INFORMATICA	2,030
2,013	ELECTRONICA CONSUMO	1,463
2,013	CASA / JARDIN	1,182
2,013	FOTOGRAFIA / VIDEO / ÓPTICA	1,053
2,013	CONSOLAS Y VIDEOJUEGOS	606

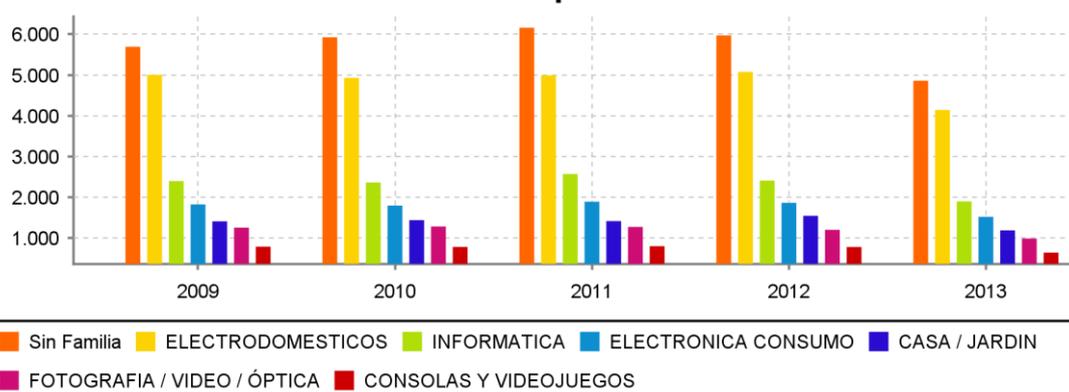
Delegación Sevilla

Año	Familia de producto	Unidades vendidas
2009	Sin Familia	5,502
2009	ELECTRODOMESTICOS	4,772
2009	INFORMATICA	2,312
2009	ELECTRONICA CONSUMO	1,778
2009	CASA / JARDIN	1,479
2009	FOTOGRAFIA / VIDEO / ÓPTICA	1,330
2009	CONSOLAS Y VIDEOJUEGOS	707
2010	Sin Familia	6,041
2010	ELECTRODOMESTICOS	4,991
2010	INFORMATICA	2,442
2010	ELECTRONICA CONSUMO	1,749
2010	CASA / JARDIN	1,535
2010	FOTOGRAFIA / VIDEO / ÓPTICA	1,333
2010	CONSOLAS Y VIDEOJUEGOS	738
2011	Sin Familia	5,851
2011	ELECTRODOMESTICOS	5,042
2011	INFORMATICA	2,419
2011	ELECTRONICA CONSUMO	1,808
2011	CASA / JARDIN	1,488
2011	FOTOGRAFIA / VIDEO / ÓPTICA	1,302
2011	CONSOLAS Y VIDEOJUEGOS	723
2012	Sin Familia	5,753
2012	ELECTRODOMESTICOS	4,915
2012	INFORMATICA	2,487
2012	ELECTRONICA CONSUMO	1,888

2012 CASA / JARDIN	1,485
2012 FOTOGRAFIA / VIDEO / ÓPTICA	1,313
2012 CONSOLAS Y VIDEOJUEGOS	776
2,013 Sin Familia	4,811
2,013 ELECTRODOMESTICOS	3,906
2,013 INFORMATICA	1,861
2,013 ELECTRONICA CONSUMO	1,509
2,013 CASA / JARDIN	1,220
2,013 FOTOGRAFIA / VIDEO / ÓPTICA	1,089
2,013 CONSOLAS Y VIDEOJUEGOS	628

Delegación Asturias

Unidades por familia

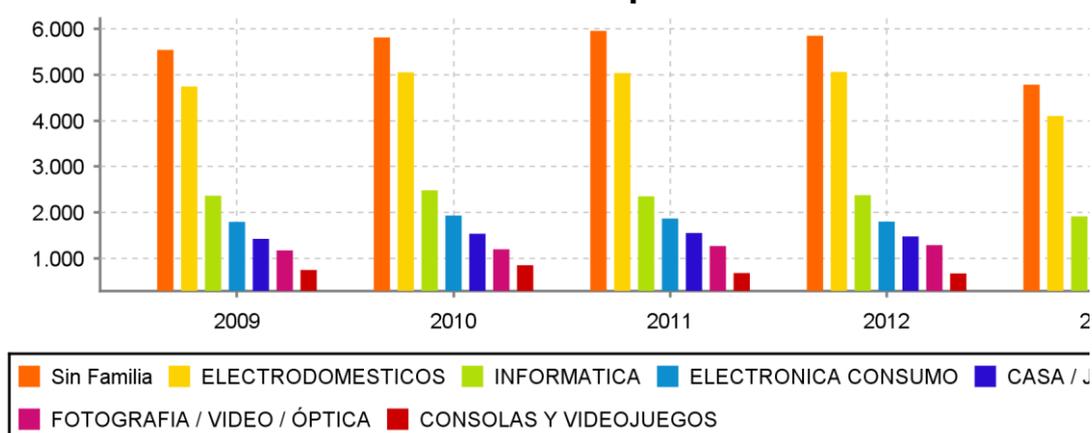


Año	Nombre artículo	Unidades vendidas
2009	Frogorrífico Balay 3KFI7861	130
2009	Aspiradora ufesa AC 6230 Bogge Potencia	118
2009	X BOX 360 TRITTON AX 180	102
2009	Televisor LG ELECTRONICS 50PZ950	101
2009	Congelador Siemens KG39EAI40	97
2009	Televisor Philips 40PFL5527H/12	93
2009	Tostadora Solac HO 6019	92
2009	Herramientas eléctricas	91
2009	X BOX 360 MadCatz MLG Pro-Circuit Contoller XBOX	89
2009	Ordenador cLenovo IdeaCentre C225 VB91CSP Negro	89
2010	Apple Nuevo iPad Wi-Fi 4G 64GB Negro	87
2010	Tablet Targus Vuscape Protective Cover	86
2010	Aspiradora Dyson DC 37 Animal	70
2010	Manguera Kärcher 2.645-179	65
2010	Home Cinema Philips HTS 3260	65
2010	Televisor LG ELECTRONICS 19LS3500	63
2010	Taladro Bosch PLL 360 0 603 663 000	63

2010 Centro planchado ALFA 7340	56
2010 Juego cuchillos BEEM F0000.150	51
2010 Equipos de música	50
2011 Ordenadores	90
2011 Cámara digital Sony NEX-7 Cuerpo	83
2011 Cámara digital Canon PowerShot G12	80
2011 Planchado y costuras	80
2011 Centro planchado Jata CP 440	79
2011 Taladro Bosch PSS 200 A 0603340000	77
2011 Plancha Philips GC 4410	73
2011 Robot iRobot Roomba 521	68
2011 Marco digital Transcend TS-PF705W	67
2011 Coseedora Jata MC 822	44
2012 Marco digital Transcend TS-PF705W	78
2012 Taladro Bosch PSS 200 A 0603340000	74
2012 Ordenadores	68
2012 Centro planchado Jata CP 440	68
2012 Estufa Olimpia Splendid Caldo Nove	67
2012 Planchado y costuras	66
2012 Plancha Philips GC 4410	63
2012 X BOX 360 Ubisoft Assassins Creed Revelations	62
2012 Coseedora Jata MC 822	53
2012 Ordenador Acer Aspire X3990 DT.SGKEB.010	48
2013 Centro planchado Polti PLEU0153	61
2013 Cafetera DeLonghi EN 450.CW	58
2013 Cafeteras	56
2013 Jardinería	55
2013 Ordenador HP TouchSmart Elite 7320 LH177EA	51
2013 Marco digital Energy Sistem Photo Frame 710 Soft White	49
2013 Lijadora Bosch PFS 65 O 603 206 100	46
2013 Olla Jata OPRP6	45
2013 Cámara digital Nikon D3200 Cuerpo	43
2013 Horno Fagor 6H185AX	41

Delegación Bilbao

Unidades por familia

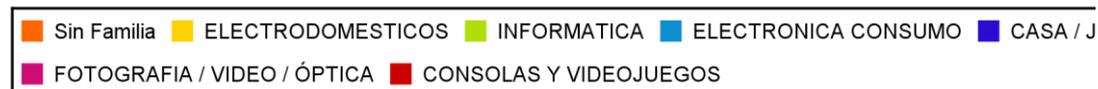
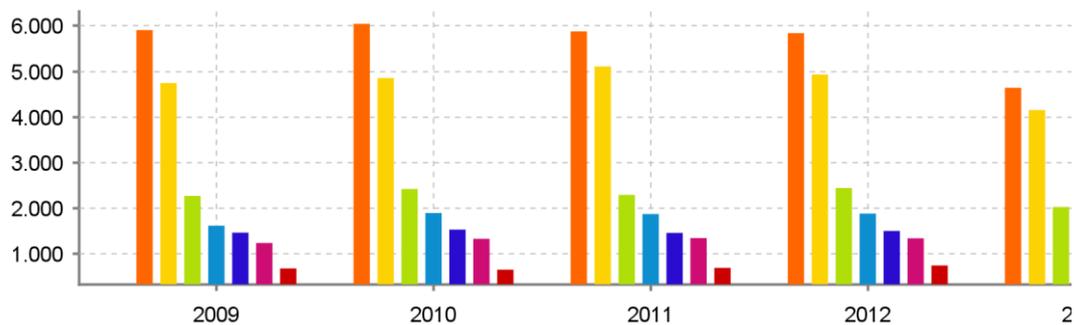


Año	Nombre artículo	Unidades vendidas
2009	Aspiradora ufesa AC 6230 Bogge Potencia	126
2009	X BOX 360 Konami XB360 PES 2013	112
2009	Frogorífico Balay 3KFI7861	108
2009	Ordenador Asus CM6730-ESCH18	99
2009	Frogorífico Edesa URBAN-F630	96
2009	Lavavajillas Indesit DFG 051 EU	93
2009	Ordenador cLenovo IdeaCentre C225 VB91CSP Negro	91
2009	Videocámara GOPRO HD HERO2 Motorsports	91
2009	Lavadora Samsung WF1602NHW	90
2009	Apple Nuevo iPad Wi-Fi 4G 16GB Negro	90
2010	Home Cinema Philips HTS 3260	90
2010	Manguera Kärcher 2.645-179	82
2010	Equipos de música	74
2010	Centro planchado ALFA 7340	73
2010	Taladro Bosch PLL 360 0 603 663 000	65
2010	Aspiradora Dyson DC 37 Animal	60
2010	Tablet Targus Vuscaped Protective Cover	60
2010	Juego cuchillos BEEM F0000.150	55
2010	Televisor LG ELECTRONICS 19LS3500	53
2010	Apple Nuevo iPad Wi-Fi 4G 64GB Negro	52
2011	Robot iRobot Roomba 521	75
2011	Cámara digital Sony NEX-7 Cuerpo	74
2011	Marco digital Transcend TS-PF705W	72
2011	Centro planchado Jata CP 440	68
2011	Cámara digital Canon PowerShot G12	67
2011	Ordenadores	62
2011	Coseedora Jata MC 822	59
2011	Taladro Bosch PSS 200 A 0603340000	58
2011	Plancha Philips GC 4410	57

2011 Planchado y costuras	55
2012 Marco digital Transcend TS-PF705W	96
2012 Ordenador Acer Aspire X3990 DT.SGKEB.010	91
2012 Taladro Bosch PSS 200 A 0603340000	84
2012 Centro planchado Jata CP 440	81
2012 Ordenadores	80
2012 Coseedora Jata MC 822	75
2012 Planchado y costuras	66
2012 X BOX 360 Ubisoft Assassins Creed Revelations	62
2012 Plancha Philips GC 4410	59
2012 Estufa Olimpia Splendid Caldo Nove	59
2013 Ordenador HP TouchSmart Elite 7320 LH177EA	64
2013 Marco digital Energy Sistem Photo Frame 710 Soft White	64
2013 Cámara digital Nikon D3200 Cuerpo	58
2013 Cafeteras	57
2013 Jardinería	57
2013 Horno Fagor 6H185AX	54
2013 Olla Jata OPRP6	49
2013 Lijadora Bosch PFS 65 0 603 206 100	48
2013 Cafetera DeLonghi EN 450.CW	38
2013 Centro planchado Polti PLEU0153	33

Delegación Levante

Unidades por familia

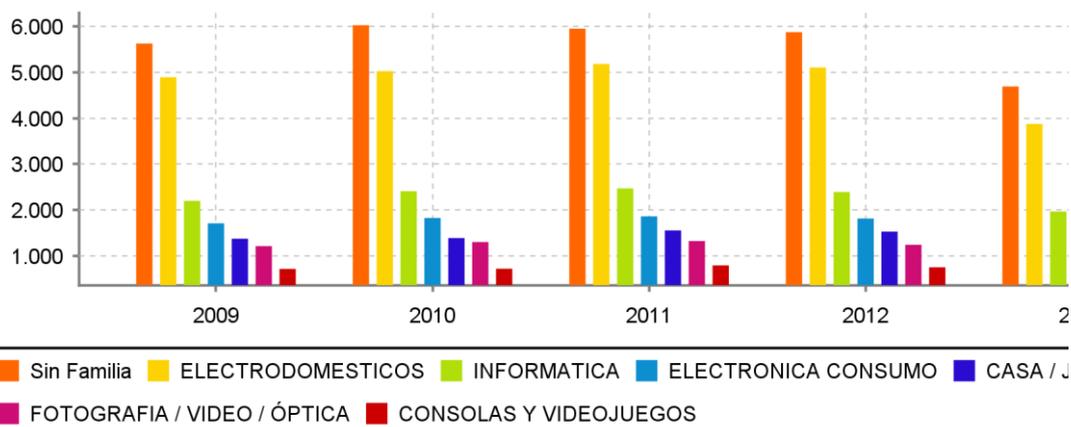


Año	Nombre artículo	Unidades vendidas
2009	Aspiradora ufesa AC 6230 Bogge Potencia	191
2009	Frogorrífico Balay 3KFI7861	112
2009	Playstation 3 Logitech diNovo Mini	97
2009	Barbacoa eléctrica Philips HD 4418/20	97
2009	Lavadora y secadora Hotpoint WMG 823 B EU	93
2009	X BOX 360 Electronic Arts XBOX360 FIFA 13	92

2009 Microondas Candy CMW 7217 DS silver	91
2009 Televisor Samsung UE40ES6800S	91
2009 Ordenador Acer Aspire X3990 DT.SGKEB.008	90
2009 Aspiradora de mano Electrolux ZB2906	89
2010 Tablet Targus Vuscupe Protective Cover	74
2010 Apple Nuevo iPad Wi-Fi 4G 64GB Negro	74
2010 Taladro Bosch PLL 360 0 603 663 000	70
2010 Equipos de música	65
2010 Centro planchado ALFA 7340	62
2010 Manguera Kärcher 2.645-179	56
2010 Home Cinema Philips HTS 3260	55
2010 Aspiradora Dyson DC 37 Animal	52
2010 Juego cuchillos BEEM F0000.150	41
2010 Televisor LG ELECTRONICS 19LS3500	40
2011 Cámara digital Sony NEX-7 Cuerpo	101
2011 Ordenadores	82
2011 Planchado y costuras	78
2011 Cámara digital Canon PowerShot G12	72
2011 Marco digital Transcend TS-PF705W	70
2011 Coseedora Jata MC 822	67
2011 Plancha Philips GC 4410	66
2011 Centro planchado Jata CP 440	62
2011 Robot iRobot Roomba 521	44
2011 Taladro Bosch PSS 200 A 0603340000	43
2012 Ordenador Acer Aspire X3990 DT.SGKEB.010	87
2012 Marco digital Transcend TS-PF705W	74
2012 Coseedora Jata MC 822	74
2012 Plancha Philips GC 4410	70
2012 Taladro Bosch PSS 200 A 0603340000	65
2012 Ordenadores	65
2012 Centro planchado Jata CP 440	62
2012 X BOX 360 Ubisoft Assassins Creed Revelations	53
2012 Planchado y costuras	52
2012 Estufa Olimpia Splendid Caldo Nove	48
2013 Cafeteras	66
2013 Centro planchado Polti PLEU0153	61
2013 Cafetera DeLonghi EN 450.CW	58
2013 Marco digital Energy Sistem Photo Frame 710 Soft White	57
2013 Cámara digital Nikon D3200 Cuerpo	57
2013 Ordenador HP TouchSmart Elite 7320 LH177EA	54
2013 Olla Jata OPRP6	51
2013 Lijadora Bosch PFS 65 0 603 206 100	50
2013 Jardinería	45

Delegación Galicia

Unidades por familia

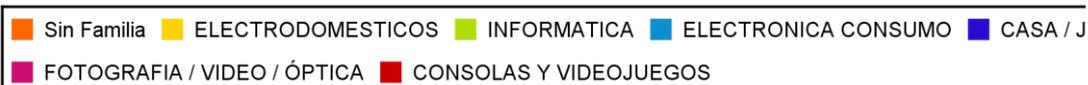
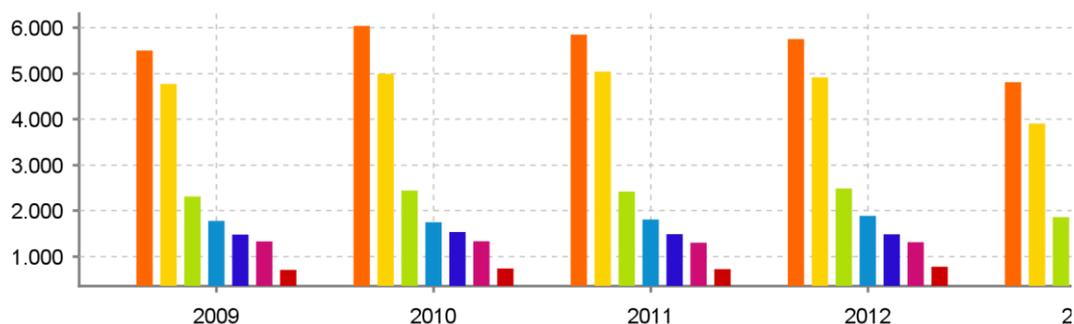


Año	Nombre artículo	Unidades vendidas
2009	Frogorífico Balay 3KFI7861	128
2009	Aspiradora ufesa AC 6230 Bogge Potencia	123
2009	X BOX 360 MadCatz MLG Pro-Circuit Contoller XBOX	93
2009	Ionizador Medisana VHW 60065	92
2009	Televisor LG ELECTRONICS 47LM671S	91
2009	Playstation 3 Logitech diNovo Mini	90
2009	Aspiradora de mano Electrolux ZB2906	89
2009	Jardinería	88
2009	Cosedora Singer Heavy Duty 4411	87
2009	Herramientas eléctricas	87
2010	Tablet Targus Vuscaped Protective Cover	73
2010	Home Cinema Philips HTS 3260	73
2010	Televisor LG ELECTRONICS 19LS3500	71
2010	Juego cuchillos BEEM F0000.150	66
2010	Equipos de música	66
2010	Apple Nuevo iPad Wi-Fi 4G 64GB Negro	61
2010	Centro planchado ALFA 7340	59
2010	Taladro Bosch PLL 360 0 603 663 000	57
2010	Manguera Kärcher 2.645-179	54
2010	Aspiradora Dyson DC 37 Animal	51
2011	Planchado y costuras	92

2011 Centro planchado Jata CP 440	81
2011 Taladro Bosch PSS 200 A 0603340000	78
2011 Ordenadores	77
2011 Coseedora Jata MC 822	74
2011 Robot iRobot Roomba 521	66
2011 Marco digital Transcend TS-PF705W	62
2011 Plancha Philips GC 4410	57
2011 Cámara digital Sony NEX-7 Cuerpo	56
2011 Cámara digital Canon PowerShot G12	52
2012 Estufa Olimpia Splendid Caldo Nove	96
2012 Ordenador Acer Aspire X3990 DT.SGKEB.010	83
2012 X BOX 360 Ubisoft Assassins Creed Revelations	81
2012 Centro planchado Jata CP 440	76
2012 Coseedora Jata MC 822	74
2012 Taladro Bosch PSS 200 A 0603340000	72
2012 Ordenadores	71
2012 Plancha Philips GC 4410	67
2012 Marco digital Transcend TS-PF705W	58
2012 Planchado y costuras	53
2013 Lijadora Bosch PFS 65 0 603 206 100	76
2013 Cafetera DeLonghi EN 450.CW	64
2013 Olla Jata OPRP6	63
2013 Jardinería	60
2013 Horno Fagor 6H185AX	59
2013 Cafeteras	56
2013 Ordenador HP TouchSmart Elite 7320 LH177EA	48
2013 Marco digital Energy Sistem Photo Frame 710 Soft White	47
2013 Cámara digital Nikon D3200 Cuerpo	44
2013 Centro planchado Polti PLEU0153	27

Delegación Sevilla

Unidades por familia



Año	Nombre artículo	Unidades vendidas
2009	Aspiradora ufesa AC 6230 Bogge Potencia	141
2009	Frogorífico Balay 3KFI7861	138
2009	Cámara digital Canon DCC-1600	93
2009	Frogorífico Bosch KGN39AI20	92
2009	Temporizador Kärcher WT 4.000 2.645-174.0	90
2009	Barbacoa Severin PG 8523	90
2009	Televisor Samsung UE40ES6800S	90
2009	Videocámara Canon LEGRIA HF M56	88
2009	Horno Fagor 6H185AX	87
2009	SBS Lápiz Capacitivo iPad / iPhone	87
2010	Juego cuchillos BEEM F0000.150	89
2010	Apple Nuevo iPad Wi-Fi 4G 64GB Negro	78
2010	Aspiradora Dyson DC 37 Animal	77
2010	Home Cinema Philips HTS 3260	71
2010	Televisor LG ELECTRONICS 19LS3500	71
2010	Tablet Targus Vuscaped Protective Cover	65
2010	Manguera Kärcher 2.645-179	65
2010	Equipos de música	60
2010	Taladro Bosch PLL 360 0 603 663 000	57
2010	Centro planchado ALFA 7340	43
2011	Marco digital Transcend TS-PF705W	93
2011	Coseedora Jata MC 822	86
2011	Cámara digital Canon PowerShot G12	81
2011	Cámara digital Sony NEX-7 Cuerpo	78
2011	Robot iRobot Roomba 521	75
2011	Ordenadores	67
2011	Taladro Bosch PSS 200 A 0603340000	67
2011	Planchado y costuras	65
2011	Plancha Philips GC 4410	63
2011	Centro planchado Jata CP 440	62
2012	Ordenador Acer Aspire X3990 DT.SGKEB.010	101
2012	Estufa Olimpia Splendid Caldo Nove	77
2012	Plancha Philips GC 4410	76
2012	Taladro Bosch PSS 200 A 0603340000	66
2012	Marco digital Transcend TS-PF705W	62
2012	Planchado y costuras	62
2012	Ordenadores	61
2012	Centro planchado Jata CP 440	59
2012	X BOX 360 Ubisoft Assassins Creed Revelations	58
2012	Coseedora Jata MC 822	54

2013 Cámara digital Nikon D3200 Cuerpo	61
2013 Horno Fagor 6H185AX	60
2013 Lijadora Bosch PFS 65 O 603 206 100	58
2013 Olla Jata OPRP6	58
2013 Marco digital Energy Sistem Photo Frame 710 Soft White	57
2013 Cafetera DeLonghi EN 450.CW	48
2013 Centro planchado Polti PLEU0153	42
2013 Cafeteras	40
2013 Ordenador HP TouchSmart Elite 7320 LH177EA	35
2013 Jardinería	31

2.4.2.1. Estructura del informe

El informe muestra inicialmente una tabla con el número de artículos vendidos por familia de productos, delegación y año. Para obtener estos datos se ha utilizado el siguiente *data set* a partir de la consulta:

```

1 SELECT d_delegacion.desc_delegacion,d_fecha.ano,d_familia.desc_familia,count(1) as unidades
2 FROM
3     public.h_ventas,
4     public.d_familia,
5     public.d_fecha,
6     public.d_delegacion
7 WHERE h_ventas.id_familia = d_familia.id_familia AND d_fecha.id_fecha=h_ventas.id_fecha
8 AND h_ventas.id_delegacion = d_delegacion.id_delegacion
9 GROUP BY d_familia.desc_familia,d_fecha.ano,d_delegacion.desc_delegacion
10 ORDER BY d_delegacion.desc_delegacion,d_fecha.ano,unidades desc

```

Ilustración 35 Data set evolución de las ventas por familia de productos

A continuación para cada una de las delegaciones se muestra un diagrama de barras que muestra el número de artículos vendidos, agrupados por año y familia. Para obtener estos datos se ha utilizado el siguiente *data set* a partir de la consulta en la que se establece la delegación:

```

SELECT d_fecha.ano,d_familia.desc_familia,count(1) as unidades
FROM
    public.h_ventas,
    public.d_familia,
    public.d_fecha,
    public.d_delegacion
WHERE h_ventas.id_familia = d_familia.id_familia AND d_fecha.id_fecha=h_ventas.id_fecha
AND h_ventas.id_delegacion = d_delegacion.id_delegacion
AND d_delegacion.desc_delegacion = 'Delegación Asturias'
GROUP BY d_familia.desc_familia,d_fecha.ano
ORDER BY d_fecha.ano,unidades desc

```

Ilustración 36 Data set artículos vendidos por familia y año

Finalmente se muestra una tabla con los 10 artículos más vendidos para cada una de las delegaciones.



Ilustración 37 Estructura informe familias y artículos más vendidos

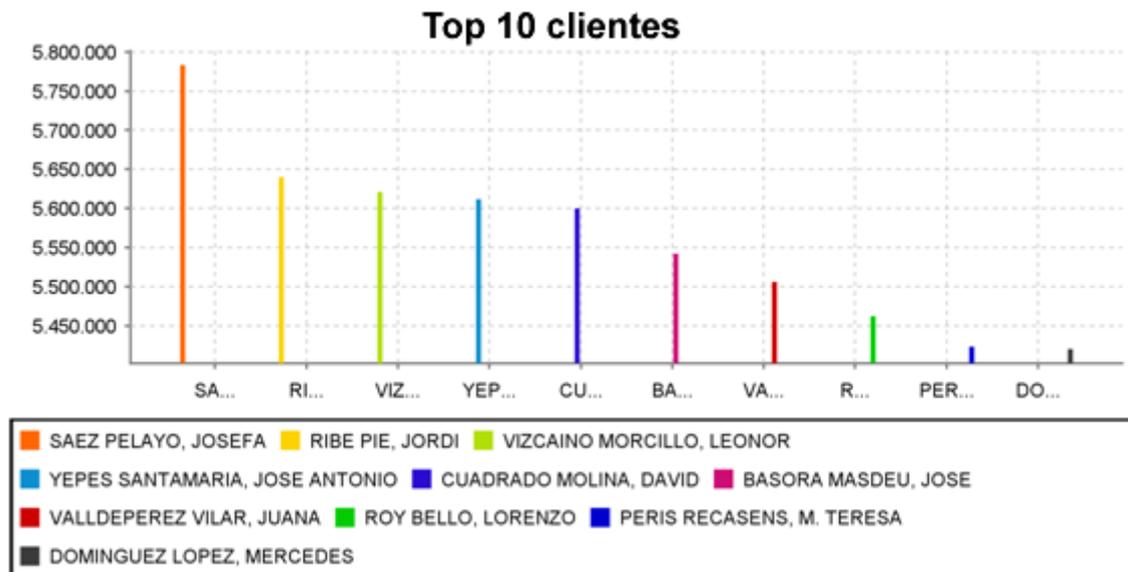
2.4.3. Top clientes

Los clientes que más compran tienen un volumen de facturación similar que se sitúa entre los 5,4 y 5,8 millones de euros. A partir del cliente número 20, el volumen de compra disminuye de forma importante, por lo que se pueden emprender acciones de fidelización para potenciar las compras.

Cliente	Unidades	Importe
SAEZ PELAYO, JOSEFA	5,838	5,783,239
RIBE PIE, JORDI	5,808	5,640,028
VIZCAINO MORCILLO, LEONOR	5,445	5,621,006
YEPES SANTAMARIA, JOSE ANTONIO	5,413	5,611,729
CUADRADO MOLINA, DAVID	5,558	5,599,865
BASORA MASDEU, JOSE	5,364	5,542,050
VALLDEPEREZ VILAR, JUANA	5,492	5,505,985
ROY BELLO, LORENZO	5,440	5,462,197

PERIS RECASENS, M. TERESA 5,250 5,423,205

DOMINGUEZ LOPEZ, MERCEDES 5,603 5,420,182



2.4.3.1. Estructura del informe

El informe muestra inicialmente una tabla junto con un diagrama de barras con los 10 mejores clientes, importe facturado y el número de artículos comprados. Para obtener estos datos se ha utilizado el siguiente *data set* a partir de la consulta:

```
1 select c.nombre_cliente as cliente, sum (h.importe) as importe, count(1) as unidades
2 from public.h_ventas h, public.d_cliente c
3 where h.id_cliente = c.id_cliente
4 group by c.nombre_cliente
5 order by importe desc
6 limit 10
```

Ilustración 38 Data set top 10 clientes



Ilustración 39 Estructura informe top 10 clientes

2.4.4. Distribución ventas según zona cliente

En la distribución de ventas por provincias hay una gran diferencia entre las primeras (Girona, La Rioja) y las provincias que tienen una facturación menor como puede ser Castellón. Puede ser interesante desplazar comerciales de la zona de Cataluña a zonas con un volumen de ventas inferior dado que Barcelona, Tarragona, Lleida y Girona tienen una facturación importante y consolidada.

Provincia cliente	Unidades	Importe
Girona	29,228	29,215,265
Rioja (La)	24,803	24,709,211
Cáceres	24,331	23,966,420
Zaragoza	22,739	22,732,878
Barcelona	20,636	20,689,917
Zamora	20,200	19,945,232
Valencia/València	19,462	19,523,411
Guadalajara	18,983	18,773,109
Burgos	18,622	18,342,228
Valladolid	17,345	17,260,838
Salamanca	15,789	15,835,250
Tarragona	14,780	14,546,945
Almería	14,327	14,518,462
Soria	13,764	14,123,848
Lleida	14,553	14,081,864
Navarra	10,379	10,413,255

Palencia	9,870	9,967,877
Cuenca	10,136	9,912,333
Avila	10,163	9,894,052
Madrid	9,729	9,867,768
Pontevedra	9,576	9,605,497
León	9,356	9,509,628
Badajoz	9,118	8,970,623
Granada	9,153	8,905,123
Vizcaya	5,838	5,783,239
Toledo	5,445	5,621,006
Balears (Illes)	5,364	5,542,050
Asturias	5,084	5,041,227
Teruel	5,062	4,972,972
Soria provincia	4,839	4,890,977
Jaén Provincia	4,852	4,868,223
Sevilla	4,960	4,840,828
Huelva	4,960	4,834,914
Huesca	4,801	4,778,091
Córdoba	4,674	4,710,384
Cantabria	4,548	4,628,082
Sin Población	4,700	4,543,398
Castellón/Castelló	3,928	3,770,134

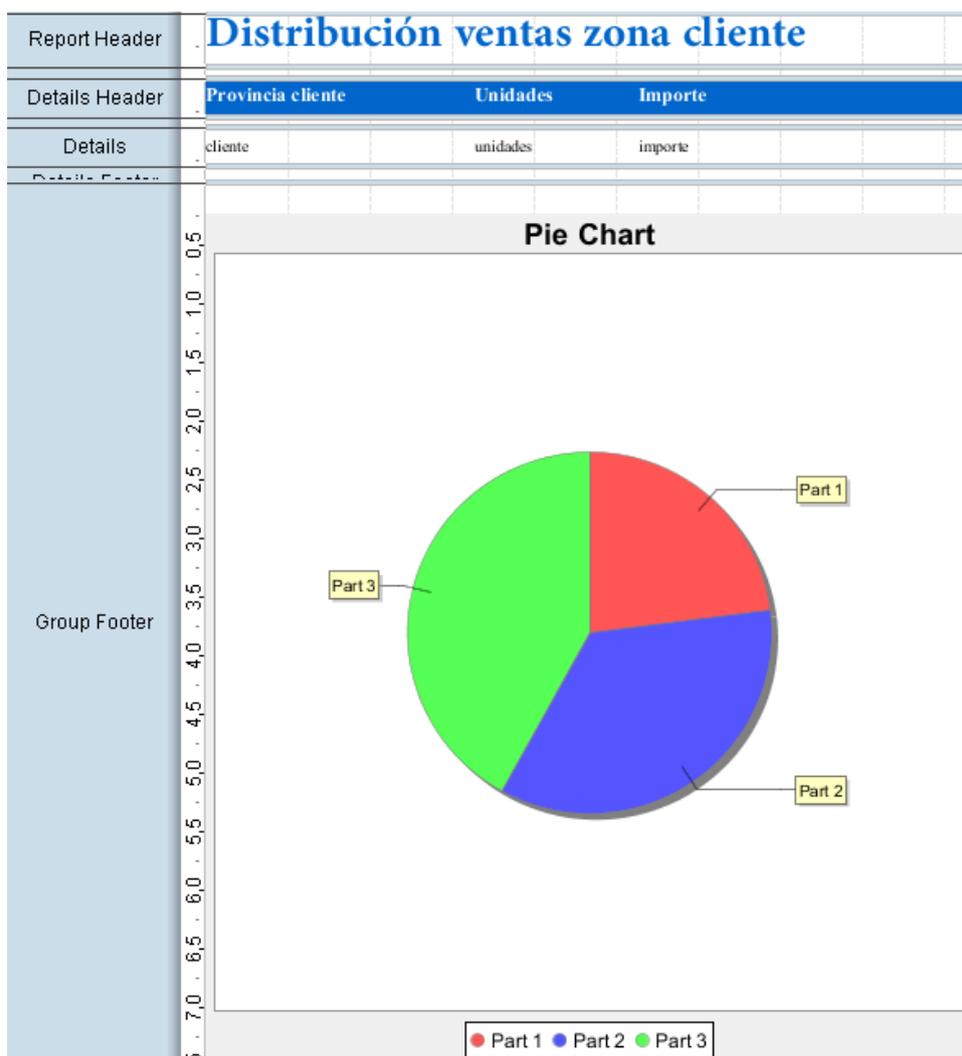


Ilustración 41 Estructura del informe ventas distribuidas por provincia

2.4.5. Evolución de las ventas en función del tiempo

En la evolución de las ventas que va desde el 2009 al 2013, se observa una clara caída en el último año, que se acentúa todavía más en el último trimestre del año. En el año 2009 se produce un aumento constante de las ventas a medida que avanza el año y que en el último trimestre se sitúa en un volumen de ventas similar a los siguientes dos años.

Exceptuando el año 2013, a nivel de ventas mensuales, todos los años presentan una evolución similar por lo que se deberían hacer campañas para aumentar las ventas en meses como septiembre y octubre en los que el volumen es sensiblemente más bajo.

Año	Trimestre	Mes	Importe
2009	Primer trimestre	Enero	6,424,805
2009	Primer trimestre	Febrero	8,001,793
2009	Primer trimestre	Marzo	7,566,086
2009	Segundo trimestre	Abril	7,535,628
2009	Segundo trimestre	Mayo	8,044,938

2009	Segundo trimestre	Junio	7,741,764
2009	Tercer trimestre	Julio	8,013,074
2009	Tercer trimestre	Agosto	8,061,392
2009	Tercer trimestre	Septiembre	7,168,228
2009	Cuarto trimestre	Octubre	7,718,502
2009	Cuarto trimestre	Noviembre	8,355,045
2009	Cuarto trimestre	Diciembre	7,570,696
2010	Primer trimestre	Enero	8,147,429
2010	Primer trimestre	Febrero	7,980,960
2010	Primer trimestre	Marzo	7,370,934
2010	Segundo trimestre	Abril	7,903,337
2010	Segundo trimestre	Mayo	8,993,399
2010	Segundo trimestre	Junio	7,680,501
2010	Tercer trimestre	Julio	8,169,726
2010	Tercer trimestre	Agosto	8,456,763
2010	Tercer trimestre	Septiembre	7,796,255
2010	Cuarto trimestre	Octubre	7,682,692
2010	Cuarto trimestre	Noviembre	8,596,813
2010	Cuarto trimestre	Diciembre	8,381,728
2011	Primer trimestre	Enero	7,641,552
2011	Primer trimestre	Febrero	7,976,528
2011	Primer trimestre	Marzo	8,858,824
2011	Segundo trimestre	Abril	7,898,273
2011	Segundo trimestre	Mayo	7,889,891
2011	Segundo trimestre	Junio	7,876,728
2011	Tercer trimestre	Julio	8,186,721
2011	Tercer trimestre	Agosto	7,503,990
2011	Tercer trimestre	Septiembre	7,743,189
2011	Cuarto trimestre	Octubre	7,961,437
2011	Cuarto trimestre	Noviembre	9,033,809
2011	Cuarto trimestre	Diciembre	8,758,116
2012	Primer trimestre	Enero	7,842,370
2012	Primer trimestre	Febrero	8,286,951
2012	Primer trimestre	Marzo	8,403,178
2012	Segundo trimestre	Abril	7,407,181
2012	Segundo trimestre	Mayo	7,683,546
2012	Segundo trimestre	Junio	8,427,995
2012	Tercer trimestre	Julio	7,771,198
2012	Tercer trimestre	Agosto	8,240,193
2012	Tercer trimestre	Septiembre	7,551,600
2012	Cuarto trimestre	Octubre	7,618,149
2012	Cuarto trimestre	Noviembre	8,466,798
2012	Cuarto trimestre	Diciembre	8,168,095
2013	Primer trimestre	Enero	5,964,450
2013	Primer trimestre	Febrero	6,603,797
2013	Primer trimestre	Marzo	6,876,825
2013	Segundo trimestre	Abril	6,190,441
2013	Segundo trimestre	Mayo	6,890,013
2013	Segundo trimestre	Junio	4,313,098
2013	Tercer trimestre	Julio	3,596,619
2013	Tercer trimestre	Agosto	4,418,986
2013	Tercer trimestre	Septiembre	4,684,048

2013	Cuarto trimestre	Octubre	4,227,916
2013	Cuarto trimestre	Noviembre	3,672,303
2013	Cuarto trimestre	Diciembre	4,139,263

2.4.5.1. Estructura del informe

El informe muestra inicialmente una tabla con el importe de las ventas desglosadas por año, trimestre y mes. Para obtener estos datos se ha utilizado el siguiente *data set* a partir de la consulta:

```

1 select f.anyo as anyo,f.desc_trimestre as trimestre,f.trimestre as num_trimestre,
2 f.desc_mes as mes,f.mes as num_mes, sum (h.importe) as importe
3 from public.h_ventas h,
4 public.d_fecha f
5 where h.id_fecha =f.id_fecha
6 group by f.anyo,f.desc_trimestre, f.trimestre,f.desc_mes,f.mes
7 order by anyo, num_trimestre,num_mes

```

Ilustración 42 Data set evolución de las ventas en el tiempo

A continuación se muestra un gráfico de barras con la evolución mensual de las ventas anuales desde el 2009 al 2013. Como fuente de datos el gráfico ha utilizado el siguiente *data set* a partir de la consulta:

```

1 select f.anyo as anyo_mes,f.desc_mes as mes,f.mes as num_mes,sum (h.importe) as importe_mes
2 from public.h_ventas h,
3 public.d_fecha f
4 where h.id_fecha =f.id_fecha
5 group by f.anyo,f.desc_mes,f.mes
6 order by anyo, num_mes

```

Ilustración 43 Data set evolución ventas mensuales

Finalmente se muestra un gráfico de barras con la evolución trimestral de las ventas para los diferentes años a partir de la consulta:

```

1 select f.anyo as anyo, f.desc_trimestre as trimestre,f.trimestre as num_trimestre, sum (h.importe) as importe
2 from public.h_ventas h,
3 public.d_fecha f
4 where h.id_fecha =f.id_fecha
5 group by f.anyo,f.desc_trimestre, f.trimestre
6 order by anyo, num_trimestre

```

Ilustración 44 Data set evolución ventas trimestral

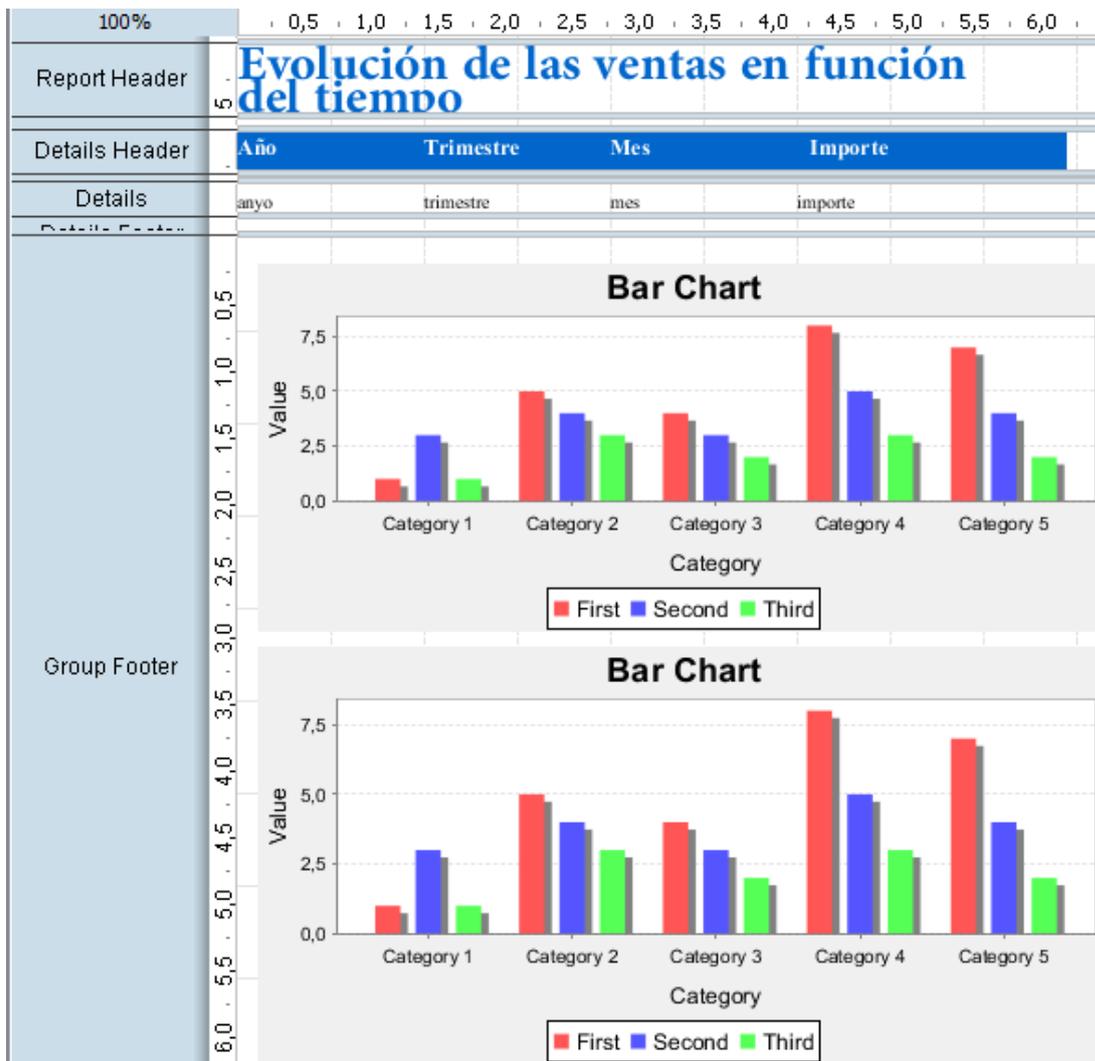


Ilustración 45 Estructura informe evolución ventas temporal

2.4.6. Comisiones a liquidar a los comerciales por año y trimestre

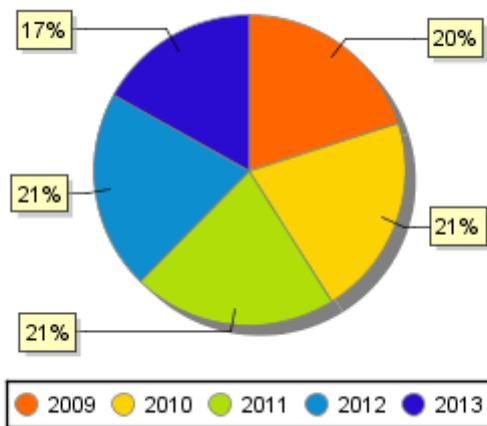
El nivel de comisiones de los comerciales es muy similar en los años comprendidos entre el 2009 y el 2012, situándose en torno al 21%. Si consideramos cada uno de estos años, a nivel trimestral las comisiones se reparten proporcionalmente con variaciones inferiores al 1%.

En cambio, en el año 2013 las comisiones a pagar a los comerciales bajan un 4% debido al descenso de las ventas. Observando el grafico trimestral del año 2013 se puede ver claramente cómo bajan las comisiones a partir del segundo trimestre, acentuándose todavía más en los dos últimos trimestres del año.

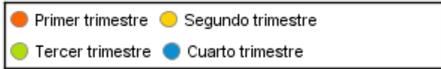
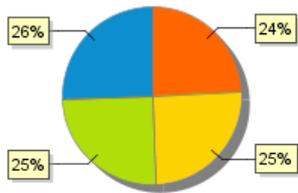
Año	Comercial	Comisión
2009	DE TORRES ESCUDERO, MARIANO	613,421
2009	DE LA CRUZ VILLARUBIA, JOSE LUIS	615,547
2009	MANCEBO LOZANO, FRANCISCO JAVIER	605,683
2009	CORCHERO LENA, ROSA MARIA	613,403
2009	GOMEZ DURAN, JOSE	603,787

2009	CABRERA CARRANZA, FELIPA	611,934
2010	DE LA CRUZ VILLARUBIA, JOSE LUIS	635,244
2010	CABRERA CARRANZA, FELIPA	637,578
2010	DE TORRES ESCUDERO, MARIANO	643,753
2010	GOMEZ DURAN, JOSE	636,196
2010	MANCEBO LOZANO, FRANCISCO JAVIER	645,475
2010	CORCHERO LENA, ROSA MARIA	629,473
2011	DE LA CRUZ VILLARUBIA, JOSE LUIS	640,977
2011	CORCHERO LENA, ROSA MARIA	647,940
2011	MANCEBO LOZANO, FRANCISCO JAVIER	639,946
2011	GOMEZ DURAN, JOSE	651,508
2011	CABRERA CARRANZA, FELIPA	644,201
2011	DE TORRES ESCUDERO, MARIANO	653,169
2012	GOMEZ DURAN, JOSE	627,086
2012	MANCEBO LOZANO, FRANCISCO JAVIER	624,782
2012	DE LA CRUZ VILLARUBIA, JOSE LUIS	627,039
2012	CABRERA CARRANZA, FELIPA	626,415
2012	CORCHERO LENA, ROSA MARIA	637,437
2012	DE TORRES ESCUDERO, MARIANO	639,099
2013	DE TORRES ESCUDERO, MARIANO	517,704
2013	DE LA CRUZ VILLARUBIA, JOSE LUIS	525,310
2013	GOMEZ DURAN, JOSE	510,645
2013	CORCHERO LENA, ROSA MARIA	500,625
2013	MANCEBO LOZANO, FRANCISCO JAVIER	510,050
2013	CABRERA CARRANZA, FELIPA	517,090

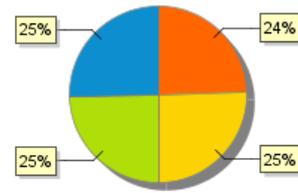
Comisión anual ventas



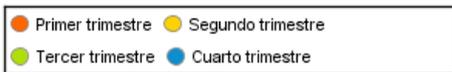
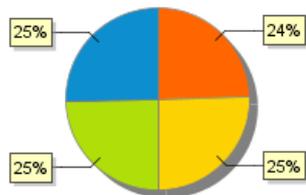
Comisión del 2009 por trimestres



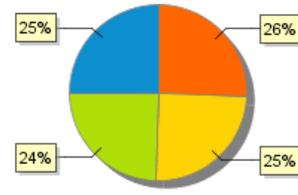
Comisión del 2010 por trimestres



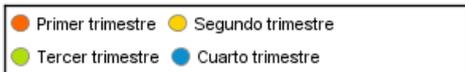
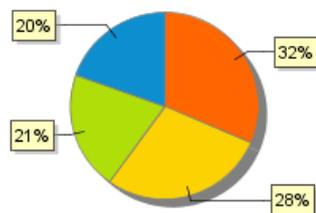
Comisión del 2011 por trimestres



Comisión del 2012 por trimestres



Comisión del 2013 por trimestres



2.4.6.1. Estructura del informe

El informe muestra inicialmente una tabla con las comisiones a pagar a cada uno de los comerciales. Para obtener estos datos se ha utilizado el siguiente *data set* a partir de la consulta:

```

1 select f.anyo as anyo,c.nombre_comercial as comercial, sum (h.comision) as comision
2 from public.h_ventas h,
3 public.d_fecha f,
4 public.d_comercial c
5 where h.id_fecha =f.id_fecha and h.id_comercial=c.id_comercial
6 group by f.anyo,c.nombre_comercial
7 order by anyo

```

Ilustración 46 Data set comisiones comercial y año

A continuación se muestra un diagrama el porcentaje de comisiones a pagar por año. Como fuente de datos el gráfico ha utilizado el siguiente *data set* a partir de la consulta:

```

1 select f.anyo as anyo, sum (h.comision) as comision
2 from public.h_ventas h,
3 public.d_fecha f
4 where h.id_fecha =f.id_fecha
5 group by f.anyo
6 order by anyo

```

Ilustración 47 Data set porcentaje comisiones por año

Adicionalmente se muestra el desglose trimestral para cada uno de los años. Se generan las diferentes gráficas a partir de consultas como la siguiente, filtrando los datos por año:

```

1 select f.anyo as anyo,c.nombre_comercial as comercial, sum (h.comision) as comision
2 from public.h_ventas h,
3 public.d_fecha f,
4 public.d_comercial c
5 where h.id_fecha =f.id_fecha and h.id_comercial=c.id_comercial
6 group by f.anyo,c.nombre_comercial
7 order by anyo

```

Ilustración 48 Data set porcentaje trimestral por año

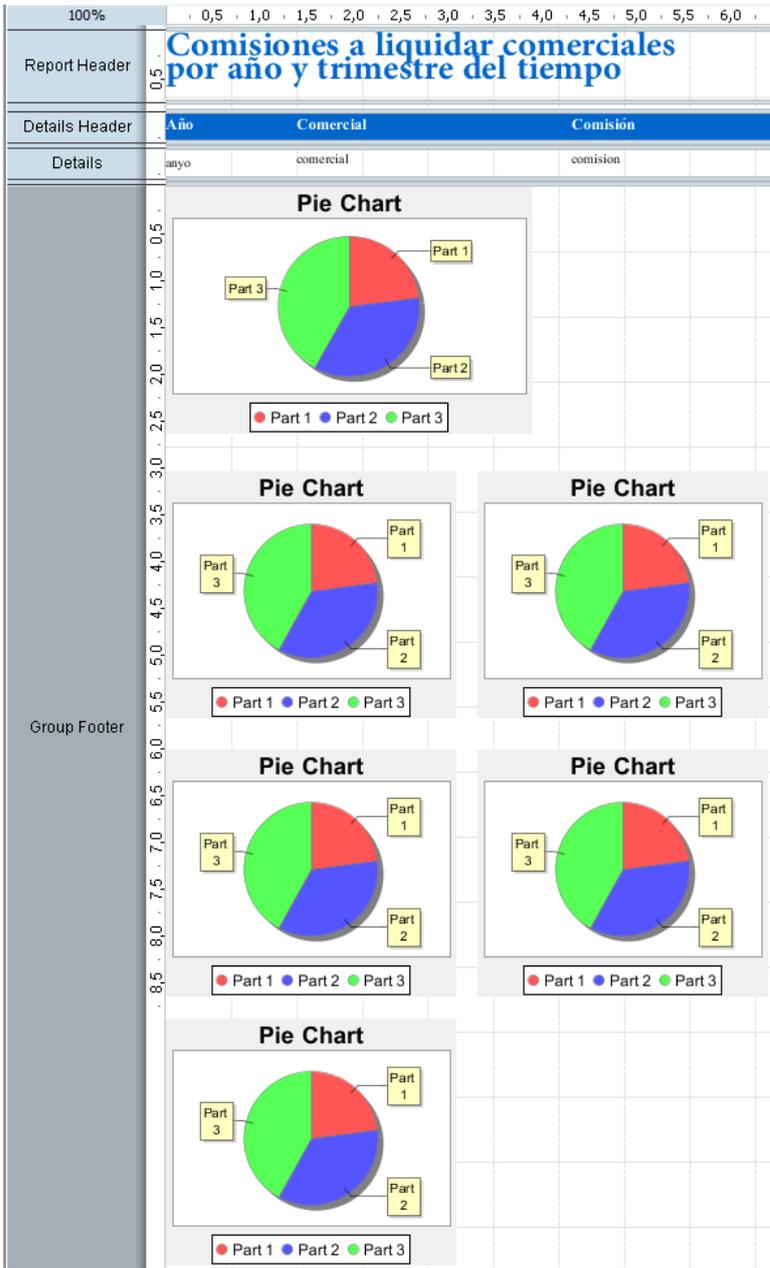


Ilustración 49 Estructura informe comisiones a liquidar a los comerciales por año y trimestre

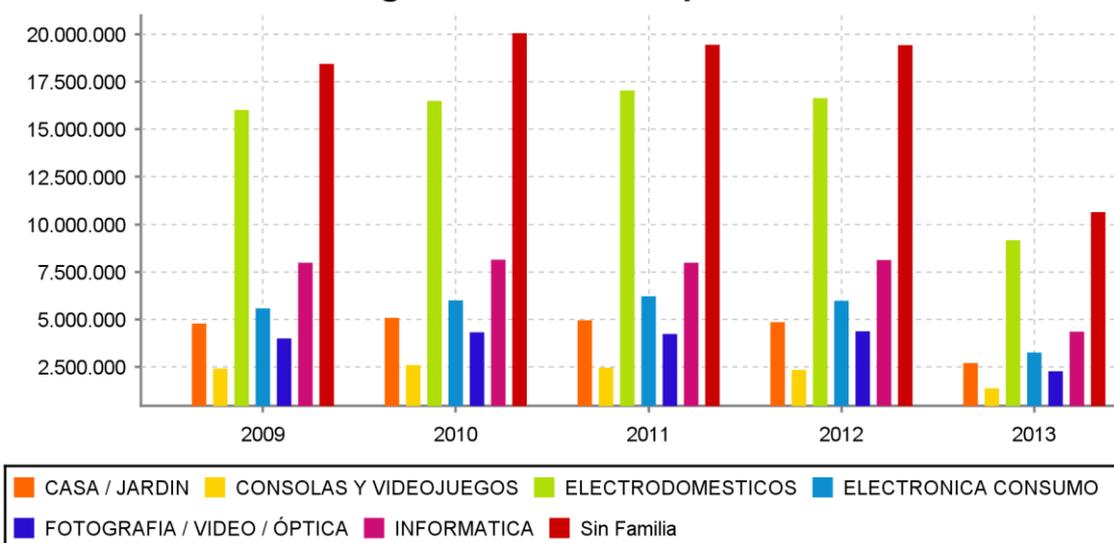
2.4.7. Margen anual obtenido por familia de productos

Proporcionalmente, el margen de beneficios se mantiene constante en el periodo que va del año 2009 al 2013 para cada una de las familias. Es importante destacar la gran cantidad de ventas que no se agrupan en ninguna familia, lo que dificulta poder hacer un análisis más detallado entre las ventas realizadas y el margen obtenido por familia.

Año	Familia	Margen
2009	CASA / JARDIN	4,791,653

2009	CONSOLAS Y VIDEOJUEGOS	2,422,066
2009	ELECTRODOMESTICOS	16,016,460
2009	ELECTRONICA CONSUMO	5,588,515
2009	FOTOGRAFIA / VIDEO / ÓPTICA	4,008,496
2009	INFORMATICA	7,987,242
2009	Sin Familia	18,438,965
2010	CASA / JARDIN	5,091,812
2010	CONSOLAS Y VIDEOJUEGOS	2,611,027
2010	ELECTRODOMESTICOS	16,493,951
2010	ELECTRONICA CONSUMO	6,011,147
2010	FOTOGRAFIA / VIDEO / ÓPTICA	4,331,808
2010	INFORMATICA	8,147,063
2010	Sin Familia	20,053,744
2011	CASA / JARDIN	4,961,311
2011	CONSOLAS Y VIDEOJUEGOS	2,471,703
2011	ELECTRODOMESTICOS	17,043,858
2011	ELECTRONICA CONSUMO	6,225,729
2011	FOTOGRAFIA / VIDEO / ÓPTICA	4,242,821
2011	INFORMATICA	7,989,087
2011	Sin Familia	19,443,097
2012	CASA / JARDIN	4,865,537
2012	CONSOLAS Y VIDEOJUEGOS	2,355,608
2012	ELECTRODOMESTICOS	16,638,629
2012	ELECTRONICA CONSUMO	5,986,912
2012	FOTOGRAFIA / VIDEO / ÓPTICA	4,384,617
2012	INFORMATICA	8,129,302
2012	Sin Familia	19,422,631
2013	CASA / JARDIN	2,713,044
2013	CONSOLAS Y VIDEOJUEGOS	1,387,894
2013	ELECTRODOMESTICOS	9,169,594
2013	ELECTRONICA CONSUMO	3,270,263
2013	FOTOGRAFIA / VIDEO / ÓPTICA	2,284,742
2013	INFORMATICA	4,364,139
2013	Sin Familia	10,645,461

Margen en las ventas por familia



2.4.7.1. Estructura del informe

El informe está formado por una tabla con el margen obtenido en las ventas agrupadas por familias de productos y año, seguido de un diagrama de barras que muestra esta información visualmente. Para obtener estos datos se ha utilizado el siguiente *data set* a partir de la consulta:

```

1 SELECT d_fecha.ano,d_familia.desc_familia,sum(h_ventas.importe - h_ventas.coste)
2 FROM
3   public.h_ventas,
4   public.d_familia,
5   public.d_fecha
6 WHERE h_ventas.id_familia = d_familia.id_familia AND d_fecha.id_fecha=h_ventas.id_fecha
7 GROUP BY d_familia.desc_familia,d_fecha.ano
8 ORDER BY d_fecha.ano,d_familia.desc_familia

```

Ilustración 50 Data set margen anual por familias de productos

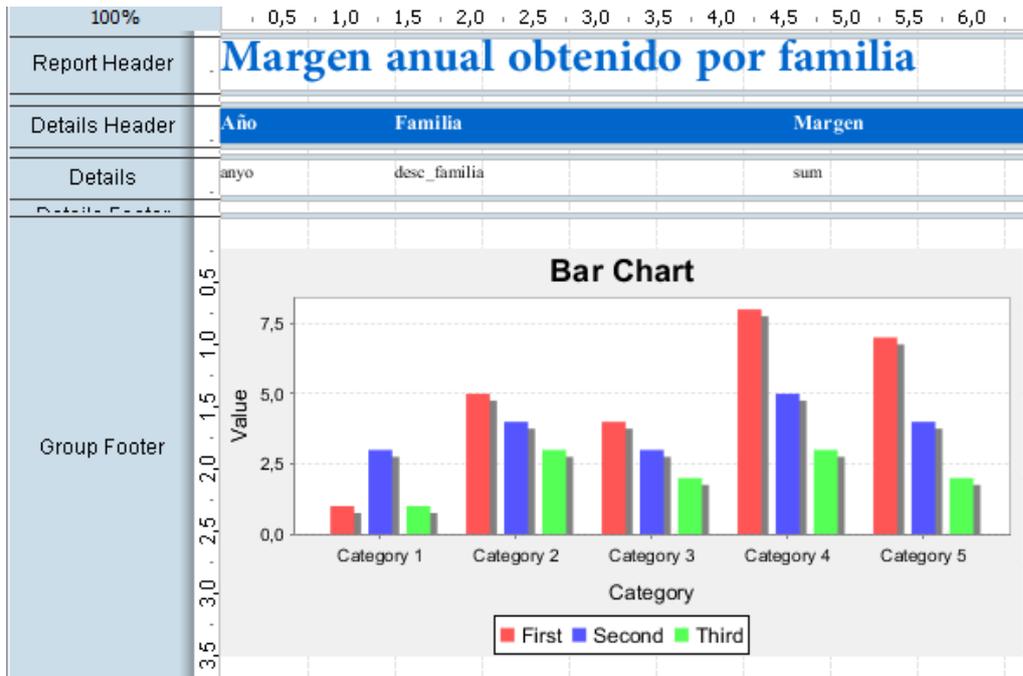


Ilustración 51 Estructura informe margen anual por familia de productos

3. Conclusiones

Una de las grandes ventajas de incorporar una plataforma BI es la posibilidad de recopilar, formatear y distribuir de manera efectiva la información de las diferentes fuentes de datos de una empresa. De esta forma, los profesionales pueden analizar información sobre las diferentes actividades de la empresa, algo que sería muy complicado hacer de otra forma dada la gran cantidad de datos y la heterogeneidad de los sistemas empresariales actuales. Disponer de este tipo de análisis proporciona una herramienta muy útil en la toma de decisiones y en la planificación estratégica de una compañía.

Este proyecto demuestra las enormes posibilidades que incorporan plataformas *open source* como *Pentaho*, que ponen al alcance de cualquier compañía un sistema BI muy útil para conocer la evolución del negocio. Aunque no incorporan todas las funcionalidad de las suites comerciales, en muchos casos pueden ser una opción muy interesante para empresas en las que el departamento de tecnologías de la información tiene un presupuesto ajustado.

Los responsables de la empresa TOTSALLES tenían la sensación de que las ventas iban bajando año tras año y, en consecuencia, se estaban viendo afectados sus márgenes de beneficio. En este sentido, el proyecto ha cumplido con los objetivos marcados, ya que no sólo se ha confirmado la percepción de los responsables, sino que además se ha generado información analítica desde diferentes puntos de vista: qué meses son los de menor actividad, cómo han evolucionado las ventas por provincia o cuáles son los productos estrella que reportan mayores beneficios. Con toda esta información van a poder tomar las decisiones oportunas para evitar seguir perdiendo cuota de mercado respecto a la competencia.

En cuanto a los hitos de planificación planteados durante el trabajo, éstos se han cumplido según lo previsto y se incluyen todos los entregables. Es importante señalar que la utilización del módulo *Pentaho Data Integration* ha requerido una dedicación extra para familiarizarse con las operaciones que permite ejecutar y cómo conseguir transformar los datos para que se ajusten al diseño del *Data Warehouse*.

Como línea de trabajo futuro destacaría explotar la funcionalidad OLAP del módulo Mondrian que incorpora *Pentaho*, ya que proporciona nuevas fuentes de datos en el momento de generar los informes finales. En este trabajo se ha optado por generar los informes directamente utilizando una conexión JDBC, dado que el volumen de los datos era razonablemente pequeño. Utilizar un modelo OLAP reduciría problemas de rendimiento en un entorno con un volumen de datos muy grande, dado que precalcula las posibles consultas multidimensionales.

4. Glosario

BI. Abreviatura de Business Intelligence.

DW. Abreviatura *Data Warehouse*

Mondrian. Servidor OLAP que incorpora Pentaho.

Middleware. Software para interconectar diferentes sistemas.

Open source. Programa informático que permiten el acceso a su código de programación permitiendo modificar el código por programadores diferentes de los que lo crearon.

Wizard. Asistente que guía al usuario en la realización de un determinado proceso.

5. Bibliografía

Se han utilizado las siguientes referencias bibliográficas a nivel de ayuda y consulta:

- **Curto Díaz, Josep (2012).** Introducción al Business Intelligence
- **Matteo Golfarelli (2010).** *open source* BI Platforms: a Functional and Architectural Comparison
- <http://www.redopenbi.com/> (Mayo 2015)
- <https://churriwifi.wordpress.com> (Mayo 2015)
- <http://wiki.pentaho.com/display/EALes/Inicio> (Mayo 2015)
- blog.espol.edu.ec/pechever/files/2009/04/manual-de-pentaho-etl-transformacion.doc (Mayo 2015)
- <http://www.pentaho.com/> (Marzo 2015)
- <http://www.spagobi.org/> (Marzo 2015)
- <https://www.jaspersoft.com/es> (Marzo 2015)

6. Anexos

6.1. Modelo físico del esquema intermedio

Modelo físico utilizado en el proceso de carga del *Data Warehouse* en el que se vuelca el contenido de los ficheros de datos. Esto permite utilizarlos como fuente de datos intermedio y realizar cómodamente la carga del *Data Warehouse*.

