# TFG – Área Data Warehouse

*Proyecto de Construcción y explotación de un almacén de datos para el análisis de información sobre alojamientos turísticos*

## MEMORIA

**Grado en Ingeniería Informática**

**17/06/2013**

**Autora: M Dolores Moyano Guerrero**

**Consultor: Carles Llorach Rius**

# DEDICATORIA

Dedico este proyecto a quien lo ha hecho posible y a lo más importante.

*Pedro, Alejandro y Esther*

# RESUMEN

En el presente documento se describe el trabajo realizado para crear un almacén de datos para el trabajo final de grado (TFG) en Ingeniería Informática.

En este documento se resumen las diferentes fases del proyecto, desde una planificación inicial donde se detallan la justificación y objetivos del proyecto, los requerimientos del cliente, pasando por el análisis y diseño, hasta la implementación del producto final en la suite Pentaho.

Los almacenes de datos proporcionan un expediente completo de una organización, más allá de la información transaccional y operacional, almacenando en una base de datos diseñada para favorecer el análisis y la divulgación eficiente de datos de forma variante en el tiempo, orientado a temas, no volátil e integrada.

PALABRAS CLAVE

Almacén de datos, Data Warehouse, MySQL, análisis multidimensional, ETL, SQL, Pentaho, Kettle, Saiku.

# TABLA DE CONTENIDOS

**1 INTRODUCCIÓN ..... 8**

1.1 Justificación del proyecto ..... 8

1.2 Objetivos del proyecto ..... 9

*1.2.1 Objetivos generales ..... 9*

*1.2.2 Objetivos específicos ..... 9*

1.3 Enfoque y método seguido ..... 10

1.4 Resultados esperados ..... 10

1.5 Planificación inicial vs final ..... 11

1.6 Análisis de riesgos ..... 14

*1.6.1 Plan de minimización de los riesgos ..... 15*

1.7 Productos obtenidos ..... 15

**2 ANÁLISIS ..... 16**

2.1 Requerimientos funcionales y no funcionales ..... 16

*2.1.1 Funcionales ..... 16*

*2.1.2 No funcionales ..... 17*

2.2 Diagrama de casos de uso ..... 19

2.3 Diagrama de la arquitectura software ..... 19

2.4 Diagrama de la arquitectura hardware ..... 20

**3 DISEÑO DE LA BASE DE DATOS ..... 21**

3.1 Modelo conceptual ..... 21

*3.1.1 Tabla de Hechos ..... 21*

*3.1.2 Dimensiones ..... 22*

**4 IMPLEMENTACIÓN ..... 23**

4.1 Proceso ETL ..... 24

*4.1.1 Primeras transformaciones ..... 24*

*4.1.2 Pentaho Data Integration ..... 26*

4.1.2.1 Automatización ..... 34

*4.1.3 Creación de tablas en el esquema ..... 35*

4.2 Modelo Multidimensional ..... 38

4.3 Informes y análisis ..... 39

**5 CONCLUSIONES ..... 54**

**6 LÍNEAS DE EVOLUCIÓN FUTURAS .....54**

**ANEXO: BIBLIOGRAFÍA .....55**

# TABLA DE ILUSTRACIONES

Ilustración 1: Hitos del proyecto ..... 11

*Ilustración 2: Diagrama de Gantt, PACs 1 y 2* ..... 13

*Ilustración 3: Diagrama de Gantt, PACs 3 y 4* ..... 14

*Ilustración 4: Diagrama de Casos de Uso* ..... 19

*Ilustración 5: Arquitectura software* ..... 20

*Ilustración 3: Arquitectura hardware* ..... 21

*Ilustración 6: Diseño final modificado* ..... 23

*Ilustración 7: Comarcas y municipios de Cataluña* ..... 25

Ilustración 8: Proceso ETL ..... 25

*Ilustración 9: Fichero establiments 2007 después de su transformación* ..... 26

*Ilustración 10: Primera Transformación* ..... 27

*Ilustración 11: Asignación del nº de habitantes* ..... 27

*Ilustración 12: Asignación del segmento y año correspondiente* ..... 28

*Ilustración 13: Unión que proporciona la comarca de "poblado"* ..... 28

*Ilustración 14: Obtención de la suma de habitantes por segmento* ..... 29

*Ilustración 15: Obtención de la suma de habitantes por comarca* ..... 29

Ilustración 16: Segunda transformación ..... 30

*Ilustración 17: Asignando valores de los campos* ..... 30

*Ilustración 18: Asignando valores fijos* ..... 31

*Ilustración 19: Transformación 3* ..... 31

*Ilustración 20: Transformación 8* ..... 32

*Ilustración 21: Transformación final* ..... 32

*Ilustración 22: Eliminación de campos no necesarios después de su tratamiento* ..... 33

*Ilustración 23: Renombrando campos de "DatComarcas"* ..... 33

*Ilustración 24: Exportación a MYSQL* ..... 34

*Ilustración 25: Estrella Oferta* ..... 38

*Ilustración 26: Estrella Equipaments* ..... 38

*Ilustración 27: Estrella habitantes* ..... 39

*Ilustración 28: Acceso a los reports* ..... 40

*Ilustración 29: Informe establecimientos y plazas (1)* ..... 41

*Ilustración 30: Informe establecimientos y plazas (2)* ..... 42

*Ilustración 31: Informe establecimientos y plazas - Detalle (por años)* ..... 43

*Ilustración 32: Informe establecimientos y plazas - Detalle (Gráfica)* ..... 43

*Ilustración 33: % Plazas con respecto a la población* ..... 44

*Ilustración 34: Oferta media de plazas* ..... 45

*Ilustración 35: Número de establecimientos / Número de equipamientos* ..... 46

*Ilustración 36: % Población por equipamiento* ..... 47

*Ilustración 37: Cantidad de plazas ofrecidas / superficie del territorio (1)* ..... 48

*Ilustración 38: Cantidad de plazas ofrecidas / superficie del territorio (2)* ..... 49

*Ilustración 39: Habitantes por población* ..... 50

*Ilustración 40: Habitantes por comarca* ..... 51

Ilustración 41: Establecimientos vs habitantes por género ..... 52

Ilustración 42: Plazas vs Personas ..... 52

Ilustración 43: Equipamientos vs Población ..... 53

Ilustración 44: Establecimientos y plazas por comarca y año ..... 53

# 1 Introducción

En el presente documento se describe el proceso de creación de un almacén de datos para el trabajo final de grado (TFG) denominado *"Proyecto de Construcción y explotación de un almacén de datos para el análisis de información sobre alojamientos turísticos".*

El Observatorio Nacional de Empleo (Ondo) quiere profundizar en la evolución de los establecimientos turísticos y sus posibles correlaciones entre alojamientos y equipamientos públicos, para ello dispone de ficheros de texto con información sobre los equipamientos, públicos, la población y los establecimientos.

En documento se resumen diferentes fases del proyecto: la primera fase comprende el **Plan de trabajo y análisis preliminar de requerimientos,** donde se incluye una planificación estimada de las diferentes tareas a realizar para llevar a cabo el proyecto, esta planificación, su concreción y desviaciones será objeto de estudio. En esta fase se incluye un análisis preliminar con la enumeración y breve descripción de los elementos de análisis identificados que estarán disponibles para los usuarios y el número de informes aproximado que se implementarán y contenido de los mismos. También se analizarán las fuentes de datos operacionales proporcionados que servirán para cargar cada uno de los elementos de análisis. La fase de **Análisis de requerimientos y diseño conceptual y técnico,** comprende un documento con el análisis detallado de requerimientos basado en el análisis preliminar realizado, además de un documento de diseño con la descripción del modelo dimensional que dará soporte a las necesidades de los usuarios. La etapa de **Implementación** consta de las tareas: Construcción del almacén de datos; configuración de la herramienta de explotación de datos y construcción de los informes y análisis de la información.

Esta **memoria** del proyecto, recopila todo el trabajo realizado.

## 1.1 Justificación del proyecto

La **actividad turística** ha tomado un papel preponderante en la economía catalana, actuando como factor de arrastre de otras actividades económicas, contrarrestando las dificultades de expansión que han sufrido los sectores primario y secundario. Tanto por la capacidad de alojamiento (aproximadamente 2,5 millones de plazas), como por el volumen

de turismo exterior que recibe (unos 14,4 millones en 2012[1]), Cataluña es la primera región turística de Europa[2].

Ante esta situación, el Observatorio Nacional de Empleo (Ondo) quiere profundizar en la evolución de este tipo de establecimientos y analizar las posibles relaciones entre alojamientos y equipamientos públicos.

Mediante las herramientas y técnicas ELT (extraer, cargar y transformar), o ETL (extraer, transformar y cargar) se extraen los datos de distintas fuentes, se depuran y preparan (homogeneización de los datos) para luego cargarlos en un almacén de datos. Esta herramienta de apoyo al estudio de la evolución de la capacidad de alojamiento en relación con los equipamientos públicos, enfocada desde la perspectiva del área de Data Warehouse será el objetivo de trabajo del presente proyecto.

En un almacén de datos se contienen datos necesarios o útiles para una organización, que se utiliza como un repositorio de datos para posteriormente transformarlos en información útil para el usuario. El almacén de datos da respuesta a las necesidades de usuarios expertos, utilizando Sistemas de Soporte a Decisiones (DSS), Sistemas de información ejecutiva (EIS) o herramientas para hacer consultas o informes, integrando datos provenientes de bases de datos heterogéneas y distribuidas por las diferentes unidades de la organización.

## 1.2 Objetivos del proyecto

### 1.2.1 Objetivos generales

Los objetivos generales de este proyecto son:

- Adquirir y mejorar conocimientos concretos en el diseño, construcción y explotación de un Data Warehouse.
- Desarrollar las habilidades para analizar, planificar y gestionar un proyecto.

### 1.2.2 Objetivos específicos

El objetivo de este TFG será aplicar una serie de métodos, técnicas y herramientas que permitan llevar a cabo de forma sistemática el proceso de creación de un almacén de datos para extraer información con los datos recibidos de un "*Proyecto de Construcción y*

[1] Fuente: Instituto de Estudios Turísticos

[2] El turismo en Cataluña JOAOUIM MOLINS I AMAT

*explotación de un almacén de datos para el análisis de información sobre alojamientos turísticos* ", bajo un alcance temporal definido.

El proceso comienza con la recepción de un conjunto de archivos del Observatorio Nacional de Empleo (Ondo) en formato texto separados por comas. A partir de este fichero y de los requerimientos de usuario, se realizará la implementación del almacén de datos corporativo por fases.

## 1.3 Enfoque y método seguido

Para realizar el sistema propuesto para el TFG se ha seguido un método del ciclo de vida, que es un enfoque por fases del análisis y diseño que sostiene que los sistemas son desarrollados de mejor manera mediante el uso de un ciclo especifico de actividades del analista y del usuario. Comprende las siguientes etapas:

- Estudio preliminar
- Determinación de los requerimientos
- Diseño del sistema
- Desarrollo
- Pruebas
- Implantación y evaluación

## 1.4 Resultados esperados

El resultado esperado es un producto que cumpla con todos los requerimientos solicitados por el cliente, en el tiempo pactado. Para controlar el desarrollo y ejecución del proyecto, se ha establecido una planificación con unos hitos fechados intermedios para controlar el desarrollo del proyecto.

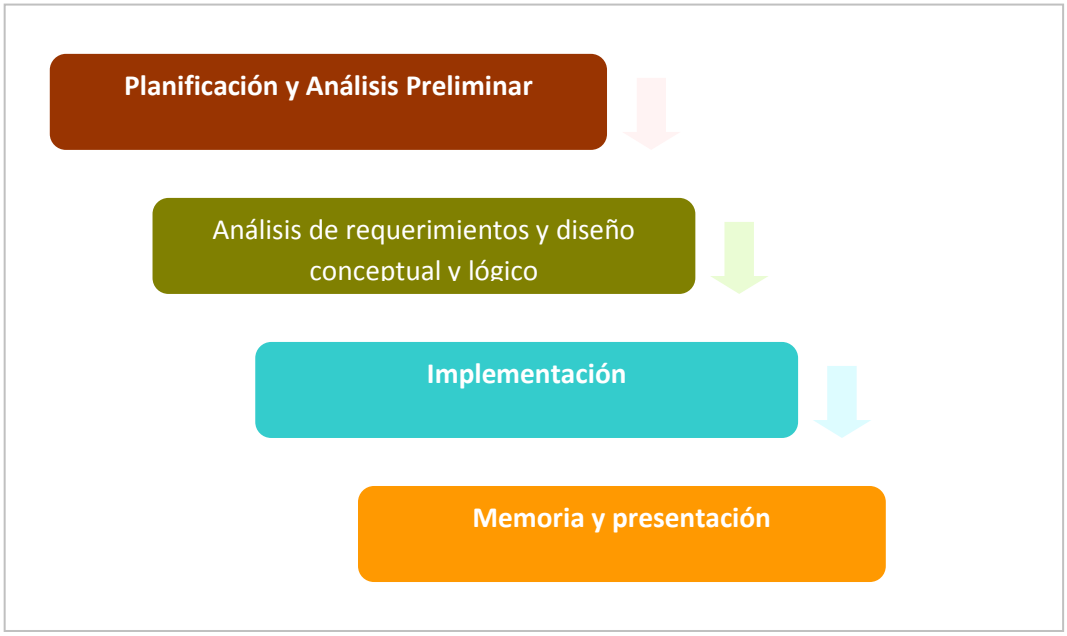



**Ilustración 1: Hitos del proyecto**

**Etapa 1: Plan de Trabajo.** El resultado esperado es un documento con la planificación y el análisis preliminar en el plazo previsto.

**Etapa 2: Análisis de requerimientos y diseño conceptual y técnico**. El resultado es un documento con el análisis detallado de requerimientos basado en el análisis preliminar realizado, con la descripción del modelo dimensional y el diseño de los procedimientos de extracción de datos a alto nivel.

**Etapa 3: Implementación.** El resultado esperado es la entrega del propio almacén de datos.

**Etapa 4: Presentación y memoria virtual.** El resultado esperado es la entrega final con un compendio de los resultados de las etapas anteriores y la entrega de una presentación en vídeo.

## 1.5 Planificación inicial vs final

A continuación se incluye la tabla de tareas en la que se han descompuesto los trabajos, dentro de la cual, se subrayan los hitos importantes del proyecto.

| CÓD | NOMBRE | COMIENZO | FIN | DÍAS |
|---|---|---|---|---|
| **1** | **PAC 1 - PLANIFICACIÓN Y ANÁLISIS PRELIMINAR** | 28/02/2013 | 12/03/2013 | 13 |
| 2 | Lectura materiales | 28/02/2013 | 28/02/2013 | 1 |
| 3 | Definición del proyecto | 28/02/2013 | 02/03/2013 | 3 |
| 4 | Estudio de los hitos a realizar | 03/03/2013 | 03/03/2013 | 1 |
| 5 | Establecimiento del plan de trabajo | 04/03/2013 | 05/03/2013 | 2 |
| 6 | Análisis de las fuentes de datos | 06/03/2013 | 07/03/2013 | 2 |
| 7 | Realización del análisis preliminar | 08/03/2013 | 09/03/2013 | 2 |
| 8 | Redacción del documento de análisis | 10/03/2013 | 10/03/2013 | 1 |
| 9 | Redacción del Plan de Trabajo | 11/03/2013 | 11/03/2013 | 1 |
| 10 | Entrega de la PAC 1 | 12/03/2013 | 12/03/2013 | 1 |
| **11** | **PAC 2 - ANÁLISIS DE REQUERIMIENTOS Y DISEÑO CONCEPTUAL Y TÉCNICO** | 13/03/2013 | 16/04/2013 | 35 |
| 12 | Lectura materiales | 13/03/2013 | 13/03/2013 | 1 |
| 13 | Incorporación recomendaciones consultor | 14/03/2013 | 14/03/2013 | 1 |
| 14 | Aprendizaje uso herramientas | 14/03/2013 | 15/03/2013 | 2 |
| 15 | Análisis de requerimientos | 15/03/2013 | 17/03/2013 | 3 |
| 16 | Diseño conceptual | 18/03/2013 | 24/03/2013 | 7 |
| 17 | Diseño lógico | 25/03/2013 | 01/04/2013 | 8 |
| 18 | Diseño de la extracción de datos de origen | 02/04/2013 | 09/04/2013 | 8 |
| 19 | Transformación datos de origen | 10/04/2013 | 14/04/2013 | 5 |
| 20 | Elaboración del documento de entrega de la PAC 2 | 15/04/2013 | 16/04/2013 | 2 |
| 21 | Entrega de la PAC 2 | 16/04/2013 | 16/04/2013 | 1 |
| **22** | **PAC 3 – IMPLEMENTACIÓN** | 17/04/2013 | 29/05/2013 | 43 |
| 23 | Letura materiales | 17/04/2013 | 17/04/2013 | 1 |
| 24 | Incorporación recomendaciones consultor | 18/04/2013 | 18/04/2013 | 1 |
| 25 | Aprendizaje uso herramientas | 18/04/2013 | 19/04/2013 | 2 |
| 26 | Configuración software | 19/04/2013 | 20/04/2013 | 2 |
| 27 | Construcción del Almacén de Datos | 21/04/2013 | 27/04/2013 | 7 |
| 28 | Elaboración del script de la carga de datos | 28/04/2013 | 05/05/2013 | 8 |
| 29 | Carga de datos | 06/05/2013 | 09/05/2013 | 4 |
| 30 | Pruebas ETL | 10/05/2013 | 13/05/2013 | 4 |
| 31 | Creación de los informes solicitados | 14/05/2013 | 23/05/2013 | 10 |
| 32 | Pruebas de los informes | 24/05/2013 | 27/05/2013 | 4 |
| 33 | Elaboración del documento de entrega de la PAC 3 | 28/05/2013 | 29/05/2013 | 2 |
| 34 | Entrega de la PAC 3 | 29/05/2013 | 29/05/2013 | 1 |
| **35** | **ENTREGA FINAL Y DEFENSA** | 30/05/2013 | 17/06/2013 | 19 |
| 36 | Incorporación recomendaciones consultor | 30/05/2013 | 30/05/2013 | 1 |
| 37 | Elaboración de la memoria | 30/05/2013 | 08/06/2013 | 10 |
| 38 | Elaboración de la presentación | 09/06/2013 | 16/06/2013 | 8 |

| | | | | |
|---|---|---|---|---|
| 39 | Entrega de la memoria, la presentación y el producto | 17/06/2013 | 17/06/2013 | 1 |

A continuación se expone el diagrama inicial de Gantt, donde se puede observar la planificación del trabajo final de grado al completo descrito anteriormente. Se realiza un cronograma en cascada.

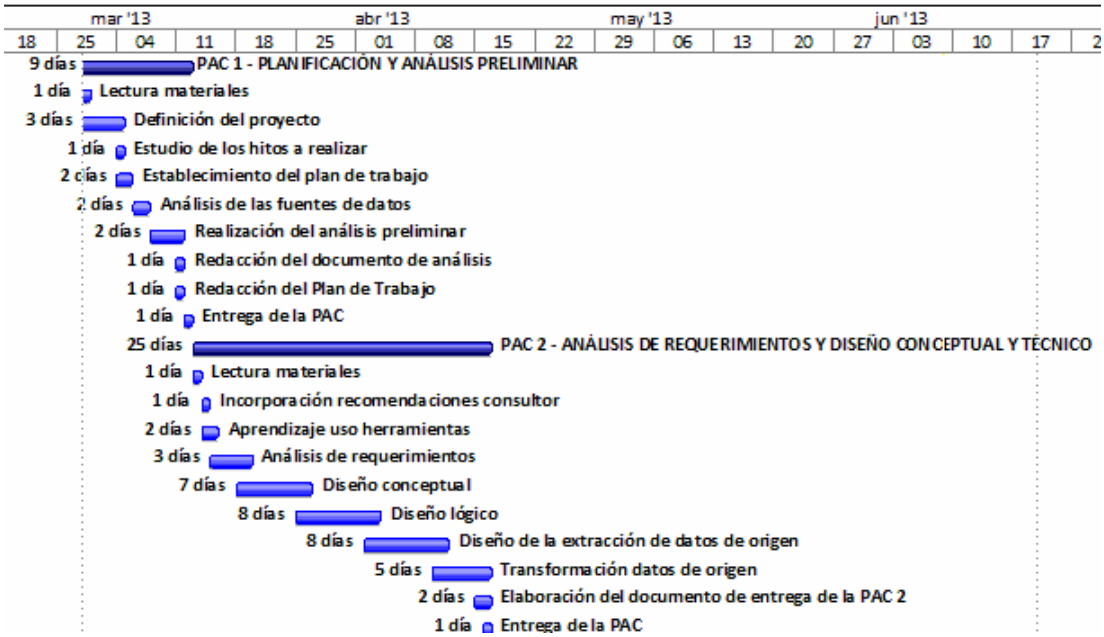



***Ilustración 2: Diagrama de Gantt, PACs 1 y 2***

Para las PACs 3 y 4 se tiene el siguiente diagrama de Gantt:

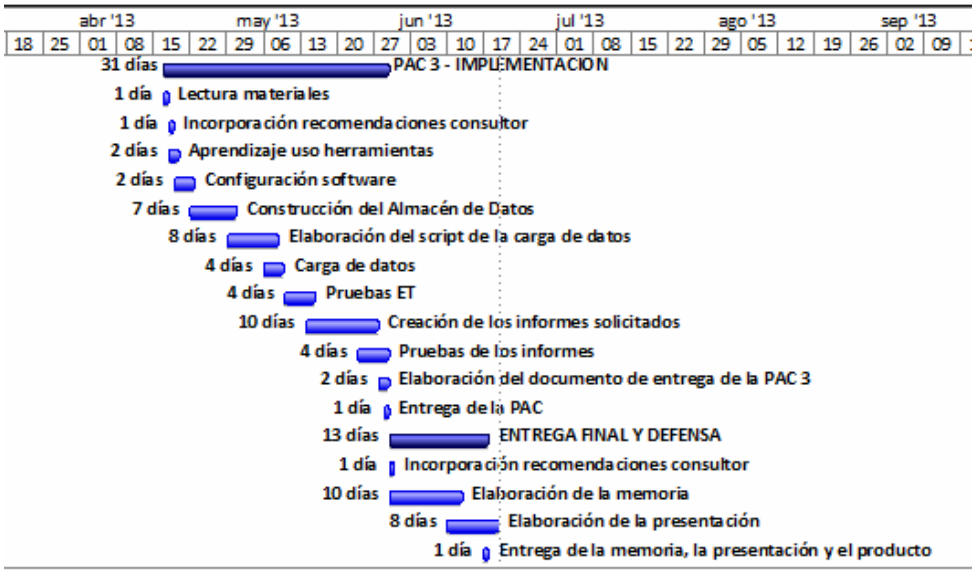



*Ilustración 3: Diagrama de Gantt, PACs 3 y 4*

La planificación se ha podido llevar a cabo en la práctica, con algunos problemas para la PAC 3, la implementación, donde se ha vuelto a analizar el producto, modificando el modelo multidimensional y realizar el proceso completo de extracción, transformación y carga de datos, que ya se presentó en la PAC2, por otro en Kettle, mucho más elegante. Todo ello se unió a un problema no contemplado en la planificación, y es la coincidencia en el tiempo con otra entrega de una asignatura diferente que también requirió mucho esfuerzo en tiempo de trabajo. Se corrigió este problema replanificando e invirtiendo más tiempo del inicialmente estimado en horas cada día.

## 1.6 Análisis de riesgos

Los riesgos más probables que se pueden encontrar en la realización del proyecto son los siguientes:

- Posibles imprevistos personales, teniendo en cuenta la situación familiar, con descendencia en edad preescolar.
- Retrasos por el aprendizaje de nuevas herramientas.
- Imprevistos laborales.
- Errores en la recogida de requisitos.
- Pérdida de los materiales por accidentes imprevistos.

### 1.6.1 Plan de minimización de los riesgos

El plan de contingencia ante los riesgos supone desarrollar las siguientes estrategias:

1. Para la planificación inadecuada para completar el proyecto se realizará un **seguimiento y replanificación diaria** pesimista ante los posibles imprevistos. Si se produjera una causa de fuerza mayor, habrá que ponerlo en conocimiento del consultor.

2. Para evitar los retrasos en el aprendizaje, se harán las **estimaciones de tiempo pesimistas** para realizar dichas labores.

3. Para evitar errores en la recogida de requisitos y que ello redunde en una mala dirección del proyecto, se consultarán las **dudas al consultor**, para minimizar el riesgo y poder reconducir las posibles desviaciones.

4. Con el fin de evitar la pérdida de los materiales se realizará un **respaldo diario en la nube** de todo el proyecto. Se ha creado un **proceso automatizado para cargar los datos** en Pentaho y crear las tablas en SQL, guardando los archivos del proceso y ejecutando la carga automatizada, se podría recomponer el almacén de datos.

## 1.7 Productos obtenidos

Los productos que se han obtenido son una serie de informes y análisis, los establecidos en los requisitos del sistema más algunos adicionales que se han considerado interesantes, junto con algunas gráficas que muestran gráficamente los datos.

Los informes obtenidos son:

- Establecimientos por comarca
- Establecimientos y equipamiento
- Establecimientos y plazas
- Establecimientos y plazas-detalle
- Habitantes por Comarca
- Habitantes
- Oferta media de plazas

- Plazas por extensión
- Plazas y población
- Población y equipamiento
- Análisis de establecimientos y plazas por año y comarca (Saiku)

A continuación se exponen las diferentes fases del proyecto: el **análisis** que comprende el análisis detallado de requerimientos basado en el análisis preliminar realizado, el **diseño**, con la descripción del modelo dimensional que dará soporte a las necesidades de los usuarios y la **Implementación** que consta de las tareas: Construcción del almacén de datos; configuración de la herramienta de explotación de datos y construcción de los informes y análisis de la información.

# 2 Análisis

En este segundo apartado se realizará un análisis detallado de los requisitos del Sistema, para ello se estudiarán las fuentes de datos proporcionados y los requisitos planteados para posteriormente realizar el diseño del modelo para el almacén de datos.

## 2.1 Requerimientos funcionales y no funcionales

Para comenzar el análisis, se van a describir los requerimientos el Data Warehouse tanto funcionales como no funcionales. Los primeros definen qué debe de hacer el sistema a desarrollar y los segundos cómo debe de ser el sistema.

### 2.1.1 Funcionales

Los requerimientos funcionales definen las operaciones que el sistema será capaz de realizar y describen las características que deberá contener la información alojada.

El sistema será capaz de, a partir de un conjunto de datos de entrada, producir una información de salida, informes e indicadores:

Los datos se reciben del Observatorio Nacional de Empleo (ONDO) vienen almacenados en ficheros csv (del inglés comma-separated values).

Los informes de salida permitirán consultar de forma agregada, por comarca / provincia, tipo de establecimiento y categoría información sobre dichos establecimientos ubicados en

poblaciones de Cataluña, además de información sobre la población de los mismos y el equipamiento de uso público con el que cuentan. La temporalidad de los datos será anual. Los informes disponibles serán los siguientes:

- Total de establecimientos
- Total de plazas
- % de plazas respecto a la población
- Oferta media de plazas
- Número de establecimientos / Número de equipamientos
- % de población por equipamiento
- Cantidad de plazas ofrecidas / superficie del territorio
- Número de habitantes
- Indicador de establecimientos vs habitantes por género
- Indicador de plazas vs personas
- Indicador de equipamientos vs población

### 2.1.2 No funcionales

Los requerimientos no funcionales son restricciones o limitaciones ofrecidas por el sistema o en el proceso de desarrollo. Según la **ISO 9126**, para la evaluación de la calidad del software, las características que definen la calidad del software son las siguientes:

**Funcionalidad** - *Las funciones son aquellas que satisfacen las necesidades implícitas o explícitas. Se pueden definir dentro de esta característica la idoneidad (la presencia y aptitud de un conjunto de funciones para tareas especificadas), la exactitud (disposición de resultados o efectos correctos o acordados), interoperabilidad (habilidad para la interacción con sistemas especificados, seguridad (habilidad para prevenir acceso no autorizado ya sea accidental o deliberado, a programas y datos) y el cumplimiento de normas*. El proyecto cumple todos los requisitos solicitados por el cliente, se ha incluido alguno más, por parecer interesantes para el análisis que se quiere realizar.

**Fiabilidad** - *Conjunto de atributos relacionados con la capacidad del software de mantener su nivel de prestación bajo condiciones establecidas durante un período establecido. Como*

*elementos de la fiabilidad, se pueden medir la madurez (relacionada con la frecuencia de falla por fallas en el software) la recuperabilidad (capacidad para restablecer su nivel de desempeño y recuperar los datos directamente afectos en caso de falla y en el tiempo y esfuerzo relacionado para ello), la tolerancia a fallos (habilidad para mantener un nivel especificado de desempeño en casos de fallas de software o de una infracción a su interfaz especificada).* En este caso la plataforma software sobre la que se realiza el análisis está desarrollada y cuenta con una comunidad de usuarios muy elevada que certifican su calidad y madurez.

**Usabilidad** - *Un conjunto de atributos relacionados con el esfuerzo necesario para su uso, y en la valoración individual de tal uso, por un establecido o implicado conjunto de usuarios. Dentro de la usabilidad se pueden evaluar el aprendizaje (relacionado con el esfuerzo de los usuarios para reconocer el concepto lógico y sus aplicaciones), la comprensión (esfuerzo de los usuarios para reconocer el concepto lógico y sus aplicaciones), la operatividad (esfuerzo de los usuarios para la operación y control del software) y la atractividad*. Como el caso anterior, la plataforma software está desarrollada y es usable.

**Eficiencia** - Conjunto de atributos relacionados con la relación entre el nivel de desempeño del software y la cantidad de recursos necesitados bajo condiciones establecidas. Dentro de la eficiencia se tiene el comportamiento en el tiempo (tiempos de respuesta y procesamiento y en las tasas de rendimientos en desempeñar su función) y el comportamiento de recursos. Se ha estudiando una estructura, con varias estrellas y tablas agregadas, que permite cumplir el requisito de eficiencia.

**Mantenibilidad** - *Conjunto de atributos relacionados con la facilidad de extender, modificar o corregir errores en un sistema software. Estabilidad (riesgo de efectos inesperados por modificaciones), facilidad de análisis (esfuerzo necesario para el diagnostico de deficiencias o causas de fallos, o identificaciones de partes a modificar), facilidad de cambio (esfuerzo necesario para la modificación, corrección de falla, o cambio de ambiente) y facilidad de pruebas (esfuerzo necesario para validar el software modificado).* El proceso de ETL es totalmente mecánico, visual con la plataforma Kettle y facilita incluir modificaciones, con lo que se facilita la mantenibilidad del producto.

**Portabilidad** - *Conjunto de atributos relacionados con la capacidad de un sistema software para ser transferido desde una plataforma a otra. Capacidad de instalación (esfuerzo necesario para instalar el software en un ambiente especificado), capacidad de reemplazamiento (oportunidad y esfuerzo de usar el software en lugar de otro software especificado en el ambiente de dicho software especificado), adaptabilidad (oportunidad para la adaptación del software a diferentes ambientes especificados sin aplicar otras acciones o medios que los proporcionados para este propósito por el software considerado) y la coexistencia.* Al ser el proceso de ETL mecánico, el sistema es fácilmente portable

copiando la estructura donde se guardan los ficheros, y los scripts que generan las tablas, realizan la ETL y los informes realizados en Pentaho Reports para publicarlos.

## 2.2 Diagrama de casos de uso

Se considera un único administrador, que administra el gestor de BBDD MySQL, y además el Sistema de Bussines Inteligence Pentaho. El usuario se encarga de ejecutar las consultas para analizar la información que proporciona el sistema.

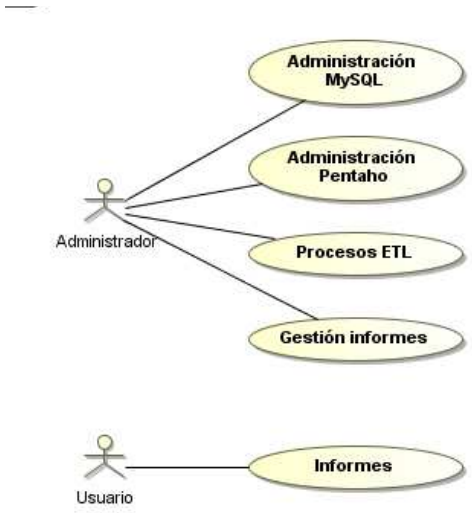



*Ilustración 4: Diagrama de Casos de Uso*

## 2.3 Diagrama de la arquitectura software

Partiendo de los datos en formato texto separados por comas (csv), se ha realizado un proceso de transformación para exportarlos al SGBD MySQL. En primer lugar, se ha realizado una pequeña transformación en Excel antes de cargarlos y transformarlos con PDI (Pentaho Data Integration), resultando como salida los datos del Esquema final del Sistema en MySQL. MySQL es el SGBD que contendrá la información sobre la que realizará los análisis el sistema de BI Pentaho. Los informes se han realizado con Pentaho Report

Designer (PRD), importando el cubo con Pentaho BI Server. El documento final se ha elaborado con Microsoft Word.

*Ilustración 5: Arquitectura software*

## 2.4 Diagrama de la arquitectura hardware

El diagrama de la arquitectura hardware del Sistema tiene los siguientes componentes:

- Servidores: Los que alojan el SGBD y el software Pentaho.
- Equipo Administrador: PC del administrador del SGBD y el Sistema de BI.
- Equipos usuarios: PCs de los usuarios finales de la aplicación.
- Red de comunicaciones: Red que comunica el sistema.

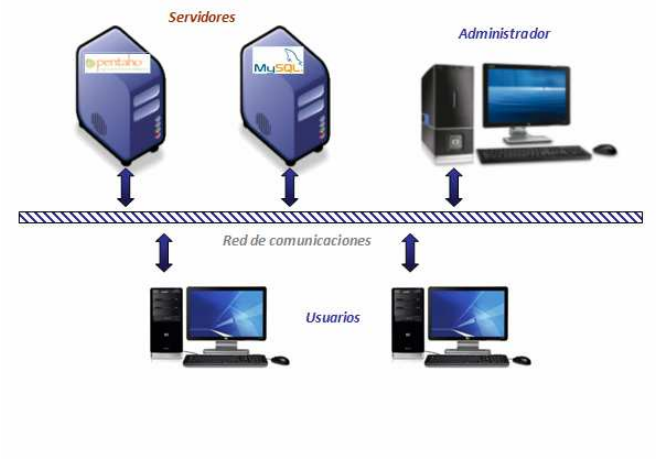



*Ilustración 6: Arquitectura hardware*

# 3 Diseño de la base de datos

En este apartado se realizará una explicación del nuevo modelo de diseño de la base de datos.

## 3.1 Modelo conceptual

El modelo conceptual multidimensional ha de representar los elementos que permitan al usuario obtener la información necesaria y suficiente para el apoyo a la toma de decisiones.

En una primera aproximación, una estructura de **constelación**, con dos tablas de hechos. Después del trabajo con los datos proporcionados, se ha considerado el modelo en **estrella** (Star Scheme) con varias estrellas de datos, para mejorar el rendimiento, independientemente al tipo de arquitectura, este tipo de esquema requiere que toda la estructura de datos este desnormalizada o semi desnormalizada, para evitar desarrollar uniones (Join) complejas para acceder a la información. Se han minimizado los join y guardado miembros calculados.

### 3.1.1 Tabla de Hechos

El hecho es la actividad que se va a centrar el objeto de estudio, y que será la tabla central del esquema dimensional. Estas tablas centrales serán:

- **Oferta**. Los datos que se tienen son el número de establecimientos y las plazas en un periodo de tiempo determinado. Cada registro se corresponde a

establecimiento, junto con el número de establecimientos y plazas de ese establecimiento en un año determinado (granularidad del sistema).

- **Habitantes.** Datos de la población a lo largo del tiempo. Los datos se proporcionan anualmente.
- **Equipamientos.** Datos de los equipamientos. Los datos no se proporcionan en ninguna dimensión temporal, se corresponden a poblaciones referenciadas por su nombre y código postal.

### 3.1.2 Dimensiones

Tras analizar el hecho, se van a describir las dimensiones de la estructura. Las dimensiones escogidas serán (3 estrellas):

Hecho Habitantes:

- **Tiempo**: Contiene los años objeto de estudio.
- **Poblacio**: Contiene los datos comunes de las poblaciones.

Hecho Oferta:

- **Comarcas**: Es una lista de comarcas de Cataluña, junto con datos calculados para mejorar el rendimiento de la base de datos, como el total de equipamientos por comarca, y la extensión de la comarca.
- **Población (Comarcas)**: Contiene los datos de las poblaciones por comarca y segmento (campo calculado). Se utiliza para mejorar el rendimiento.
- **Tiempo**: Contiene datos de los años de los que hay datos.

Hecho Equipamientos:

- **Tiempo**: Contiene los años objeto de estudio.
- **Comarcas**: Es una lista de comarcas de Cataluña, junto con datos calculados para mejorar el rendimiento de la base de datos, como el total de equipamientos por comarca, y la extensión de la comarca.
- **Población (Comarcas)**: Contiene los datos de las poblaciones por comarca y segmento (campo calculado). Se utiliza para mejorar el rendimiento.

  Se estudia como una mejora, el poder conectar la tabla de hecho “Equipaments” con la poblacio (dimensión del hecho Habitantes). No se ha podido realizar en esta

primera versión porque sólo se dispone del código postal, dato distinto al Código INE, y no es inmediata y automática el poder realizar un enlace de las tablas. Además, el nombre de las poblaciones de "Equipaments" no están normalizadas y no se corresponden con los nombres de poblaciones de la tabla "Poblacio".

Se observa la jerarquía Comarca-Población en la estrella de habitantes. Para obtener los datos de las tablas agregadas DatComarcas y PobComarcas, se utiliza una estrategia de Drill-Up, totalizando los datos por comarcas dentro de esa jerarquía.

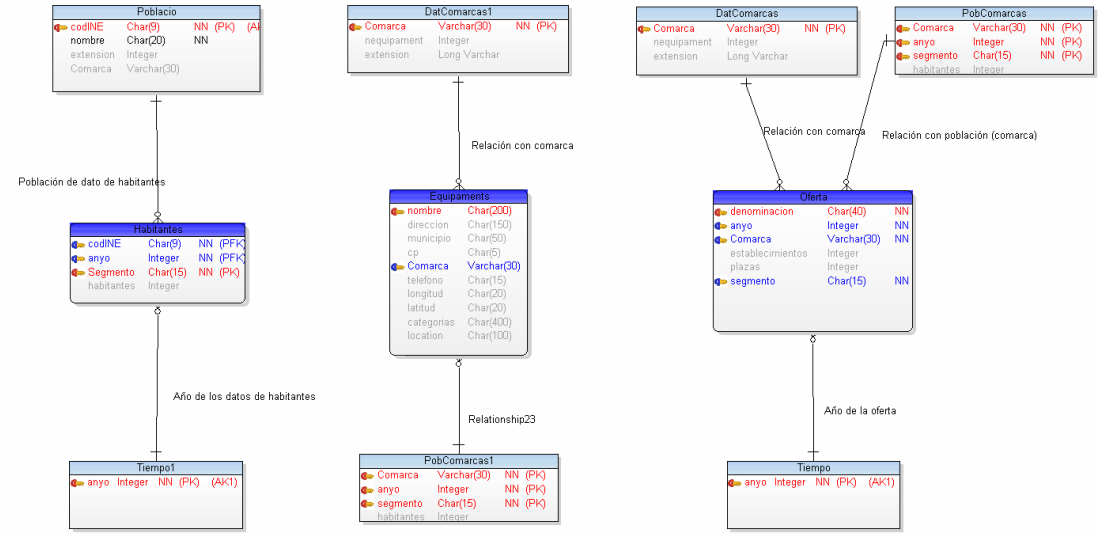



*Ilustración 7: Diseño final modificado*

# 4 Implementación

Una vez comentados los cambios en el diseño del sistema, se concretará la arquitectura hardware y software y el proceso ETL (**E**xtract, **T**ransform, **L**oad) de los datos para convertirlos en información.

## 4.1 Proceso ETL

En el proceso ETL, primero se adecúan con Excel los archivos en texto plano para ser cargados en PDI (Pentaho Data Integration), resultando finalmente la carga en para ser cargados finalmente en el Esquema de MySQL.

### 4.1.1 Primeras transformaciones

En el proceso de ETL, en primer lugar se ha trabajado con los ficheros de datos proporcionados:

- Equipaments.csv: Que contiene los equipamientos con los datos de localización.
- establimentsXXXX.txt: Donde XXXX es el año. Contiene los datos de los establecimientos turísticos por años. **Se han eliminado las cabeceras, incluido nombres de campo y convertido los archivos en texto separados por comas, Además se han eliminado unos registros que hacían agrupaciones de comarcas.**
- Poblacio.csv: Contiene los datos de población por años de las localidades de Cataluña, **se han eliminado los caracteres "·". En algún registro con dato (nd), se ha modificado el valor por "0", para evitar problemas con las conversiones a numérico**.

En cuanto a la calidad del dato, los distintos ficheros se han guardado en formato UTF-8 para evitar problemas con los caracteres especiales (`´ñç…), no se ha desechado ningún dato por erróneo, y se ha utilizado además un fichero llamado "comarcas_municipios_catalunya.txt", que relaciona los códigos INE con las comarcas de Cataluña. "comarcas.txt" es un listado de las comarcas de Cataluña y "tiempo.txt" son los años sobre los que se realiza el estudio.

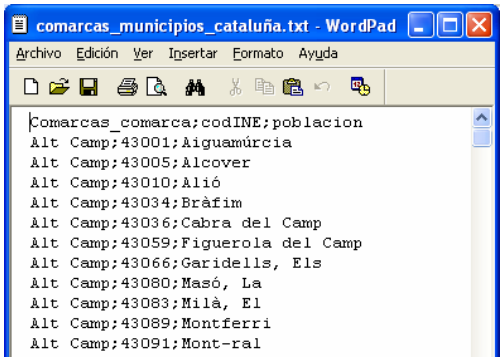



*Ilustración 8: Comarcas y municipios de Cataluña*

Así, los archivos de entrada que se van a utilizar en el proceso en PDI (Kettle o Pentaho Data Integration) son:

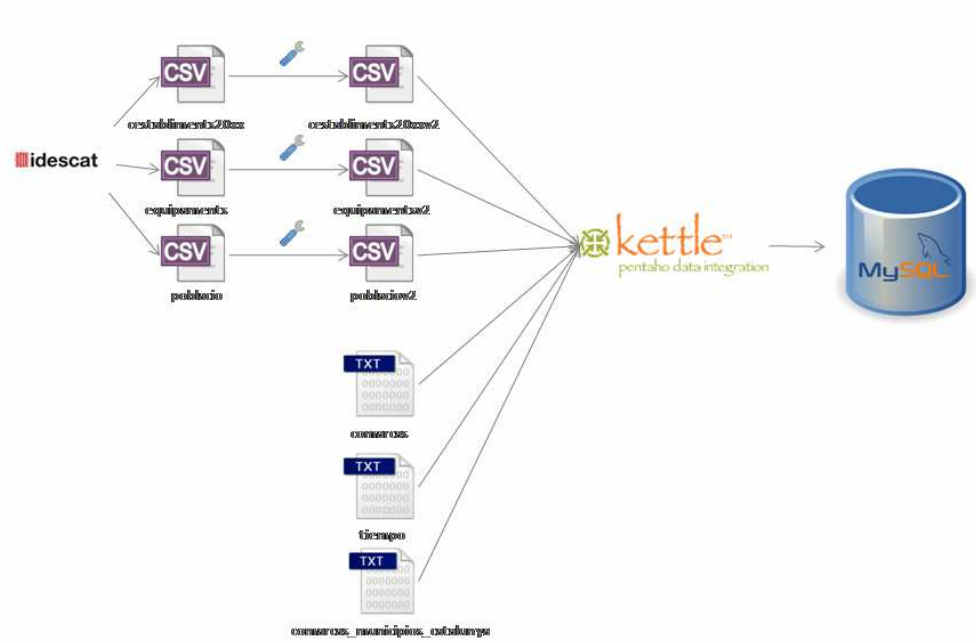



**Ilustración 9: Proceso ETL**

- Proporcionados por el cliente
  - cestabliments xxxx.csv
  - equipaments.csv
  - poblacio.csv

- Otros archivos necesarios para tratar los datos
    - comarcas.txt
    - comarcas_municipios_catalunya.txt
    - tiempo.txt

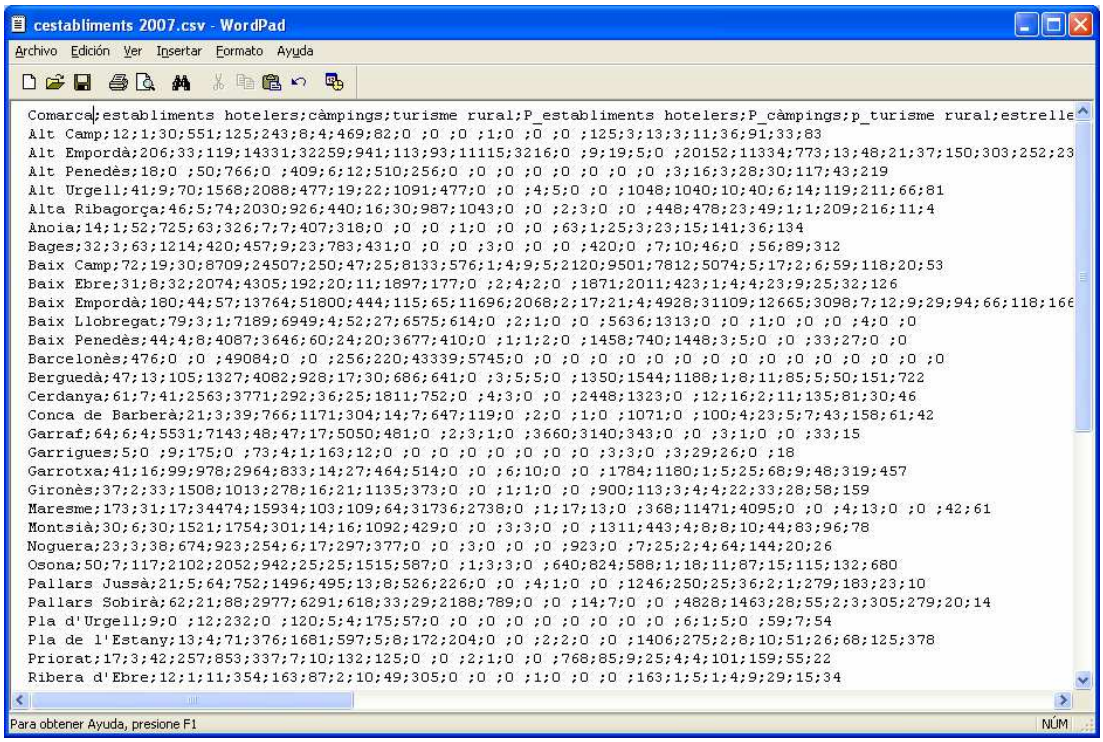



*Ilustración 10: Fichero establiments 2007 después de su transformación*

### 4.1.2 Pentaho Data Integration

Spoon es el diseñador gráfico de transformaciones y trabajos del sistema de ETTLs de Pentaho Data Integration (PDI), también conocido como Kettle (acrónimo recursivo: "Kettle Extraction, Transformation, Transportation, and Load Environment").

Las credenciales para acceder a PDI son:

Repository: ETL, User Name admin y Password admin

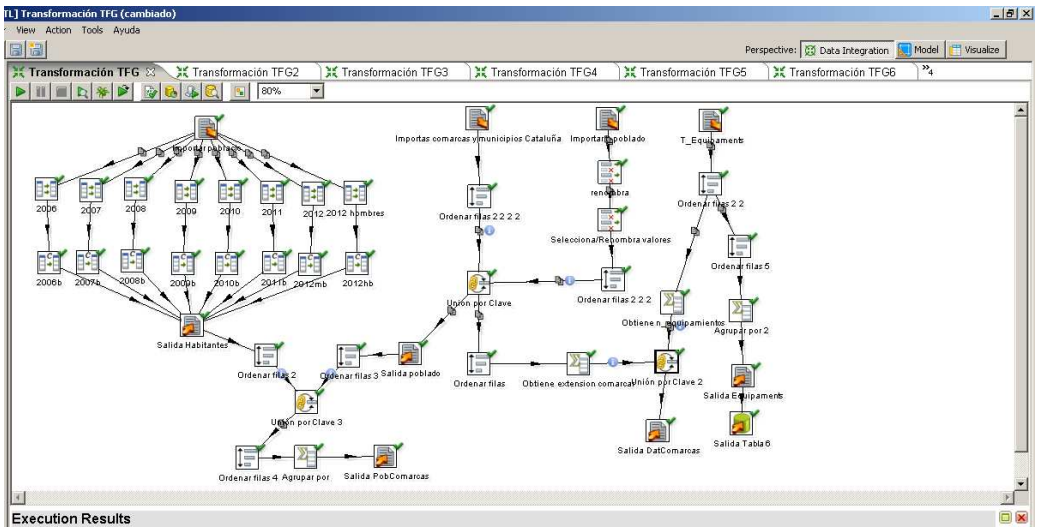



*Ilustración 11: Primera Transformación*

En la primera transformación se realizan diferentes procesos:

1. Los registros se tratan según el año, añadiendo el campo “anyo” y el dato de habitantes que según el año se encuentra en la tabla inicial en una columna distinta, resultando finalmente agrupada en el campo “habitantes”. Como resultado se obtiene la tabla “Habitantes.txt” en un archivo de texto temporal.

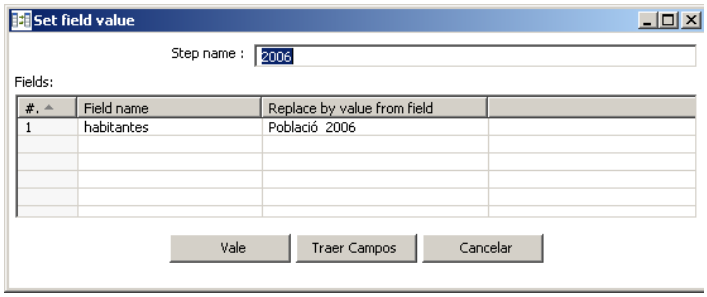



*Ilustración 12: Asignación del nº de habitantes*

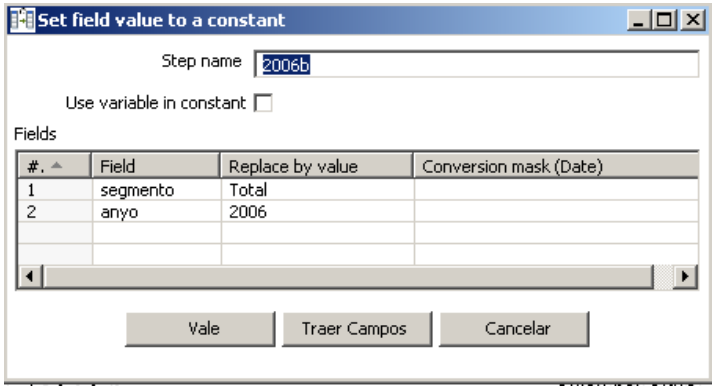



*Ilustración 13: Asignación del segmento y año correspondiente*

2. Por otro lado, con los datos obtenidos en (1) y la tabla que relaciona las comarcas y municipios de Cataluña, se obtienen las comarcas de las poblaciones, que junto a la importación de la tabla poblado, proporciona como resultado la tabla "poblado.txt", que contiene todas las poblaciones.

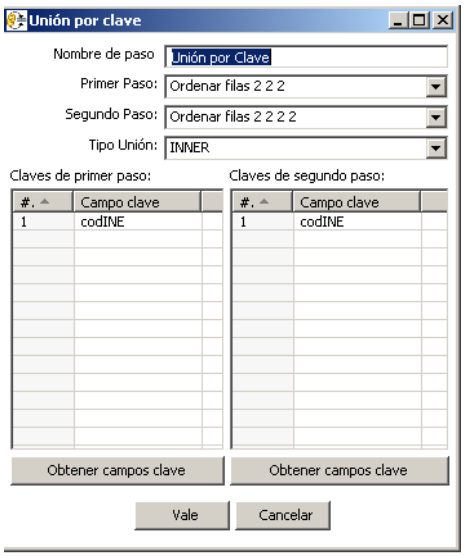



*Ilustración 14: Unión que proporciona la comarca de "poblado"*

3. De (1) y (2), haciendo la suma por comarcas, se obtiene la tabla de suma de población por comarcas, con los datos por segmentos.

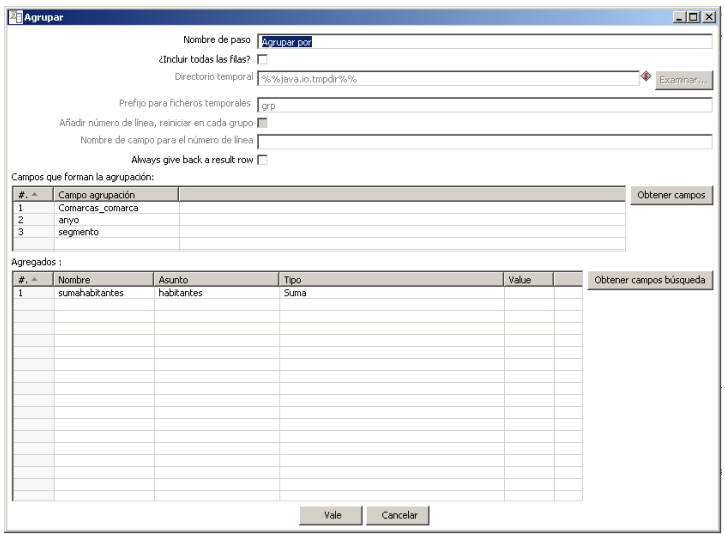



*Ilustración 15: Obtención de la suma de habitantes por segmento*

4. De la tabla equipaments, se obtiene eliminando los registros duplicados con una transformación la salida hacia MySQL de "Equipaments", y con la suma por comarcas, la tabla temporal "DatComarcas.txt".

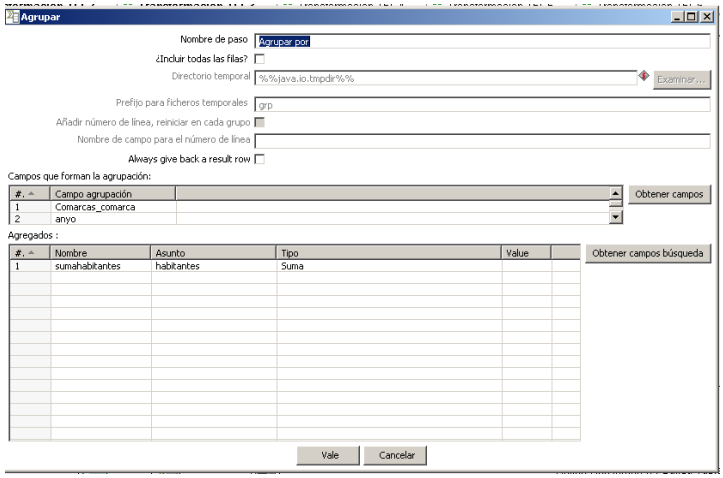



*Ilustración 16: Obtención de la suma de habitantes por comarca*

Para la segunda transformación se han realizado las siguientes operaciones:

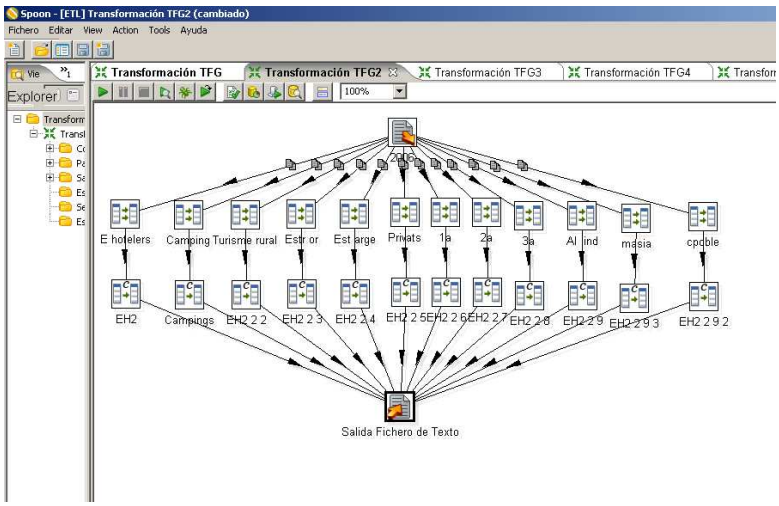



Ilustración 17: Segunda transformación

Para obtener la tabla de Hecho "Oferta" con los datos de establecimientos de 2006, se realiza la Transformación 2. Se crea un registro nuevo por cada tipo de establecimiento, con el número de plazas y equipamientos de cada tipo. Como resultado se obtiene un fichero de texto que será posteriormente tratado para su inclusión en el SGBD.

El procedimiento comienza con la asignación del valor de los campos establecimientos y plazas de un determinado tipo de establecimiento a los campos que contienen el valor:

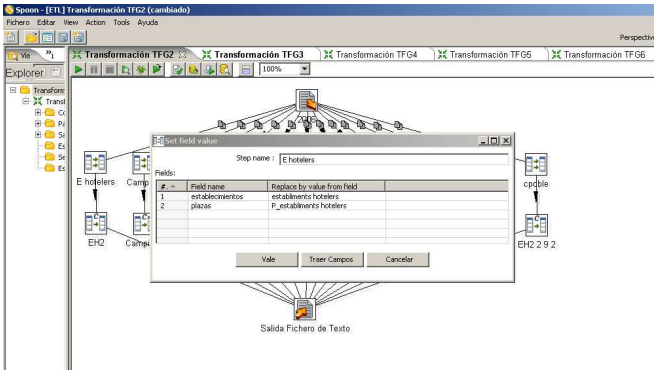



*Ilustración 18: Asignando valores de los campos*

Y acaba con la asignación de los valores fijos del tipo y del año correspondiente:

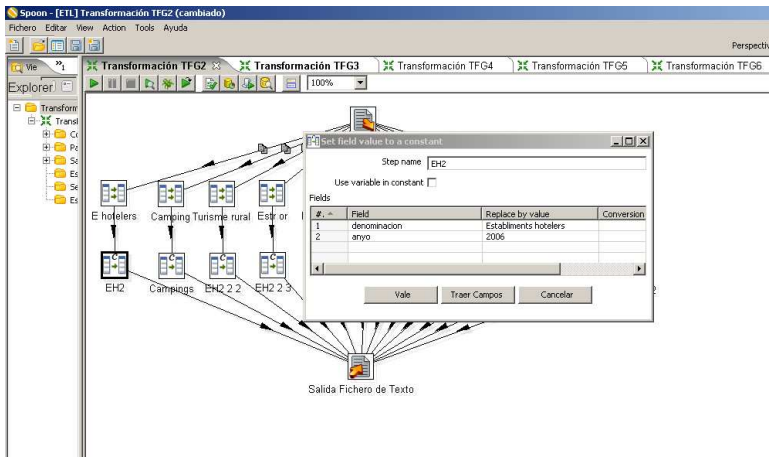



*Ilustración 19: Asignando valores fijos*

De la misma forma se trabaja con la Transformación 3 con los datos de 2007:

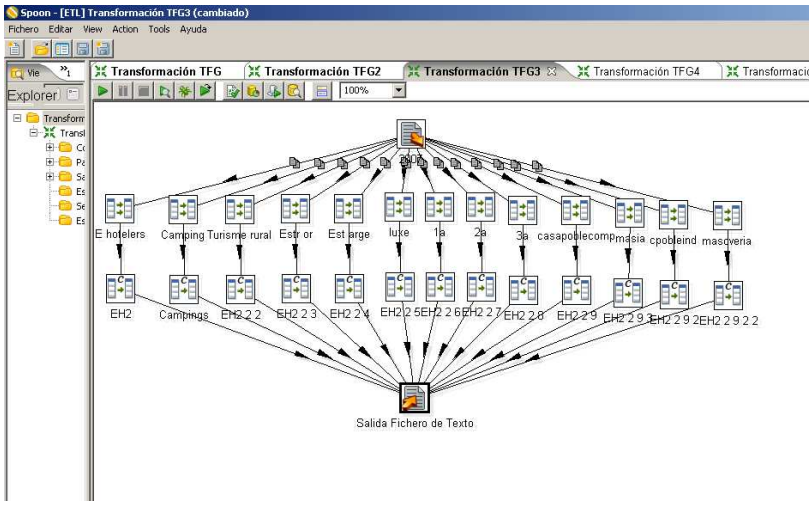



*Ilustración 20: Transformación 3*

Hasta terminar con los datos de 2012, la Transformación 8:

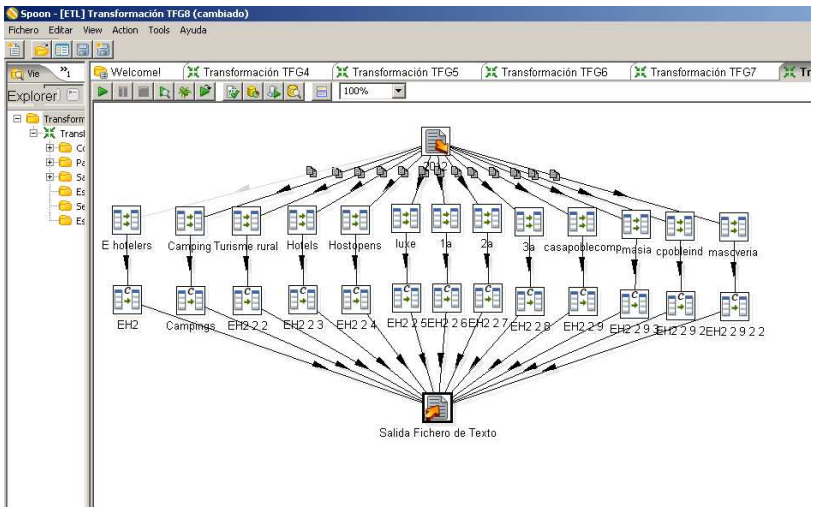



*Ilustración 21: Transformación 8*

Finalmente, se trabaja con la transformación final, que enviará los datos al SGBD:

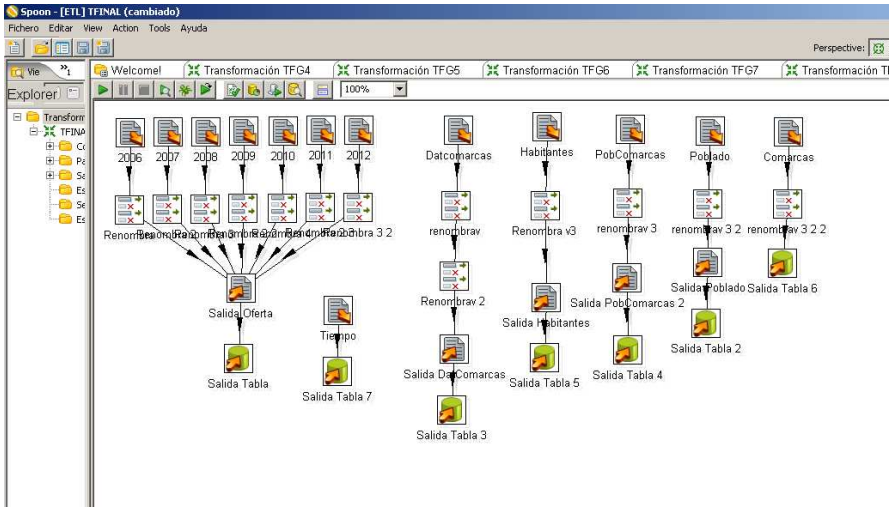



*Ilustración 22: Transformación final*

En esta transformación se han realizado las siguientes operaciones (de derecha a izquierda):

1. Unir, renombrar y eliminar los campos no necesarios para enviar los datos a las tabla “Oferta” de MySQL.

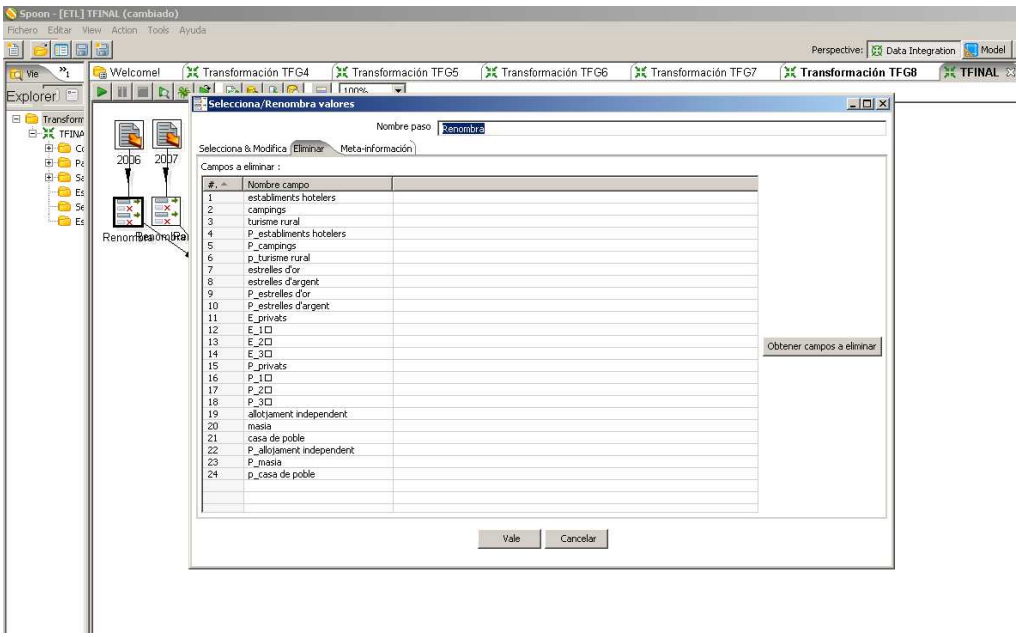



*Ilustración 23: Eliminación de campos no necesarios después de su tratamiento*

2. Exportar los datos de la tabla de dimensión “Tiempo” a MySQL.

3. Renombrar y eliminar los campos no necesarios para enviar los datos a las tabla “DatComarcas” de MySQL.

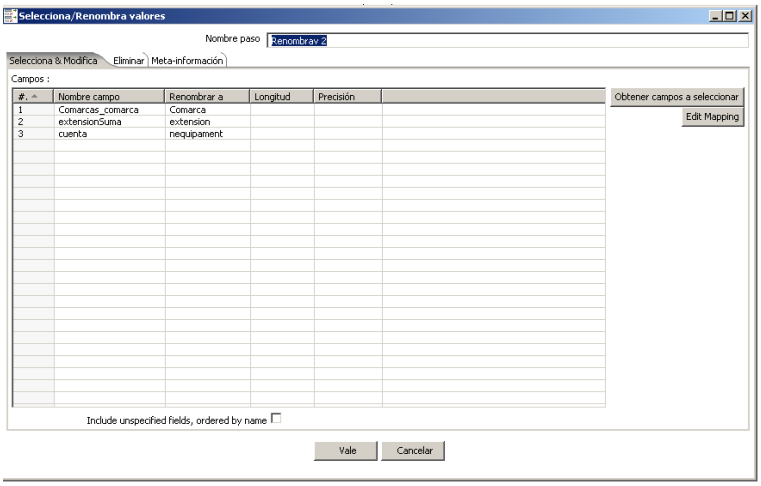



*Ilustración 24: Renombrando campos de "DatComarcas"*

4. Igual que (3) con “Habitantes”.

5. Igual que (3) con “PobComarcas”.

6. Igual que (3) con “Poblado”.

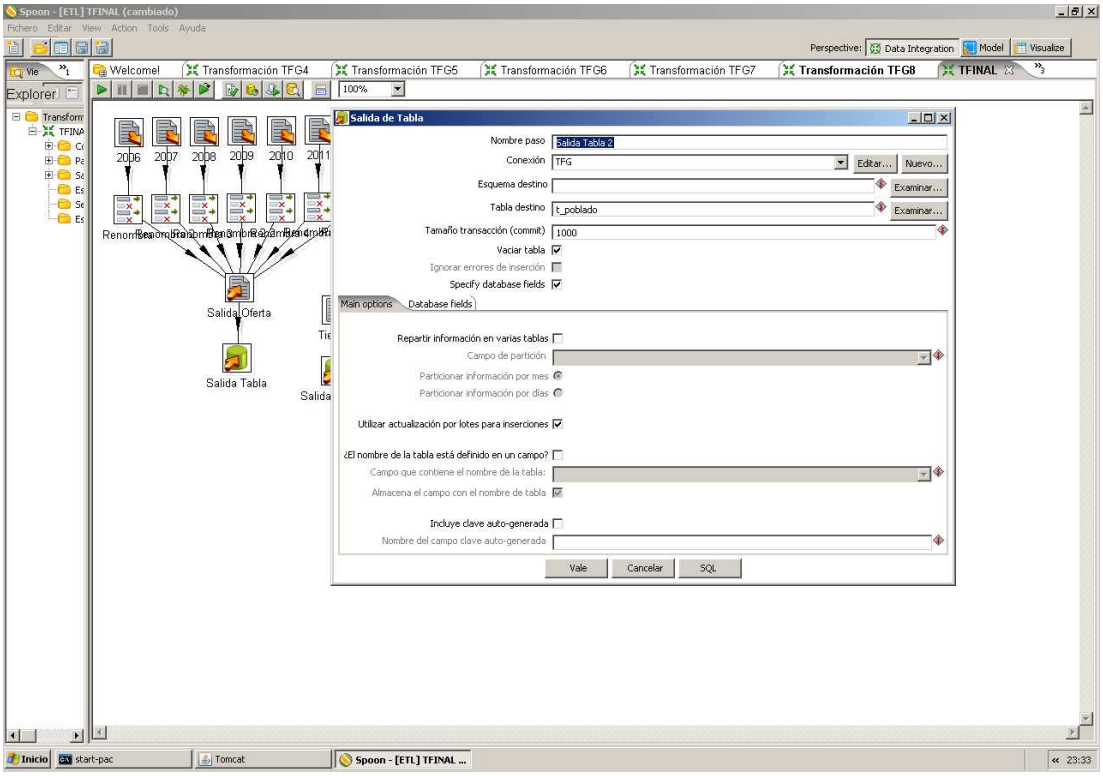



*Ilustración 25: Exportación a MYSQL*

7. Igual que (3) con “Comarcas”.

#### 4.1.2.1 Automatización

Para automatizar el proceso ETL, se ha recurrido a “Pan”, que es Pan es una aplicación que permite ejecutar transformaciones diseñadas con Spoon y almacenadas como XML.

Para programar la ejecución de las transformaciones en el orden correspondiente, se ha creado un fichero bat, llamado “ProcesoETL.bat” que contiene los siguientes comandos:

*call pan.bat /file:"c:\user-pentaho\TFG1.ktr" /level:Basic*

*call pan.bat /file:"c:\user-pentaho\TFG2.ktr" /level:Basic*

*call pan.bat /file:"c:\user-pentaho\TFG3.ktr" /level:Basic*

*call pan.bat /file:"c:\user-pentaho\TFG4.ktr" /level:Basic*

*call pan.bat /file:"c:\user-pentaho\TFG5.ktr" /level:Basic*

```
call pan.bat /file:"c:\user-pentaho\TFG6.ktr" /level:Basic
call pan.bat /file:"c:\user-pentaho\TFG7.ktr" /level:Basic
call pan.bat /file:"c:\user-pentaho\TFG8.ktr" /level:Basic
call pan.bat /file:"c:\user-pentaho\TFINAL.ktr" /level:Basic
Echo PROCESO FINALIZADO
PAUSE
```

La ejecución de este archivo, incluyendo los datos *.ktr y las tablas en txt que se encuentran en el directorio, permitiría reproducir todo el proceso ETL de forma automatizada. Si fuera necesaria una carga periódica de los datos, se podría programar la ejecución del "ProcesoETL.bat" en unos días y fechas determinados, para hacer el refresco en el Almacén de Datos.

### 4.1.3 Creación de tablas en el esquema

Antes de la ejecución de la transformaciones en Kettle, se han de crear las tablas en MySQL, el script diseñado para ello es el siguiente:

El script de creación de las tablas en el esquema TFG de MySQL es el siguiente:

```
CREATE TABLE T_POBLADO (
   nombre Char(50),
   codINE Char(9),
   extension Integer,
   Comarca VarChar(40),
   PRIMARY KEY (codINE)
);
CREATE TABLE T_EQUIPAMENTS (
   nombre VarChar(200),
   direccion VarChar(150),
   municipio VarChar(50),
```

```
    cp VarChar(5),
    Comarca VarChar(40),
    telefono VarChar(15),
    longitud Char(20),
    latitud Char(20),
    categorias VarChar(400),
    location Char(100),
    PRIMARY KEY (nombre, Comarca)

);
CREATE TABLE T_DATCOMARCAS (
    comarca VarChar(40),
    extension Integer,
    nequipament Integer,
    PRIMARY KEY (comarca)
);
CREATE TABLE T_POBCOMARCAS (
    comarca VarChar(40),
    Anyo Integer,
    segmento VarChar(20),
    poblacion Integer,
    PRIMARY KEY (comarca, anyo, segmento)
);

CREATE TABLE T_HABITANTES (
    anyo Integer,
    codINE Char(9),
```

```
   Segmento Char(15),
   habitantes Integer,
   PRIMARY KEY (anyo, codINE, Segmento)
);
CREATE TABLE T_TIEMPO (
   anyo Integer,
   PRIMARY KEY (anyo)
);

CREATE TABLE T_OFERTA (
   Comarca VarChar(40),
   anyo Integer,
   establecimientos Integer,
   plazas Integer,
   denominacion VarChar(40),
   PRIMARY KEY (Comarca, anyo, denominacion)

);
CREATE TABLE T_COMARCAS (
   IdComarca Integer,
   comarca VarChar(40),
   PRIMARY KEY (Comarca)
);

ALTER TABLE T_OFERTA
ADD CONSTRAINT PK_OF_TIEM FOREIGN KEY (anyo) references T_TIEMPO (anyo);
```

*ALTER TABLE T_HABITANTES*

*ADD CONSTRAINT PK_HA_POBL FOREIGN KEY (codINE) references T_POBLADO (codINE),*

*ADD CONSTRAINT PK_HA_TIEM FOREIGN KEY (anyo) references T_TIEMPO (anyo);*

Como resultado del proceso ETL, se tienen los datos cargados en el SGBD MySQL.

## 4.2 Modelo Multidimensional

Para crear el modelo Multidimensional se ha utilizado el Pentaho BI Server, que se conecta a la base de datos creada previamente para capturar la información.

La primera estrella es la estrella que conecta la Oferta de equipamientos con los datos de Comarca y población de comarca:

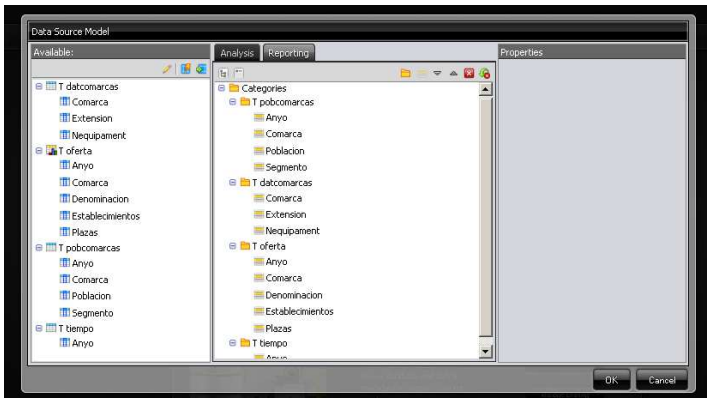



*Ilustración 26: Estrella Oferta*

La segunda conecta “Equipaments” con los datos de comarca y de población de comarca:

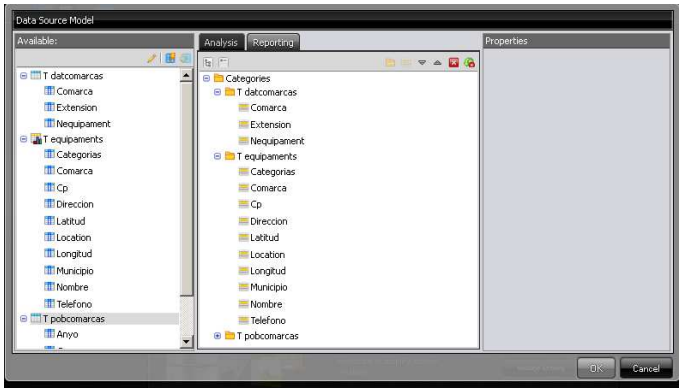



*Ilustración 27: Estrella Equipaments*

La tercera contiene el detalle de la población por municipio:

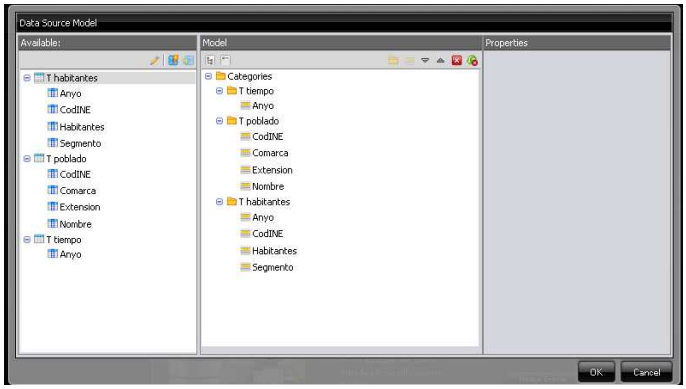



*Ilustración 28: Estrella habitantes*

Una vez introducido el modelo, se procede a la creación de informe y análisis de los datos.

## 4.3 Informes y análisis

El acceso a los reports se realiza por la consola de Pentaho (Pentaho User Console) a localhost:8080, con el usuario tfcmmd y la contraseña tfcmmd. Los reports están el TFG/reports. Se han creado con **Pentaho Reporting**, además se ha trabajado con **Saiku**. Para una posterior versión, se desarrollará un Dashboard de Pentaho para crear la interfaz de usuario.

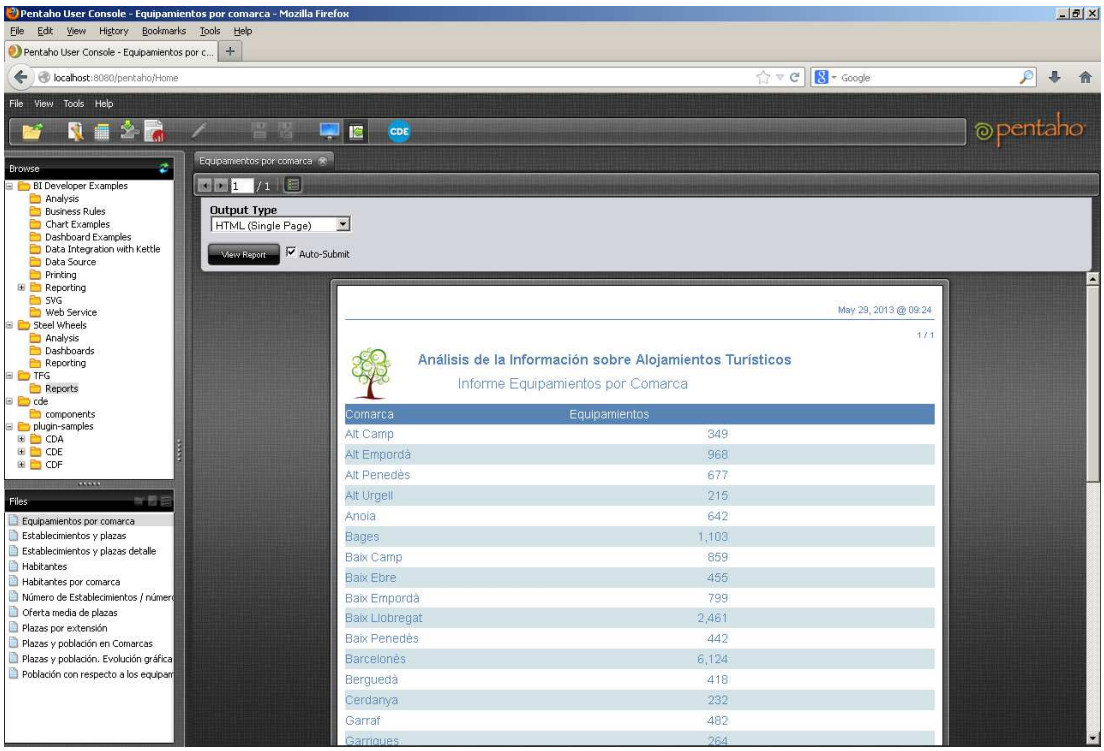



*Ilustración 29: Acceso a los reports*

**Total de establecimientos y Total de plazas**

El informe establecimientos se ha fundido con el de plazas, puesto que el acceso a los datos y las consultas son las mismas, además se ofrece una visión de los datos mucho más completa, además permite elegir el año y la comarca de la que se ofrecen los datos:

En este caso, para 2006 y la comarca Alt Camp, se observa un mayor número de plazas de establecimientos hoteleros que del resto de tipos de establecimientos.

May 27, 2013 @ 03:11

1 / 2

**Análisis de la Información sobre Alojamientos Turísticos**

Informe Establecimientos y plazas

Año 2,006
Comarca Alt Camp

| Tipo Establecimiento | Nº Establecimientos | Nº Plazas |
|---|---|---|
| Establiments hotelers | 10 | 488 |
| Turisme rural | 28 | 222 |
| Estrelles d'argent | 5 | 90 |
| 1a. | 0 | 0 |
| 3a. | 0 | 0 |
| Masia | 4 | 43 |
| Campings | 0 | 0 |
| Privats | 0 | 0 |
| Allotjament independent | 20 | 133 |
| Estrelles d'or | 5 | 398 |
| Casa de poble | 4 | 46 |
| 2a. | 0 | 0 |
| Total | 76 | 1,420 |

*Ilustración 30: Informe establecimientos y plazas (1)*

La segunda página del informe incorpora una representación gráfica de los datos:

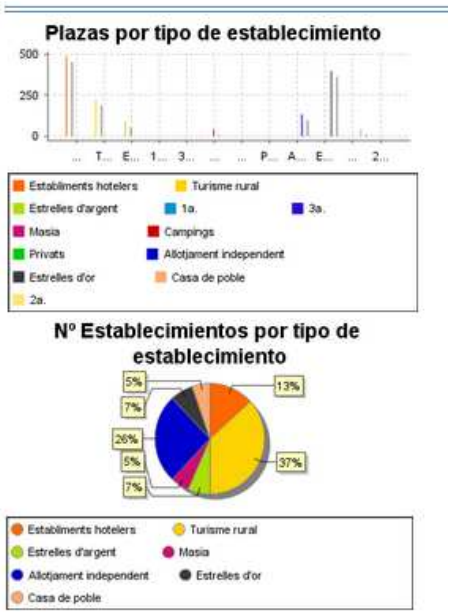



*Ilustración 31: Informe establecimientos y plazas (2)*

Por otro lado, hay otra posibilidad de visualizar los datos: viendo la totalización por comarcas para todos los años:

Análisis de la Información sobre Alojamientos Turísticos

Informe Establecimientos y plazas - Detalle

Año 2,006

| Comarca | Establecimientos | Plazas |
|---|---|---|
| Alt Camp | 76 | 1,420 |
| Alt Empordà | 692 | 96,892 |
| Alt Penedès | 122 | 2,832 |
| Alt Urgell | 230 | 8,250 |
| Alta Ribagorça | 236 | 6,736 |
| Anoia | 124 | 2,128 |
| Bages | 192 | 4,150 |
| Baix Camp | 234 | 66,860 |
| Baix Ebre | 138 | 12,052 |
| Baix Empordà | 954 | 132,624 |
| Baix Llobregat | 154 | 25,808 |
| Baix Penedès | 112 | 16,378 |
| Barcelonès | 928 | 96,626 |
| Berguedà | 300 | 13,018 |
| Cerdanya | 200 | 13,366 |
| Conca de Barberà | 118 | 4,392 |
| Garraf | 148 | 25,608 |
| Garrigues | 26 | 236 |
| Garrotxa | 300 | 9,258 |
| Gironès | 126 | 5,622 |
| Maresme | 496 | 105,602 |
| Moianès | 136 | 8,682 |
| Noguera | 126 | 3,846 |

*Ilustración 32: Informe establecimientos y plazas - Detalle (por años)*

Al final del informe se pueden visualizar dos gráficas totalizando los resultados:

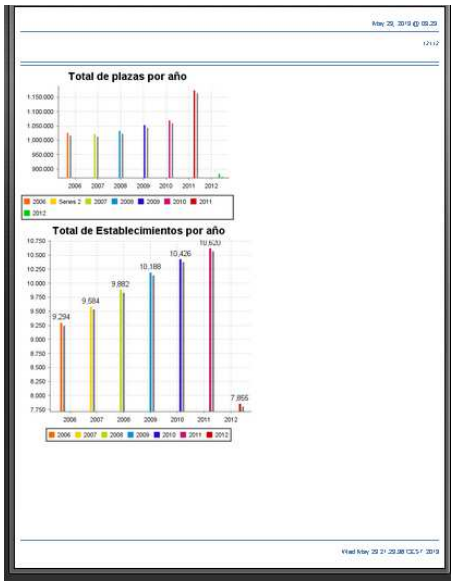



*Ilustración 33: Informe establecimientos y plazas - Detalle (Gráfica)*

**% Plazas respecto a la población**

Seleccionando el año, el informa muestra el porcentaje de plazas respecto a la población de una comarca. Hay datos que llaman la atención, como la existencia de comarcas muy turísticas con más plazas que habitantes:

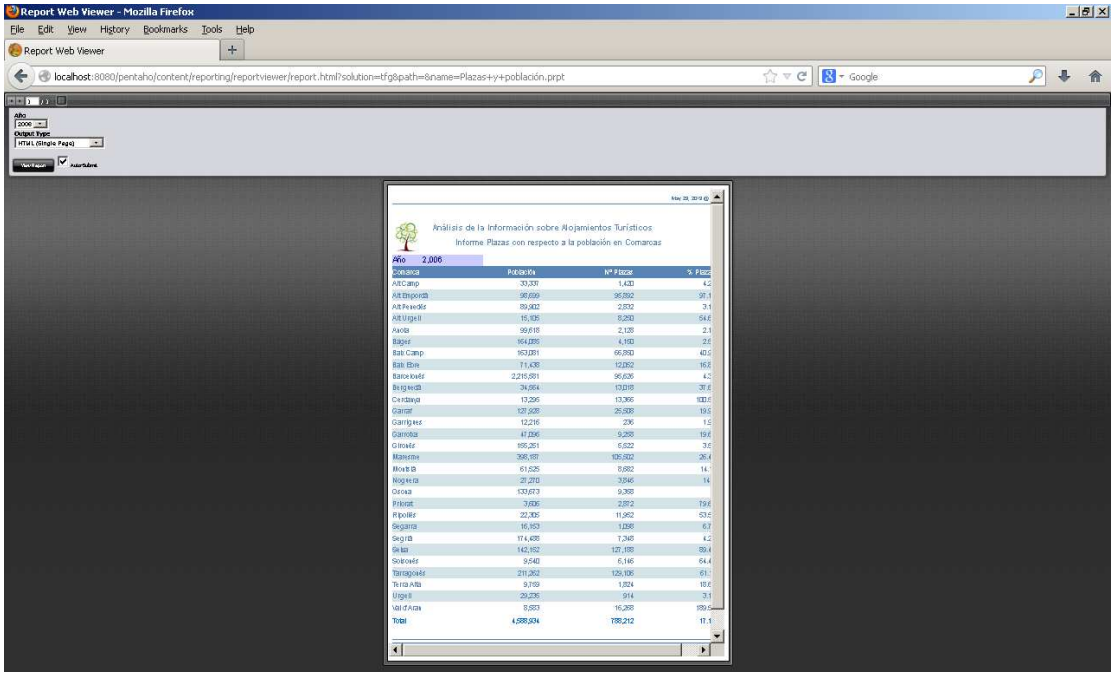



*Ilustración 34: % Plazas con respecto a la población*

**Oferta media de plazas**

La oferta media de plazas es el valor medio por comarca de las plazas. Hay comarcas con un valor muy bajo como “Alt Penedés” y otras con un valor muy alto como “Baix Llobregat”

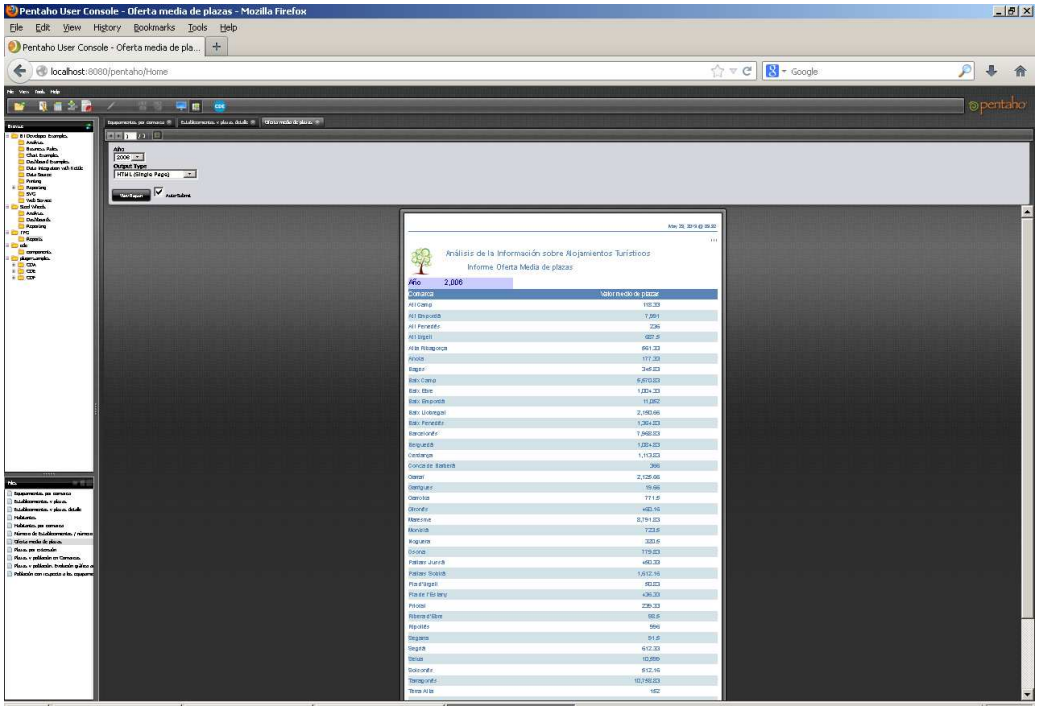



*Ilustración 35: Oferta media de plazas*

**Número de establecimientos / Número de equipamientos**

Este informe ofrece una relación entre número de establecimientos (de una comarca) y el número de equipamientos. Hay comarcas con buenos equipamientos en relación con los establecimientos, como “Alt Urgell” y otras con un ratio más desfavorable como “Pla d’Urgell”

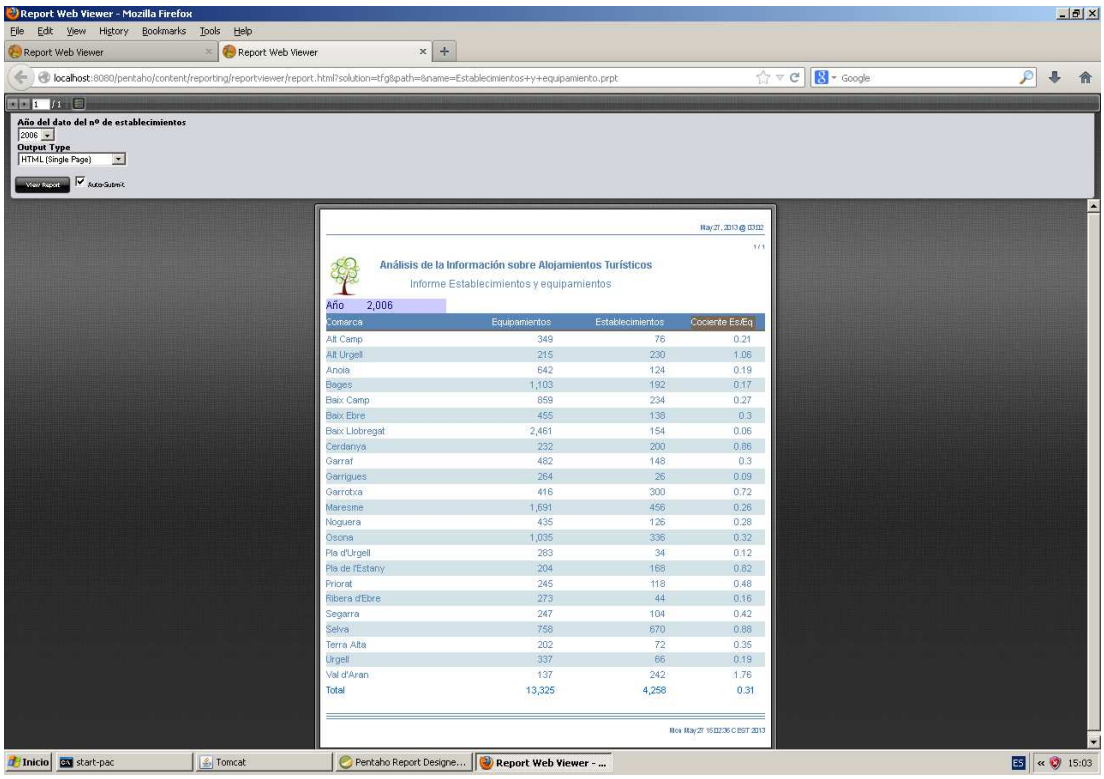
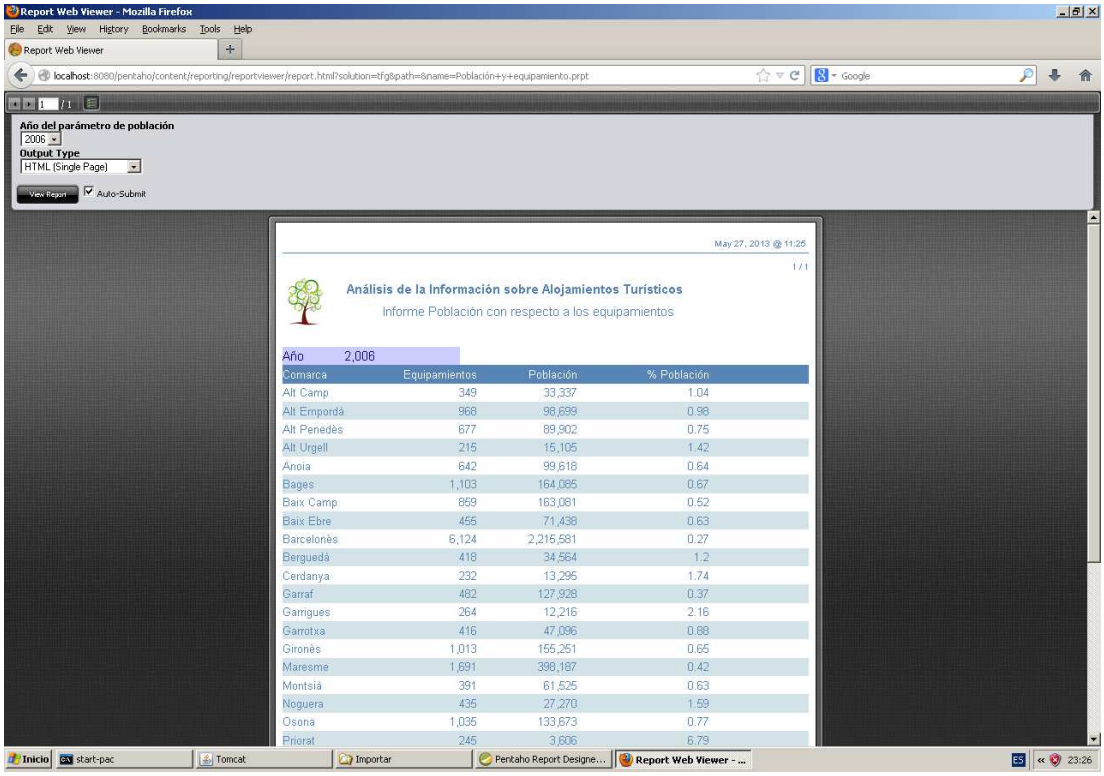



*Ilustración 36: Número de establecimientos / Número de equipamientos*

**% Población por equipamiento**

Este informe ofrece la relación entre los equipamientos y la población por comarca. Hubiera sido interesante haber podido linkar los equipamientos con las poblaciones por municipio, pero por la estructura de los datos, como se comentó anteriormente ha sido imposible. Como se puede observar, "Cerdanya" tiene un alto ratio de equipamientos por 100 habitantes, en comparación con otras comarcas.




| Comarca | Equipamientos | Población | % Población |
|---|---|---|---|
| Alt Camp | 349 | 33,337 | 1.04 |
| Alt Empordà | 968 | 98,699 | 0.98 |
| Alt Penedès | 677 | 89,902 | 0.75 |
| Alt Urgell | 215 | 15,105 | 1.42 |
| Anoia | 642 | 99,618 | 0.64 |
| Bages | 1,103 | 164,085 | 0.67 |
| Baix Camp | 859 | 163,081 | 0.52 |
| Baix Ebre | 455 | 71,438 | 0.63 |
| Barcelonès | 6,124 | 2,215,581 | 0.27 |
| Berguedà | 418 | 34,564 | 1.2 |
| Cerdanya | 232 | 13,295 | 1.74 |
| Garraf | 482 | 127,928 | 0.37 |
| Garrigues | 264 | 12,216 | 2.16 |
| Garrotxa | 416 | 47,096 | 0.88 |
| Gironès | 1,013 | 155,251 | 0.65 |
| Maresme | 1,691 | 398,187 | 0.42 |
| Montsià | 391 | 61,525 | 0.63 |
| Noguera | 435 | 27,270 | 1.59 |
| Osona | 1,035 | 133,673 | 0.77 |
| Priorat | 245 | 3,606 | 6.79 |

*Ilustración 37: % Población por equipamiento*

**Cantidad de plazas ofrecidas / superficie del territorio**

En este informe se refleja el número de plazas de una comarca que se ofrece en relación con la superficie de esa comarca, es decir, número de plazas por cada 100 $km^2$ , agrupado en años.

Llaman la atención las comarcas de “Baix Camp” y “Maresme” por la cantidad de plazas por km2 que ofrecen.

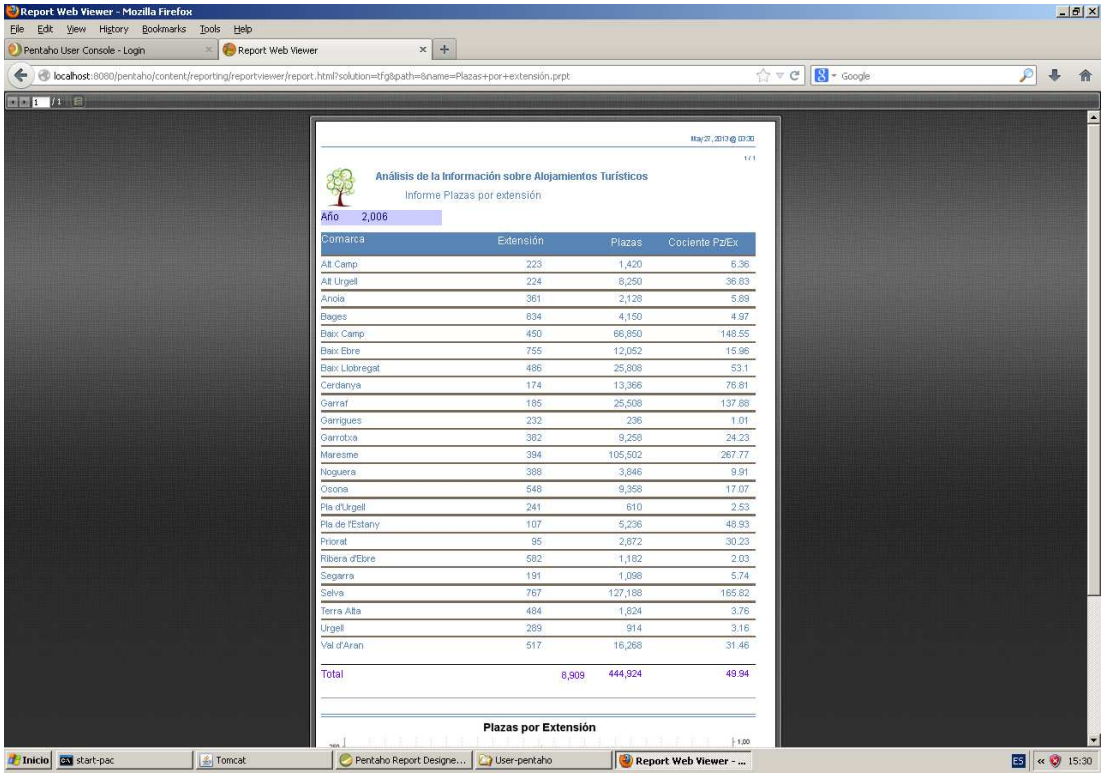



*Ilustración 38: Cantidad de plazas ofrecidas / superficie del territorio (1)*

Al final del informe se incluye un gráfico con el ratio.

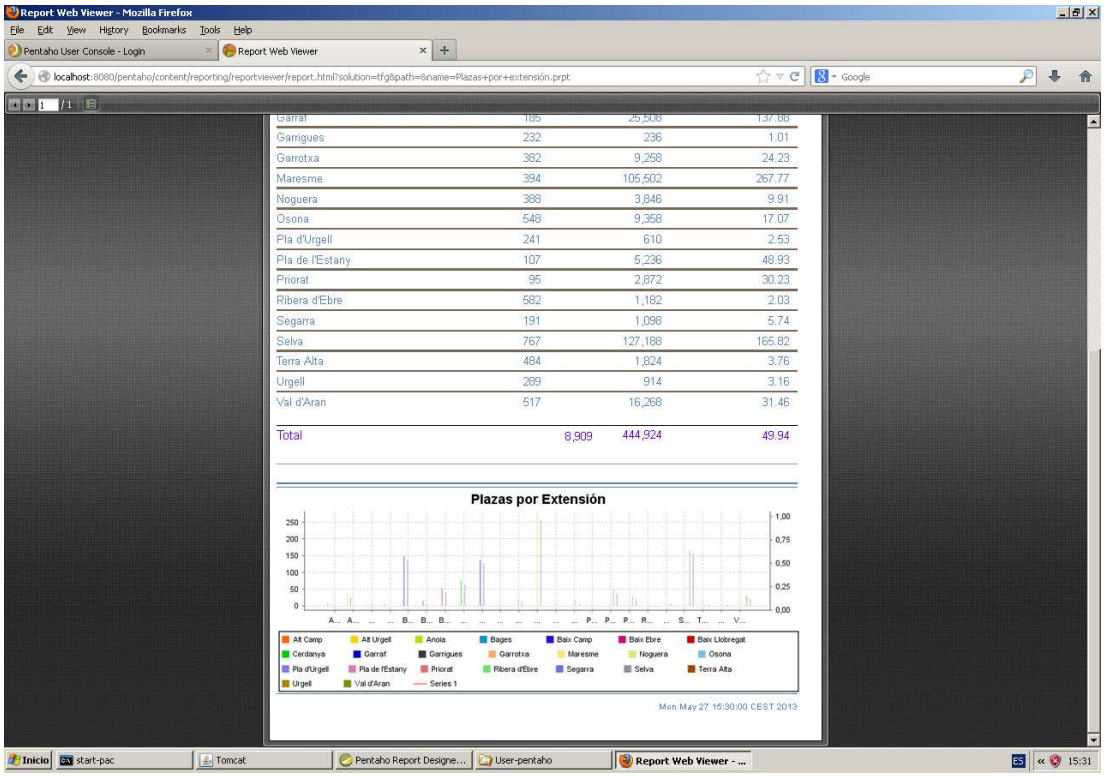



*Ilustración 39: Cantidad de plazas ofrecidas / superficie del territorio (2)*

**Habitantes**

El informe ofrece una relación de habitantes por comarca año y segmento (hombre, mujeres, total). Permite elegir el año y la comarca. Dentro de la comarca "Alt Empordá", destaca el número de habitantes de "Figueres".

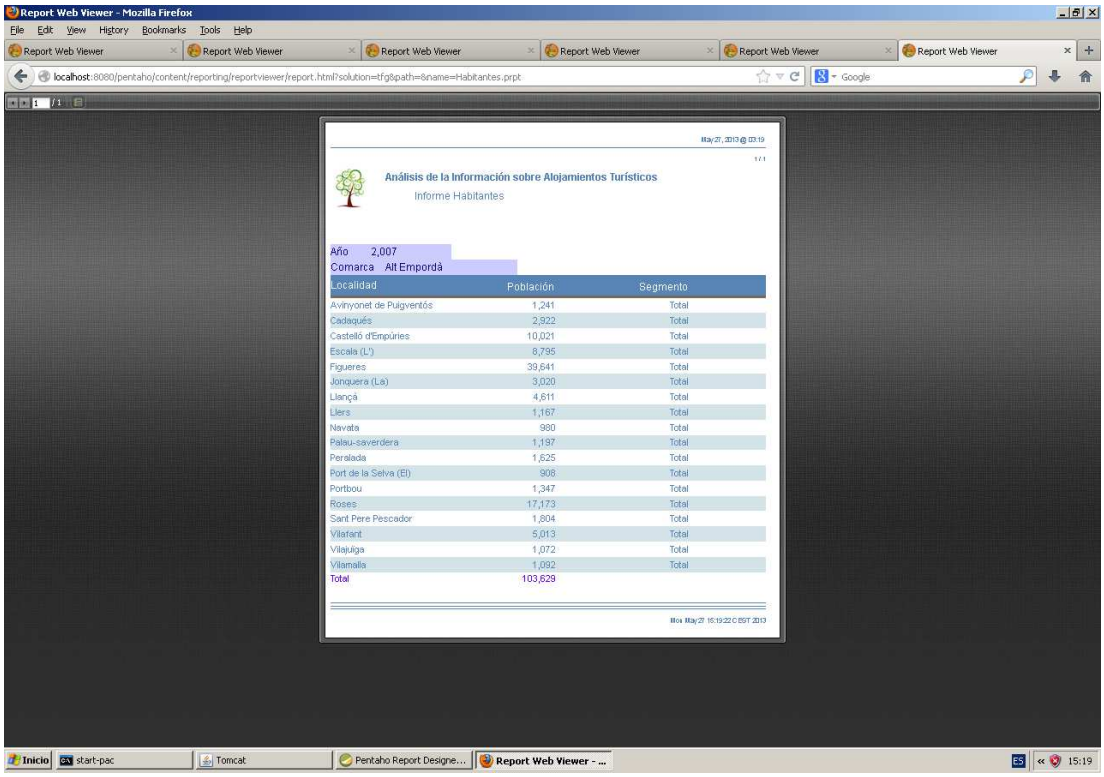



*Ilustración 40: Habitantes por población*

Se ha creado otro reporte con el número de habitantes, totalizado por comarca, permitiendo seleccionar un año determinado:

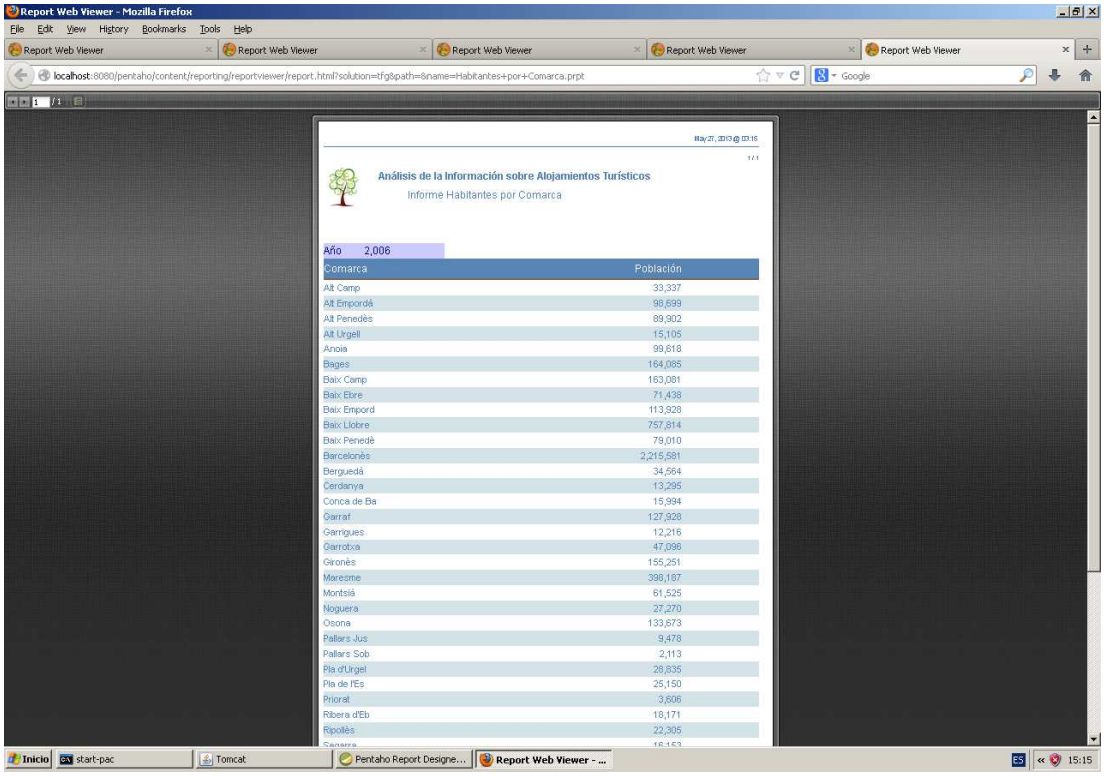



*Ilustración 41: Habitantes por comarca*

**Establecimientos vs habitantes por género**

Se ha incluido la relación entre los habitantes y los establecimientos, se aprecia una relación que tiende a crecer a partir de un nº determinado de establecimientos.

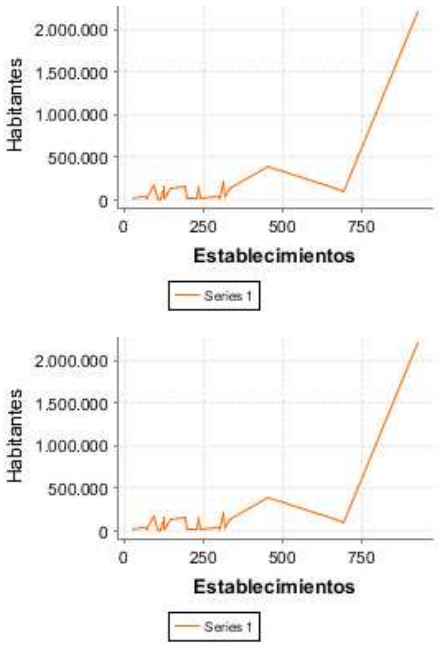



**Ilustración 42: Establecimientos vs habitantes por género**

**Plazas vs Personas**

No se aprecia una relación lineal en esta relación entre plazas y personas, como se puede constatar en la gráfica.

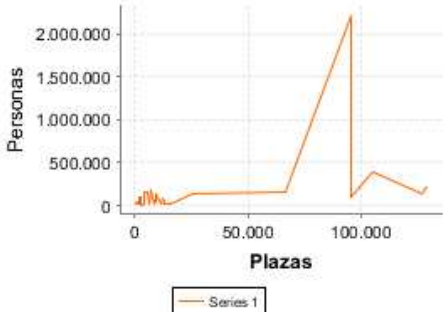



**Ilustración 43: Plazas vs Personas**

### Equipamientos vs población

La relación entre equipamientos y población es lineal

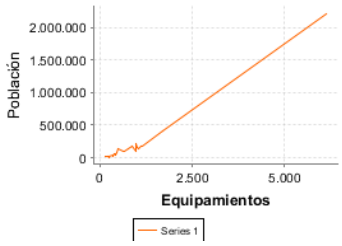
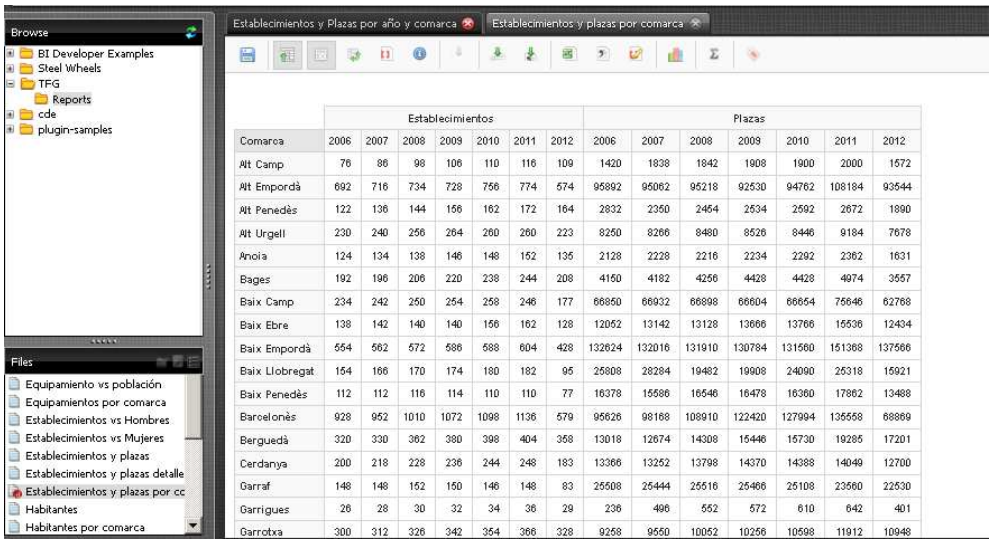



**Ilustración 44: Equipamientos vs Población**

### Análisis con Saiku

Saiku es un visor OLAP que proporciona una herramienta para realizar análisis.

Se ha considerado interesante realizar un análisis sobre los establecimientos y plazas en las comarcas, y su evolución a lo largo del tiempo. Se observa un crecimiento continuo a lo largo de los años tanto de los establecimientos, como de las plazas, para en el último año del que se disponen datos (2012), se sufra una caída tanto en establecimientos como en plazas en todas las comarcas.




| | Establecimientos | | | | | | | Plazas | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Comarca | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
| Alt Camp | 76 | 86 | 98 | 106 | 110 | 116 | 109 | 1420 | 1838 | 1842 | 1908 | 1900 | 2000 | 1572 |
| Alt Empordà | 692 | 716 | 734 | 728 | 756 | 774 | 574 | 95892 | 95062 | 95218 | 92530 | 94762 | 108184 | 93544 |
| Alt Penedès | 122 | 136 | 144 | 156 | 162 | 172 | 164 | 2832 | 2350 | 2454 | 2534 | 2592 | 2672 | 1890 |
| Alt Urgell | 230 | 240 | 256 | 264 | 260 | 260 | 223 | 8250 | 8266 | 8480 | 8526 | 8446 | 9184 | 7678 |
| Anoia | 124 | 134 | 138 | 146 | 148 | 152 | 135 | 2128 | 2228 | 2216 | 2234 | 2292 | 2362 | 1631 |
| Bages | 192 | 196 | 206 | 220 | 238 | 244 | 208 | 4150 | 4182 | 4256 | 4428 | 4428 | 4974 | 3557 |
| Baix Camp | 234 | 242 | 250 | 254 | 258 | 246 | 177 | 66850 | 66932 | 66898 | 66604 | 66654 | 75646 | 62768 |
| Baix Ebre | 138 | 142 | 140 | 140 | 156 | 162 | 128 | 12052 | 13142 | 13128 | 13666 | 13766 | 15536 | 12434 |
| Baix Empordà | 554 | 562 | 572 | 586 | 588 | 604 | 428 | 132624 | 132016 | 131910 | 130784 | 131560 | 151368 | 137566 |
| Baix Llobregat | 154 | 166 | 170 | 174 | 180 | 182 | 95 | 25808 | 28284 | 19482 | 19908 | 24090 | 25318 | 15921 |
| Baix Penedès | 112 | 112 | 116 | 114 | 110 | 110 | 77 | 16378 | 15586 | 16546 | 16478 | 16360 | 17862 | 13488 |
| Barcelonès | 928 | 952 | 1010 | 1072 | 1098 | 1136 | 579 | 95626 | 98168 | 108910 | 122420 | 127994 | 135558 | 68869 |
| Berguedà | 320 | 330 | 362 | 380 | 398 | 404 | 358 | 13018 | 12674 | 14308 | 15446 | 15730 | 19285 | 17201 |
| Cerdanya | 200 | 218 | 228 | 236 | 244 | 248 | 183 | 13366 | 13252 | 13798 | 14370 | 14388 | 14049 | 12700 |
| Garraf | 148 | 148 | 152 | 150 | 146 | 148 | 83 | 25508 | 25444 | 25516 | 25466 | 25108 | 23560 | 22530 |
| Garrigues | 26 | 28 | 30 | 32 | 34 | 36 | 29 | 236 | 496 | 552 | 572 | 610 | 642 | 401 |
| Garrotxa | 300 | 312 | 326 | 342 | 354 | 366 | 328 | 9258 | 9550 | 10052 | 10256 | 10598 | 11912 | 10948 |

**Ilustración 45: Establecimientos y plazas por comarca y año**

# 5 CONCLUSIONES

La realización del Trabajo de Fin de Grado ha requerido un trabajo intenso y largo, que ha sobrepasado la planificación, porque ciertas tareas han llevado más tiempo del inicialmente previsto. Además, como ya se ha indicado en la planificación inicial vs planificación final, la PAC 3 supuso una carga extra de trabajo, puesto que se tuvo que volver a realizar el análisis y el diseño del producto, además de la implementación.

En cualquier caso, el funcionamiento de la Suite Pentaho ha sido un agradable descubrimiento, que permite tratar datos visualmente con una gran potencia y flexibilidad.

Elegí este área en concreto, para profundizar en el aprendizaje de unas herramientas que van a ser necesarias en mi futuro laboral a corto plazo, experiencia que me ha realizado muy enriquecedora.

# 6 LÍNEAS DE EVOLUCIÓN FUTURAS

Como línea de evolución futura, se propone crear un Dashboard, para que los usuarios puedan interactuar de forma cómoda con la información, además de la generación de nuevos informes que permitan realizar análisis de los datos con la herramienta Saiku.

## ANEXO: BIBLIOGRAFÍA


- Data Warehousing Review - http://www.dwreview.com/
- Data Warehousing - http://www.1keydata.com/datawarehousing/datawarehouse.html
- MySQL – www.mysql.com
- VirtualBox - https://www.virtualbox.org
- Dataprix, portal de referencia sobre Tecnologías de la Información - http://www.dataprix.com
- Wikipedia - www.wikipedia.org/
- Wikilearning - http://www.wikilearning.com