



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

Alumno

Fernando Turón Manzanares  
Curso 2012-2013 2º Semestre  
Ingeniería Técnica Informática de Gestión

Consultor

Carles Llorach Rius



*“A los mayores les gustan las cifras.  
Cuando se les habla de un nuevo amigo, jamás preguntan sobre lo esencial del mismo.  
Nunca se les ocurre preguntar:  
“¿Qué tono tiene su voz? ¿Qué juegos prefiere? ¿Le gusta coleccionar mariposas?”  
Pero en cambio preguntan:  
“¿Qué edad tiene? ¿Cuántos hermanos? ¿Cuánto pesa? ¿Cuánto gana su padre?”  
Solamente con estos detalles creen conocerle.”  
El Principito, Capítulo IV*

A los que en algún momento me han soportado durante estos años.  
A mi familia y a mis amigos.

## Índice de contenido

Dedicatoria.....	2
1.- Introducción.....	5
2.- Desarrollo Teórico.....	5
2.1.- Almacenes de Datos.....	5
2.2.- Aplicaciones OLAP.....	6
2.3.- Arquitectura general de los almacenes de datos.....	6
3.- Objetivos, Métodos y Planificación.....	7
3.1.- Objetivos de la asignatura.....	7
3.2.- Objetivos de la tecnología Datawarehouse.....	7
3.3.- Objetivos del TFC.....	8
3.3.1.- Objetivos primarios.....	8
3.3.2.- Objetivos específicos.....	8
3.4.- Método del TFC.....	9
3.5.- Disponibilidades.....	10
3.6.- Planificación.....	10
4.- Análisis.....	12
4.1.- Análisis de las fuentes proporcionadas.....	13
4.2.- Análisis de las características de los datos de las fuentes de información.....	14
4.3.- Análisis de los datos.....	15
4.3.1.- Fichero Poblacio.csv.....	16
4.3.2.- Fichero Equipaments.csv.....	16
4.3.3.- Ficheros EstablimentsXXXX.txt.....	16
4.4.- Diagrama de Casos de Uso.....	17
4.5.- Modelo Conceptual.....	18
4.5.1.- Identificación del Hecho.....	18
4.5.2.- Granularidad.....	18
4.5.3.- Dimensiones a establecer para el análisis.....	18
4.5.4.- Identificación de los atributos de las Tablas de Dimensiones.....	19
4.5.5.- Identificación de las medidas a emplear.....	20
4.5.6.- Identificación de los atributos de la tabla de Hechos.....	21
4.6.- Estudio de Viabilidad.....	22
4.7.- Esquema Conceptual y Lógico.....	23
5.- Diseño.....	24
5.1.- Tabla de Zonas.....	24
5.2.- Tabla de Equipamientos.....	25
5.3.- Tabla de Establecimientos.....	25
5.4.- Tabla de Ofertas.....	26
5.5.- Esquema Físico.....	27
5.6.- Diseño de Arquitectura.....	28
5.6.1.- Arquitectura del Software.....	28
5.6.3.- Arquitectura del Hardware.....	29
6.- Implementación.....	29
6.1.- Entorno de trabajo.....	29
6.2.- Construcción de la ETL.....	30



6.2.1.- Carga inicial.....	33
6.2.2.- Carga de la tabla de Hechos.....	38
6.2.3.- Construcción del Esquema.....	39
7.- Informes.....	41
7.1.- Acceso a la herramienta de Reporting.....	41
7.2.- Descripción de los informes propuestos.....	41
7.3.- Total de Establecimientos.....	43
7.4.- Total de Plazas.....	44
7.5.- % de Plazas Respecto a Población.....	45
7.6.- Oferta Media de Plazas.....	46
7.7.- Número de Establecimientos / Número de Equipamientos.....	47
7.8.- % de Población por Equipamiento.....	48
7.9.- Indicador de Establecimientos vs por Habitantes por Género.....	49
7.10.- Indicador de Plazas vs Personas.....	50
7.11.- Indicador de Equipamientos vs Población.....	51
7.12.- Cantidad de Plazas Ofertadas / Superficie del Territorio.....	52
7.13.- Informe de Establecimientos.....	53
8.- Conclusiones.....	53
9.- Líneas de Evolución Futura.....	54
10.- Glosario.....	54
11.- Bibliografía.....	56
12.- Webgrafía.....	56
13.- Relación de Tablas e Imágenes.....	57
12.1.- Tablas.....	57
12.2.- Imágenes.....	57
Anexo A.- Script de creación de la Base de Datos.....	59
Anexo B.- Instalación y ejecución de Scripts en la máquina virtual.....	59
Anexo C.- Estructura de carpetas definidas en la máquina virtual.....	59

## 1.- Introducción

Con este proyecto empieza la última etapa de mi formación en los estudios de Ingeniería Técnica Informática De Gestión. La elección de este área no ha sido casual, sino que he ido madurando durante el tiempo en el que he ido avanzando en mi carrera. Varias son las implicaciones:

- *Ámbito Personal.* El mundo de las Bases de Datos me ha interesado desde mucho antes de comenzar mis estudios en la UOC. Desde las primeras bases de datos que tuve que utilizar en la Universidad, como dbase III+, comprendí el alcance de los sistemas que permitieran procesar y consultar datos organizados de forma estructurada.
- *Ámbito Académico.* He tratado de cursar todas las asignaturas ofertadas en los estudios relacionadas con Bases de Datos. Me ha parecido un área muy interesante, con un tratamiento muy bueno en el conjunto de asignaturas ofertadas, y en el que he obtenido unos resultados muy satisfactorios por mi parte.
- *Ámbito Profesional.* Considero que las Bases de Datos me han ayudado a comprender la forma en que se puede organizar la información en cualquier compañía. Entender, al menos en parte, como funcionan los motores de bases de datos para, a su vez, entender como procesan la información los paquetes de software de gestión empresarial. Desde los primeras aplicaciones de gestión contable hasta los ERP que existen en la actualidad, y en particular, Navision, del que soy administrador en la compañía para la que trabajo. Conocer todos los puntos de vista de un cliente, administrador, usuario convencional y programador, me ha permitido obtener un sentido global.
- *Expectativas de futuro.* Mi interés en el tema me lleva a plantearme continuar mis estudios en este ámbito.

El estudio de Almacenes de Datos o Datawarehouse constituye en mi caso una necesidad de culminar las asignaturas superadas previamente, aportando una perspectiva diferente que las complementa y proporciona un sentido práctico, en la medida en que tratare de aplicar los conocimientos que adquiriera en mi actividad profesional habitual.

Ciñéndonos al proyecto en concreto se va a partir de la información suministrada por el Observatorio Nacional de Ocupación, ONdO, a partir de los datos estadísticos de los que dispone el IDESCAT, el Instituto de Estadística de Catalunya sobre población y equipamientos, y de los datos de la Federación Catalana de Alojamientos Turísticos, sobre alojamientos turísticos, plazas ofertadas y características de los mismos.

A partir de la información reunida se desea obtener una serie de datos estadísticos que permitan mostrar la relación entre diferentes variables, aparentemente independientes.

## 2.- Desarrollo Teórico

A continuación realizaremos una aproximación a los conceptos fundamentales en el ámbito de los almacenes de datos.

### 2.1.- Almacenes de Datos

A principios de los años 90 del pasado siglo Bill Inmon proporciona una definición de Almacén de Datos que sigue vigente hoy en día:

*“Un almacén de datos es una colección de datos orientados por tema, integrados, variables en el tiempo y no volátiles que se emplea como apoyo a la toma de decisiones estratégicas”*

En esta definición podemos observar los siguientes elementos:

- *Orientado por temas.* El diseño del almacén de datos está orientado más a la respuesta eficaz de consultas estratégicas y no a procesar transacciones, como en las bases de datos tradicionales o transaccionales. Se organiza por temas, los más adecuados para responder a las consultas estratégicas, que pueden ser:

- **Actividades de Interés** para el análisis: compras, ventas, producción,...
- **Contexto** de análisis para estas actividades: proveedores, clientes, máquinas, productos,...

El modelo multidimensional se estructura en dos conceptos:

- **Hechos**, que son actividades de interés para la empresaria.
- **Dimensiones**, referidas al contexto sobre el que se desea analizar estas actividades.

- *Integrados.* Los datos pueden proceder de fuentes de datos heterogéneas dentro de la organización, o bien por comparación con datos externos a ella. En cualquiera de los dos casos se plantean problemas por la existencia de formatos, definiciones o esquemas diferentes. Se plantean diferencias a la hora de cargar los datos en el almacén de datos.
- *Variables en el tiempo.* Todos los datos se cargan con una referencia temporal para la que son válidos. La referencia ha de ser puntual y concreta.
- *No volátiles.* Los datos no se insertan, modifican o eliminan en tiempo real, sino que se refrescan de manera periódica a partir de las fuentes de datos, añadiendo datos nuevos a los ya existentes.

En resumen se puede afirmar que los almacenes de datos ofrecen información histórica para la toma de decisiones integrando la información procedente de distintas fuentes de datos. En la siguiente tabla se muestran las ventajas e inconvenientes de este tipo de tecnología:

Ventajas	Inconvenientes
Decisiones soportadas por datos fiables, coherentes y homogéneos.	Infravalorar los recursos necesarios para su puesta en construir, poblar y mantener el almacén de datos.
Rentabilidad de Inversiones (ROI, Return Of Investment). Muchos informes incluyen referencia a la rentabilidad de un almacén de datos.	No planificar el esfuerzo necesario para lograr diseños correctos.
Aumenta la competitividad en entornos hostiles, al disponer de datos más fiables.	
Entornos amigables, con interfaces fáciles de utilizar para usuarios no expertos en herramientas informáticas.	

Tabla 1. Ventajas e inconvenientes de los Almacenes de Datos. Fuente: Propia y Bibliografía Básica.

## 2.2.- Aplicaciones OLAP

Se basan en considerar el tratamiento de los datos multidimensionales haciendo uso de la tecnología relacional. A partir de finales de los 90 aparecen dos tipos de arquitectura:

MOLAP (Multidimensional OLAP)	Utiliza bases de datos optimizadas para la consulta mediante el almacenamiento de las consultas que se pueden ejecutar.	Están optimizadas para la recuperación de datos, para mejorar la velocidad de acceso. Al menos contiene una dimensión temporal.	Es recomendada en bases de datos mas pequeñas, la velocidad de respuesta es menor y el espacio requerido es también menor.
ROLAP (Relational OLAP)	Trabaja directamente sobre bases de datos relacionales- Las dimensiones se almacenan en tablas relacionales.	Guardan la información agregada en nuevas tablas.	Es más escalable. Los procesos de consulta en volúmenes muy grandes de datos no resultan óptimos.
HOLAP (Hybrid OLAP)	Utiliza tablas relacionales para guardar los volúmenes grandes de datos y el almacén multidimensional para volúmenes más pequeños.		Presenta ventajas en el proceso de grandes volúmenes a la vez que presenta niveles aceptables de escalabilidad
DOLAP (Desktop OLAP) O In-memory OLAP	Representan las tendencias actuales y tratan de aprovechar la potencia de terminales de escritorio y la disminución en los costes de sus componentes		

## 2.3.- Arquitectura general de los almacenes de datos

En el presente apartado voy a comentar las características básicas de la arquitectura de un almacén de datos, de acuerdo a la bibliografía de referencia utilizada.

Consta de 5 partes principalmente:

- Fuentes de datos externas, sobre las que realizar los procesos de *ETL (Extraction, Transformation and Loading)*.
- Repositorio del almacén de datos y *datamarts*, o almacenes departamentales.
- Metadatos, datos sobre datos, que contienen información sobre todo lo relativo al almacén de datos.
- Herramientas de consulta del usuario final, o *front-end*.
- Servidores *OLAP, On-Line Analytical Processing*, que procesa las consultas de las herramientas de los usuarios finales y las lanza al servidor del almacén de datos.

La arquitectura se apoya en un conjunto de tecnologías que se pueden resumir en:

- El propio sistema de repositorio, la base de datos donde se encuentra el almacén de datos. Normalmente se trata de un *SGBDR, Sistema Gestión de Base de Datos Relacional*.
- Los servidores *OLAP*, que pueden ser *ROLAP*, basados en *OLAP* relacional, o *MOLAP*, basados en *OLAP* multidimensional.
- Las herramientas de consulta de los clientes para generar informes, gráficos, ficheros u otras para su procesamiento posterior con herramientas *OLAP* o de minería de datos.

De una manera gráfica se puede reflejar en la siguiente imagen:

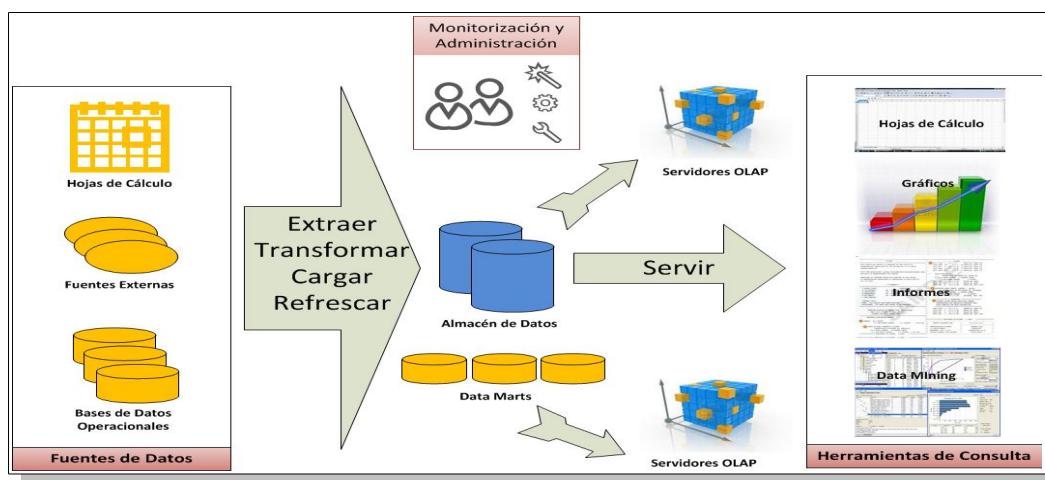


Imagen 1. Arquitectura general de un almacén de datos. Fuente: Propia y Bibliografía Básica.

### 3.- Objetivos, Métodos y Planificación

El objetivo del Trabajo Final de Carrera, TFC en adelante, consiste en “... proporcionar un conjunto prefijado de informes donde se muestre la información solicitada y cualquier otro que creamos que pueda ser útil para ONdO.”. En particular se trata de elaborar una serie de informes con unas características determinadas que relacione actividad turística disponible y equipamientos disponibles.

Vamos a tratar de diferenciar los objetivos, agrupándolos por intereses:

#### 3.1.- Objetivos de la asignatura

Los objetivos del TFC en el área de Datawarehouse, en adelante DW, vienen definidos explícitamente en el plan docente de la asignatura:

- Análisis de técnicas existentes para proyectar la base de datos en un DW.
- Diseño orientado a un almacén físico ROLAP.
- Crear las dimensiones, atributos y hechos necesarios.
- Considerar factores del tipo: desnormalización de tablas, inclusión de información agregada e historificación de la información.

#### 3.2.- Objetivos de la tecnología Datawarehouse

De acuerdo a las características generales de los almacenes de datos se pueden establecer los siguientes objetivos generales:

- Mantener histórico de datos, incluso si el sistema de origen no lo mantiene
- Integrar datos de múltiples sistemas de origen, habilitando una consulta centralizada para toda la organización.
- Mejorar la calidad de los datos, aportando código consistente y marcando datos erróneos e incluso corrigiéndolos cuando sea posible.
- Presentar información de la organización de manera consistente, válida en diferentes momentos del tiempo y coherente en la información mostrada a lo largo del tiempo
- Proveer un modelo de datos simple y común para los datos de interés, que sea intuitiva, de uso fácil y atractiva para el usuario final.
- Reestructurar los datos para que tengan sentido para los usuarios del negocio, aportando valor a los diferentes intereses de los miembros de la organización y que se pueden corresponder tanto a las áreas funcionales, los departamentos en sentido tradicional, como organizacionales, los diferentes grados de responsabilidad en la organización.
- Reestructurar los datos para que ofrezcan consultas de alto rendimiento, tanto en la ejecución de las mismas como en la utilidad proporcionada por los resultados generados en la consulta.
- Añadir valor a las aplicaciones de negocio transaccional más habituales en entornos empresariales: ERP, CRM,SCM,...

### 3.3.- Objetivos del TFC

Se pueden establecer dos tipos de objetivos, según el interés. En el primer grupo se incluirán aquellos que se constituyen en requisitos básicos para lograr los objetivos últimos del TFC – en general definición de la base de datos, elección del motor de base de datos, ... - mientras que en el segundo se identificarían aquellos objetivos concretos propuestos por el enunciado de la PEC – fundamentalmente la elaboración de informes concretos -. En algunos casos estos objetivos se corresponden con los expuestos anteriormente.

#### 3.3.1.- Objetivos primarios

- Creación de una base de datos en Oracle, de acuerdo a un modelo multidimensional. Incluye una serie de subtarefas propias del diseño de una base de datos: definición de los atributos y sus tipos, definición de relaciones entre tablas, establecimiento de reglas de normalización y desnormalización, ...
- Proporcionar una herramienta de consultas orientada a un usuario no experto en el manejo de herramientas informáticas pero con un conocimiento avanzado del modelo de negocio.
- Utilizar una interfaz de usuario sencilla e intuitiva.
- Proporcionar un modelo útil a las diferentes áreas del negocio, de forma que compartan estructuras comunes, se reduzcan los costes de aprendizaje y se puedan aplicar economías de experiencia y economías de escala.
- Realizar proyecciones, basándose en información histórica, que permitan desarrollar acciones eficaces y eficientes a plazo.
- Ha de reflejar los intereses de los diferentes usuarios de la organización, manteniendo los criterios de coherencia.

#### 3.3.2.- Objetivos específicos

Todos los informes que se proponen han de permitir agregar la información en base a diferentes criterios:

- Por comarca.
- Por provincia.
- Por tipo de establecimiento.
- Por categoría del establecimiento.

Los informes propuestos son los siguientes:

- Elaborar un informe que indique el número total de establecimientos.
- Elaborar un informe que indique el número total de plazas.
- Elaborar un informe que indique el número de plazas respecto a la población.
- Elaborar un informe que indique la oferta media de plazas.



- Elaborar un informe que indique la relación entre el número de establecimientos y el número de equipamientos.
- Elaborar un informe que indique el porcentaje de población por equipamiento.
- Elaborar un informe que indique la relación entre establecimientos y habitantes por género.
- Elaborar un informe que relacione plazas disponibles y población.
- Elaborar un informe que relaciones las variables equipamiento y población.
- Elaborar un informe que indique la relación entre cantidad de plazas ofertadas y la superficie del territorio.

### 3.4.- Método del TFC

Se pueden establecer en dos las bases para desarrollar el método a aplicar en el TFC:

- Por un lado los conocimientos adquiridos a lo largo de la carrera. Los conocimientos generales, fundamentalmente en el ámbito de la programación y de las bases de Datos, y los conocimientos específicos adquiridos en las asignaturas optativas y de libre elección, también en el ámbito de las bases de datos. La realización de asignaturas con un planteamiento similar a un proyecto, Técnicas de Desarrollo del Software, proporciona una perspectiva muy útil a la hora de enfocar el presente trabajo.
- Por otro lado la manera de enfocar las diferentes PECs, prácticas,... a lo largo de la carrera ha ido determinando un estilo que trato de utilizar aquí.
- Finalmente la experiencia profesional a la hora de redactar memorandos, informes, recopilar información de diversas áreas,... hace que pueda aplicar una perspectiva crítica al planteamiento general del proyecto.

Respecto al método, adoptaré propuesto en el enunciado de la PEC pero incluiré alguna modificación al mismo.

Fase	Subfase
Plan de Trabajo	Definición del Proyecto. Aproximación a la técnica de Datawarehouse. Definición de Objetivos. Establecimiento de Plazos.
Análisis Preliminar de Requerimientos	Identificación de las fuentes de información. Búsqueda de información complementaria interna. Búsqueda de información complementaria externa. Identificación de la estructura de las fuentes de información. Identificación de las entidades básicas y de las relaciones entre ellas. Identificación de los atributos y los tipos de datos.
Análisis de Requerimientos	Feedback de la fase anterior. Definición de las entidades. Definición de las relaciones entre entidades.
Fase de Requisitos	Instalación y Configuración de Oracle Database Express Edition 11 g. Test de creación de tablas, de importación de datos, ...
Diseño Conceptual y Técnico	Creación del modelo dimensional. Creación del diseño procedimental. Elaboración de la documentación del proyecto.
Implementación	Construcción de la Base de Datos: entidades, relaciones, índices, ... Procesos ETL (Extraction, Transformation and Loading), Carga de datos a partir de la información suministrada por el Observatorio Nacional de Ocupación y el Instituto de Estadística de Catalunya en formato no relacional. Instalación de herramientas de Datawarehouse. Familiarización con las herramientas. Diseño y Ejecución de los informes.

Análisis de los resultados obtenidos.
---------------------------------------

Tabla 2. Fases y Subfases principales del TFC. Fuente: Propia y Bibliografía Básica.

### 3.5.- Disponibilidades

Una vez he definido las actividades y tareas del proyecto a continuación voy a detallar la disponibilidad en mi horario para desarrollarlo:

*Disponibilidad General.* Se ha establecido una dedicación media diaria de 2-3 horas, que en la mayor parte de los casos se ha cumplido e incluso superado ampliamente. Han existido algunas excepciones: viajes de trabajo, reuniones familiares, reuniones,... que en general no han supuesto un retraso de mas de 48 horas respecto al calendario planificado inicialmente. Los fines de semana se han reservado una media de 4-5 horas diarias, pero con holgura suficiente como para recuperar jornadas o retrasos generados a lo largo de la semana. En algunas ocasiones el tiempo dedicado en fin de semana ha superado las 8 horas diarias.

Excepcionalmente se han utilizado días extra: Semana Santa, fiestas locales y un período extra de 2 días, 23 y 24 de Mayo, ya pactados con mi empresa para cubrir posibles incidencias extraordinarias que finalmente no han ocurrido.

*Excepciones Programadas.* Se han limitado a la asistencia de un curso de ingles, durante todos los miércoles del curso, la realización de un curso sobre arte que ocupo 4 jornadas en el mes de Marzo y un par de jornadas de formación-reunión con un colega de trabajo

*Excepciones No Programadas.* Por mi trabajo me he tenido que desplazar fuera de mi domicilio habitual en un par de ocasiones. EN cualquier caso han sido desplazamientos en el día y no han afectado más que en un ligero retraso a determinadas actividades.

*Excepciones Por Motivos de Salud.* No han tenido lugar.

*Excepciones por Motivos Técnicos.* Salvo un par de incidencias que dieron lugar a una degeneración de la máquina virtual, provocada por la realización repetida de pruebas, no se han producido incidencias que pudieran afectar gravemente a la realización del proyecto. En cualquier caso se realizaba copia de la información cada 2-3 días y conforme se acercaba el plazo de entrega aumentaba la frecuencia.

### 3.6.- Planificación

Las fases en las que se ha dividido el TFC están en función de las PECs planteadas en el Plan Docente de la asignatura y resumo en el siguiente cuadro:

PEC	Contenido	Fecha Inicio	Fecha Fin
<b>Actividades Previas</b>	Adquisición de Bibliografía básica sobre Data Warehouse. Instalación de la Máquina Virtual. Instalación de Oracle Express Edition. Realización de pruebas de definición de una base de datos, creación de tablas con índices y restricciones. Carga de datos de prueba.	30/12/99	28/02/13
<b>PEC 1</b>	Análisis Preliminar de Requerimientos. Plan de Trabajo. CALIFICACION: 24- Marzo-2013.	28/02/13	12/03/13
<b>PEC 2</b>	Creación del Modelo de Datos. Diseño Conceptual y Técnico. CALIFICACION: 28- Abril-2013.	13/03/13	16/04/13
<b>PEC 3</b>	Carga de Datos (ETL), Análisis de la Información, Creación de Informes y Análisis de Resultados. CALIFICACION: 09- Junio-2013	17/04/13	29/05/13
<b>Memoria Final</b>	Elaboración y Presentación de la Memoria. Debate Final. CALIFICACION: 09- Julio-2013	30/05/13	17/06/13

Tabla 3. Plazos definidos para las entregas del proyecto. Fuente: Propia.

En la imagen siguiente se muestra la planificación detallada de cada una de las fases en que se ha desarrollado el proyecto:

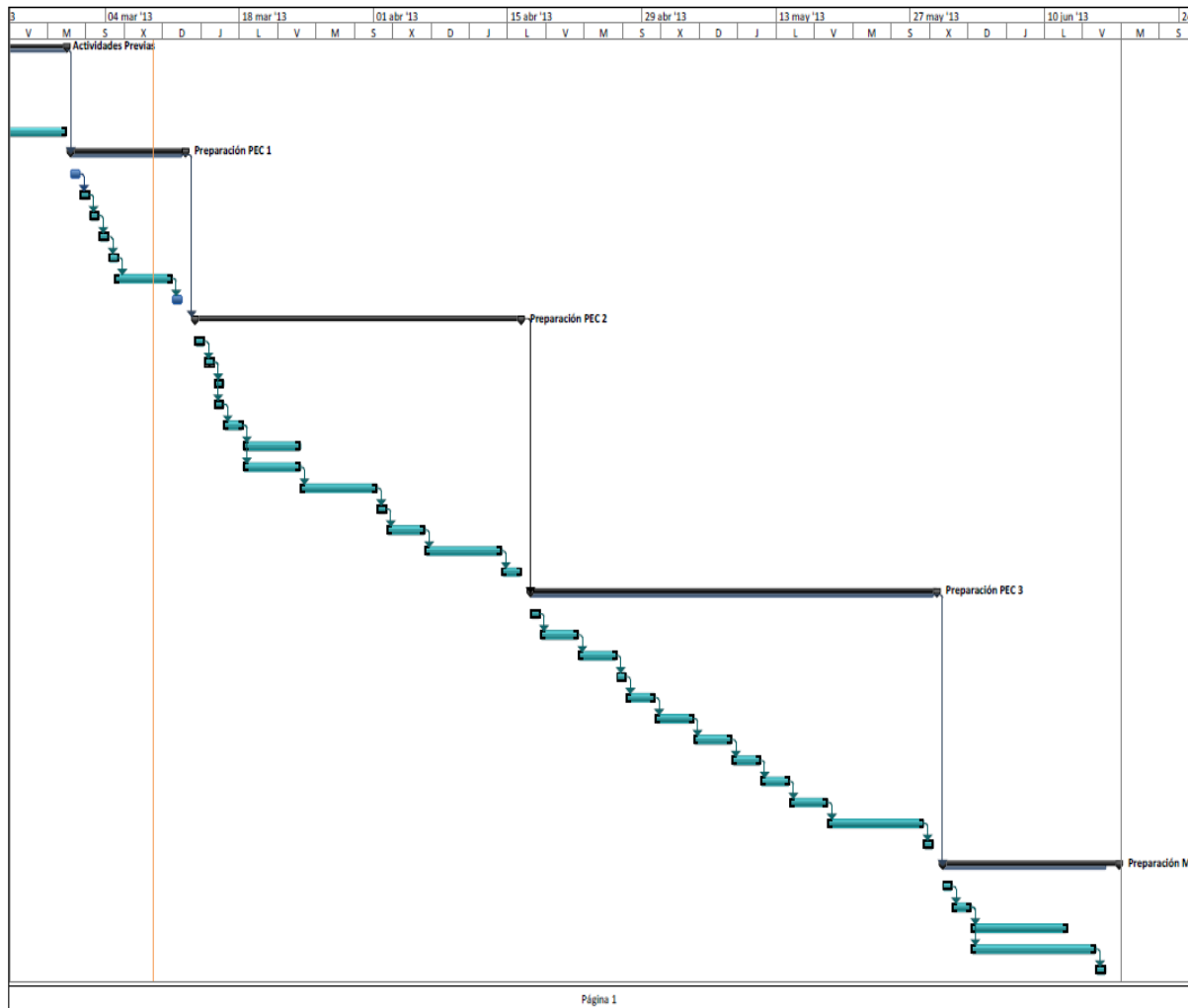


Imagen 2. Cronograma de cada una de las actividades. Fuente: Propia.

En la tabla siguiente se muestra la enumeración detallada de cada una de de las tareas:

Tarea	Fecha de Inicio	Fecha de Fin
<b>Actividades Previas</b>	<b>mar01/ene 13</b>	<b>mié 27/ene 13</b>
Recopilación de Información general	sáb 29/12/12	lun 07/01/13
Descarga, Instalación y Configuración de la máquina Virtual	mar 08/01/13	dom 13/01/13
Descarga, Instalación y Configuración de la Base de Datos	lun 14/01/13	dom 20/01/13
Realización de Pruebas Creación de Bases de Datos, de Tablas, de procesos ETL,... Descarga de documentación de la Biblioteca de la UOC	lun 21/01/13	mié 27/02/13
<b>Preparación PEC 1: Análisis</b>	<b>jue 28/feb 13</b>	<b>mié 13/mar 13</b>
Definición de Objetivos	jue 28/feb 13	vie 01/mar 13
Análisis de los Datos de Origen	vie 01/mar 13	sáb 02/mar 13
Análisis de la Exportación de Datos de Origen	sáb 02/mar 13	dom 03/mar 13
Valoración de los Datos de Origen	dom 03/mar 13	lun 04/mar 13
Análisis Preliminar de los Elementos Identificados	lun 04/mar 13	mar 05/mar 13
Elaboración del documento PEC 1	mar05/mar 13	lun 11/mar 13
Revisión PEC 1 y entrega	lun 11/mar 13	mar 12/mar 13
<b>Preparación PEC 2: Diseño</b>	<b>mié 13/mar 13</b>	<b>sáb 30/dic 99</b>
Revisión PEC 1	mié 13/03/13	jue 14/03/13

Tarea	Fecha de Inicio	Fecha de Fin
Instalación Máquina Virtual	jue 14/03/13	vie 15/03/13
Configuración del Entorno de Trabajo	vie 15/03/13	sáb 16/03/13
Lectura Documentación	vie 15/03/13	sáb 16/03/13
Análisis de Requerimientos	sáb 16/03/13	lun 18/03/13
Definición del Modelo	lun 18/03/13	dom 24/03/13
Diseño	lun 18/03/13	dom 24/03/13
Pruebas en el Entorno de Trabajo	dom 24/03/13	lun 01/04/13
Revisión de la documentación de la PEC 2 a fecha	lun 01/04/13	mar 02/04/13
Holgura para desviaciones	mar 02/04/13	sáb 06/04/13
Elaboración del documento PEC 2	sáb 06/04/13	dom 14/04/13
Revisión PEC 2 y entrega	dom 14/04/13	Mar 16/04/13
<b>Preparación PEC 3: Implementación</b>	<b>mié 17/abr 13</b>	<b>mié 29/may 13</b>
Revisión PEC 1	mié 17/04/13	jue 18/04/13
Diseño, Construcción de Scripts y Ejecución de la Base de Datos	jue 18/04/13	lun 22/04/13
Carga de Datos	lun 22/04/13	vie 26/04/13
Instalación de la Herramienta de BI	vie 26/04/13	sáb 27/04/13
Implementación del Sistema	sáb 27/04/13	mar 30/04/13
Elaboración de los Informes requeridos	mar 30/04/13	sáb 04/05/13
Ejecución de los informes requeridos	sáb 04/05/13	mié 08/05/13
Análisis de los Resultados	mié 08/05/13	sáb 11/05/13
Feedback de los Análisis. Revisión de diseño e informes	sáb 11/05/13	mar 14/05/13
Holgura para desviaciones	mar 14/05/13	sáb 18/05/13
Elaboración del documento PEC 3	sáb 18/05/13	mar 28/05/13
Revisión PEC 3 y entrega	mar 28/05/13	mié 29/05/13
<b>Memoria Final</b>	<b>jue 30/may 13</b>	<b>lun 17/jun 13</b>
Revisión PEC 3	mié 29/05/13	vie 31/05/13
Conclusiones	vie 31/05/13	dom 02/06/13
Elaboración de la Memoria	dom 02/06/13	mié 12/06/13
Elaboración Presentación Para el debate	dom 02/06/13	sáb 15/06/13
Revisión Memoria y entrega	sáb 15/06/13	Dom 16/06/13

Tabla 4. Plazos definidos para las entregas del proyecto. Fuente: Propia.

#### 4.- Análisis

En general en todo proceso de identificación de los datos hay que tratar de localizar fuentes fiables, válidas, documentadas y útiles para el proceso que tratamos de establecer un modelo. En este sentido, y de manera genérica, en el siguiente gráfico se muestra las distintas fuentes que se pueden considerar en un proyecto de este tipo:

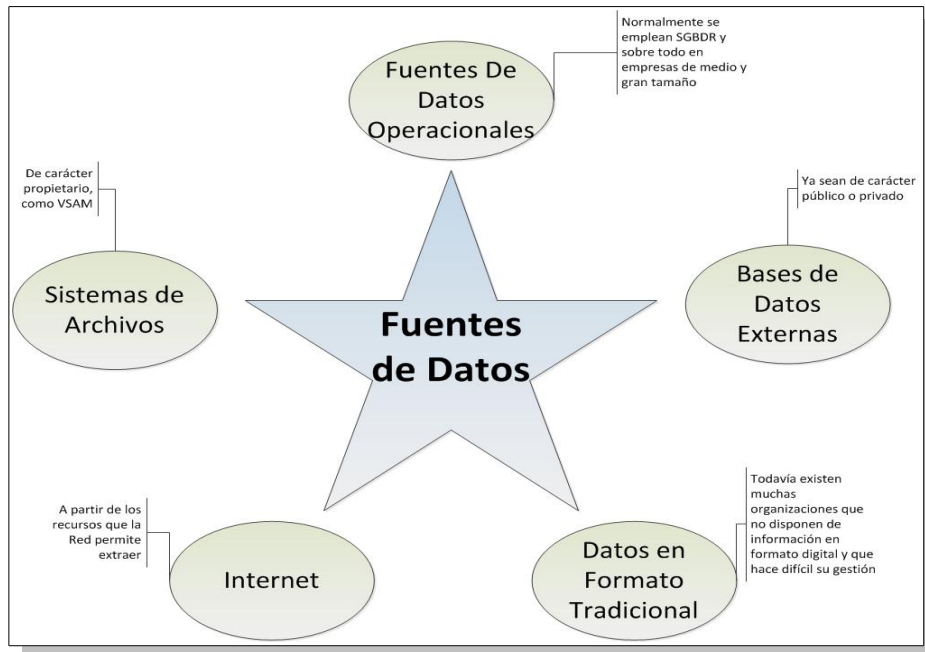


Imagen 3. Orígenes de las distintas fuentes de datos. Fuente: Propia y Bibliografía Básica.

#### 4.1.- Análisis de las fuentes proporcionadas

A continuación se analizarán los datos fuente proporcionados.

Si bien en la PEC 1 ya se llevó a cabo una aproximación, en esta etapa se va a presentar un análisis diferenciado del contenido de cada uno de los ficheros, de forma que nos sirve para la definición de las estructuras necesarias que implantaremos posteriormente.

Vamos a tratar de resumir toda la información de las tablas en unos formatos tipo ficha, para ver las semejanzas entre unos y otros:

Fichero	Elemento	Valores
<b>Poblacio.csv</b>	Descripción general	Proporciona estadísticas de una serie de municipios de la comunidad autónoma de Catalunya.
	Formato	Texto plano.
	Extensión	csv
	Organización del fichero	Se organiza en filas , con estructura determinada. Cada línea contiene información de un municipio. Esta organizado alfabéticamente por nombre del municipios.
	Separador de registro	Los registros están separado por los caracteres [CR]: retorno de carro. [LF]: salto de línea.
	Separador de campo	Los campos están separados por el carácter “;”.

Tabla 5. Estructuras del fichero Poblacion.csv. Fuente: Propia.

Fichero	Elemento	Valores
<b>Equipaments.csv</b>	Descripción general	Proporciona información sobre infraestructuras de los municipios de Catalunya así como datos específicos sobre los mismos: municipio, comarca, dirección, teléfono, comarca, código postal y datos para el geoposicionamiento.
	Formato	Texto plano.
	Extensión	csv

Fichero	Elemento	Valores
	Organización del fichero	Se organiza en filas , con estructura determinada. Cada línea contiene información de un equipamiento. La información detallada de los equipamientos está consolidada en único campo.
	Separador de registro	Los registros están separado por el carácter [LF]: salto de línea.
	Separador de campo	Los campos están separados por el carácter “;”. Existe un campo, el segundo, que hace referencia a la dirección y que puede contener el carácter “;”. Para evitar que se confunda con el carácter separador de campo dicho campo va incluido entre comillas dobles (“”) para evitar saltos de campo no deseados.

Tabla 6. Estructuras del fichero Equipaments.csv. Fuente: Propia.

Fichero	Elemento	Valores
<b>Establiments NNN.txt</b>	Descripción general	Proporciona información sobre establecimientos turísticos y las zonas asociadas, aportando información redundante en muchos casos, totales de otras columnas, y en otros casos la información es inconsistente debido a la desestructuración de la información del fichero.
	Formato	Texto plano.
	Extensión	txt
	Organización del fichero	Se organiza en filas , con estructura diferente de acuerdo a la siguiente distribución: Estructura tipo A: 2006. Estructura tipo B: 2007, 2008, 2009, 2010. Estructura tipo C: 2011. Estructura tipo D: 2012. Cada línea contiene información de un equipamiento. La información detallada de los equipamientos está consolidada en único campo.
	Separador de registro	Los registros están separado por el carácter [LF]: salto de línea.
	Separador de campo	Los campos están separados por el tabulador, [TAB].

Tabla 7. Estructuras del fichero EstablimentsNNN.txt. Fuente: Propia.

#### 4.2.- Análisis de las características de los datos de las fuentes de información

Independientemente de la estructura de cada uno de los ficheros y sus diferentes formatos se puede establecer el siguiente cuadro por el que podemos establecer el tipo de campo y el tamaño de los mismos:

Fichero	Atributo	Tipo	Valores
<b>Poblacio.csv</b>	[Municipio]	Texto.	Longitud Máxima: 36.
	[Codigo Ine]	Texto.	Longitud Máxima: 34. ERROR. Correspondiente a un dato incoherente en la fila 142. Habitual: 4 o 5.
	[Poblacion2012], [Poblacion2011], [Poblacion2010], [Poblacion2009],	Entero, con separador de miles (“.”).	Máximo: 1621537.

Fichero	Atributo	Tipo	Valores
	[Poblacion2008], [Poblacion2008], [Poblacion2007], [Poblacion2006]		
	[Poblacion2012Homes], [Poblacion2012Dones]	Entero, con separador de miles (“.”).	Máximo: 847533.
	Extensió (km2)	Entero.	Máximo: 631.
<b>Equipments.csv</b>	[nom]	Texto.	Longitud Máxima: 149.
	[adreca]	Texto, entrecomillada.	Longitud Máxima: 113.
	[municipi]	Texto.	Longitud Máxima: 45.
	[cp]	Texto.	Longitud Máxima: 21. ERROR: Correspondiente a un dato incoherente en la fila 15135. Habitual 4 o 5
	[comarca]	Texto	Longitud Máxima: 24
	[telefon]	Texto	Longitud Máxima: 12
	[longitud]	Decimal Carácter “.” como separador de decimales	Número máximo de decimales 18. (Fila 15891) Habitual 10 o 12 decimales
	[latitud]	Decimal Caracter “.” como separador de decimales	Número máximo de decimales 18. Habitual 10 o 12 decimales
	[categories]	Texto	Longitud Máxima: 382
	[Location]	Texto	Longitud Máxima: 83
<b>EstablimentsNNN.txt</b>	[Zona]	Texto	Longitud Máxima: 19
	[Establiments XXX]	Entero, con separador de miles (“.”)	Valor Máximo: 588814 *
	[Plazas XXX]	Entero, con separador de miles (“.”)	Valor Máximo: 5346 *

Tabla 8. Tipos de datos para los ficheros proporcionados. Fuente: Propia.

\* Esos valores se corresponden con campos calculados, que responden a la suma de campos simples.

#### 4.3.- Análisis de los datos

Se observan ciertas anomalías en los ficheros que constituyen las fuentes de información. Ante la presencia de errores caben tres posibles actuaciones:

- Tratar de determinar los valores correctos, en base a series históricas, registros similares, valores medios o estimaciones por procedimientos estadísticos (inferencia, estimadores MCO,..).
- Omitir esos datos en el proceso de importación de forma que la información traspasada sea el resultado de mediciones ciertas y no por otro procedimientos aunque sean matemáticamente aceptables.
- Solicitar a ONdO la comprobación de los datos proporcionados, la corrección de los mismos, si procede, y la autorización para omitir datos, establecer hipótesis de trabajo y restricciones al modelo-

A partir de un análisis visual se pueden comprobar una serie de incidencias para cada uno de los ficheros fuente que pueden dar lugar a dificultades en la generación de las tablas de dimensiones y de la propia tabla de

hechos.

#### 4.3.1.- Fichero Poblacio.csv

Las incidencias observadas en el fichero con datos relativos a a evolución de la población por zonas quedan resumidas en:

- Línea 142. El campo [Codi INE] no es de tipo numérico, como debería, sino que hace referencia a un texto. *Solución:* se puede acudir a los datos del Instituto Nacional de Estadística y obtener el valor correcto. De esta forma a buscar el Código INE del municipio de Cruilles se comprueba que el municipio tiene por nombre completo “Cruilles Monells i Sant Sadurn de l'Heura”. *Explicación:* la coma que incluye el nombre puede ser considerada como un salto de campo. La forma más adecuada de solucionar este problema es la de omitir la coma y sustituirla por otro carácter, por ejemplo un guión. A la hora de generar estos datos habría que considerar la posibilidad de entrecorillarlos con comillas dobles.
- Línea 99. Canonja(La). Los campos [Poblacio XXXX] tienen valor nd para los años 2007 al 2011 y tienen valor 0 para el año 2006 y valores coherentes para el año 2012. *Solución:* asignar valor 0. *Explicación:* Se trata de un ayuntamiento que se constituyó en 2010, por lo que no existía como tal anteriormente.

#### 4.3.2.- Fichero Equipaments.csv

Respecto al fichero que contiene la información relativa a las infraestructuras disponibles y sus características encontramos las siguientes incidencias:

- Se observa información imprecisa o errónea pero que no afecta de manera relevante al modelo de BI que se quiere implementar:
  - Equipamientos con CP asignado “0”. *Solución:* asignar el código postal de otro equipamiento que haya en el mismo domicilio. Si no es posible se asignará valor cero. Existen 16 líneas de este tipo.
  - Equipamientos con CP claramente erróneo. Existen dos casos: la línea 3313 y la 11551, con valores 32 y 9862. *Solución:* asignar el código postal de otro equipamiento que haya en el mismo domicilio. Si no es posible se asignará valor cero. Existen 16 líneas de este tipo.
  - Equipamientos con CP igual a “-”. Existen 349 casos. *Solución:* asignar el código postal de otro equipamiento que haya en el mismo domicilio. Si no es posible se asignará valor cero.
  - Equipamientos con CP con datos de un tipo no permitido. No incluyen un valor numérico, sino un texto, correspondiente al nombre del municipio. Se trata de 6 registros. *Solución:* asignar el código postal de otro equipamiento que haya en el mismo domicilio. Si no es posible se asignará valor cero. Existen 16 líneas de este tipo.
  - Equipamientos con CP correspondiente a otras provincias, considerando los dos primeros dígitos del código postal. Existen 15 registros de este tipo. *Solución:* asignar el código postal de otro equipamiento que haya en el mismo domicilio. Si no es posible se asignará valor cero.
- Longitud con valores no coherentes. La mayor parte de las coordenadas de Longitud están en el rango [0,4]. Existen 60 registros que no tienen valores coherentes. De estos 60 tenemos 49 registros que comienzan con el carácter “.”, por lo que podemos asimilar que son realmente “0.”. *Solución:* habrá que realizar el cambio “.” por “0”.
- Latitud con valores no coherentes. La mayor parte de las coordenadas de Latitud están en el rango [32.49], el resto son realmente incoherentes por lo que no es posible interpretar los valores correctos. *Solución:* ignorar esos registros o asignarles un valor nulo.
- Location. Hay que establecer una regla que permita verificar que los valores coinciden con los valores de longitud y latitud.

#### 4.3.3.- Ficheros EstablimentsXXXX.txt

Finalmente en la información contenida en los ficheros relativos a datos de establecimientos turísticos y que están diferenciados por años encontramos las siguientes anomalías:

- Aparentemente no hay inconsistencia en los datos que forma parte de la serie, para todos y cada uno de los ficheros.
- En los datos el año 2011 hay una nueva Comarca, Penedes, lo que altera el número de zonas a



considerar en el listado.

- En los datos del año 2012 el fichero aparece desplazado respecto a la fila en la que están referenciados. Así entre las columnas 18 y 27, relativas a los datos de Campings, las filas están desplazadas 2 posiciones hacia arriba. Por otro lado entre las columnas 29 y 40, relativas a información de turismo rural, se observa un desplazamiento hacia arriba de 1 fila, considerando la posición en la que debería estar situada.

#### 4.4.- Diagrama de Casos de Uso

En el análisis que nos ocupa podemos distinguir dos tipos de actores o perfiles que llevarán cabo funciones claramente diferenciadas:

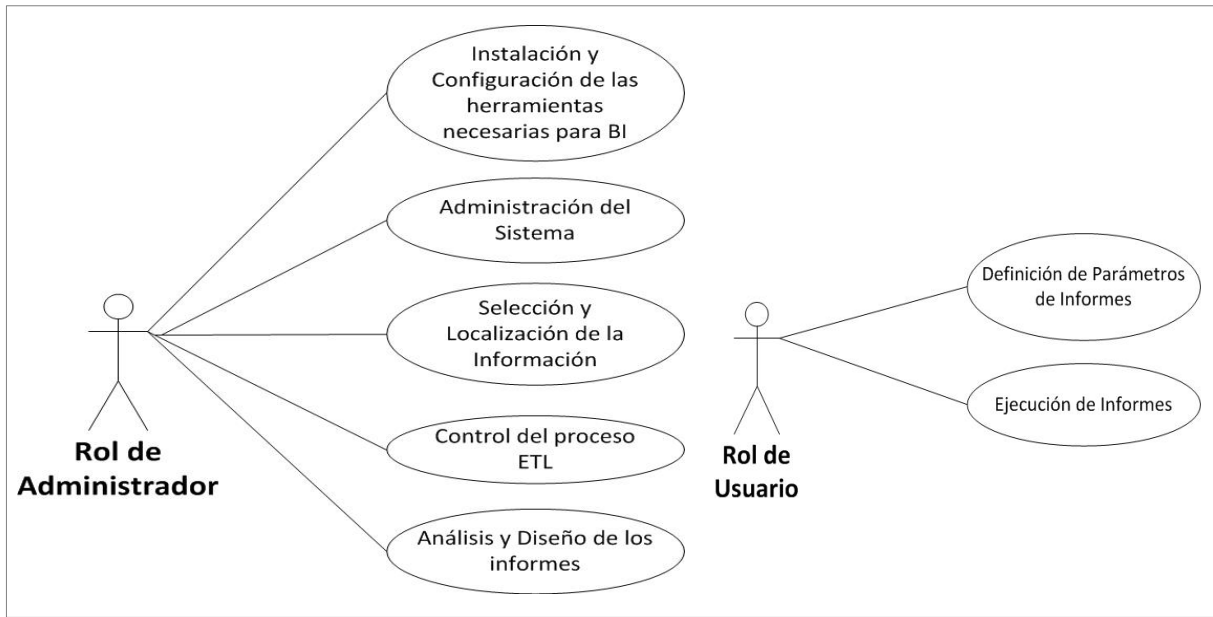


Imagen 4. Casos de Uso del modelo. Fuente: Propia y Referencias. Fuente: Propia.

El rol de Administrador tendrá a su cargo las siguientes funciones:

- *Instalación y Configuración de las Herramientas Necesarias* para desarrollar el modelo de BI. Entre ellas se encuentran: la instalación y configuración de la máquina virtual, descarga; instalación y configuración de las herramientas de software libre explicadas más adelante; instalación o configuración de todas aquellas herramientas que considere necesarias para la consecución del objetivo final: herramientas ofimáticas, editores de texto,...
- *Administración del Sistema*. Ha de velar por el correcto funcionamiento del sistema diseñando y aplicando una política de seguridad adecuada, gestionando las actualizaciones del software instalado y realizando las copias de seguridad mediante un plan diseñado e implantando de manera correcta.
- *Selección y localización de la Información*. Principalmente refiriéndose a las fuentes de información, tanto las proporcionadas por el cliente como aquellas que pudiera ser necesario instalar, indicando siempre su origen, disponibilidad y fiabilidad.
- *Control del Proceso ETL (Extracción, Transformación y Carga)*. Con las herramientas disponibles y las fuentes de información disponibles ha de asegurar la correcta ejecución de los procesos de extracción de la información, transformación de la misma y carga en la estructura de la base de datos diseñada al efecto.
- *Análisis y Diseño de los informes*. A partir de la información, conveniente cargada y transformada si fuera necesario, y de acuerdo a las especificaciones del cliente se diseñarán los informes, tanto en forma de tabla, como gráfico o incluso formato de fichero externo (PDF, XLS, JPG, CSV,...).

Indicar así mismo que ocasionalmente el Administrador puede desempeñar funciones de Usuario, aunque no se recomienda esta opción para garantizar la seguridad en los permisos de acuerdo a la función que desempeñe el usuario.

Análogamente el Rol de Usuario dispone de una serie de funciones mucho más limitadas, encaminadas fundamentalmente a preservar la seguridad de la información y acceder a la información en los términos acordados. De este modo las funciones asignadas son:

- *Definición de parámetros de Informes.* Del tipo características del establecimientos, echas, zonas o tipos de equipamientos.
- *Ejecución de Informes.* A partir de la parametrización establecida previamente.

#### 4.5.- Modelo Conceptual

El modelo conceptual pasa por identificar los dos elementos principales en cualquier modelo de Business Intelligence: las dimensiones y la tabla de hechos.

##### 4.5.1.- Identificación del Hecho

En el caso que nos ocupa ONDO, ella institución que quiere realizar el estudio, ha proporcionado los datos a relacionar: zonas, equipamientos y datos de ocupación de establecimientos turísticos.

A partir de está información el hecho más simple es la oferta turística de una unidad administrativa de Catalunya. La oferta turística viene determinada por un período, en nuestro caso ejercicios completos, un tipo de establecimiento y una categoría determinada. Además será necesario considerar los equipamientos ofertados para esa zona. Por zona entendemos distintos niveles de agrupaciones administrativas: municipios, comarcas, veguerías, provincias y comunidad autónoma.

##### 4.5.2.- Granularidad

La granularidad del modelo condicionará la eficiencia del modelo propuesto, La granularidad supone determinar el grado de detalle en las consultas que se van a ejecutar y que puede condicionar la velocidad de ejecución de las mismas. Una granularidad poco detallada puede que no sea adecuada a las necesidades de información de ONDO.

En el caso que nos ocupa y a partir de la información que es suministrada se considera que es suficiente para la obtención de informes. Explícitamente se establece en el enunciado que “... la temporalidad de los datos será el año.”.

La granularidad para la dimensión temporal considerada será de carácter anual.

Respecto a la granularidad de los datos relativos a establecimientos la información suministrada es a nivel de establecimiento por municipio, en el caso más detallado, y por zona en el caso de la información por establecimientos. En principio la información solicitada por ONDO es a un nivel superior, a nivel de comarca, lo cual nos va a exigir una granularidad no tan detallada: la proporcionada a nivel de municipio frente a la solicitada a nivel de comarca.

##### 4.5.3.- Dimensiones a establecer para el análisis

Las dimensiones determinan los ejes en base a los cuales se van a extender los datos y por tanto podemos obtener los informes solicitados. También se denominan dimensiones del cubo, puesto que solo podemos imaginar una representación tridimensional.

De acuerdo a la granularidad establecida en el apartado anterior las dimensiones establecidas para e modelos son:

- *Fecha.* Esta dimensión, de obligada inclusión en cualquier modelo multidimensional se constituye en la unidad temporal del modelo, ya que no existe, ni se considera otro tipo de agrupación temporal.

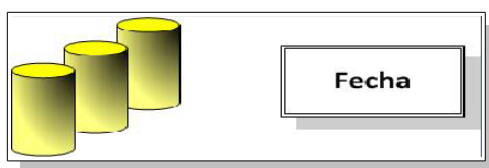


Imagen 5. Dimensión Fecha. Fuente: Propia.

- *Zona.* Se constituye en una de las dimensiones más importantes, ya que va a recoger los datos agregados para la mayor parte de los informes y va a ser el criterio en base al cual se va a consolidar la información de los establecimientos.

- Jerarquía: Comunidad Autónoma → Provincia → Comarca
- Por otro lado tenemos un conjunto de zonas, denominadas veguerías, que viene a corresponderse como una unidad administrativa intermedia entre la provincia y la comarca. No se ha podido disponer de la composición exacta de las veguerías, incluso se plantean dificultades legales, por lo que, aunque se incorporarán a la tabla de hechos y las dimensionales correspondientes no se puede asegurar la fiabilidad de la mayoría de los datos como los relativos a equipamientos, desde la perspectiva de las veguerías. por ejemplo.

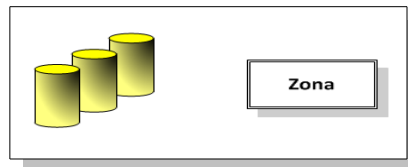


Imagen 6. Dimensión Zona. Fuente: Propia.

- *Equipamiento* es una dimensión que permitirá obtener información por el tipo de equipamiento quede acuerdo a otras dimensiones.

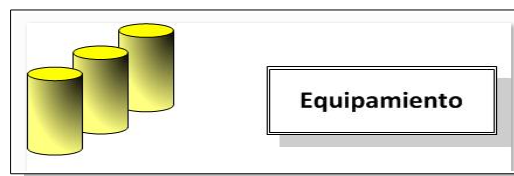


Imagen 7. Dimensión Equipamiento. Fuente: Propia.

- *Establecimiento*. Esta dimensión hace referencia al tipo y la categoría de los establecimientos que se quieren analizar. Esta dimensión viene determinada por la granularidad, tal y como nos ha sido proporcionada en ella fuente de datos, de manera que los datos proporcionados son al máximo nivel de detalle: por municipio. Observamos también que necesitamos un nivel de detalle superior, por lo que podemos establecer la siguiente jerarquía
  - Jerarquía: tipo → subtipo

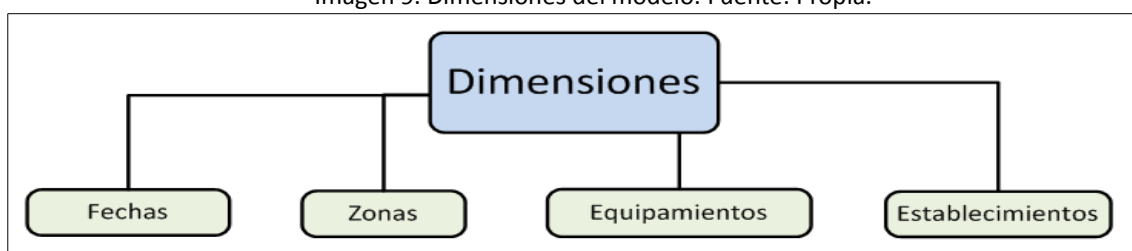


Imagen 8. Dimensión Establecimiento. Fuente: Propia.

De igual manera para la información referida a establecimientos se va a considerar una granularidad por zona, tipo de establecimiento y subtipo, o categoría, aprovechando la que proporcionan las fuentes de información suministradas.

Gráficamente la dimensiones propuestas darían lugar a un modelo bastante diferente del propuesto:

Imagen 9. Dimensiones del modelo. Fuente: Propia.



#### 4.5.4.- Identificación de los atributos de las Tablas de Dimensiones

A partir de las dimensiones definidas anteriormente el siguiente paso será establecer los atributos y las jerarquías, si procede, para cada uno de los registros de la tabla de Hechos.

- *Dimensión Fechas.* Está dimensión solamente consta de un único atributo. En ningún caso se puede omitir la dimensión Fecha, ya que se constituye en el pivote del análisis y de acuerdo a la bibliografía básica consultada se indica que “... la definición de una dimensión temporal es obligatoria en un modelo multidimensional,...”<sup>1</sup>
- *Dimensión Zonas.* Referida a la Comarca en la que se encuentra el dato sobre establecimiento. Nos determinará la granularidad que permitirá el máximo detalle. Se establecerán los siguientes atributos:
  - *Id.* De tipo entero. Clave Primaria. Se trata de una clave autogenerada, conforme vayamos rellenando la tabla a partir de la información de las fuentes.
  - *tipoZona.* De tipo Texto y no nula.
  - *nombreZona.* De tipo Texto y no nulo.
  - *superficieZona.* De tipo Entero y no nulo.
- *Dimensión Establecimientos.* Se utilizará para identificar los tipos de establecimientos que considera el modelo. Se cargará también en el momento de crear el modelo en el motor de base de datos. Consta de los siguientes atributos:
  - *Id.* De tipo Entero. Clave Primaria. Se trata también de una clave autogenerada.
  - *tipoEstablecimiento.* De tipo Texto.
  - *categoria.* De tipo Texto y no nulo.
- *Dimensión Equipamientos.* Se utilizará para almacenar la información referida a los diferentes equipamientos proporcionados en la fuente de información, fichero Equipaments.csv. Se genera conforme se lee la información del fichero fuente. Consta de los siguientes atributos:
  - *Id.* De tipo Entero. Clave Primaria. Se trata también de una clave autogenerada.
  - *tipoEquipamiento.* De tipo Texto y no nulo.
  - *SubTipoEquipamiento.* De tipo Texto y no nulo.

#### 4.5.5.- Identificación de las medidas a emplear

Las medidas que se van a considerar para responder a las necesidades de ONdO serán las siguientes:

- Población, que denominaré censo para evitar confusión con el fichero poblacio.csv. A considerar que en la mayor parte de los datos suministrados la población no se diferencia por sexos, siendo sólo en la última serie de información cuando se distingue entre hombres y mujeres. Para solucionar este problema, voy a realizar una estimación de la distribución de la población en mujeres y hombres. Para ello aplicaré una sencilla fórmula que me permitirá obtener la proporción de unos y otros para el año 2012, que es cuando se dispone de esa información. Así:

$$\%Población Femenina = \frac{Població 2012 dones \times 100}{Població 2012}$$

$$\%Población Masculina = \frac{Població 2012 homes \times 100}{Població 2012}$$

A partir de estos porcentajes y dentro de la fase de ELT se propondrá una estimación de la distribución de la población para cada uno de los años de los que se dispone información agregada. En la práctica y para evitar los efectos del redondeo consideraremos la población femenina como el complemento a la Población Total en un Año de la cifra de Población Masculina.

$$Población NNNN homes = \frac{Població NNNN * Població 2012 homes}{Població 2012}$$

$$Población NNNN dones = Població NNNN - Población NNNN homes$$

- Establecimientos. Son dos las variables que se van procesar: el número de establecimientos y las plazas ofertadas para cada uno de ellos.
- Equipamientos. Se expresa en número de equipamientos ofertados, diferenciándose por tipo de equipamiento, de manera que un único establecimiento ofrezca diferentes equipamientos. En el enunciado aparecen referenciados a fecha 31-12-12. Son varias las opciones las que se pueden establecer para incorporar esta información a la tabla de hechos:
  - Considerar una cifra constante para cada uno de los ejercicios de los que se dispone

<sup>1</sup> Diseño y Explotación de Almacenes de Datos. Juan Carlos Trujillo y otros. Página 44

información. Sería la hipótesis más sencilla.

- Realizar una estimación, con un criterio objetivo y verosímil, que permita asociar una estructura de equipamientos diferenciada por años.

Aunque inicialmente se había propuesto realizar una estimación deflactando por el PIB, se ha considerado omitir ese tratamiento, dado que los equipamientos responden a individuos concretos y reales con un conjunto de equipamientos, variable para cada uno de ellos, de manera que la omisión o creación de equipamientos según ese deflactor para cada año considerado podría alterar sustancialmente los resultados.

Todas estas medidas tendrán su reflejo en la hoja de Hechos, en forma de atributos. De su combinación con diferentes criterios de selección y agrupándola convenientemente ha de ser suficiente para poder elaborar los informes que se plantean por parte de ONdO.

La información solicitada se podrá mostrar en formato de gráfico, complementando la información que se presenta en los formatos tradicionales de informe o en hoja de cálculo.

#### 4.5.6.- Identificación de los atributos de la tabla de Hechos

De acuerdo al modelo propuesto en el gráfico anterior, Imagen 9, tenemos claramente identificadas las dimensiones, por lo que la tabla de hechos nos ha de proporcionar toda la información para cada una de las combinaciones de los atributos de la clave primaria de dicha tabla, considerando que el atributo temporal, en nuestro caso "Año" ha de ejercer de pivote, de forma que la no inclusión de la misma en una consulta puede dar lugar a resultados desvirtuados.

En este sentido muestro algunos ejemplos de consultas que han de ser respondidas con la tabla de Hechos:

- Número de establecimientos de una determinada zona en una fecha determinada. Hay que recordar que fecha se corresponde con una agrupación de fechas, pudiendo ser un trimestre, aunque en la práctica no ocurrirá puesto que no disponemos de esa granularidad tan detallada.
- Número de plazas bajo los mismos condiciones que el número de equipamientos.
- Número de equipamientos disponibles para un determinado tipo de establecimiento en una fecha determinada para una zona concreta.

En cambio estas otras consultas, que no incluyen el pivote temporal, pueden proporcionar información errónea:

- Número de establecimientos de tipo hotel. Esta consulta carece de sentido, a priori, porque devolvería la suma de plazas ofertadas de tipo hotel para todo el período histórico del que se tuviera información.
- Población de una veguería. Esta consulta nos devolvería la suma de todos los habitantes para todos los períodos de los que se tiene información. No tiene sentido en ningún caso.

En cambio existen otras consultas que pueden tener sentido sin considerar, explícitamente, la dimensión temporal:

- Población media. Se obtendría de acuerdo a la siguiente fórmula:

$$Poblacion = \frac{\sum Poblacion_i}{Número de Años}$$

Implícitamente considera el número de valores diferentes de la dimensión temporal.

En función de la información solicitada por ONdO la tabla de Hechos contendrá información que es necesaria para elaborar los informes solicitados. La tabla de Hechos tiene por objetivo maximizar el rendimiento de las consultas realizadas aun a costa de no cumplir los criterios de normalización definidos por Boyce-Codd. De acuerdo con la bibliografía básica consultada dos son las ventajas:

- Mejorar el rendimiento eliminando las uniones entre tablas, ya que está es la operación que tiene un coste mayor en el ámbito de las bases de datos.
- Manejo intuitivo. Para el usuario no avanzado la información principal se encuentra en una única tabla y ha de olvidarse de las relaciones y restricciones de un modelo relacional clásico.

La propuesta de atributos de la tabla de Hechos queda como sigue:

Tipo	Atributo	Descripción
	idFecha	Identificador del Año al que se refiere la información.
	idZona	Identificador único de la comarca.

Tipo	Atributo	Descripción
Clave Primaria	idEstablecimiento	Identificador único de la categoría del establecimiento.
	idEquipamiento	Identificador único del equipamiento considerado.
Clave Primaria	censoHombres	Censo de población masculina para la comarca considerada en el año referido. Expresado en unidades.
	censoMujeres	Censo de mujeres para la comarca considerada en el año referido. Expresado en unidades.
	numeroEstablecimientos	Número de establecimientos para la comarca considerada con la categoría definida y en el año referido. Expresado en unidades.
	numeroPlazas	Número de plazas para la comarca considerada con la categoría definida y en el año referido. Expresado en unidades.
	numeroEquipamientos	Número de Equipamientos para la comarca considerada y el tipo de equipamiento referido. Expresado en unidades

Tabla 9. Estructura de la Tabla de Hechos. Fuente: Propia.

#### 4.6.- Estudio de Viabilidad

El estudio de viabilidad de cualquier análisis tiene por objeto determinar el tamaño aproximado de los datos que se van a gestionar, en base a las consideraciones previas del modelo que se han ido detallando en este estudio.

Para determinar el tamaño de la información es necesario determinar correctamente las el tamaño de los datos de las dimensiones, de acuerdo a las especificaciones de MySQL<sup>2</sup>:

Tipo	Dimensión	Tamaño Aproximado / Valor Máximo	Tipo de Dato / Tamaño Unitario
Clave Primaria	fecha	7	YEAR, Máx 65535, tamaño 1 byte
	idZona	54	Entero, TINYINT, Máx 250, tamaño 1 byte
	idEstablecimiento	28	Entero, TINYINT, Máx 250, tamaño 1 byte
	idEquipamiento	219	Entero, SMALLINT, Máx 65535, tamaño 2 byte
	censoHombres	7,363,891	Entero, MEDIUMINT, Máx 16777215, tamaño 3 byte
	censoMujeres	7,363,891	Entero, MEDIUMINT, Máx 16777215, tamaño 3 byte
	numeroEstablecimientos	22,696	Entero SMALLINT, Máx 65535, tamaño 2 byte
	numeroPlazas	2,358,216	Entero, MEDIUMINT, Máx 16777215, tamaño 3 byte
	numeroDeEquipamientos	31,771	Entero SMALLINT, Máx 65535, tamaño 2 byte

Tabla 10. Dimensionamiento de la Tabla de Hechos. Fuente: Propia.

Los datos estimados para esta tabla se han obtenido considerando los siguientes supuestos:

- Se han establecido datos para 7 años, de 2006 a 2012, ha partir de los datos disponibles de establecimientos.
- Se estiman 54 zonas diferentes, a partir de los datos de establecimientos para el año 2012.
- Se estiman 15 establecimientos diferentes: tres tipos de alojamientos – hoteles, camping y turismo rural- y entre 4 y 5 categorías para cada uno de ellos.
- Se estiman en 255 equipamientos diferentes a partir de los datos del fichero equipments.csv.

<sup>2</sup> <http://dev.mysql.com/doc/refman/5.0/es/storage-requirements.html>

- Se estima una población máxima de 7,363,891 en el año 2012 de acuerdo
- El número de equipamientos se ha obtenido a partir de los datos del fichero equipaments.csv.

De aquí podemos establecer el siguiente volumen de datos:

Variable	Tamaño Aproximado / Valor Máximo
Número de Registros	$7 * 54 * 15 * 255 = 1.445.850$ registros
Tamaño de la clave primaria	$2 \text{ byte} * 1 \text{ atributos} + 1 \text{ byte} * 3 \text{ atributos} = 5 \text{ bytes}$
Tamaño de campos que no son clave <sup>(3)</sup>	$3 \text{ byte} * 4 \text{ atributos} + 2 \text{ byte} * 2 \text{ atributos} = 16 \text{ bytes}$
Tamaño de un registro	$5 + 16 = 21 \text{ bytes}$
Tamaño de la Base de Datos	$21 * 1.445.850 = 30.362.850 \text{ bytes} = 28,956 \text{ Mb}$

Tabla 11. Dimensionamiento de la Base de Datos. Fuente: Propia.

#### 4.7.- Esquema Conceptual y Lógico

Finalmente el esquema propuesto se muestra en el siguiente cuadro:

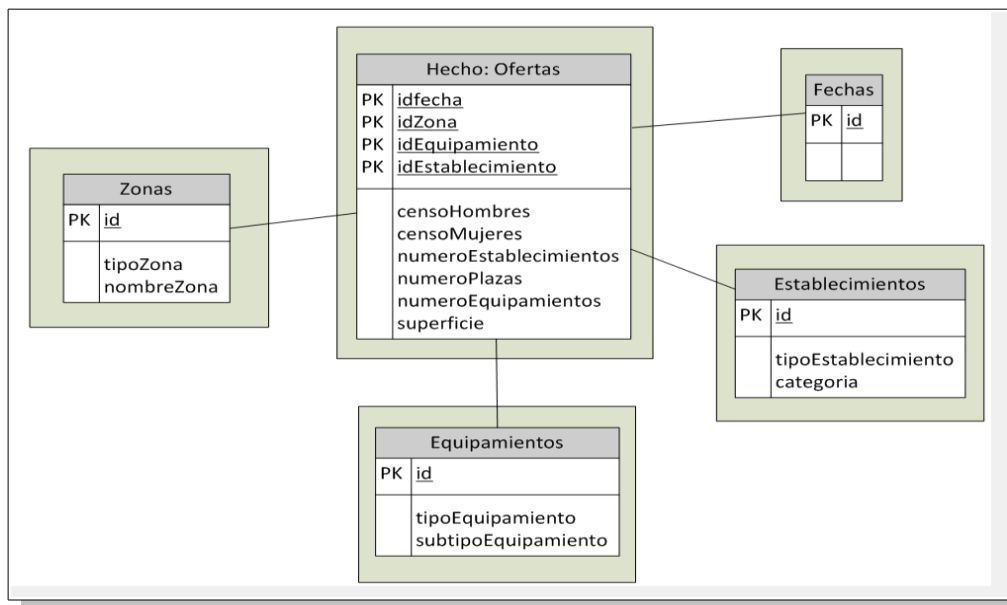


Imagen 10. Esquema conceptual. Fuente: Propia.

El esquema lógico se puede definir considerando el objetivo de acelerar al máximo la ejecución de las consultas todas las claves serán numéricas y adecuadas al tamaño más adecuado, de acuerdo al análisis de viabilidad que se ha realizado anteriormente. En todo caso se ha de generar consultas lo más eficientes posible.

El modelo lógico propuesto queda como sigue:

- *Ofertas* (idFecha, idZona, idEstablecimiento, idEquipamiento, censoHombres, censoMujeres, numeroEstablecimientos, numeroPlazas, numeroEquipamientos, superficieZona)
- *Establecimientos* (id, tipoEstablecimiento, categoria)
- *Zonas* (id, tipoZona, nombreZona, superficieZona)
- *Equipamientos* (id, tipoEquipamiento, subtipoEquipamiento)

Gráficamente el esquema lógico quedaría de la siguiente manera:

3 Asumimos que no se producen cambios en las características de los campos que no forman parte de la clave primaria. Ver PEC2.

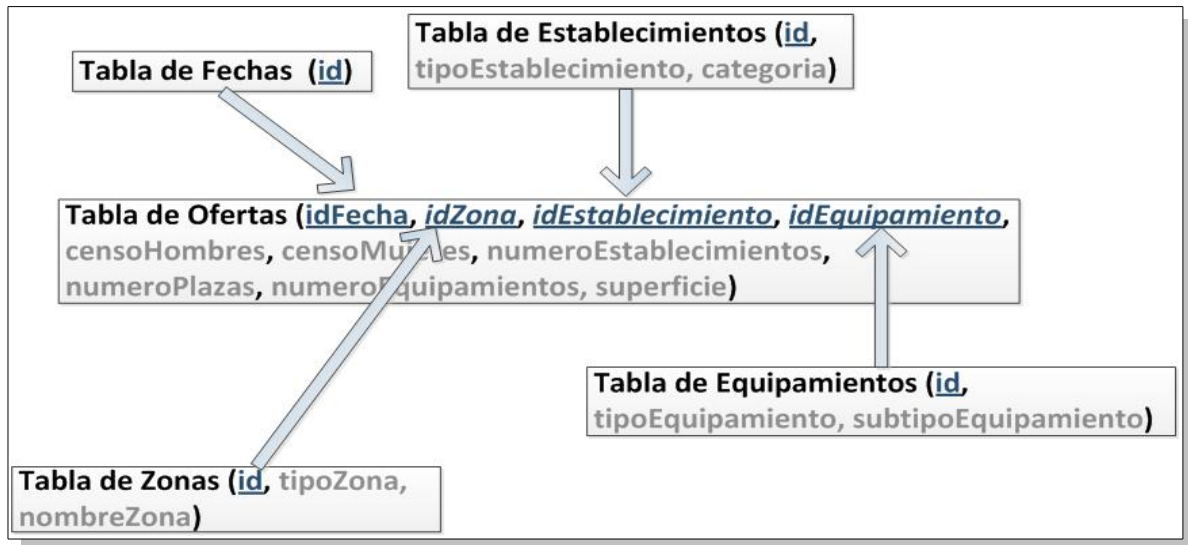


Imagen 11. Esquema lógico. Fuente: Propia.

## 5.- Diseño

A partir de la información expuesta hasta este momento el esquema físico consta de una tabla de hechos, que sería la tabla Oferta, vinculada a las tablas dimensionales de Fechas, Zonas, Equipamientos y Establecimientos. La estructura de cada tabla queda como sigue:

### 5.1.- Tabla de Zonas

Esta tabla hace referencia a la dimensión Zonas y se genera a partir de la información contenida en las fuentes relativas a establecimientos. El proceso de importación genera una clave autonumérica, que podemos afirmar será de un volumen de datos relativamente pequeño.

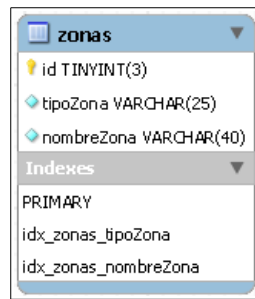


Imagen 12. Tabla Zonas. De tipo dimensional. Fuente: Propia.

La descripción de los atributos se detalla a continuación:

Nombre	Tipo	Descripción
Id (PRIMARY KEY)	TINYINT	Identificador único de los registros de la tabla zona. Constituye la clave primaria.
TipoZona	VARCHAR(25)	Descripción de la zona. Se refiere a Comarcas, Provincias, Veguerías y Comunidad Autónoma.
NombreZona	VARCHAR(40)	Nombre de la zona que constituye la dimensión.

Tabla 12. Estructura tabla Zonas. Fuente: Propia.

Para acelerar el rendimiento general y la velocidad de la ejecución de las consultas se ha establecido los siguientes índices de la tabla:

Nombre	Descripción
idx_zonas_tipoZona	Indexa la tabla por los tipos de zona disponibles, para facilitar el acceso en el



Nombre	Descripción
	momento de generar la tabla de hechos.
idx_zonas_nombreZona	Indexa la tabla por el nombre de la zona, para facilitar el acceso en el momento de generar la tabla de hechos.

Tabla 13. Estructura de índices de la tabla Zonas. Fuente: Propia.

### 5.2.- Tabla de Equipamientos

Se refiere a la dimensión equipamiento. Se genera a partir de la información contenida en la fuente equipamientos, es de carácter estático, se refiere a diciembre de 2012.

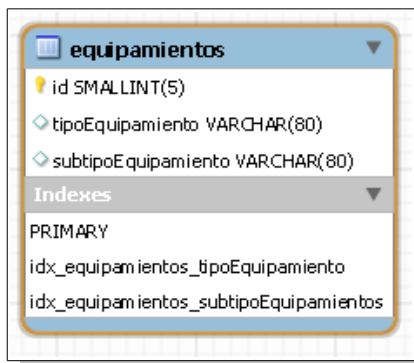


Imagen 13. Tabla Equipamientos. De tipo dimensional. Fuente: Propia.

Los campos que aparecen en la tabla se explican a continuación:

Nombre	Tipo	Descripción
Id (PRIMARY KEY)	TINYINT	Identificador único de los registros de la tabla equipamiento. Constituye la clave primaria.
TipoEquipamiento	VARCHAR(80)	Tipo de Equipamiento. Se refiere a un primer agrupamiento de los tipos de equipamientos de la fuente de datos.
SubTipoEquipamiento	VARCHAR(80)	SubTipo de Equipamiento. Se refiere a un segundo nivel de entre los tipos de equipamientos de la fuente de datos.

Tabla 14. Estructura de la tabla Equipamientos. Fuente: Propia.

Respecto a los índices creados se explican seguidamente:

Nombre	Descripción
idx_equipamientos_tipoEquipamiento	Indexa la tabla por los tipos de equipamientos disponibles, para facilitar el acceso en el momento de generar la tabla de hechos.
idx_equipamientos_subtipoEquipamiento	Indexa la tabla por subtipos de equipamientos disponibles, para facilitar el acceso en el momento de generar la tabla de hechos.

Tabla 15. Estructura de índices de la tabla Equipamientos. Fuente: Propia.

### 5.3.- Tabla de Establecimientos

Contiene información sobre los diferentes tipos de establecimientos y las categorías disponibles para cada uno de ellos. El modelo de la tabla se comprueba en la siguiente imagen:

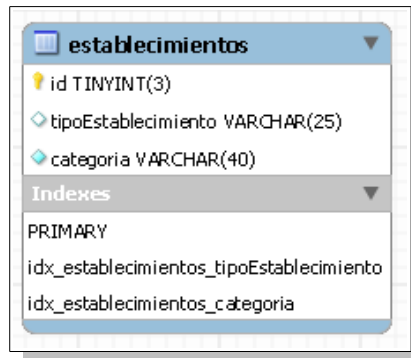


Imagen 14. Tabla Establecimientos. De tipo dimensional. Fuente: Propia.

Los campos empleados en la definición de la tabla son los siguientes:

Nombre	Tipo	Descripción
Id (PRIMARY KEY)	TINYINT	Identificador único de los registros de la tabla equipamiento. Constituye la clave primaria.
TipoEstablecimiento	VARCHAR(25)	Tipo de Establecimiento. Se refiere al tipo de establecimiento obtenido a partir de las fuentes.
categoria	VARCHAR(40)	Hace referencia a la categorías disponibles para cada uno de los tipos de establecimientos.

Tabla 16. Estructura de la tabla Establecimientos. Fuente: Propia.

Respecto a los índices creados se explican seguidamente:

Nombre	Descripción
idx_establecimientos_tipoEstablecimiento	Indexa la tabla por los tipos de equipamientos disponibles, para facilitar el acceso en el momento de generar la tabla de hechos
idx_establecimientos_categoria	Indexa la tabla por subtipos de equipamientos disponibles, para facilitar el acceso en el momento de generar la tabla de hechos

Tabla 17. Estructura de índices de la tabla Establecimientos. Fuente: Propia.

#### 5.4.- Tabla de Ofertas

Constituye la tabla de hechos dentro del modelo de BI que queremos desarrollar para nuestro cliente. Permite para una fecha, una zona, un tipo de establecimiento y un tipo de equipamiento obtener información sobre datos de censo, establecimientos y equipamientos.

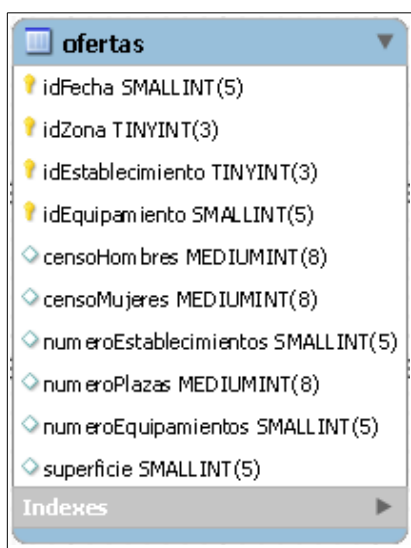


Imagen 15. Tabla Ofertas. De tipo Hecho. Fuente: Propia.

Respecto a los atributos considerados se detallan como sigue:

Nombre	Tipo	Descripción
Fecha (PRIMARY KEY)	SMALLINT	Identificador de fecha de la tabla de Hechos. En la práctica se refiere a un año.
IdZona (PRIMARY KEY)	TINYINT	Identificador de la zona de la tabla de Hechos. Es clave foránea de la tabla dimensional Zonas.
IdEstablecimiento (PRIMARY KEY)	TINYINT	Identificador de la establecimiento de la tabla de Hechos. Es clave foránea de la tabla dimensional Establecimientos.
IdEquipamiento (PRIMARY KEY)	SMALLINT	Identificador de la zona de la tabla de Hechos. Es clave foránea de la tabla dimensional Equipamientos.
censoHombres	MEDIUMINT	Hace referencia a la población masculina para una fecha en una zona determinada.
censoMujeres	MEDIUMINT	Hace referencia a la población total para una fecha en una zona determinada.
numeroEstablecimientos	SMALLINT	Hace referencia al número de establecimientos en una fecha, zona y tipo determinado.
numeroPlazas	MEDIUMINT	Hace referencia al número de plazas en una fecha, zona y tipo determinado.
numeroEquipamientos	SMALLINT	Hace referencia al número de equipamientos en una fecha, zona y tipo determinado.
superficie	SMALLINT	Hace referencia a la superficie de la zona referenciada.

Tabla 18. Estructura de la tabla de Hechos: Ofertas. Fuente: Propia.

### 5.5.- Esquema Físico

A partir de la estructura de tablas, las de dimensiones y la de hechos, podemos representar gráficamente el esquema físico:

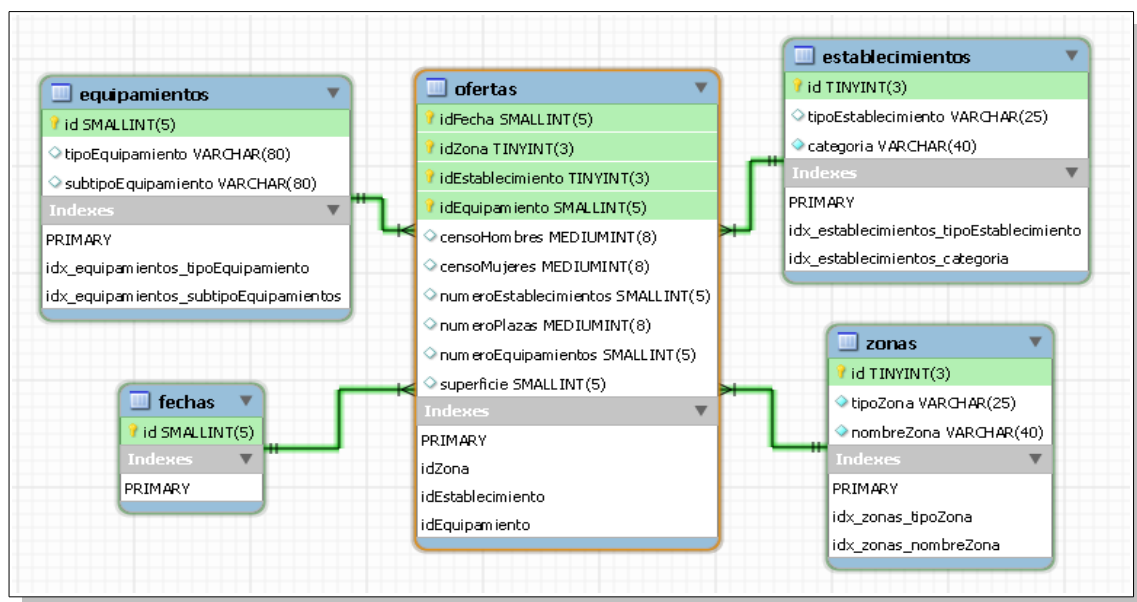


Imagen 16. Esquema físico. Fuente: Propia.

Dentro del esquema físico destacar que la elección de tipos de datos ha quedado justificado por criterios de optimización de las consultas. Así

- Para el caso de las claves primarias se ha utilizado el tipo TINYINT, para aquellas tablas con un número relativamente pequeño de datos, como Zonas o Establecimientos; o el tipo SMALLINT cuando los valores a alcanzar son altos, por ejemplo en el caso de Equipamientos u Ofertas, donde el número de registros es relativamente alto, o en el caso de la tablas Fechas, donde el índice se corresponde con el ejercicio.
- Para campos de tipo texto hemos utilizado VARCHAR, en la medida que optimiza el espacio ocupado por las cadenas de texto de longitud inferior a la definida para el campo por defecto.
- En el caso de campo de tipo numérico la elección de un tipo u otro ha dependido de los valores observados en las fuentes. Des este modo se ha elegido el tipo MEDIUMINT para datos de población o SMALLINT para datos de superficie o numero de equipamientos, por ejemplo.

### 5.6.- Diseño de Arquitectura

A la hora de definir la arquitectura necesaria para poner en marcha el modelo de Business Intelligence es necesario diferenciar los roles de usuario, por un lado, y además la arquitectura de software y la de hardware, sujeto a las necesidades del primero. Por otro lado se diferenciará el entorno empleado para la realización de la PEC, simplificado, y un entorno real.

Desde un punto de vista global podemos considerar dos subsistemas: uno relativo a los procesos ETL, en los que la función del Administrador es predominante, y otro relativo al diseño y generación de informes, en los que también intervienen los Usuarios. Gráficamente se muestra en las siguientes imágenes:

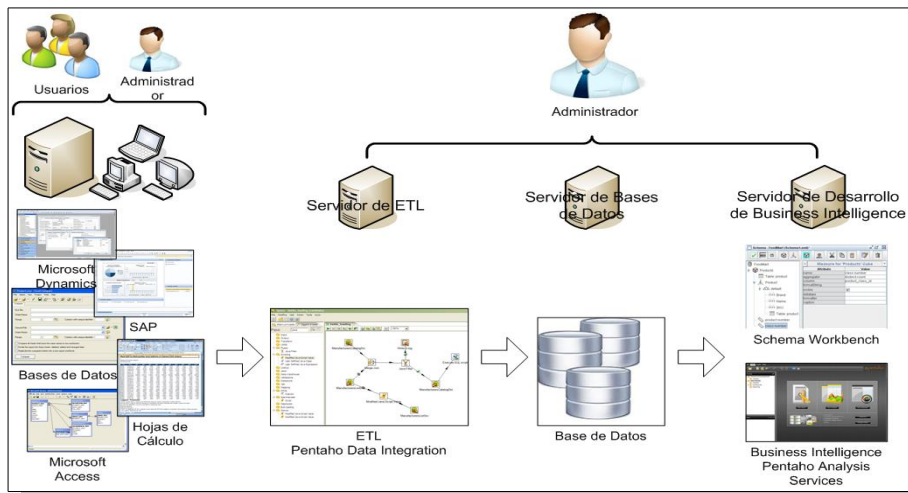


Imagen 17. Arquitectura sistema de desarrollo. Fuente: Propia.

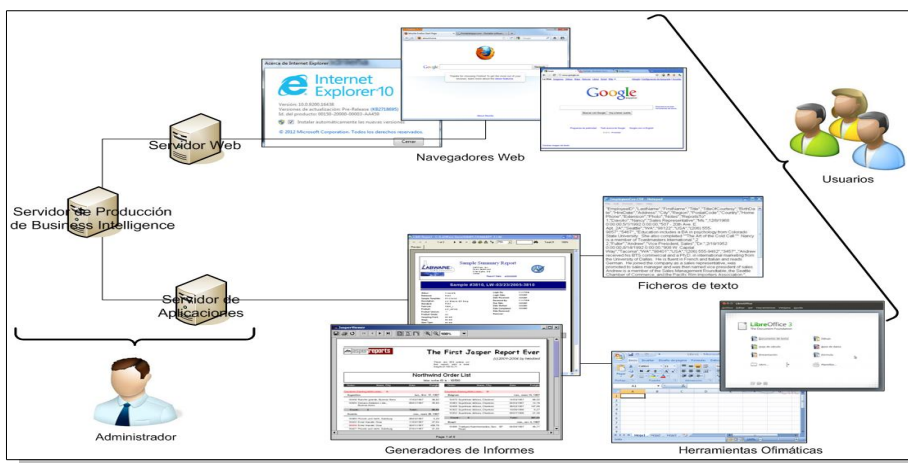


Imagen 18. Arquitectura sistema en producción. Fuente: Propia.

#### 5.6.1.- Arquitectura del Software

Desde el punto de la arquitectura del Software las herramientas empleadas pasan por:

- *Software* empleado para generar las diferentes fuentes de información. Se trata de un un conjunto de



aplicaciones heterogéneo, desubicado, ... que pueden generar información por parte de usuarios, administradores, procesos automatizados,... ya sean de la propia organización o ajenos a ella. El formato de la información es diverso: ficheros Excel, ficheros de texto plano (csv, txt,...), XML, propietarios de bases de datos,...

- *Pentaho Data Integration*. Proporciona una serie de utilidades para transformar las fuentes de información en formatos que puedan ser utilizados directamente por nuestro modelo. Incluye tareas de formateo de la información, conversión de tipos, normalización / desnormalización de datos, scripts para la creación / modificación / eliminación de la información, captura de datos, automatización de tareas,...
- *Base de datos*. El SGBD elegido ha sido MySQL, dentro de la filosofía general del proyecto para utilizar herramientas basadas en software libre. La versión utilizada es la 5.5 release 28.
- *Schema Workbench*. Permite generar código en formato MDX de una manera intuitiva. Permite crear fórmulas, miembros calculados en su terminología, y publicar en el repositorio correspondiente los esquemas creados. Un esquema es un conjunto de dimensiones, métricas y tabla de hechos que se combinan para la creación de cubos que puedan ser desarrollados con una herramienta de BI.
- *Pentaho Analysis Services*. Proporciona, también bajo la característica de software libre, un conjunto de herramientas para diseñar reportes, cubos,... que puedan ser utilizados por los diferentes usuarios.
- *Servidor Web*. El servidor utilizado, bajo la misma filosofía es Apache. Permite que los usuarios visualicen gráficos o informes y que exporten la información en diferentes formatos.
- *Software* empleado para utilizar la información generada. Al igual que ocurría con las fuentes de información se trata de un conjunto de herramientas que son proporcionadas por el administrador para que el usuario las utilice según la política de la empresa o las preferencias del propio usuario.

### 5.6.3.- Arquitectura del Hardware

Desde el punto de vista de arquitectura del hardware hemos de distinguir entre el entorno utilizado para el desarrollo del proyecto y un entorno idealizado.

- *Entorno del proyecto*. Se utiliza un máquina virtual que contiene todos los servicios necesarios para la realización de la práctica. Requiere unos 6 Gb de espacio disponible en la máquina física y hay un espacio disponible entorno a 1 Gb en la máquina virtual. Este espacio disponible es necesario para las tablas temporales que ha de emplear el motor de bases de datos para la ejecución de las consultas.
- *Entorno idealizado*. El esquema propuesto es independiente de la utilización de servidores físicos o virtualizados. Desde el punto de vista de optimización de recursos y minimización de costes se podrían agrupar servicios y funcionalidades en un único servidor. En este sentido distinguiríamos entre:
  - Servidores, equipos de escritorio, portátiles, consolas,...
  - Servidor para la ejecución de los procesos ETL.
  - Servidor de Bases de Datos.
  - Servidor de Desarrollo, para realizar la programación de informes, gráficos,...
  - Servidor de Producción. Contiene los informes verificados de acuerdo a los criterios de calidad de software establecidos por la organización.
  - Servidor Web, para la publicación y acceso a la información desde un navegador.
  - Servidor de Aplicaciones, para integración de la información generada en otras herramientas: software de gestión, repositorio de archivos,...

## **6.- Implementación**

La fase de implementación se puede descomponer en varias subfases, diferenciadas y secuenciales, que tiene por objeto la preparación de los datos para la creación de los informes solicitados por el cliente, ONDO.

### **6.1.- Entorno de trabajo**

El entorno de trabajo exige un conocimiento relativamente avanzado de una serie de aplicaciones que lo componen. Todas ellas están integradas en una máquina virtual, tal y como hemos explicado anteriormente: MySQL, Pentaho Data Integration, Schema Workbench, Pentaho Analysis Services,... Por otro lado es necesario conocer herramientas estándares bajo el entorno Windows: Ofimática, Sistema Operativo, Notepad++, Microsoft Visio,

Microsoft Project, Microsoft PowerPoint,...

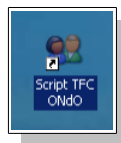
La estructura de carpetas responde a la siguiente estructura:

- C:\Trabajos\Eschema. Contiene la información relativa al esquema que contiene los cubos empleados para la generación de los cubos y gráficos
- C:\Trabajos\Fuentes De Datos. Aquí se incluyen los ficheros de datos que van a ser transformados y a los que se les han aplicado algunas transformaciones de manera manual.
- C:\Reportes. Esta carpeta contiene los informes, gráficos,... generados

## 6.2.- Construcción de la ETL

Las acciones necesarias para lanzar los procesos de ETL se pueden enumerar como sigue:

- 1.- *Importación de la máquina virtual.* Es la opción más segura y recomendada para la instalación de la máquina virtual.
- 2.- *Ejecución del Script de creación de la BBDD y de las operaciones de ETL.* Podemos acceder desde escritorio, haciendo doble click en el icono "Script TFC ONdo".



Otra opción consiste en ejecutar directamente el archivo, en la máquina virtual, a través de la ruta "C:\Trabajos\Script" ejecutar el archivo "Script.bat", de acuerdo a las imágenes siguientes:

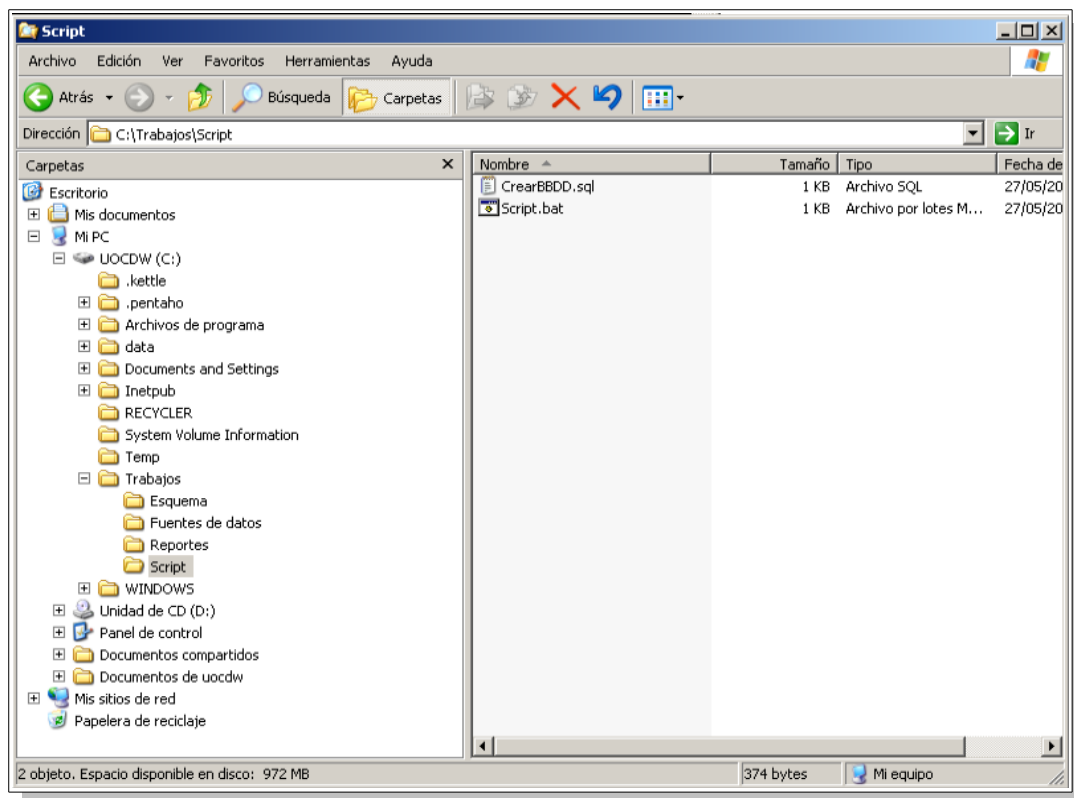


Imagen 19. Estructura de la carpeta C:\Trabajos\Scripts. Fuente: Propia.

El contenido del fichero Script.bat, en formato de texto plano, ejecuta el archivo CrearBBDD.sql, que elimina la base de datos del repositorio de bases de datos disponible y lo vuelve a crear, de acuerdo a las sentencia:

```
mysql -utfcmmd -ptfcmmd < CrearBBDD.sql
```

con el contenido del fichero "CrearBBDD.sql" según muestra la imagen:

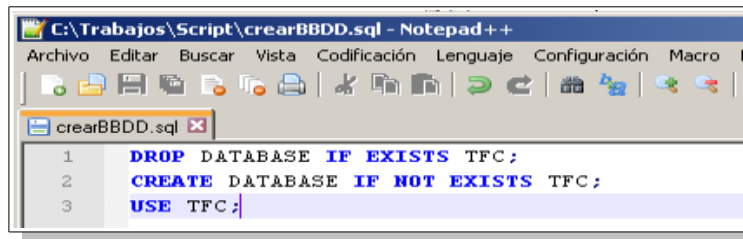


Imagen 20. Estructura del script de creación de la BBDD. Fuente: Propia.

Las instrucciones que afectan a la ejecución de los procesos ETL, y siempre dentro del fichero Script".sql", se detallan a continuación:

```
"C:\Archivos de programa\pdi-ce-4.4.0-stable\data-integration\kitchen" /rep:ETL /user:admin /pass:admin /job:"Trabajo TFC" /level:Minimal
```

```
"C:\Archivos de programa\pdi-ce-4.4.0-stable\data-integration\pan" /rep:ETL /user:admin /pass:admin /trans:"Generacion Tabla de Hechos" /level:Minimal
```

### 3.- Carga de Schema WorkBench.



4.- Carga del archivo EsquemaFTM.xml con la definición del esquema propuesto para el TFC, desde Fichero\Open y de acuerdo a la ruta que se muestra en la imagen, C:\Trabajos\Eschema se abre el fichero SchemaFTM.xml.

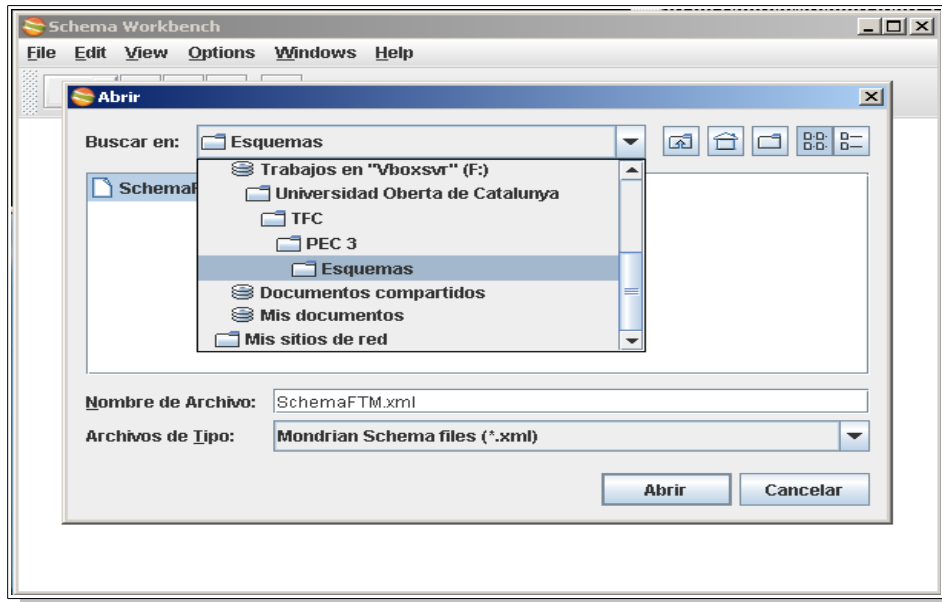


Imagen 21. Esquema del modelo para el repositorio (I). Fuente: Propia.

5.- Publicación del esquema en el repositorio de esquemas

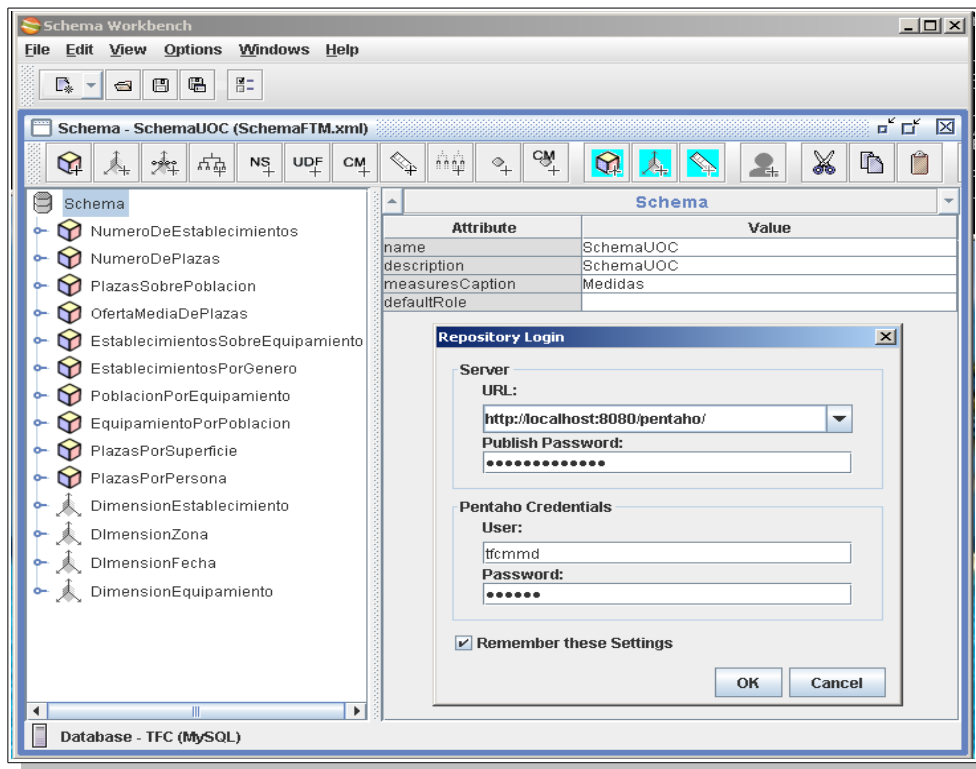


Imagen 22. Esquema del modelo para el repositorio (II). Fuente: Propia.

6.- Carga de la consola de Pentaho Analysis Services. Ya están disponibles tanto los cubos definidos como el repositorio por si se quisieran crear nuevos informes.



**Usuario:** tfcmmmd  
**Contraseña:** tfcmmmd

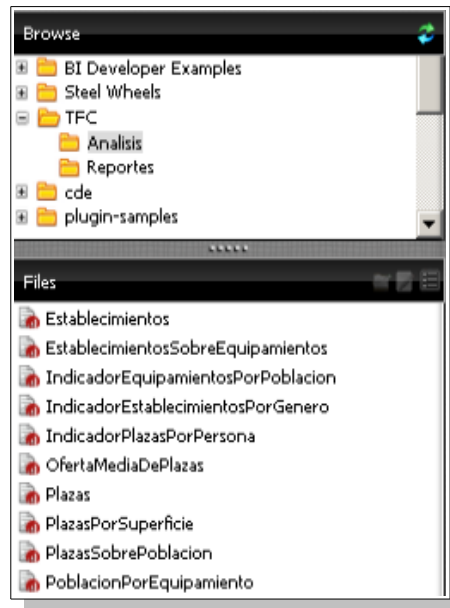
Imagen 23. Acceso a Pentaho Analysis Services. Fuente: Propia.



Imagen 24. Acceso a Pentaho Analysis Service. Fuente: Propia.

7.- Ejecución de los diferentes cubos creados. Se encuentran en la carpeta TFC\Analysis





Comentaremos en las siguientes secciones las transformaciones y trabajos ejecutadas anteriormente.

#### 6.2.1.- Carga inicial

Sobre los ficheros proporcionados por el cliente ONdO se aplican los ajustes manuales propuestos en el punto 4.3. De manera que los archivos de origen son los siguientes:

- Equipaments00.csv
- Establiments 2006.txt, Establiments 2007.txt, Establiments 2008.txt, Establiments 2009.txt, Establiments 2010.txt, Establiments 2011.txt, Establiments 2012 ver 01.txt
- poblacio01.csv

El trabajo que hemos denominado “Trabajo TFC” se puede descomponer en varias subfases que la primera de las cuales tratan de importar los datos en una serie de tablas intermedias sobre las que aplicaremos posteriormente una serie de transformaciones para generar la tabla de hechos y las tablas de dimensiones.

De esta primera subfase podemos destacar el script inicial para la creación de tablas, que no sólo afecta a las tablas del modelo, propiamente dicho, sino que además se crean las tablas para la importación inicial de datos, que son:

- IN\_Equipamientos
- IN\_Poblacion
- IN\_Establecimientos
- IN\_Fechas

Para cada una de estas tablas se ha definido una transformación y, especialmente, en el caso de datos de establecimientos, considerando la estructura dispar de los ficheros fuente se han realizado 3 procesos diferentes de importación, agrupando por estructuras comunes: para el año 2006, para los años 2007 a 2011 y para el año 2011.

Las transformaciones que se llevan a cabo en esta fase son las siguientes:

- *Crear BBDD y Tablas.* Contiene el script para crear la BBDD y las tablas necesarias para la implementación del modelo. Se definen tanto las tablas necesarias para importar los datos directamente desde las fuentes como las tablas dimensionales y la tabla de hecho. Igualmente se definen los índices que optimizarán los procesos de búsqueda en las tablas por atributos diferentes a los de la clave primaria.
- *Importar Población.* En esta importación se realiza una verificación de la consistencia de la información para el año 2012, de manera que la población total ha de ser la suma de la población masculina más la femenina.

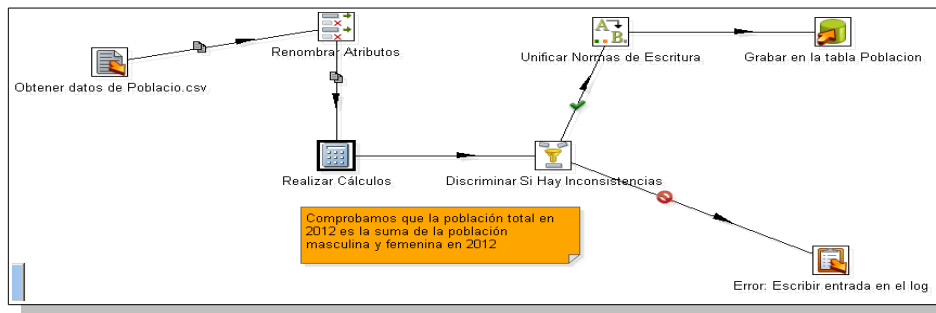


Imagen 25. Transformación Importar Poblacion. Fuente: Propia.

- **Importar Fechas.** En este caso realizamos una asignación directa de valores, ya que se trata de una serie con un número reducido de valores.

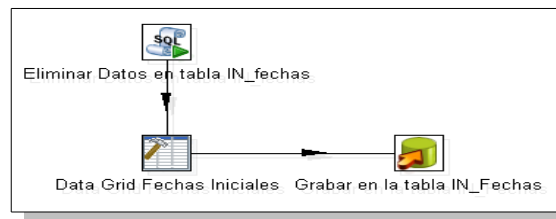


Imagen 26. Transformación Importar Fechas. Fuente: Propia.

- **Importar Equipamientos.** En este caso la fuente de datos asigna a cada registro, a cada equipamiento ofertado, el tipo y el conjunto de subtipos a los que se puede asignar, separados por un carácter de división. Es necesario ejecutar un script que genere tantos nuevos registros como pares tipo/subtipo de equipamiento dispongamos. Finalmente se lleva a cabo la eliminación de caracteres en blanco. Por otro lado se omite la importación de aquellos datos que no tengan un subtipo asignado, es decir, que sea nulo o con el carácter espacio en blanco.

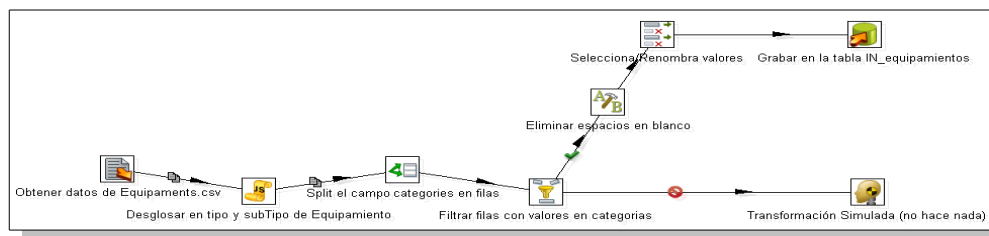


Imagen 27. Transformación Importar Equipamientos. Fuente: Propia.

- **Importar Establecimientos 2006.** En esta transformación y las dos posteriores hemos de incorporar un nuevo atributo que haga referencia a la fecha, en este caso 2006. Omitimos los datos de columnas que son redundantes, innecesarios o en blanco e ignoramos la importación de filas sin datos.

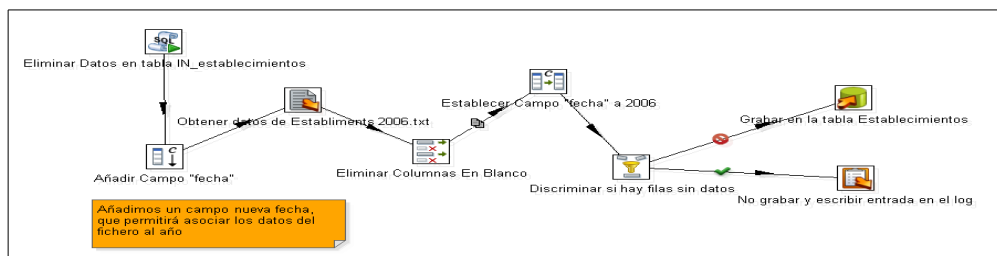
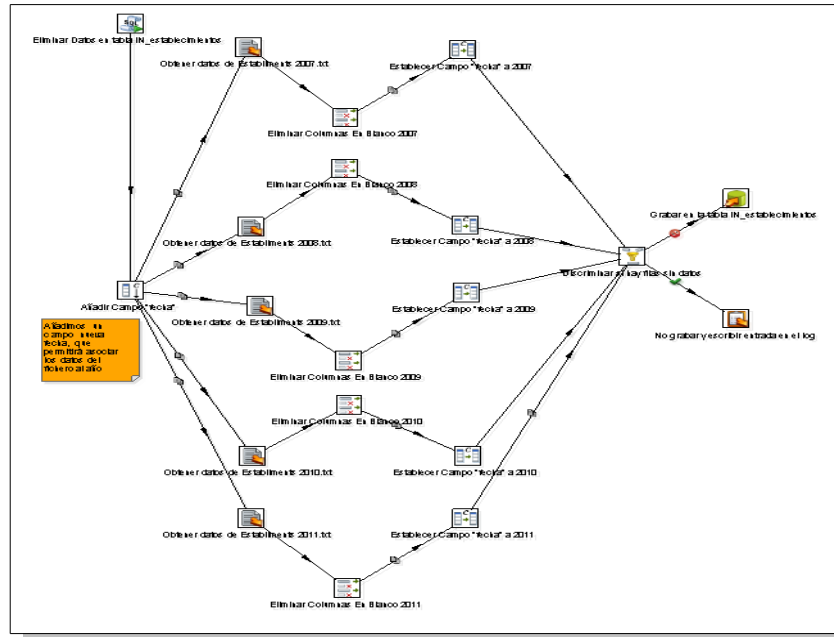


Imagen 28. Transformación Importar Establecimientos 2006. Fuente: Propia



- **Importar Establecimientos 2007-2008-2009-2010-2011.** El procedimiento de importación es idéntico al explicado anteriormente pero considerando la estructura de columnas a omitir, que es diferente. Imagen 29. Transformación Importar Establecimientos 2007 a 2011. Fuente: Propia.
- **Importar Establecimientos 2012.** Al igual que los anteriores en este caso la estructura es ligeramente diferente y por eso requiere una transformación diferente. La transformación queda como sigue:

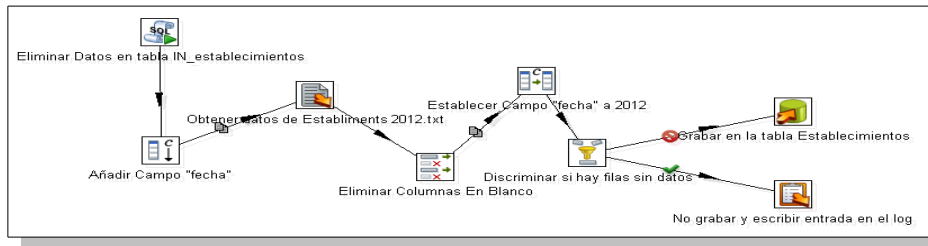


Imagen 30. Transformación Importar Establecimientos 22. Fuente: Propia.

La integración de todas las transformaciones anteriores aparece en las primeras transformaciones del trabajo "Trabajo TFC":

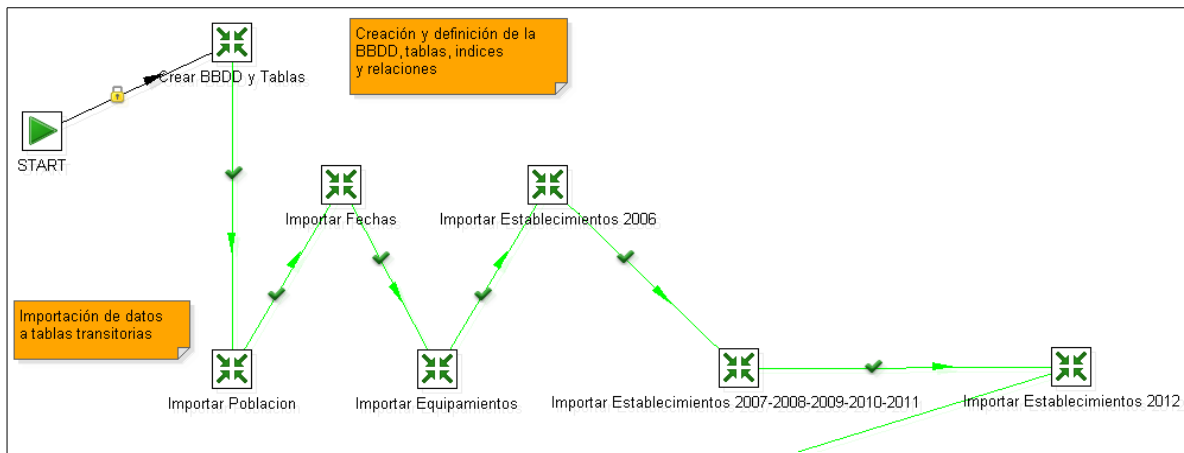


Imagen 31. Creación de tablas, índices y carga inicial de datos de población, equipamientos, fechas y establecimientos. Fuente: Propia.

En la siguiente subfase se lleva a cabo una actualización de datos sobre las tablas completadas en la subfase anterior, y que se refieren fundamentalmente a datos de comarcas, provincias y asignación de población por sexo.

Las transformaciones que se han llevado a cabo se resumen a continuación:

- **Actualizar Comarcas.** En esta transformación se actualizan las comarcas de las poblaciones en las que se localizan los diferentes equipamientos. Así mismo se observa que algunas comarcas han cambiado el nombre y es necesario hacer una asignación manual – para el caso de Cabrera d'Anoia o Vimbodí i Poblet entre otros -. Finalmente se registra en un log los municipios no encontrados. Gráficamente tenemos:

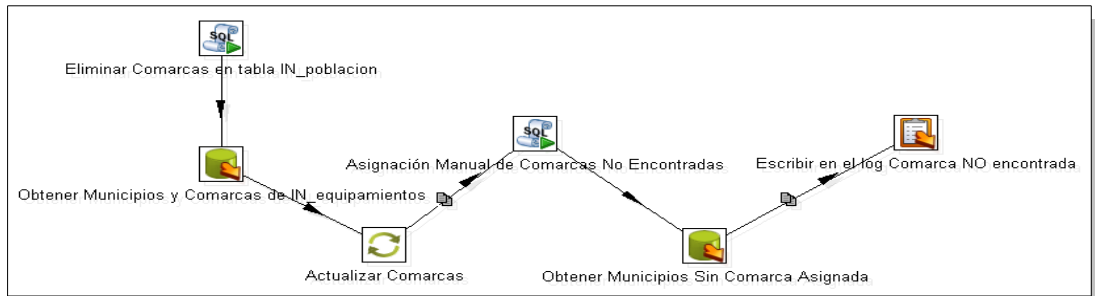


Imagen 32. Actualización de Comarcas. Fuente: Propia.

- **Actualizar Provincias en IN\_Poblacion.** En esta transformación se actualiza el campo provincia, en IN\_Poblacion, a partir del código INE proporcionado, asignando nombres a los valores que comiencen por 8 para Barcelona, 17 para Girona, 25 para Lleida y 43 para Tarragona. Vemos la gráfica a continuación:

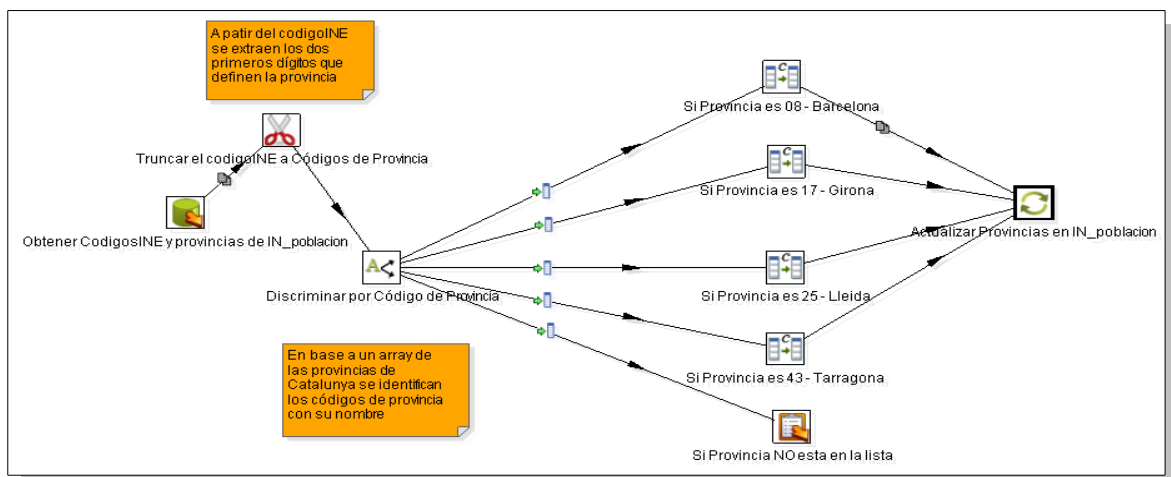


Imagen 33. Actualización de Provincias en IN\_Poblacion. Fuente: Propia.

- **Actualizar Provincias en IN\_Equipamiento.** Esta transformación se trata de un script que permite asignar a mano todas aquellas municipios a los que no se les puede asignar una provincia a partir de los municipios contenidos en la tabla IN\_Poblacion ni tampoco a partir de los primeros dígitos del código postal, como habíamos hecho en el caso anterior.
- **Actualizar Poblacion Por Sexo.** En si mismo es un trabajo que contiene las siguientes transformaciones:
  - Actualizar Población Por Sexos 2006
  - Actualizar Población Por Sexos 2007
  - Actualizar Población Por Sexos 2008
  - Actualizar Población Por Sexos 2009
  - Actualizar Población Por Sexos 2010
  - Actualizar Población Por Sexos 2011
  - Actualizar Población Por Sexos 2012

En todas ellas se utilizan el mismo procedimiento de cálculo: considerando los datos de distribución de población por sexo que tenemos para el año 2012 aplicamos esa distribución para cada uno de los años en los que es desconocida. Igualmente se ha incluido un deflactor, pero por simplificar el modelo se ha

establecido su valor a 1.

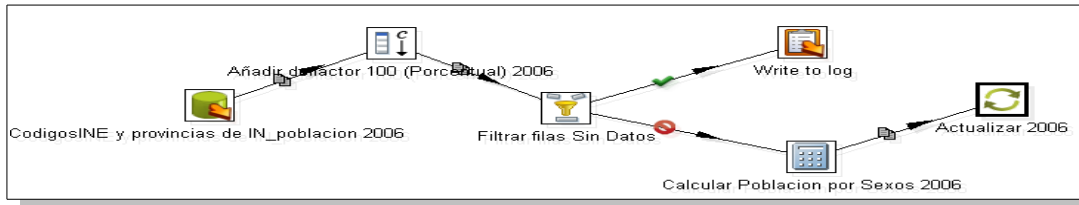


Imagen 34. Actualización de Población por Sexos. Fuente: Propia.

Considerando todo el proceso de las actualizaciones en su conjunto obtendremos la siguiente subfase:



Imagen 35. Actualización de datos. Fuente: Propia.

En la última subfase y a partir de las importaciones y transformaciones realizadas en las subfases anteriores procedemos a obtener los datos necesarios para completar la información de las tablas dimensionales, que hemos explicado anteriormente. Las transformaciones están encaminadas a completar la información de las tablas dimensionales: fechas, equipamientos, zonas y establecimientos. Las detallamos a continuación:

- **Rellenar Fecha.** Es la transformación más sencilla de esta subfase y supone un mero trasvase de la tabla IN\_Fechas hacia la tabla Fechas.
- **Rellenar Equipamientos.** Tomando como referencia la información importada sobre equipamientos se realiza una ordenación y una selección por registros únicos de pares equipamientos por tipos y subtipos. Finalmente se renombran los campos para una correcta asignación de datos al dimensional.

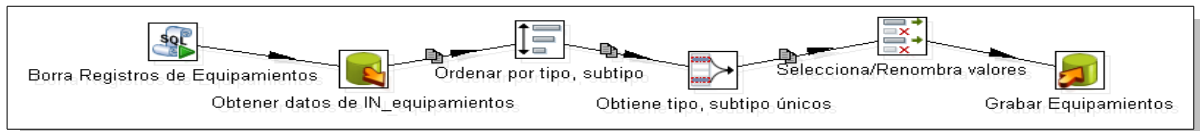


Imagen 36. Rellenar Equipamientos. Fuente: Propia.

- **Rellenar Zonas.** En este caso la asignación de zonas por tipos y subtipos se realiza aplicando una discriminación por el tipo de zona: Comarca, Veguería, Provincia y Comunidad Autónoma, y la asignación a cada una de las zonas, tomadas de los datos de establecimientos, del tipo y subtipo correspondiente. De este modo gráficamente tenemos:

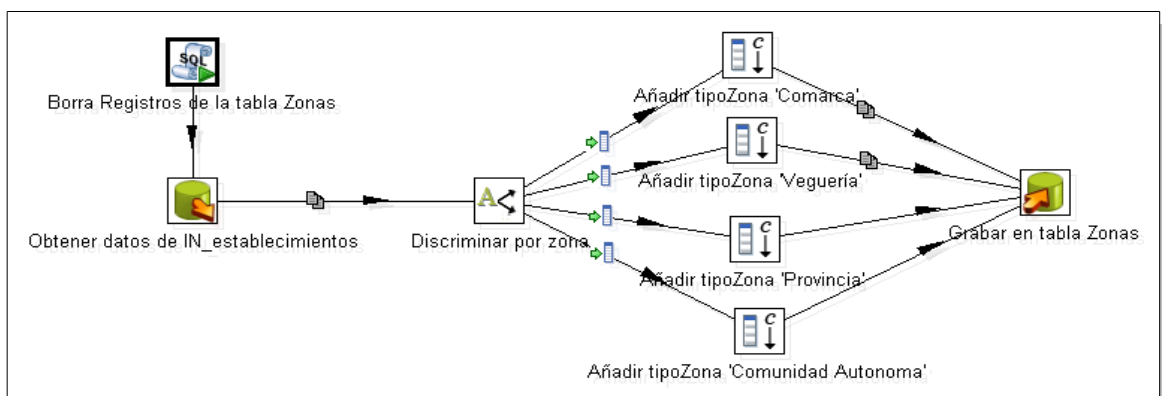


Imagen 37. Rellenar Zonas. Fuente: Propia.

- **Rellenar Establecimientos.** En este caso no cabe más que la definición manual de todos los tipos y categorías del modelo que queremos contemplar.

Gráficamente la secuencia de transformaciones de esta subfase queda como sigue:



Imagen 38. Asignación de información a las tablas dimensionales. Fuente: Propia.

### 6.2.2.- Carga de la tabla de Hechos

Para la carga de la tabla de Hechos hemos tenido que partir de una correcta carga de las tablas anteriormente explicadas y de las transformaciones propuestas, así como de la carga de las tablas dimensionales correspondientes.

El procedimiento seguido para cargar la tabla de hechos ha consistido en definir una serie de flujos de información que se van consolidando conforme se ejecutan en unas tablas temporales intermedias para finalmente cargar la tabla de hechos manteniendo la integridad referencial del modelo, acorde con las claves, primarias y ajenas, de las tablas correspondientes.

Como en el apartado anterior vamos a detallar el proceso dividiéndolo en subfases.

En la primera subfase se realiza una agrupación de zonas y equipamientos, conforme a los valores de la tabla dimensional de equipamientos y zonas a partir de los datos de la tabla IN\_Equipamientos. Se realiza un proceso de normalización y ordenación para conseguir esta agrupación. Gráficamente podemos observar la secuencia a continuación:



Imagen 39. Agrupación por Zonas y Equipamientos y . Fuente: Propia.

En la siguiente subfase a partir de los datos de la tabla IN\_Poblacion se hace un triple proceso de normalización, por fecha, zona y por género, de manera que obtenemos la información agrupada por esas tres dimensiones, de acuerdo a la siguiente imagen:



Imagen 40. Agrupación por Fechas y Zonas . Fuente: Propia.

En la subfase tercera realizamos una normalización por zonas encaminada a obtener la superficie de la zona correspondiente a partir de la superficie de los municipios que la componen. En la imagen siguiente se observa la secuencia:



Imagen 41. Agrupación por Fechas y Zonas. Fuente: Propia.

Para estas tres primeras subfases se realiza una consolidación, considerando además que hemos tenido que agregar manualmente la zona Comunidad Autónoma de manera que obtenemos una estructura de registros normalizada por fechas, zonas y equipamientos. Esta estructura se combinará posteriormente para cargar la tabla de Hechos.

En la siguiente subfases se realiza una doble normalización, también, de cara a obtener una estructura ordenada por fechas, zonas y establecimientos de tipo camping, hotel y turismo rural respectivamente. En este caso cabe considerar la normalización de establecimientos, para obtener la relación adecuada por tipos y categorías de acuerdo a los valores de la tabla dimensional de Establecimientos. Gráficamente:

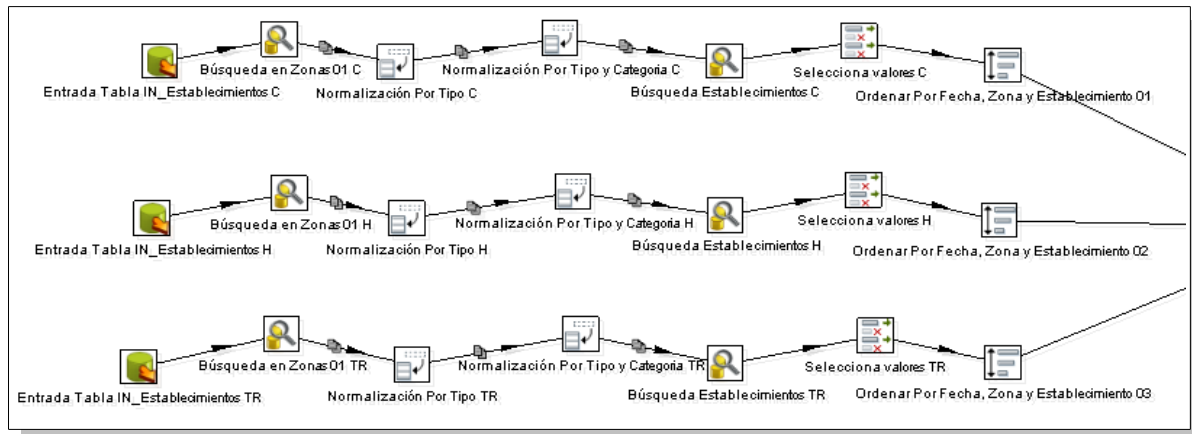


Imagen 42. Agrupación por Fechas y Establecimientos – Camping, Hotel y Turismo Rural-. Fuente: Propia.

Del conjunto de registros generadas con las tres primeras subfases: zonas, equipamientos, población y superficie y la información resultante de agrupar establecimientos por zonas y fechas, de las tres últimas subfases, aplicando las uniones y las ordenaciones correspondientes, obtenemos la estructura necesaria para completar la tabla de Hechos. Las transformaciones necesarias aparecen en la imagen siguiente:

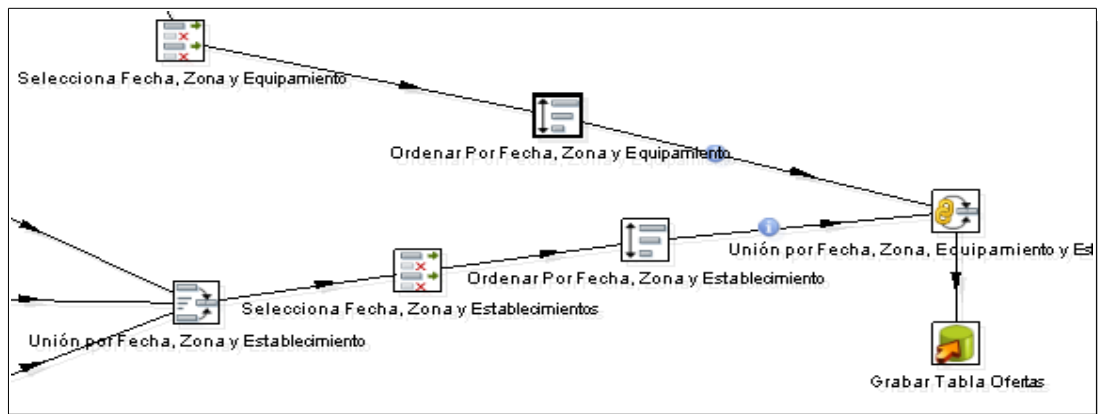


Imagen 43. Uniones y Ordenaciones para generar datos de la tabla de Hechos. Fuente: Propia.

### 6.2.3.- Construcción del Esquema

Una vez definidos tanto la tabla de Hechos como la tabla de Dimensiones el paso siguiente consiste en definir las estructuras que permitirán diseñar los informes. Para ello se utiliza la herramienta Workbench Schema, mediante la que una interfaz gráfica se define el código MDX necesario.

En el esquema es necesario definir cada una de las dimensiones, y dentro de ella hay que definir una jerarquía. Para cada jerarquía se definen los niveles correspondientes así como las tablas a las que se refieren. De esta manera las dimensiones definidas en el esquema son las siguientes:

Dimensión	Jerarquía	Miembros	Tabla
Zona	JerarquiaZona	tipoZona nombreZona	Zona
Fecha	JerarquíaFecha	fecha	Fecha
Establecimiento	JerarquíaEstablecimiento	tipoEstablecimiento categoria	Establecimiento
Equipamiento	JerarquíaEquipamiento	tipoEquipamiento subTipoEquipamiento	Equipamiento

Tabla 19. Estructura de la tabla de Hechos: Ofertas. Fuente: Propia.

A continuación procederemos a explicar todos los cubos que se han definido dentro del esquema y que serán empleados en la creación de los informes que ha solicitado nuestro cliente, ONdO. Como elemento común a

todos ellos tenemos la tabla de Hechos, por lo que la omitiremos el análisis siguiente:

- *NumeroDeEstablecimientos*. Incluye las dimensiones Establecimiento, Zona y Fecha. Tiene una métrica, numeroDeEstablecimientos, sobre la columna numeroEstablecimientos de la tabla de Hechos, de tipo numérico y con Average como tipo de agregador. No se asigna un formato especial.
- *NumeroDePlazas*. Incluye las dimensiones Establecimiento, Zona y Fecha. Tiene una métrica, numeroDePlazas, sobre la columna nombrePlazas de la tabla de Hechos, de tipo numérico y con Average como tipo de agregador. No se asigna un formato especial.
- *PlazasSobrePoblacion*. Incluye las dimensiones Establecimiento, Zona y Fecha. Tiene tres métricas, poblacionHombres, poblacionMujeres y numeroDePlazas, que se corresponde con las columnas censoHombres, censoMujeres y numeroPlazas de la tabla de Hechos y con Average como tipo de agregador. Del mismo modo existe dos miembros calculados: poblacionTotal, como suma de poblacionHombres y poblacionMujeres, de tipo Entero, y PlazasSobrePoblacion, definida como el cociente numeroDePlazas sobre poblacionTotal, expresada en formato porcentual.
- *OfertaMediaDePlazas*. Incluye las dimensiones Establecimiento, Zona y Fecha. Tiene una métrica, numeroDePlazas, sobre la columna nombrePlazas de la tabla de Hechos, de tipo numérico y con Average como tipo de agregador. No se asigna un formato especial. Considerando el agregador podemos afirmar que este cubo es idéntico al definido anteriormente en el cubo NumeroDePlazas.
- *EstablecimientosSobreEquipamientos*. Incluye las dimensiones Equipamiento, Establecimiento, Zona y Fecha. Dispone de dos métricas: numeroDeEquipamientos y numeroDeEstablecimientos, que se corresponden con las columnas numeroEquipamientos y numeroEstablecimientos respectivamente de la tabla de Hechos y con Average como tipo de agregador. No se les asigna un formato especial. Así mismo dispone de un miembro calculado: EstablecimientosSobreEquipamientos que responde al cociente numeroDeEstablecimientos entre numeroDeEquipamientos y se expresa en formato porcentual.
- *PoblacionPorEquipamiento*. Incluye las dimensiones Fecha, Equipamiento, y Zona. Tiene tres métricas, poblacionHombres, poblacionMujeres y numeroDeEquipamientos, que se corresponde con las columnas censoHombres, censoMujeres y numeroEquipamientos de la tabla de Hechos con Average como tipo de agregador. Dispone de un miembro calculado PoblacionPorEquipamientos que se define como el ratio entre la suma de poblacionHombres más poblacionMujeres sobre numeroDeEquipamientos. No se aplica ningún formato especial.
- *IndicadorEstablecimientosPorGenero*. Incluye las dimensiones Equipamiento, Establecimiento, Zona y Fecha. Se establecen tres métricas, poblacionHombres, poblacionMujeres y numeroDeEstablecimientos, que se corresponde con las columnas censoHombres, censoMujeres y numeroEstablecimientos de la tabla de Hechos con Average como tipo de agregador. Se establecen dos miembros calculados: IndicadorEstablecimientosSobreHombres y IndicadorEstablecimientosSobreMujeres, como cocientes entre poblacionHombres y poblacionMujeres sobre numeroDeEstablecimientos. No se utiliza ningún formato predeterminado.
- *IndicadorPlazasPorPersona*. Incluye las dimensiones Fecha, Establecimiento y Zona. Se establecen tres métricas, poblacionHombres, poblacionMujeres y numeroDePlazas, que se corresponde con las columnas censoHombres, censoMujeres y numeroPlazas de la tabla de Hechos con Average como tipo de agregador. Igualmente disponemos de un miembro calculado: IndicadorPlazasPorPersona, que responde al cociente de la suma de poblacionHombres más poblacionMujeres sobre numeroDePlazas. No se especifica ningún formato especial para el resultado.
- *IndicadorEquipamientoPorPoblacion*. Incluye las dimensiones Fecha, Equipamiento, y Zona. Tiene tres métricas, poblacionHombres, poblacionMujeres y numeroDeEquipamientos. Tiene tres métricas, poblacionHombres, poblacionMujeres y numeroDeEquipamientos, que se corresponde con las columnas censoHombres, censoMujeres y numeroEquipamientos de la tabla de Hechos con Average como tipo de agregador. Dispone de un miembro calculado IndicadorEquipamientoPorPoblacion que se expresa como el cociente entre la suma de poblacionHombres más poblacionMujeres sobre numeroEquipamientos.
- *PlazasPorSuperficie*. Incluye las dimensiones Fecha, Establecimiento y Zona. Se establecen dos métricas, numeroDePlazas y superficie, que se corresponde con las columnas numeroPlazas y superficie de la tabla de Hechos con Average como tipo de agregador. Se define un miembro calculado



PlazasPorSuperficie como el cociente entre numeroDePlazas y superficie. No se aplica ningún formato específico.

## 7.- Informes

### 7.1.- Acceso a la herramienta de Reporting

Una vez analizados la información disponible, los requerimientos del cliente y las herramientas a utilizar; diseñadas la estructura de la base de datos y los procesos de ETL; generadas las tablas dimensionales y la tabla de hechos la fase siguiente supone la utilización de esa información, con las herramientas adecuadas, de manera que se puedan obtener informes y extraer conclusiones de la manera más sencilla, sin perder la potencia de la herramienta. Para ello se propone la utilización del software gratuito “Pentaho Analysis Services”, que en base a la utilización del esquema diseñado anteriormente, permite diseñar, almacenar y ejecutar un conjunto de informes de acuerdo a las necesidades establecidas por el cliente.

### 7.2.- Descripción de los informes propuestos

A continuación pasaremos a detallar la estructura de los informes propuestos. Una vez detallados se mostrará una captura de pantalla de todos y cada uno combinando diferentes criterios a efectos de comprobar las diferentes posibilidades que ofrece cada uno de ellos.

- *Total de Establecimientos.* Este informe permite obtener información sobre el número de establecimientos considerando las dimensiones Zonas, Establecimientos y Fechas. Permite obtener información por comarca, por provincia y por comunidad autónoma y dentro de estas por cada uno de sus valores. También permite obtener información por camping, hotel y turismo rural e igualmente por cada uno de sus valores. La dimensión Fecha se propone como filtro pero puede ser considerado como fila o columna. El valor devuelto es “NumeroEstablecimientos” de la tabla de hechos. El cubo utilizado dentro de los disponibles en el esquema se denomina “NumeroDeEstablecimientos”. [Acceso al informe](#)
- *Total de Plazas.* Este informe permite obtener información sobre el número total de plazas asociada a los establecimientos considerados en función de las dimensiones asociadas: Zonas, Establecimientos y Fechas. Permite obtener información por comarca, por provincia y por comunidad autónoma y por cada uno de sus valores. También permite obtener información por camping, hotel y turismo rural y para cada uno de sus valores. La dimensión Fecha se propone como filtro pero puede ser considerado como fila o columna. El valor devuelto es “NumeroPlazas” de la tabla de hechos. El cubo utilizado dentro de los disponibles en el esquema se denomina “NumeroDePlazas”. [Acceso al informe](#)
- *% Plazas respecto a Población.* Este informe proporciona información sobre el ratio plazas sobre población total considerando las dimensiones Zonas, Establecimientos y Fechas. Permite obtener información por comarca, por provincia y por comunidad autónoma y para cada uno de sus valores. También permite obtener información por camping, hotel y turismo rural y también para cada uno de sus valores. La dimensión Fecha se propone como filtro pero puede ser considerado como fila o columna. El valor devuelto es un miembro calculado, “PlazasSobrePoblacion” del cubo “PlazasSobrePoblacion”. [Acceso al informe](#)
- *Oferta Media de Plazas.* Este informe muestra información sobre el número de plazas asociada a los establecimientos considerados en función de las dimensiones asociadas: Zonas, Establecimientos y Fechas. Permite obtener información por comarca, por provincia y por comunidad autónoma y cada uno de sus valores. También permite obtener información por camping, hotel y turismo rural y también para cada uno de sus valores. La dimensión Fecha se propone como filtro pero puede ser considerado como fila o columna. El valor devuelto es “NumeroPlazas” de la tabla de hechos. El cubo utilizado dentro de los disponibles en el esquema se denomina “OfertaMediaDePlazas”. El agregador utilizado en la consulta ha sido el de Average, por lo que la información proporcionada ya se refiere a datos medios. [Acceso al informe](#)
- *Número de Establecimientos / Número de Equipamientos.* Este informe proporciona el ratio número de Establecimientos sobre Número de Equipamientos, considerando las dimensiones Establecimientos, Equipamientos y Fechas. Permite obtener información camping, hotel y turismo rural y sus correspondientes valores. También por tipo y categoría de Equipamiento y cada uno de sus valores. La

dimensión Fecha se propone como filtro pero puede ser considerado como fila o columna. El valor devuelto es un miembro calculado, "EstablecimientosSobreEquipamientos" del cubo "EstablecimientosSobreEquipamiento". [Acceso al informe](#)

- *% Población por Equipamiento*. Este informe muestra el ratio de población por equipamiento, considerando las dimensiones Zonas, Equipamientos y Fechas. También permite obtener información por comarca, por provincia y por comunidad autónoma y para cada uno de sus valores. También por tipo y categoría de Equipamiento y cada uno de sus valores. La dimensión Fecha se propone como filtro pero puede ser considerado como fila o columna. El valor devuelto es un miembro calculado, "PoblacionPorEquipamiento" que indica la relación entre la suma de "CensoHombres" más "censoMujeres" sobre "numeroDeEquipamientos". [Acceso al informe](#)
- *Indicador de Establecimientos vs Habitantes por Género*. Este indicador expresa el número de habitantes por cada establecimiento considerando las dimensiones Zonas, Establecimientos y Fechas. Permite obtener información camping, hotel y turismo rural y sus correspondientes valores y también obtener información por comarca, por provincia y por comunidad autónoma y para cada uno de sus valores. La dimensión Fecha se propone como filtro pero puede ser considerado como fila o columna. El valor devuelto es un miembro calculado, "IndicadorEstablecimientosSobreHombres" e "IndicadorEstablecimientosSobreMujeres" que indica la relación entre la suma de "CensoHombres" y "censoMujeres" respectivamente sobre "numeroDeEstablecimientos". [Acceso al informe](#)
- *Indicador de Plazas vs Personas*. Este indicador muestra el número de personas por cada plaza de establecimiento turístico ofertado considerando las dimensiones Zonas, Establecimientos y Fechas. Permite obtener información camping, hotel y turismo rural y sus correspondientes valores y también obtener información por comarca, por provincia y por comunidad autónoma y para cada uno de sus valores. La dimensión Fecha se propone como filtro pero puede ser considerado como fila o columna. El valor devuelto es un miembro calculado, "IndicadorPlazasPorPersona", que expresa el ratio entre la suma de "censoHombres" y "censoMujeres" sobre "numeroDePlazas". [Acceso al informe](#)
- *Indicador de Equipamientos vs Población*. Este indicador muestra la población por cada equipamiento considerando las dimensiones Zonas, Equipamientos y Fechas. También permite obtener información por comarca, por provincia y por comunidad autónoma y para cada uno de sus valores. También por tipo y categoría de Equipamiento y cada uno de sus valores. La dimensión Fecha se propone como filtro pero puede ser considerado como fila o columna. El valor devuelto es un miembro calculado, "IndicadorEquipamientoPorPoblacion" del cubo "IndicadorEquipamientoPorPoblacion" obtenido a partir del ratio. [Acceso al informe](#)
- *Cantidad de Plazas Ofertas / Superficie del Territorio*. Este informe muestra el ratio entre el número de plazas ofertadas por los establecimientos turísticos y la superficie de la zona a la que se refieren considerando las dimensiones Zonas, Establecimientos y Fechas. Permite obtener información camping, hotel y turismo rural y sus correspondientes valores y también obtener información por comarca, por provincia y por comunidad autónoma y para cada uno de sus valores. La dimensión Fecha se propone como filtro pero puede ser considerado como fila o columna. El valor devuelto es un miembro calculado, "PlazasPorSuperficie" a partir de "numeroDePlazaas" y "Superficie". [Acceso al informe](#)
- *Informe sobre Establecimiento*. Este informe presenta diferencias sustanciales respecto a los anteriores, en la medida que ha sido creado con Pentaho Report Designer, una herramienta más avanzada que permite diseñar informes, en formato tradicional, en base a la selección de un conjunto de criterios. Exigen al existencia de consulta relacional previamente establecida, parametrizada, y con un nivel mayor de control por el programador: control de errores, validación de datos, campos calculados, scripts,... En el informe se incluye un objeto que permite generar un conjunto de gráficos de acuerdo a la estructura del informe y los filtros aplicados.



### 7.3.- Total de Establecimientos.

Statistics	Camping / De 1a	Camping / De 2a	Camping / De 3a	Camping / De Luxe	Camping / Privats	Hotel / Estrella De Oro	Hotel / Estrella De Plata	Hotel / Hostal o Pensión	Hotel / Hotel	Turismo Rural / Alojamiento Independiente	Turismo Rural / Casa de Pueblo	Turismo Rural / Casa de Pueblo Compartida	Turismo Rural / Casa de Pueblo Independiente	Turismo Rural / Masia	Turismo Rural / Masovería
Min	65.000	177.000	92.000	3.000	12.000	1434.000	1130.000	1760.000	1077.000	1307.000	187.000	187.000	590.000	175.000	797.000
Max	71.000	187.000	100.000	3.000	12.000	1729.000	1190.000	1760.000	1077.000	1307.000	187.000	202.000	695.000	251.000	1008.000
Sum	474.000	1276.000	673.000	18.000	12.000	9584.000	6930.000	1760.000	1077.000	1307.000	187.000	1167.000	3887.000	1526.000	5465.000
Average	67.714	182.286	96.143	3.000	12.000	1597.333	1155.000	1760.000	1077.000	1307.000	187.000	194.500	647.833	218.000	910.833
Std. Deviation	2.050	3.844	2.416	0.000	0.000	105.590	21.840	0.000	0.000	0.000	0.000	4.537	36.735	23.391	75.050

Tipo Zona	Nombre Zona	Fecha	MeasuresLevel	Camping					Hotel				Turismo Rural					
				De 1a	De 2a	De 3a	De Luxe	Privats	Estrella De Oro	Estrella De Plata	Hostal o Pensión	Hotel	Alojamiento Independiente	Casa de Pueblo	Casa de Pueblo Compartida	Casa de Pueblo Independiente	Masia	Masovería
Comunidad Autónoma	Catalunya	2006	Establecimientos	65	180	97	0	12	1.434	1.190	0	0	1.307	187	0	0	175	0
		2007	Establecimientos	65	178	98	3	0	1.496	1.175	0	0	0	0	187	590	203	797
		2008	Establecimientos	67	177	100	3	0	1.580	1.155	0	0	0	0	193	614	207	845
		2009	Establecimientos	68	182	94	3	0	1.650	1.149	0	0	0	0	195	644	223	886
		2010	Establecimientos	69	185	96	3	0	1.695	1.131	0	0	0	0	197	662	229	946
		2011	Establecimientos	69	187	96	3	0	1.729	1.130	0	0	0	0	193	682	238	983
		2012	Establecimientos	71	187	92	3	0	0	0	1.760	1.077	0	0	202	695	251	1.008

Imagen 44. Evolución Establecimientos por Años. Fuente: Propia.

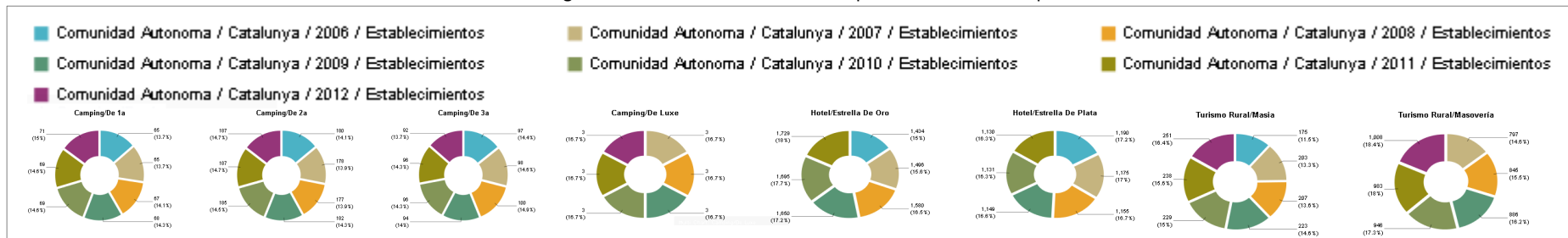


Imagen 45. Gráfico. Evolución Establecimientos por Años (Solo datos representativos). Fuente: Propia.

### 7.4.- Total de Plazas

Statistics	Camping / De 1a / Plazas	Camping / De 2a / Plazas	Camping / De 3a / Plazas	Camping / De Luxe / Plazas	Camping / Privats / Plazas	Hotel / Estrella De Oro / Plazas	Hotel / Estrella De Plata / Plazas	Hotel / Hostal o Pensión / Plazas	Hotel / Hotel / Plazas	Turismo Rural / Alojamiento Independiente / Plazas	Turismo Rural / Casa de Pueblo / Plazas	Turismo Rural / Casa de Pueblo Compartida / Plazas	Turismo Rural / Casa de Pueblo Independiente / Plazas	Turismo Rural / Masia / Plazas	Turismo Rural / Masovería / Plazas
Min	99339.000	98207.000	23369.000	7048.000	11934.000	2004.000	31071.000	265568.000	29209.000	8723.000	2004.000	2040.000	3585.000	2138.000	5574.000
Max	119189.000	122071.000	28935.000	8457.000	11934.000	2004.000	34181.000	265568.000	29209.000	8723.000	2004.000	2170.000	4440.000	3051.000	7307.000
Sum	744367.000	742574.000	180814.000	44968.000	11934.000	2004.000	194508.000	265568.000	29209.000	8723.000	2004.000	12679.000	24397.000	18590.000	39070.000
Average	106338.143	106082.000	25830.571	7494.667	11934.000	2004.000	32418.000	265568.000	29209.000	8723.000	2004.000	2113.167	4066.167	2655.714	6511.667
Std. Deviation	8243.264	9966.980	1846.958	632.937	0.000	0.000	1155.841	0.000	0.000	0.000	0.000	39.935	304.102	287.772	606.016

Tipo Zona	Nombre Zona	Fecha	Camping					Hotel				Turismo Rural					
			De 1a	De 2a	De 3a	De Luxe	Privats	Estrella De Oro	Estrella De Plata	Hostal o Pensión	Hotel	Alojamiento Independiente	Casa de Pueblo	Casa de Pueblo Compartida	Casa de Pueblo Independiente	Masia	Masovería
Comunidad Autónoma	Catalunya	2006	103.235	100.608	25.206	0	11.934	2.004	34.181	0	0	8.723	2.004	0	0	2.138	0
		2007	103.490	99.545	25.755	7.048	0	0	33.588	0	0	0	0	2.040	3.585	2.430	5.574
		2008	100.638	98.207	25.132	7.048	0	0	32.485	0	0	0	0	2.112	3.809	2.536	5.999
		2009	99.339	99.635	23.369	7.048	0	0	32.012	0	0	0	0	2.119	4.022	2.726	6.331
		2010	99.377	101.008	24.336	7.048	0	0	31.171	0	0	0	0	2.140	4.144	2.799	6.776
		2011	119.119	122.071	28.935	8.457	0	0	31.071	0	0	0	0	2.098	4.397	2.910	7.083
		2012	119.169	121.500	28.081	8.319	0	0	0	265.568	29.209	0	0	2.170	4.440	3.051	7.307

Imagen 46. Evolución Plazas por Años. Fuente: Propia.

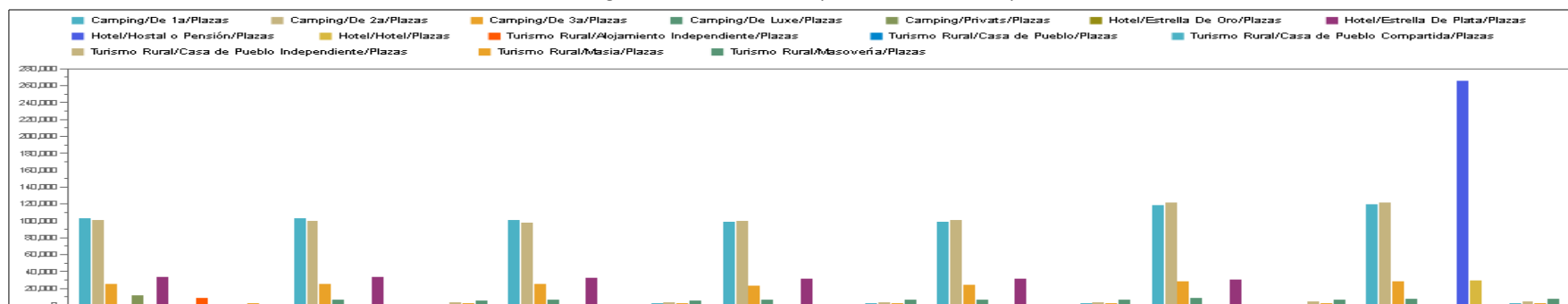


Imagen 47. Gráfico. Evolución Plazas por Años. Fuente: Propia.

### 7.5.- % de Plazas Respecto a Población

Camping / De 1a / Plazas Sobre Poblacion	Camping / De 2a / Plazas Sobre Poblacion	Camping / De 3a / Plazas Sobre Poblacion	Camping / De Luxe / Plazas Sobre Poblacion	Camping / Privats / Plazas Sobre Poblacion	Hotel / Estrella De Oro / Plazas Sobre Poblacion	Hotel / Estrella De Plata / Plazas Sobre Poblacion	Hotel / Hostal o Pensión / Plazas Sobre Poblacion	Hotel / Hotel / Plazas Sobre Poblacion	Turismo Rural / Alojamiento Independiente / Plazas Sobre Poblacion	Turismo Rural / Casa de Pueblo / Plazas Sobre Poblacion	Turismo Rural / Casa de Pueblo Compartida / Plazas Sobre Poblacion	Turismo Rural / Casa de Pueblo Independiente / Plazas Sobre Poblacion	Turismo Rural / Masia / Plazas Sobre Poblacion	Turismo Rural / Masovería / Plazas Sobre Poblacion
0.014	0.014	0.003	0.001	0.002	0.000	0.004	0.036	0.004	0.001	0.000	0.000	0.001	0.000	0.001
0.016	0.017	0.004	0.001	0.002	0.000	0.005	0.036	0.004	0.001	0.000	0.000	0.001	0.000	0.001
0.104	0.104	0.025	0.006	0.002	0.000	0.027	0.036	0.004	0.001	0.000	0.002	0.003	0.003	0.005
0.015	0.015	0.004	0.001	0.002	0.000	0.005	0.036	0.004	0.001	0.000	0.000	0.001	0.000	0.001
0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Nombre Zona	Fecha	Camping					Hotel				Turismo Rural					
		De 1a	De 2a	De 3a	De Luxe	Privats	Estrella De Oro	Estrella De Plata	Hostal o Pensión	Hotel	Alojamiento Independiente	Casa de Pueblo	Casa de Pueblo Compartida	Casa de Pueblo Independiente	Masia	Masovería
		Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion	Plazas Sobre Poblacion
Catalunya	2006	1,51%	1,48%	0,37%	0,00%	0,18%	0,03%	0,50%	0,00%	0,00%	0,13%	0,03%	0,00%	0,00%	0,03%	0,00%
	2007	1,49%	1,43%	0,37%	0,10%	0,00%	0,00%	0,48%	0,00%	0,00%	0,00%	0,00%	0,03%	0,05%	0,03%	0,08%
	2008	1,43%	1,40%	0,36%	0,10%	0,00%	0,00%	0,46%	0,00%	0,00%	0,00%	0,00%	0,03%	0,05%	0,04%	0,09%
	2009	1,38%	1,39%	0,33%	0,10%	0,00%	0,00%	0,45%	0,00%	0,00%	0,00%	0,00%	0,03%	0,06%	0,04%	0,09%
	2010	1,36%	1,39%	0,33%	0,10%	0,00%	0,00%	0,43%	0,00%	0,00%	0,00%	0,00%	0,03%	0,06%	0,04%	0,09%
	2011	1,63%	1,67%	0,40%	0,12%	0,00%	0,00%	0,42%	0,00%	0,00%	0,00%	0,00%	0,03%	0,06%	0,04%	0,10%
	2012	1,62%	1,65%	0,38%	0,11%	0,00%	0,00%	0,00%	3,61%	0,40%	0,00%	0,00%	0,03%	0,06%	0,04%	0,10%

Imagen 48. %Plazas respecto a Población por Años. Fuente: Propia.

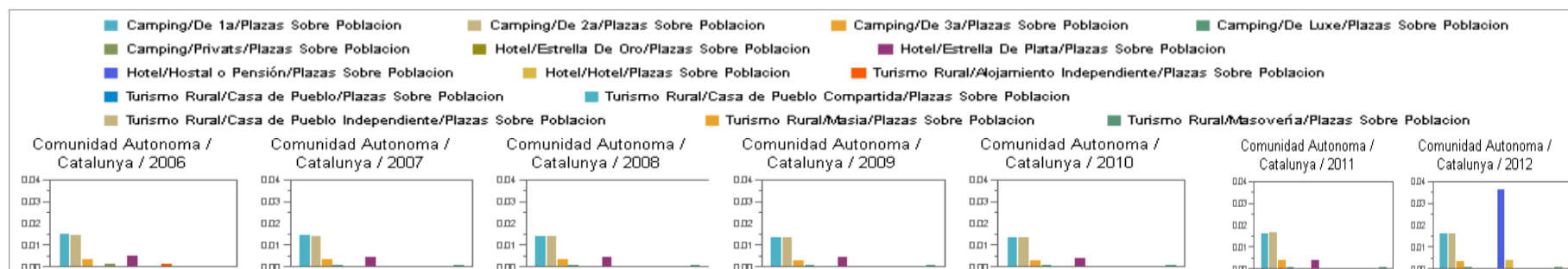


Imagen 49. Gráfico. %Plazas respecto a Población por Años. Fuente: Propia.

### 7.6.- Oferta Media de Plazas

Statistics	Camping / De 1a / Numero De Plazas	Camping / De 2a / Numero De Plazas	Camping / De 3a / Numero De Plazas	Camping / De Luxe / Numero De Plazas	Camping / Privats / Numero De Plazas	Hotel / Estrella De Oro / Numero De Plazas	Hotel / Estrella De Plata / Numero De Plazas	Hotel / Hostal o Pensión / Numero De Plazas	Hotel / Hotel / Numero De Plazas	Turismo Rural / Alojamiento Independiente / Numero De Plazas	Turismo Rural / Casa de Pueblo / Numero De Plazas	Turismo Rural / Casa de Pueblo Compartida / Numero De Plazas	Turismo Rural / Casa de Pueblo Independiente / Numero De Plazas	Turismo Rural / Masia / Numero De Plazas	Turismo Rural / Masoveria / Numero De Plazas
Min	99339.000	98207.000	23369.000	7048.000	11934.000	2004.000	31071.000	265568.000	29209.000	8723.000	2004.000	2040.000	3585.000	2138.000	5574.000
Max	119169.000	122071.000	28935.000	8457.000	11934.000	2004.000	34181.000	265568.000	29209.000	8723.000	2004.000	2170.000	4440.000	3051.000	7307.000
Sum	744367.000	742574.000	180814.000	44968.000	11934.000	2004.000	194508.000	265568.000	29209.000	8723.000	2004.000	12679.000	24397.000	18590.000	39070.000
Average	108338.143	106082.000	25830.571	7494.667	11934.000	2004.000	32418.000	265568.000	29209.000	8723.000	2004.000	2113.167	4066.167	2655.714	6511.867
Std. Deviation	8243.264	9966.980	1846.958	632.937	0.000	0.000	1155.841	0.000	0.000	0.000	0.000	39.935	304.102	287.772	606.016

Imagen 50. Oferta Media de Plazas por Años. Fuente: Propia.



Imagen 51. Gráfico. Oferta Media de Plazas por Años por Años (Solo datos representativos). Fuente: Propia.

### 7.7.- Número de Establecimientos / Número de Equipamientos

Statistics	Camping / De 1a	Camping / De 2a	Camping / De 3a	Camping / De Luxe	Camping / Privats	Hotel / Estrella De Oro	Hotel / Estrella De Plata	Hotel / Hostal o Pensión	Hotel / Hotel	Turismo Rural / Alojamiento Independiente	Turismo Rural / Casa de Pueblo	Turismo Rural / Casa de Pueblo Compartida	Turismo Rural / Casa de Pueblo Independiente	Turismo Rural / Masia	Turismo Rural / Masoveria
Min	0.243	0.662	0.344	0.011	0.046	5.366	4.227	6.584	4.029	4.890	0.700	0.700	2.207	0.655	2.982
Max	0.266	0.700	0.374	0.011	0.046	6.468	4.462	6.584	4.029	4.890	0.700	0.756	2.600	0.939	3.771
Sum	1.773	4.774	2.518	0.067	0.046	35.855	25.926	6.584	4.029	4.890	0.700	4.366	14.542	5.709	20.445
Average	0.253	0.682	0.360	0.011	0.046	5.976	4.321	6.584	4.029	4.890	0.700	0.728	2.424	0.816	3.408
Std. Deviation	0.008	0.014	0.009	0.000	0.000	0.395	0.082	0.000	0.000	0.000	0.000	0.017	0.137	0.088	0.281

Tipo Zona	Nombre Zona	MeasuresLevel	Fecha	Camping					Hotel				Turismo Rural						
				De 1a	De 2a	De 3a	De Luxe	Privats	Estrella De Oro	Estrella De Plata	Hostal o Pensión	Hotel	Alojamiento Independiente	Casa de Pueblo	Casa de Pueblo Compartida	Casa de Pueblo Independiente	Masia	Masoveria	
Comunidad Autónoma	Catalunya	Establecimientos Sobre Equipamientos	2006	24,32%	67,34%	36,29%	0,00%	4,49%	536,48%	445,20%	0,00%	0,00%	488,97%	69,96%	0,00%	0,00%	0,00%	65,47%	0,00%
			2007	24,32%	66,59%	36,66%	1,12%	0,00%	559,67%	439,58%	0,00%	0,00%	0,00%	0,00%	69,96%	220,73%	75,95%	298,17%	
			2008	25,07%	66,22%	37,41%	1,12%	0,00%	591,10%	432,10%	0,00%	0,00%	0,00%	0,00%	72,20%	229,71%	77,44%	316,13%	
			2009	25,44%	68,09%	35,17%	1,12%	0,00%	617,29%	429,86%	0,00%	0,00%	0,00%	0,00%	72,95%	240,93%	83,43%	331,47%	
			2010	25,81%	69,21%	35,91%	1,12%	0,00%	634,12%	423,12%	0,00%	0,00%	0,00%	0,00%	73,70%	247,66%	85,67%	353,91%	
			2011	25,81%	69,96%	35,91%	1,12%	0,00%	646,84%	422,75%	0,00%	0,00%	0,00%	0,00%	72,20%	255,15%	89,04%	367,75%	
			2012	26,56%	69,96%	34,42%	1,12%	0,00%	0,00%	0,00%	658,44%	402,92%	0,00%	0,00%	75,57%	260,01%	93,90%	377,11%	

Imagen 52. Número de Establecimientos / Número de Equipamientos por Años. Fuente: Propia.

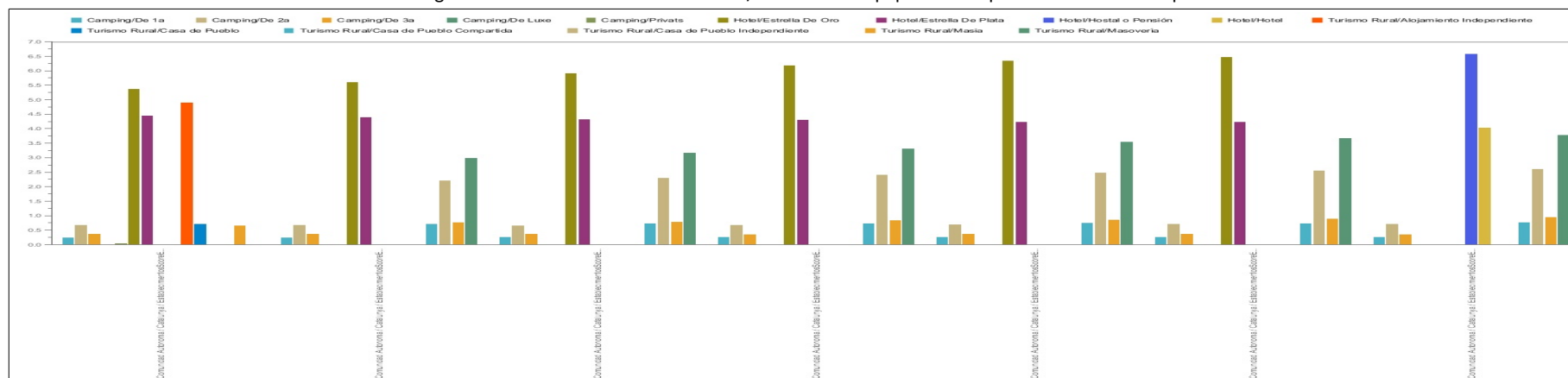


Imagen 53. Gráfico. Número de Establecimientos / Número de Equipamientos por Años. Fuente: Propia

### 7.8.- % de Población por Equipamiento

Notas: Considerando el elevado número en los tipos y subtipos de equipamientos sólo mostraremos resultados parciales, a modo de ejemplo. Se omite el % Población por Equipamiento ya que proporciona valores muy bajos. Se muestra el indicador de población por Equipamiento.

			Comunidad Autónoma						
			Catalunya						
TipoEquipamiento	SubtipoEquipamiento	MeasuresLevel	2006	2007	2008	2009	2010	2011	2012
Universitats	Escoles	Poblacion Por Equipamientos	79.245,349	80.841,814	81.678,64	83.424,349	84.697,07	85.121,512	85.433,395
	Estructures_universitarias	Poblacion Por Equipamientos	6.815.100	6.952.396	7.024.363	7.174.494	7.283.948	7.320.450	7.347.272
	Facultats	Poblacion Por Equipamientos	59.781,579	60.985,93	61.617,219	62.934,158	63.894,281	64.214,474	64.449,754
	Facultats_escoles	Poblacion Por Equipamientos	34.075,5	34.761,98	35.121,815	35.872,47	36.419,74	36.602,25	36.736,36
	No_presencials	Poblacion Por Equipamientos	6.815.100	6.952.396	7.024.363	7.174.494	7.283.948	7.320.450	7.347.272
	Recerca	Poblacion Por Equipamientos	6.815.100	6.952.396	7.024.363	7.174.494	7.283.948	7.320.450	7.347.272

Imagen 54. Población por Equipamiento, filtrado por Comunidad Autónoma y Equipamiento tipo Universitat por Años. Fuente: Propia.

			Salut						
Tipo Zona	Nombre Zona	Fecha	Centre_sanitari	Centres_atencio_seguiment_drogodependencies	Centres_d_atencio Primaria	Centres_de_salut_mental	Centres_sociosanitaris	Hospitals	centres_seguretat_salut_Laboral
			Poblacion Por Equipamientos	Poblacion Por Equipamientos	Poblacion Por Equipamientos	Poblacion Por Equipamientos	Poblacion Por Equipamientos	Poblacion Por Equipamientos	Poblacion Por Equipamientos
Provincia	Barcelona	2008	7.102,405	105.967,88	12.379,425	37.051,706	58.871,044	151.382,686	5.298.394
		2009	7.213,886	107.631,18	12.573,736	37.633,28	59.795,1	153.758,829	5.381.559
		2010	7.309,221	109.053,58	12.739,904	38.130,622	60.585,322	155.790,829	5.452.679
		2011	7.339,996	109.512,74	12.793,544	38.291,168	60.840,411	156.446,771	5.475.637
		2012	7.364,016	109.871,12	12.835,411	38.416,476	61.039,511	156.958,743	5.493.556
	Girona	2008	2.622,42	109.267,5	3.182,549	59.600,455	46.828,929	50.431,154	655.605
		2009	2.719,464	113.311	3.300,32	61.806	48.561,857	52.297,385	679.886
		2010	2.778,812	115.783,833	3.372,345	63.154,818	49.621,643	53.438,692	694.703
		2011	2.797,88	116.578,333	3.395,485	63.588,182	49.962,143	53.805,385	699.470
		2012	2.811,132	117.130,5	3.411,588	63.889,364	50.198,786	54.060,231	702.783
	Lleida	2008	1.479,059	44.002	1.833,417	16.762,667	32.001,465	58.669,333	352.016
		2009	1.530,845	45.542,625	1.897,609	17.348,571	33.121,909	60.723,5	364.341
		2010	1.570,399	46.719,375	1.946,641	17.797,857	33.977,727	62.292,5	373.755
		2011	1.585,529	47.169,5	1.965,396	17.969,333	34.305,091	62.892,667	377.356
		2012	1.596,468	47.494,625	1.978,943	18.093,19	34.541,545	63.326,167	379.957
	Tarragona	2008	2.861,944	179.587	3.453,596	44.896,75	51.310,571	79.816,444	718.348
		2009	2.982,98	187.182	3.599,654	46.795,5	53.480,571	83.192	748.728
		2010	3.039,088	190.702,75	3.667,361	47.675,688	54.486,5	84.756,778	762.811
		2011	3.059,709	191.996,75	3.692,245	47.999,188	54.856,214	85.331,889	767.987
		2012	3.071,618	192.744	3.706,615	48.186	55.069,714	85.664	770.976

Imagen 55. Población por Equipamiento, filtrado por Provincias y Equipamiento tipo Salut de 2008 a 2012. Fuente: Propia.



### 7.9.- Indicador de Establecimientos vs por Habitantes por Género

Tipo Zona	Nombre Zona	MeasuresLevel	Fecha	Camping					Hotel				Turismo Rural					
				De 1a	De 2a	De 3a	De Luxe	Privats	Estrella De Oro	Estrella De Plata	Hostal o Pensión	Hotel	Alojamiento Independiente	Casa de Pueblo	Casa de Pueblo Compartida	Casa de Pueblo Independiente	Masia	Masoveria
Comunidad Autónoma	Catalunya	Indicador Establecimientos Sobre Hombres	2006	51.218,954	18.495,733	34.321,979	∞	277.436	2.321,64	2.797,674	∞	∞	2.547,232	17.803,38	∞	∞	19.024,183	∞
			2007	52.267,2	19.086,337	34.667,02	1.132.466	∞	2.270,968	2.891,377	∞	∞	∞	∞	18.167,743	5.758,251	16.735,803	4.262,695
			2008	51.248,985	19.399,333	34.336,82	1.144.560,667	∞	2.173,216	2.972,885	∞	∞	∞	∞	17.791,098	5.592,316	16.587,836	4.063,529
			2009	51.587,588	19.274,484	37.318,661	1.169.318,667	∞	2.126,034	3.053,051	∞	∞	∞	∞	17.989,518	5.447,137	15.730,744	3.959,318
			2010	51.625,493	19.254,914	37.105,823	1.187.386,333	∞	2.101,569	3.149,566	∞	∞	∞	∞	18.082,025	5.380,905	15.555,279	3.765,496
			2011	51.889,609	19.146,433	37.295,656	1.193.461	∞	2.070,783	3.168,481	∞	∞	∞	∞	18.551,207	5.249,828	15.043,826	3.642,302
			2012	51.153,085	19.421,759	39.476,837	1.210.623	∞	∞	∞	2.063,562	3.372,209	∞	∞	17.979,55	5.225,711	14.469,598	3.603,045
	Indicador Establecimientos Sobre Mujeres	2006	53.628,738	19.365,933	35.936,784	∞	290.489	2.430,87	2.929,301	∞	∞	2.667,076	18.641,005	∞	∞	19.919,246	∞	
		2007	54.692,738	19.972,067	36.275,796	1.185.009,333	∞	2.376,356	3.025,556	∞	∞	∞	∞	19.010,845	6.025,471	17.512,453	4.460,512	
		2008	53.592,254	20.286,333	35.906,81	1.196.893,667	∞	2.272,583	3.108,815	∞	∞	∞	∞	18.604,565	5.848,015	17.346,285	4.249,327	
		2009	53.919,676	20.145,813	39.005,723	1.222.179,333	∞	2.222,144	3.191,069	∞	∞	∞	∞	18.802,759	5.693,382	16.441,874	4.138,305	
		2010	53.938,971	20.117,778	38.768,635	1.240.596,333	∞	2.195,746	3.290,706	∞	∞	∞	∞	18.892,33	5.622,038	16.252,354	3.934,238	
		2011	54.203,87	20.000,358	38.959,031	1.246.689	∞	2.163,139	3.309,794	∞	∞	∞	∞	19.378,585	5.483,969	15.714,567	3.804,748	
		2012	52.329,62	19.868,465	40.384,815	1.238.467,667	∞	∞	∞	2.111,024	3.449,771	∞	∞	18.393,084	5.345,904	14.802,402	3.685,916	

Imagen 56. Indicador de Establecimientos vs Habitantes por Género por Años. Fuente: Propia.

Tipo Zona	Nombre Zona	MeasuresLevel	Camping					Hotel				Turismo Rural					
			De 1a	De 2a	De 3a	De Luxe	Privats	Estrella De Oro	Estrella De Plata	Hostal o Pensión	Hotel	Alojamiento Independiente	Casa de Pueblo	Casa de Pueblo Compartida	Casa de Pueblo Independiente	Masia	Masoveria
Provincia	Barcelona	Indicador Establecimientos Sobre Hombres	245.203,545	72.898,351	96.329,964	∞	∞	∞	∞	3.777,646	5.993,864	∞	∞	224.769,917	28.096,24	39.665,279	7.250,642
		Indicador Establecimientos Sobre Mujeres	254.210,636	75.576,135	99.868,464	∞	∞	∞	∞	3.916,41	6.214,038	∞	∞	233.026,417	29.128,302	41.122,309	7.516,981
	Girona	Indicador Establecimientos Sobre Hombres	8.624,195	4.980,169	11.406,194	176.796	∞	∞	∞	656,015	1.091,333	∞	∞	8.418,857	2.699,176	2.828,736	883,98
		Indicador Establecimientos Sobre Mujeres	8.516,854	4.918,183	11.264,226	174.595,5	∞	∞	∞	647,85	1.077,75	∞	∞	8.314,071	2.665,58	2.793,528	872,978
	Lleida	Indicador Establecimientos Sobre Hombres	96.175,5	4.691,488	10.123,737	∞	∞	∞	∞	815,047	1.028,615	∞	∞	1.732,892	577,631	8.363,087	1.373,936
		Indicador Establecimientos Sobre Mujeres	93.803	4.575,756	9.874	∞	∞	∞	∞	794,941	1.003,241	∞	∞	1.690,144	563,381	8.156,783	1.340,043
	Tarragona	Indicador Establecimientos Sobre Hombres	22.863,941	10.228,605	27.763,357	388.687	∞	∞	∞	1.434,269	3.350,75	∞	∞	10.505,054	2.879,163	11.105,343	4.048,823
		Indicador Establecimientos Sobre Mujeres	22.487,588	10.060,237	27.306,357	382.289	∞	∞	∞	1.410,661	3.295,595	∞	∞	10.332,135	2.831,77	10.922,543	3.982,177

Imagen 57. Indicador de Establecimientos vs Habitantes por Género por Provincias para el Año 2012. Fuente: Propia.



### 7.10.- Indicador de Plazas vs Personas

Tipo Zona	Nombre Zona	MeasuresLevel	Fecha	Camping					Hotel				Turismo Rural						
				De 1a	De 2a	De 3a	De Luxe	Privats	Estrella De Oro	Estrella De Plata	Hostal o Pensión	Hotel	Alojamiento Independiente	Casa de Pueblo	Casa de Pueblo Compartida	Casa de Pueblo Independiente	Masia	Masoveria	
Provincia	Barcelona	Indicador Plazas Por Persona	2008	664,292	298,855	663,543	∞	∞	∞	436,513	∞	∞	∞	∞	∞	70.645,263	10.268,205	7.623,588	2.165,261
			2009	674,719	286,573	797,268	∞	∞	∞	443,109	∞	∞	∞	∞	∞	71.754,12	9.199,246	7.252,775	2.080,232
			2010	683,636	263,071	839,519	∞	∞	∞	451,68	∞	∞	∞	∞	∞	60.585,322	9.012,693	7.358,541	1.960,69
			2011	568,956	220,197	657,024	∞	∞	∞	445,282	∞	∞	∞	∞	∞	52.650,356	8.609,492	7.339,996	1.889,454
	2012	487,277	220,971	660,998	∞	∞	∞	∞	45,102	481,933	∞	∞	∞	∞	43.256,346	8.651,269	7.025,008	1.865,384	
	Girona	Indicador Plazas Por Persona	2008	10,461	15,824	99,545	133,037	∞	∞	56,856	∞	∞	∞	∞	∞	1.583,587	849,229	496,67	298,138
			2009	10,995	16,323	107,95	137,96	∞	∞	60,8	∞	∞	∞	∞	∞	1.592,192	828,095	486,661	293,806
			2010	11,267	16,937	95,492	140,971	∞	∞	64,45	∞	∞	∞	∞	∞	1.611,84	839,013	482,098	279,109
			2011	9,461	14,246	87,031	118,294	∞	∞	65,261	∞	∞	∞	∞	∞	1.533,925	790,362	464,147	268,099
	2012	9,465	14,454	89,413	121,694	∞	∞	∞	∞	9,909	69,104	∞	∞	∞	1.422,638	773,99	443,117	256,771	
	Lleida	Indicador Plazas Por Persona	2008	212,698	26,38	100,547	∞	∞	∞	65,056	∞	∞	∞	∞	∞	284,572	208,912	1.796	441,123
			2009	242,894	26,546	104,068	∞	∞	∞	69,201	∞	∞	∞	∞	∞	294,774	209,753	1.481,061	426,13
2010			249,17	27,232	106,757	∞	∞	∞	72,786	∞	∞	∞	∞	∞	301,172	212,12	1.604,099	425,205	
2011			209,642	22,924	89,847	∞	∞	∞	74,18	∞	∞	∞	∞	∞	332,18	198,087	1.578,895	404,888	
2012	211,087	23,082	90,466	∞	∞	∞	∞	25,066	79,158	∞	∞	∞	∞	329,824	197,175	1.423,06	396,201		
Tarragona	Indicador Plazas Por Persona	2008	25,352	27,947	101,749	338,843	∞	∞	210,969	∞	∞	∞	∞	∞	1.861,005	871,782	2.210,302	1.294,321	
		2009	26,769	29,385	109,784	353,174	∞	∞	218,926	∞	∞	∞	∞	∞	1.965,165	851,795	2.195,683	1.302,136	
		2010	27,007	29,76	109,411	359,817	∞	∞	239,501	∞	∞	∞	∞	∞	2.018,019	803,805	1.986,487	1.216,604	
		2011	22,746	24,271	91,821	301,882	∞	∞	258,669	∞	∞	∞	∞	∞	1.910,415	790,924	1.837,289	1.192,526	
2012	24,208	24,431	99,997	303,057	∞	∞	∞	∞	13,367	271,47	∞	∞	∞	1.942,005	794,821	1.853,308	1.157,622		

Imagen 58. Indicador de Establecimientos vs Habitantes por Género por Provincias entra 2008 y 2012. Fuente: Propia.

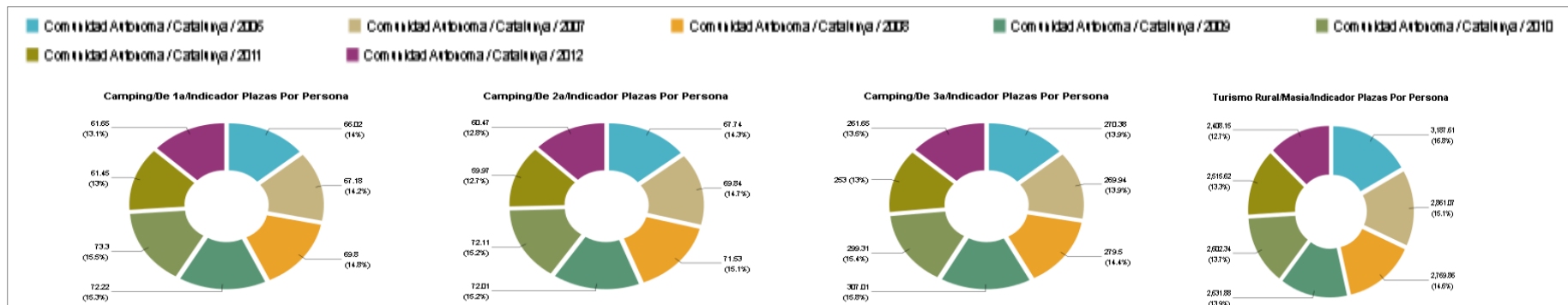


Imagen 59. Indicador de Establecimientos vs Habitantes por Género por CCAA por Años (Solo datos representativos). Fuente: Propia.

### 7.11.- Indicador de Equipamientos vs Población

Statistics	Comunidad Autonoma / Catalunya / 2006	Comunidad Autonoma / Catalunya / 2007	Comunidad Autonoma / Catalunya / 2008	Comunidad Autonoma / Catalunya / 2009	Comunidad Autonoma / Catalunya / 2010	Comunidad Autonoma / Catalunya / 2011	Comunidad Autonoma / Catalunya / 2012
Min	37861.667	38624.422	39024.239	39858.300	40466.378	40669.167	40818.178
Max	681510.000	695239.600	702436.300	717449.400	728394.800	732046.000	734727.200
Sum	908680.000	926986.133	936581.733	956599.200	971193.067	976060.000	979636.267
Average	302893.333	308995.378	312193.911	318866.400	323731.022	325353.333	326545.422
Std. Deviation	274768.943	280304.398	283205.940	289258.872	293671.802	295143.477	296224.878

			Comunidad Autonoma Catalunya						
TipoEquipamiento	SubtipoEquipamiento	MeasuresLevel	2006	2007	2008	2009	2010	2011	2012
Turisme	Albergs_de_Joventut	Indicador Equipamiento Por Poblacion	189.308,333	193.122,111	195.121,194	199.291,5	202.331,889	203.345,833	204.090,889
	Oficines_de_turisme	Indicador Equipamiento Por Poblacion	681.510	695.239,6	702.436,3	717.449,4	728.394,8	732.046	734.727,2
	oficines_municipals_turisme	Indicador Equipamiento Por Poblacion	37.861,667	38.624,422	39.024,239	39.858,3	40.466,378	40.669,167	40.818,178

Imagen 60. Indicador de Equipamientos vs Población por CCAA para el equipamiento tipo Turisme. Fuente: Propia.

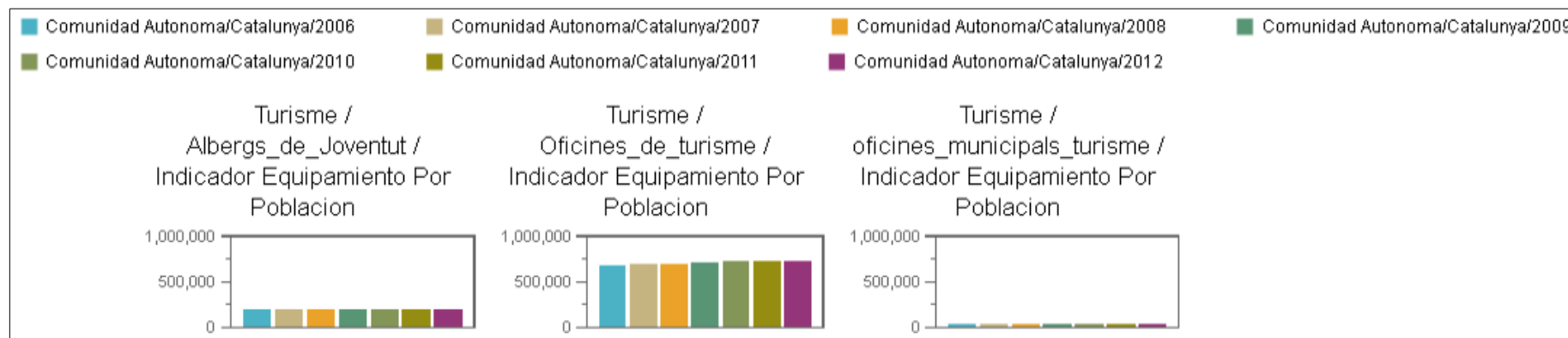


Imagen 61. Indicador de Equipamientos vs Población por CCAA para el equipamiento tipo Turisme. Fuente: Propia.

### 7.12.- Cantidad de Plazas Ofertadas / Superficie del Territorio

Statistics	Provincia / Barcelona / Plazas Por Superficie	Provincia / Girona / Plazas Por Superficie	Provincia / Lleida / Plazas Por Superficie	Provincia / Tarragona / Plazas Por Superficie
Min	0.025	0.166	0.065	0.099
Max	24.336	24.957	4.009	14.395
Sum	36.391	75.067	11.379	34.097
Average	4.043	7.507	1.264	3.410
Std. Deviation	7.327	9.584	1.425	4.699

Tipo Establecimiento	Categoria	Fecha	Provincia			
			Barcelona	Girona	Lleida	Tarragona
Camping	De 1a	2012	2,253	24,957	0,438	7,948
	De 2a	2012	4,967	16,343	4,009	7,875
	De 3a	2012	1,661	2,642	1,023	1,924
	De Luxe	2012	0	1,941	0	0,635
	Privats	2012	0	0	0	0
Hotel	Estrella De Oro	2012	0	0	0	0
	Estrella De Plata	2012	0	0	0	0
	Hostal o Pensión	2012	24,336	23,841	3,692	14,395
	Hotel	2012	2,278	3,418	1,169	0,709
Turismo Rural	Alojamiento Independiente	2012	0	0	0	0
	Casa de Pueblo	2012	0	0	0	0
	Casa de Pueblo Compartida	2012	0,025	0,166	0,281	0,099
	Casa de Pueblo Independiente	2012	0,127	0,305	0,469	0,242
	Masia	2012	0,156	0,533	0,065	0,104
	Masovería	2012	0,588	0,92	0,234	0,166

Imagen 62. Plazas Ofertadas / Superficie por Provincias para el Año 2012. Fuente: Propia.

Tipo Zona	Nombre Zona	MeasuresLevel	Camping					Hotel				Turismo Rural					
			De 1a	De 2a	De 3a	De Luxe	Privats	Estrella De Oro	Estrella De Plata	Hostal o Pensión	Hotel	Alojamiento Independiente	Casa de Pueblo	Casa de Pueblo Compartida	Casa de Pueblo Independiente	Masia	Masovería
Comunidad Autónoma	Catalunya	Plazas Por Superficie	2012	2012	2012	2012	2012	2012	2012	2012	2012	2012	2012	2012	2012	2012	2012
			7,405	7,55	1,745	0,517	0	0	0	16,502	1,815	0	0	0,135	0,276	0,19	0,454

Imagen 63. Plazas Ofertadas / Superficie por CCAA para el Año 2012. Fuente: Propia.

### 7.13.- Informe de Establecimientos

Imagen 64. Informe sobre Establecimientos de Turismo Rural en Barcelona en el Año 2006 (I). Fuente: Propia.

TFC - Construcción y explotación de un almacén de datos para el análisis de información sobre alojamientos turísticos  
 Alumno: Fernando Turón Manzanares  
 Curso: 2012-2013. 2º Semestre

## Informe sobre Establecimientos

Año	Tipo de Zona	Zona	Tipo de Establecimiento	Categoría	Número de Plazas	Número de Establecimientos
2006	Provincia	Barcelona	Turismo Rural	Alojamiento Independiente	2,572	350
2006	Provincia	Barcelona	Turismo Rural	Masia	555	48
2006	Provincia	Barcelona	Turismo Rural	Casa de Pueblo	96	11
2006	Provincia	Barcelona	Turismo Rural	Casa de Pueblo Compartida	0	0
2006	Provincia	Barcelona	Turismo Rural	Casa de Pueblo Independiente	0	0
2006	Provincia	Barcelona	Turismo Rural	Masovería	0	0

1 de 3

Imagen 65. Informe sobre Establecimientos de Turismo Rural en Barcelona en el Año 2006 (II). Fuente: Propia.

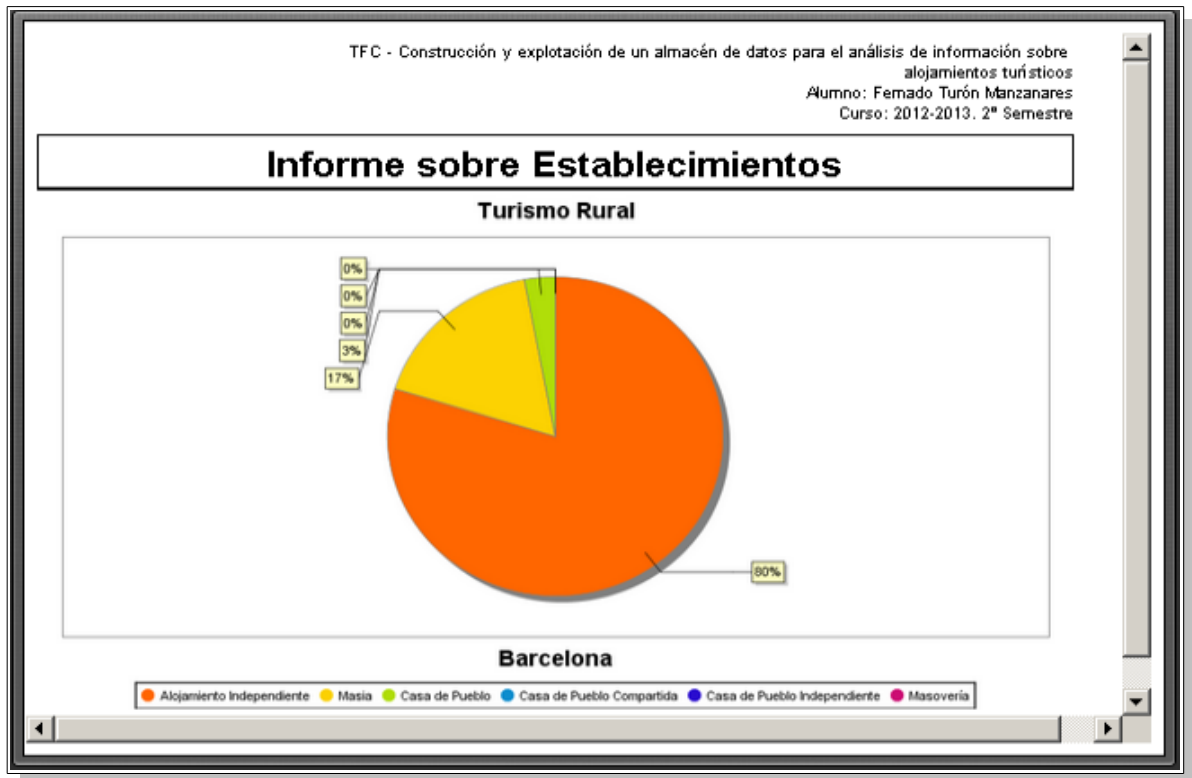


Imagen 66. Informe sobre Establecimientos de Turismo Rural en Barcelona en el Año 2006 (III). Fuente: Propia.

## 8.- Conclusiones

El objetivo previsto para el presente proyecto era doble, por un lado conocer el funcionamiento de un almacén de datos y los principios de Business Intelligence y por otro lado la construcción de un modelo, en este caso aplicado a la oferta de servicios turísticos y equipamientos disponibles.

Desde el punto de vista del Análisis ha sido relativamente fácil. Las fuentes de información estaban claramente identificadas, al ser proporcionadas por el cliente, así como los objetivos pretendidos por el mismo cliente. Las primeras fases del proyecto que exigían un análisis exhaustivo de los ficheros sacaron a la luz una serie de inconsistencias que han ocasionado ciertas complicaciones. En algunos casos las dificultades vienen dadas por cambios legales o administrativos, como en el caso de la falta de información sobre las veguerras o segregación o unificación de municipios e incluso los cambios en la clasificación de categorías. En otros casos por el diseño ineficiente en la elaboración de los propios ficheros, con información no estructurada, redundante o innecesarias. En fases más avanzadas del análisis no se han encontrado especiales dificultades.

Respecto al diseño y centrándonos en el diseño propio de la base de datos se ha tratado de seguir un criterio de simplicidad en la estructura, minimizando el número de campos, optimizando el tipo de datos a las características y volumen de la información proporcionada. En este sentido se ha optado por un modelo no basado en el crecimiento, sino en la adecuación a las necesidades del cliente. Se considera que las fuentes de información son tan heterogéneas en el contenido y dinámicas en el tiempo que es difícil establecer una estructura estable para un horizonte temporal razonablemente amplio. En la información considerada se observan contenidos muy cambiantes de un año a otro.

Considerando la utilización de las herramientas de ETL si bien inicialmente se ha mostrado cierta desconfianza – opacidad de algunos procesos, desconocimiento previo de este tipo de herramientas – frente a un posible diseño basado en código, finalmente se acepta la utilización comprobándose las ventajas del producto utilizado: facilidad en el uso, gestión de errores, diversidad de objetos ya programados, flexibilidad – se puede programar código Java o Javascript -, estabilidad, ... La modularidad en el diseño de transformaciones y la capacidad de agregación, para la creación de trabajos, hacen de la herramienta un producto muy potente, de fácil implementación incluso en entornos profesionales y críticos.

Respecto a la implementación, en particular, a la hora de diseñar el conjunto de reportes que permiten

visualizar la información, se puede considerar que ha sido la fase más cómoda, en la medida que el esfuerzo en las fases de análisis y diseño proporcionan un conocimiento del conjunto de datos que facilita la consecución de los objetivos del cliente. La herramienta es relativamente sencilla, tanto Workbench Schema como Pentaho Analysis Service, con una interfaz gráfica simple y potente que permite ejecutar las consultas en un tiempo relativamente corto, considerando el volumen de registros disponibles y el amplio número de restricciones y combinaciones de datos que se pueden presentar.

Considerando la planificación prevista y la ejecución de las distintas fases no se han observado desviaciones importantes. En general se puede afirmar que las realizaciones han ido adelantadas respecto a las estimaciones, lo que ha permitido ahondar en algunos aspectos y testear aquellas acciones – transformaciones, informes, instalaciones,... - que se han considerado más críticas. Tan solo en dos momentos se produjo un retraso respecto a las tareas previstas: en las primeras fases del uso de Pentaho Data Integration, donde la gran cantidad de objetos disponibles exigía probar algunos antes de seleccionar el más adecuado y por otro lado en las primeras fases de diseño de los informes, fue necesario instalar Workbench Schema que no venía con la máquina – y no estaba clara la herramienta de reporting a utilizar. En cualquier caso, la utilización de los foros en Internet sobre Pentaho, tanto el oficial como las innumerables entradas en blogs, ha permitido solventar la mayor parte de las dudas que han ido surgiendo durante la vida del proyecto.

En definitiva se puede afirmar que el proyecto considerado en su conjunto proporciona una visión relativamente amplia de las posibilidades del ámbito de los almacenes de datos. Se trata sin duda de una aproximación, basada en herramientas de software libre, pero que permite establecer las vías de acción de un proyecto profesional, más aún considerando que los datos de partida son reales, públicos, de fácil acceso y ampliables, tanto desde el punto de vista temporal como geográfico o temático.

## 9.- Líneas de Evolución Futura

Varias son las vías de ampliación del presente estudio. A continuación se va a reseñar algunas de ellas:

- Desarrollar la información referida a veguerras. En los datos proporcionados no ha sido posible, en la mayoría de los casos obtener información relativa a veguerras. Quizás las trabas legales al reconocimiento de este tipo de entidades ha contribuido a esta dificultad.
- Desde el punto de vista geográfico se podría ampliar el análisis a otras comunidades autónomas o incluso países. Las dificultades vendrían por la consolidación de datos, por la homogeneización de la información, en la medida que no existe, ni siquiera a nivel nacional, uniformidad en los parámetros de clasificación. En este sentido se podría incluir unidades de carácter no administrativo y basadas en zonas geográficas: sistemas montañosos, cuencas hidrográficas, parques naturales,...
- Ampliar los tipos de establecimientos turísticos a analizar. Así se podría incluir información sobre balnearios, albergues, restaurantes, bares, teatros,...
- Llevar a cabo un análisis de costes de las consultas realizadas tanto en los procesos ETL como en la ejecución de los informes propuestos. Si bien el esquema lógico del modelo es relativamente simple, bajo la premisa de minimizar el número de relaciones entre tablas, todas las consultas son susceptibles de este análisis. En especial la elección de índices adecuados y la utilización de vistas convenientemente indexadas, puede acelerar extraordinariamente los procesos ETL.
- Ampliar el horizonte temporal de los datos a analizar, recuperando información de años anteriores al 2006.
- Utilización de herramientas profesionales de análisis, especialmente en lo relativo a la elaboración de informes. Soportan más posibilidades de cálculo: más funciones, estadísticas, accesibilidad a la información...

## 10.- Glosario

**Código INE.** Hace referencia a la relación de municipios y sus códigos por provincias actualizados anualmente a fecha 1 de Enero, de acuerdo a los municipios y provincias que figuran en el Registro de Entidades Locales (REL) Del Ministerio de Hacienda y Administraciones Públicas. Se compone de 5 dígitos, de los que los dos primeros corresponden al código de la provincia y los tres restantes al municipio dentro de esta.

Fuente: Instituto Nacional de Estadística.

<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t20/e245/codmun&file=inebase>



**CSM.** Sistema informático de apoyo a la gestión de la compañía como parte integrante de la estrategia adoptada para la mejora de todos los flujos físicos y administrativos que permiten planificar, implantar, minimizar costes y controlar la entrega a los clientes de bienes y servicios de calidad demandados por el mercado, a partir de productos obtenidos de los proveedores de la compañía.

*Fuente: UOC, Apuntes de la asignatura Fundamentos de Sistemas de Información*

**ERP.** Enterprise Resource Planning o Plan de Recursos Empresariales. Conjunto de aplicaciones relacionadas con los procesos de negocio de una empresa, que permite lograr los objetivos estratégicos definidos por los directivos de una organización.

*Fuente: UOC, Asignatura de Fundamentos de Sistemas de Información, Grado de Informática.*

**ETL.** En el contexto de Almacenes de Datos nos referimos a los procesos responsables de extraer los datos de las fuentes de datos transaccionales, realizar las transformaciones necesarias, cargarlos en el almacén de datos una vez hayan sido tratados y realizar los refrescos o cargas sucesivas de datos durante la vida del almacén de datos.

*Fuente: Diseño y Explotación de Almacenes de Datos. Conceptos básicos de modelado multidimensional (Editorial Club Universitario)*

**MDX.** Acrónimo para MultiDimensional eXpressions. Es un lenguaje de consulta para bases de datos multidimensionales sobre cubos OLAP basados en datos históricos., con la posibilidad de intercambiar los ejes del cubo. Devuelve un conjunto de celdas del cubo original. Se basa en las relaciones de jerarquías.

Fuentes: Wikipedia y Otros

[http://es.wikipedia.org/wiki/Expresiones\\_multidimensionales](http://es.wikipedia.org/wiki/Expresiones_multidimensionales)

[http://www.iccube.com/support/documentation/mdx\\_tutorial/gentle\\_introduction.html](http://www.iccube.com/support/documentation/mdx_tutorial/gentle_introduction.html)

**Normalización de Bases de Datos.** Se refiere al conjunto de reglas encaminadas a eliminar redundancias e inconsistencias de dependencia en el diseño de las tablas. Supone una redefinición del diseño encaminada a la obtención de estructuras mas simples y más estables. Además se produce una optimización de los recursos disponibles: tanto almacenamiento como capacidad de proceso. En la práctica se traduce una serie de reglas en las que cada una se basa en la anterior para su implantación. Fundamentalmente las reglas establecidas inicialmente por Boyce y Codd en 1974, se pueden resumir en:

- 1FN, Primera Forma Normal. Las columnas repetidas debe eliminarse y colocarse en tablas separadas.
- 2FN, Segunda Forma Normal. Las dependencias parciales deben eliminarse y separarse dentro de sus propias tablas. Una dependencia hace referencia a aquel conjunto de datos que no depende de la clave principal para su interpretación.
- 3FN, Tercera Forma Normal. Exige que las columnas que no sean llave sean funcionalmente dependientes de la clave primaria y además no dependen de de otra columna que tampoco sea llave.

*Fuentes:*

<http://www.eet2mdp.edu.ar/alumnos/MATERIAL/MATERIAL/info/infonorma.pdf>

[http://ocw.uc3m.es/ingenieria-informatica/diseno-de-bases-de-](http://ocw.uc3m.es/ingenieria-informatica/diseno-de-bases-de-datos/teoria/Tema4%28Normalizacion%29.pdf/view)

[datos/teoria/Tema4%28Normalizacion%29.pdf/view](http://ocw.uc3m.es/ingenieria-informatica/diseno-de-bases-de-datos/teoria/Tema4%28Normalizacion%29.pdf/view)

[http://en.wikipedia.org/wiki/Boyce%E2%80%93Codd\\_normal\\_form](http://en.wikipedia.org/wiki/Boyce%E2%80%93Codd_normal_form)

**OLAP.** Acrónimo de On Line Analytical Processing. Desarrolla análisis multidimensional de datos de negocio y ofrece la capacidad de cálculos complejos, análisis de tendencia y sofisticado modelado de datos.

*Fuente: OLAP Education Wiki*

**SGBDR.** Sistema Gestión de Bases de Datos Relacional. Conjunto de aplicaciones que permiten el almacenamiento, modificación y extracción de la información en una Base de Datos. De igual manera proporciona herramientas para añadir, modificar, consultar y eliminar los datos así como una interface para su acceso.

*Fuente: Propia y Wikipedia*

**Veguería.** Se refiere a una distribución administrativa que se aplica en Catalunya, que es de origen medieval. Se establecen 7 veguerías y se contempla la posibilidad de que sustituyan a las actuales diputaciones provinciales. No está clara la demarcación en algunos casos, puesto que algunos municipios han pedido pertenecer a unidades diferentes de las establecidas. En cualquier caso existe una sentencia del Tribunal Constitucional que ha impedido que se constituyan de pleno derecho.

Fuentes: <http://ca.wikipedia.org/wiki/Vegueria>

<http://www.vegeries.com/indexESP.htm>



## 11.- Bibliografía

### Básica

**Diseño y Explotación de Almacenes de Datos. Conceptos Básicos de Modelado Multidimensional.**

Editorial: Club Universitario. Alicante

Autores: Juan Carlos Trujillo Mondéjar, José Norberto Mazón López y Jesús Pardillo Vela

Año 2011

**Fundamentos de Bases de Datos.**

Editorial: Mc Graw Hill

Autores: Silberschatz, Kortz, Sudarsahan

Año 2002

**Introducción al Business Intelligence**

Editorial: UOC

Autores: Jordi Conesa Caralt, Josep Curto Diaz

Año 2010

### Complementaria

**Building the Data Warehouse**

Editorial: Winley

Autor: W H Inmon

Año 1996

**The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouse**

Editorial: Winley

Autor: R. Kimbal

Año 1996

**Introducción a la Minería de Datos**

Editorial: Pearson – Prentice Hall

Autores: José Hernández Orallo, M<sup>a</sup> José Ramírez Quintana, Cesar Ferri, Ramírez

Año 2005

## 12. -Webgrafía

**IDESCAT**, Instituto de Estadística de Catalunya. Se trata del organismo especializado en estadística de Catalunya. Planifica, coordina y normaliza la información estadística con el objeto de preparar y elaborar el programa anual de información estadística de Catalunya. <http://www.idescat.cat/es/>

**OALP.com**, Wiki para el desarrollo de la tecnología OLAP

[http://olap.com/w/index.php/OLAP\\_Education\\_Wiki](http://olap.com/w/index.php/OLAP_Education_Wiki)

**Lucentia**, Grupo de investigación en aplicaciones y soluciones de negocio y en particular Business Intelligence.

<http://www.lucentia.es/>

**Oracle, Data Warehouseing**

<http://www.oracle.com/us/products/database/datawarehousing/overview/index.html>

**Mysql, Data Warehouse**

<http://www.mysql.com/why-mysql/data-warehouse.html>

**SQL Server, Data Warehouse**

<http://www.microsoft.com/es-es/sqlserver/editions/data-warehouse.aspx>

**SAP, Data Warehouse**

<http://www54.sap.com/solutions/tech/data-warehousing.html>

**Microsoft Dynamics, Data Warehouse**

[http://www.microsoft.com/casestudies/Case\\_Study\\_Detail.aspx?casestudyid=390000000081](http://www.microsoft.com/casestudies/Case_Study_Detail.aspx?casestudyid=390000000081)

**Pentaho Business Analytics.** Se refiere al conjunto de aplicaciones de carácter libre que integran tantas soluciones de integración de datos como análisis de negocios en una única plataforma que se caracteriza por la innovación, la agilidad y el ahorro en costes. Sus pilares básicos se basan en la integración en las empresas, grandes volúmenes de datos, integración con el software ya existente e integración de datos.

<http://wiki.pentaho.com/resources/2/pentaho-business-analytics-solution-brief/>

**Schema Workbench.** Se trata de una interface de diseño que permite la creación de cubos OLAP de manera gráfica. El motor de consultas de Modrian procesa peticiones MDX con esquemas ROLAP, que en definitiva son modelos XML creados en una estructura usada por el motor de Mondrian.

<http://mondrian.pentaho.com/documentation/workbench.php>

**Mondrian.** Se trata de un servidor OLAP que permite a los usuarios de negocios analizar grandes cantidades de información en tiempo real. La última versión disponible data del 29-11-2012. Esta desarrollada en java y no requiere de la sintaxis típica del lenguaje SQL

<http://mondrian.pentaho.com/>

## 13.- Relación de Tablas e Imágenes

### 12.1.- Tablas

*Tabla 1. Ventajas e inconvenientes de los Almacenes de Datos. Fuente: Propia y Bibliografía Básica.*

*Tabla 2. Fases y Subfases principales del TFC. Fuente: Propia y Bibliografía Básica.*

*Tabla 3. Plazos definidos para las entregas del proyecto. Fuente: Propia.*

*Tabla 4. Plazos definidos para las entregas del proyecto. Fuente: Propia.*

*Tabla 5. Estructuras del fichero Poblacion.csv. Fuente: Propia.*

*Tabla 6. Estructuras del fichero Equipaments.csv. Fuente: Propia.*

*Tabla 7. Estructuras del fichero EstablimentsNNN.txt. Fuente: Propia.*

*Tabla 8. Tipos de datos para los ficheros proporcionados. Fuente: Propia.*

*Tabla 9. Estructura de la Tabla de Hechos. Fuente: Propia.*

*Tabla 10. Dimensionamiento de la Tabla de Hechos. Fuente: Propia.*

*Tabla 11. Dimensionamiento de la Base de Datos. Fuente: Propia.*

*Tabla 12. Estructura tabla Zonas. Fuente: Propia.*

*Tabla 13. Estructura de indices de la tabla Zonas. Fuente: Propia.*

*Tabla 14. Estructura de la tabla Equipamientos. Fuente: Propia.*

*Tabla 15. Estructura de indices de la tabla Equipamientos. Fuente: Propia.*

*Tabla 16. Estructura de la tabla Establecimientos. Fuente: Propia.*

*Tabla 17. Estructura de indices de la tabla Establecimientos. Fuente: Propia.*

*Tabla 18. Estructura de la tabla de Hechos: Ofertas. Fuente: Propia.*

*Tabla 19. Estructura de la tabla de Hechos: Ofertas. Fuente: Propia.*

### 12.2.- Imágenes

*Imagen 1. Arquitectura general de un almacén de datos. Fuente: Propia y Bibliografía Básica.*

*Imagen 2. Cronograma de cada una de las actividades. Fuente: Propia.*

*Imagen 3. Orígenes de las distintas fuentes de datos. Fuente: Propia y Bibliografía Básica.*

*Imagen 4. Casos de Uso del modelo. Fuente: Propia y Referencias.*

*Imagen 5. Dimensión Fecha. Fuente: Propia.*

*Imagen 6. Dimensión Zona. Fuente: Propia.*

*Imagen 7. Dimensión Equipamiento. Fuente: Propia.*

*Imagen 8. Dimensión Establecimiento. Fuente: Propia.*

*Imagen 9. Dimensiones del modelo. Fuente: Propia.*

*Imagen 10. Esquema conceptual. Fuente: Propia.*

*Imagen 11. Esquema lógico. Fuente: Propia.*

*Imagen 12. Tabla Zonas. De tipo dimensional. Fuente: Propia.*

*Imagen 13. Tabla Equipamientos. De tipo dimensional. Fuente: Propia.*

*Imagen 14. Tabla Establecimientos. De tipo dimensional. Fuente: Propia.*

*Imagen 15. Tabla Ofertas. De tipo Hecho. Fuente: Propia.*

*Imagen 16. Esquema físico. Fuente: Propia.*

*Imagen 17. Arquitectura sistema de desarrollo. Fuente: Propia.*

- Imagen 18. Arquitectura sistema en producción. Fuente: Propia.
- Imagen 19. Estructura de la carpeta C:\Trabajos\Scripts. Fuente: Propia.
- Imagen 20. Estructura del script de creación de la BBDD. Fuente: Propia.
- Imagen 21. Esquema del modelo para el repositorio (I). Fuente: Propia.
- Imagen 22. Esquema del modelo para el repositorio (II). Fuente: Propia.
- Imagen 23. Acceso a Pentaho Analysis Services. Fuente: Propia.
- Imagen 24. Acceso a Pentaho Analysis Service. Fuente: Propia.
- Imagen 25. Transformación Importar Población. Fuente: Propia.
- Imagen 26. Transformación Importar Fechas. Fuente: Propia.
- Imagen 27. Transformación Importar Equipamientos. Fuente: Propia.
- Imagen 28. Transformación Importar Establecimientos 2006. Fuente: Propia.
- Imagen 29. Transformación Importar Establecimientos 2007 a 2011. Fuente: Propia.
- Imagen 30. Transformación Importar Establecimientos 22. Fuente: Propia.
- Imagen 31. Creación de tablas, índices y carga inicial de datos de población, equipamientos, fechas y establecimientos. Fuente: Propia.
- Imagen 32. Actualización de Comarcas. Fuente: Propia.
- Imagen 33. Actualización de Provincias en IN\_Poblacion. Fuente: Propia.
- Imagen 34. Actualización de Población por Sexos. Fuente: Propia.
- Imagen 35. Actualización de datos. Fuente: Propia.
- Imagen 36. Rellenar Equipamientos. Fuente: Propia.
- Imagen 37. Rellenar Zonas. Fuente: Propia.
- Imagen 38. Asignación de información a las tablas dimensionales. Fuente: Propia.
- Imagen 39. Agrupación por Zonas y Equipamientos y . Fuente: Propia.
- Imagen 40. Agrupación por Fechas y Zonas . Fuente: Propia.
- Imagen 41. Agrupación por Fechas y Zonas. Fuente: Propia.
- Imagen 42. Agrupación por Fechas y Establecimientos – Camping, Hotel y Turismo Rural-. Fuente: Propia.
- Imagen 43. Uniones y Ordenaciones para generar datos de la tabla de Hechos. Fuente: Propia.
- Imagen 44. Evolución Establecimientos por Anos. Fuente: Propia.
- Imagen 45. Trafico. Evolución Establecimientos por Anos (Solo datos representativos). Fuente: Propia.
- Imagen 46. Evolución Plazas por Anos. Fuente: Propia.
- Imagen 47. Trafico. Evolución Plazas por Anos. Fuente: Propia.
- Imagen 48. %Plazas respecto a Población por Años. Fuente: Propia.
- Imagen 49. Gráfico. %Plazas respecto a Población por Anos. Fuente: Propia.
- Imagen 50. Oferta Media de Plazas por Anos. Fuente: Propia.
- Imagen 51. Gráfico. Oferta Media de Plazas por Anos por Anos (Solo datos representativos). Fuente: Propia.
- Imagen 52. Numero de Establecimientos / Numero de Equipamientos por Anos. Fuente: Propia.
- Imagen 53. Gráfico. Numero de Establecimientos / Numero de Equipamientos por Anos. Fuente: Propia.
- Imagen 54. Población por Equipamiento, filtrado por Comunidad Autónoma y Equipamiento tipo Universitat por Anos. Fuente: Propia.
- Imagen 55. Población por Equipamiento, filtrado por Provincias y Equipamiento tipo Salut de 2008 a 2012. Fuente: Propia.
- Imagen 56. Indicador de Establecimientos vs Habitantes por Genero por Anos. Fuente: Propia.
- Imagen 57. Indicador de Establecimientos vs Habitantes por Genero por Provincias para el Año 2012. Fuente: Propia.
- Imagen 58. Indicador de Establecimientos vs Habitantes por Genero por Provincias entra 2008 y 2012. Fuente: Propia.
- Imagen 59. Indicador de Establecimientos vs Habitantes por Genero por CCAA por Anos (Solo datos representativos). Fuente: Propia.
- Imagen 60. Indicador de Equipamientos vs Población por CCAA para el equipamiento tipo Turisme. Fuente: Propia.
- Imagen 61. Indicador de Equipamientos vs Población por CCAA para el equipamiento tipo Turisme. Fuente: Propia.
- Imagen 62. Plazas Ofertadas / Superficie por Provincias para el Año 2012. Fuente: Propia.
- Imagen 63. Plazas Ofertadas / Superficie por CCAA para el Año 2012. Fuente: Propia.
- Imagen 64. Informe sobre Establecimientos de Turismo Rural en Barcelona en el Año 2006 (I). Fuente: Propia.
- Imagen 65. Informe sobre Establecimientos de Turismo Rural en Barcelona en el Año 2006 (II). Fuente: Propia.
- Imagen 66. Informe sobre Establecimientos de Turismo Rural en Barcelona en el Año 2006 (III). Fuente: Propia.



**Anexo A.- Script de creación de la Base de Datos**

**Anexo B.- Instalación y ejecución de Scripts en la máquina virtual**

**Anexo C. - Estructura de carpetas definidas en la máquina virtual**