



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

E. Pitton

ETIS

Consultor: Carles Llorach Rius

17 junio 2013

RESUMEN

Este documento corresponde a la memoria de Proyecto de Fin de Carrera de los estudios de Ingeniería Informática de Sistemas de la Universitat Oberta de Catalunya. El proyecto sigue el formato de un encargo de una solución de almacén de datos por parte de un cliente hipotético llamado ONdO que necesita poder analizar datos de establecimientos turísticos, población y equipamientos del territorio Catalán.

La primera parte de la memoria se centra en los aspectos de organizativos y de gestión del proyecto, detallando los recursos, la planificación de tareas, el plan de comunicación y la gestión de riesgos, así como presenta el enfoque y la metodología seguidos, basados en los de Ralph Kimball.

La segunda corresponde al desarrollo en sí de la solución, con sus etapas de análisis, diseño e implementación. El análisis basado en los requerimientos del cliente y los datos de partida conducen a un diseño en esquema de estrella que se plasma en un modelo conceptual, lógico y físico.

En esta fase se expone la estructura de la solución identificando grandes módulos y sus relaciones para entrar luego en el diseño detallado de los mismos. Junto con la arquitectura técnica de la solución se presenta y justifica la elección de la plataforma Microsoft SQL Server 2012 utilizada en el proyecto.

Asimismo se presenta el diseño general del sistema de ETL, su arquitectura modular y el funcionamiento de los procesos de extracción, transformación y carga que lo componen.

Finalmente se presenta y justifica el entorno de Acceso a datos elegido, el modelo analítico con las medidas analíticas creadas y todos los informes desarrollados sobre el mismo para cumplir con los requerimientos de partida.

PALABRAS CLAVE

Gestión de proyectos, Almacén de datos, BI, Business Intelligence, Modelado Dimensional, ETL, Microsoft SQL Server 2012, PowerPivot, DAX

ÍNDICE DE CONTENIDOS

1	INTRODUCCIÓN	8
1.1	Introducción y objetivos.....	8
1.2	Justificación y contexto del proyecto	8
1.3	Enfoque y método seguido	9
1.4	Organización y planificación del proyecto.....	10
1.4.1	Calendario laboral y recursos	10
1.4.2	Relación de actividades y cronograma	10
1.4.3	Fases e hitos destacables.....	11
1.4.4	Diagrama de Gantt.....	11
1.4.5	Organización del proyecto.....	12
1.4.6	Plan de comunicación.....	12
1.4.7	Gestión de riesgos	13
1.4.8	Planificación inicial vs final	13
1.5	Productos obtenidos	14
1.6	Breve descripción otros capítulos.....	14
2	ANÁLISIS.....	15
2.1	Análisis de Requerimientos Funcionales	15
2.2	Casos de uso	15
2.3	Requerimientos no funcionales.....	16
2.4	Análisis datos de partida	17
2.4.1	Fuente Datos Equipamientos	18
2.4.2	Fuente Datos Población.....	19
2.4.3	Fuente Datos Establecimientos	20
2.4.4	Fuente Datos Municipios.....	22
3	DISEÑO.....	23
3.1	Arquitectura general	23
3.2	Diseño de la arquitectura del Software.....	23
3.3	Diagrama de la arquitectura del Hardware	24
3.4	Selección de productos	24
3.5	Selección de productos de la solución	25
3.6	Selección de productos entorno de desarrollo.....	25
3.7	Modelo conceptual	26

3.7.1	Tablas de hechos y granularidad	26
3.7.2	Dimensiones de análisis	26
3.7.3	Descripción y Atributos de cada Dimensión	27
3.7.4	Jerarquías de Agregación.....	27
3.7.5	Medidas.....	28
3.7.6	Restricciones de integridad	28
3.7.7	Viabilidad.....	28
3.8	Diagrama E-R del diseño	29
3.9	Diseño Lógico.....	30
3.10	Diseño Físico	31
3.11	Diseño procesos ETL.....	32
3.11.1	Diseño zona o área de staging	32
3.11.1	Diseño Registro.....	33
3.11.2	Proceso principal	34
3.11.3	Fase Extracción	35
3.11.4	Fase Transformación	36
3.12	Diseño Informes	39
3.12.1	Sobre la herramienta.....	39
3.12.2	Modelo Analítico Acceso a Datos	40
3.12.1	Medidas creadas	40
4	INFORMES.....	43
4.1	Informe de indicadores principales – usuarios analistas.....	43
4.2	Informe equipamientos y servicios – usuarios analistas	44
4.3	Informe establecimientos – usuarios analistas	45
4.4	Informe población – usuarios analistas.....	46
4.5	Informe indicadores principales	47
4.6	Informe Plazas turísticas por tipo	48
4.7	Informe comparativa equipamientos por comarca	49
4.8	Informe comparativa de población por comarca	50
4.9	Informe evolución comparativa de población por comarca	51
4.10	Informe comparativa número y tipo de establecimientos por comarca	52
4.11	Informe evolución comparativa número y tipo de establecimientos por comarca	53
4.12	Informe comparativa plazas y equipamientos relativos a la población por comarca	54
4.13	Informe evolución comparativa plazas y equipamientos relativos a la población	55
4.14	Informe geoespacial Población y Equipamientos por tipo	56

4.15	Informe geoespacial de Plazas Turísticas y Población en valores absolutos	57
4.16	Informe geoespacial de plazas turísticas y equipamientos relativos a la población	58
5	Líneas de evolución futura	59
6	Conclusiones	59
7	Bibliografía	60

INDICE DE FIGURAS

Ilustración 1 Bill Inmon, padre de los almacenes de datos	9
Ilustración 2 Ralph Kimball, pionero de los almacenes de datos	9
Ilustración 3 Metodología de Desarrollo de Kimball	9
Ilustración 4 Diagrama de Gantt	11
Ilustración 5 Organigrama Proyecto	12
Ilustración 6 Diagramas Casos de Uso	16
Ilustración 7 Diagrama de la Arquitectura del Software	23
Ilustración 8 Arquitectura del Hardware	24
Ilustración 9 Productos del stack Microsoft Business Intelligence	25
Ilustración 10 Diagrama Entidad Relación	29
Ilustración 11 Diagrama Diseño Lógico	30
Ilustración 12 Diagrama Diseño Físico	31
Ilustración 13 Tablas Zona Staging	32
Ilustración 14 Rutina Escritura Log	33
Ilustración 15 Registro log ejecución	33
Ilustración 16 Proceso ETL de control	34
Ilustración 17 Ejemplo Proceso Extracción Equipamientos – Flujo Control	35
Ilustración 18 Ejemplo Proceso Extracción Equipamientos – Flujo Datos	35
Ilustración 19 Ejemplo Proceso Transformación Equipamientos	36
Ilustración 20 Ejemplo fase carga- procesado medidas equipamientos	37
Ilustración 21 Ejemplo proceso carga – Medidas Equipamientos	37
Ilustración 22 Ejemplo fase carga- carga dimensión TipoCategoría	38
Ilustración 23 Ejemplo fase carga- carga dimensión Año	38
Ilustración 24 Ejemplo fase carga- carga dimensión Municipio	38
Ilustración 25 Modelo analítico de acceso a datos	40

INDICE DE TABLAS

Tabla 1 Relación de actividades y Cronograma	10
Tabla 2 Fases e Hitos requeridos.....	11
Tabla 3 Plan de Comunicación	12
Tabla 4 Gestión de Riesgos.....	13
Tabla 5 Entregables del Proyecto	14
Tabla 6 Perfiles de usuario.....	15
Tabla 7 Requerimientos no funcionales.....	16
Tabla 8 Origen de datos Equipamientos	18
Tabla 9 Análisis origen de datos Equipamientos	18
Tabla 10 Origen de datos Población.....	19
Tabla 11 Análisis Origen de datos Población	19
Tabla 12 Origen de datos Establecimientos.....	20
Tabla 13 Análisis Origen de datos Establecimientos	21
Tabla 14 Origen de datos Municipios	22
Tabla 15 Análisis Origen de datos Municipios	22
Tabla 16 Selección de productos software de plataforma.....	25
Tabla 17 Selección de productos software de desarrollo	25
Tabla 18 Tablas de hechos y granularidades	26
Tabla 19 Tablas de Hechos y Dimensiones	26
Tabla 20 Dimensiones y Atributos.....	27
Tabla 21 Tablas de hechos y Medidas.....	28
Tabla 22 Estimación cotas superiores volumen de datos requerido.....	29

1 INTRODUCCIÓN

1.1 Introducción y objetivos

Este documento corresponde a la memoria de Proyecto de Fin de Carrera de los estudios de Ingeniería Informática de Sistemas de la Universitat Oberta de Catalunya. El proyecto consiste en el desarrollo de un caso práctico con el fin de aplicar y evidenciar los conocimientos y las competencias adquiridas a lo largo de los estudios, así como adquirir los necesarios en el ámbito de los almacenes de datos. Sigue el formato de un encargo de una solución de almacén de datos por parte de un cliente hipotético.

El desarrollo del mismo ha servido para ahondar en la disciplina de gestión de proyectos informáticos y como introducción en desarrollo de soluciones de un ámbito en pleno auge en el sector de las TIC como es el de la *Business Intelligence*.

1.2 Justificación y contexto del proyecto

El Observatorio Nacional de Ocupación (de ahora en adelante ONdO) es una entidad adscrita a la Consejería de Economía y Finanzas de la Generalitat de Catalunya, encargada de la fijación de los criterios, la regulación de las condiciones y la ejecución y el control de las líneas públicas de ayuda y promoción al turismo.

Debido al gran peso histórico del sector turístico en la economía Catalana, y al retroceso en otros sectores por la actual crisis de la zona euro, ONdO tiene mayor necesidad que nunca tanto de conseguir la máxima eficacia y eficiencia en las políticas que implementa, como de poder justificar las mismas ante el creciente escrutinio popular.

Dado que ciertas posiciones muy influyentes en Dirección de ONdO provienen de la industria de las Tecnologías de la Información, la Dirección está muy sensibilizada con el papel positivo que éstas pueden jugar en la consecución de sus objetivos. En particular es conocedora del poder transformador que las herramientas de *Business Intelligence* pueden tener en su proceso de toma de decisiones y en el desarrollo de su actividad, así como del nivel de transparencia y trazabilidad que pueden ayudar a proyectar hacia el exterior.

Convencida de la idoneidad de poner en marcha una iniciativa de estas características pero consciente de la dificultad en la obtención de financiación para grandes proyectos y de los riesgos inherentes a los mismos, ONdO ha optado por desarrollarla como un Programa, o una sucesión de proyectos interrelacionados menores en alcance pero de autorización y ejecución más ágil, con casos de uso muy concretos y con un aporte de valor a corto plazo que permitan justificar y acometer el siguiente con garantías.

Este trabajo consiste en el primero de esos proyectos.

1.3 Enfoque y método seguido

El proyecto se ha realizado en base a la aproximación al desarrollo de almacenes de datos abogada por Ralph Kimball consistente en el desarrollo de almacenes de datos departamentales más o menos independientes especializados en procesos específicos de la organización y formando una colección a partir de la cual los datos fluyen hasta el almacén de datos corporativo.

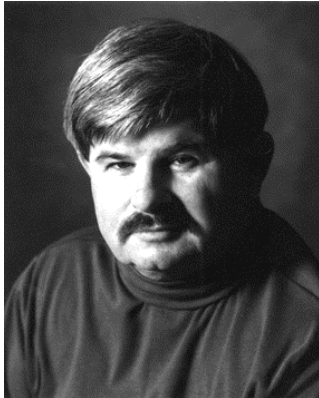


Ilustración 1 Bill Inmon, padre de los almacenes de datos

Se trata la aproximación alternativa a la arriba-abajo formulada por Bill Inmon – conocido como padre de los Almacenes de Datos- que concibe el almacén de datos en base a un modelo unificado de datos que abarca todos los procesos de la organización y como repositorio central de datos desde donde fluye la información a los almacenes departamentales.



Ilustración 2 Ralph Kimball, pionero de los almacenes de datos

Se escogió la primera alternativa por considerarse mucho más acorde a las características de este proyecto en particular y a la realidad de las dinámicas y los condicionantes que rigen en la mayoría de organizaciones en la actualidad.

La metodología usada como base es la correspondiente al Ciclo de Vida en Cascada de Desarrollo, con las fases clásicas de Análisis, Diseño e Implementación pero aplicando los conceptos específicos de desarrollo de almacenes de datos de la Metodología de Kimball

En esta se identifica en el flujo de desarrollo la secuenciación de tareas en 3 dimensiones distintas – Tecnología, Datos y Aplicación- resaltando las actividades que deberían realizarse en paralelo ya que tienen dependencia entre sí.

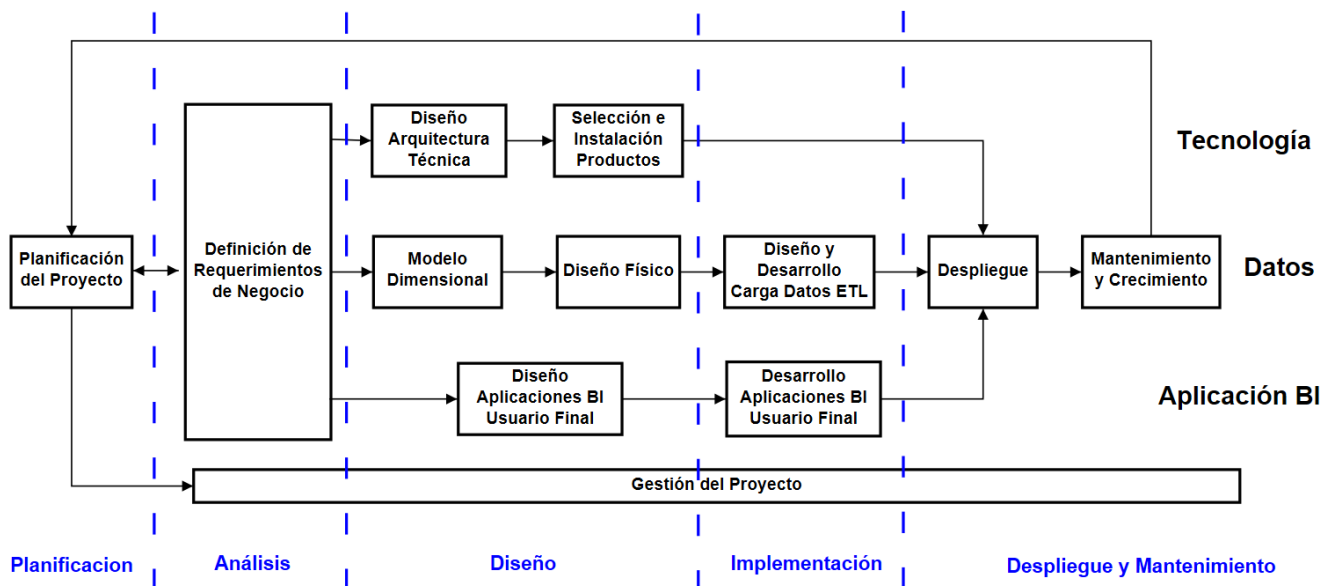


Ilustración 3 Relación Metodología de Desarrollo Clásico y Metodología de Desarrollo de Kimball

1.4 Organización y planificación del proyecto

1.4.1 Calendario laboral y recursos

El proyecto se ha realizado entre los meses de febrero y junio de 2013, con un solo recurso de una disponibilidad media de 20 horas/semana. Aunque no se desglose explícitamente, el recurso ha desempeñado los diferentes roles de consultor, arquitecto, analista programador, probador, ingeniero de sistemas, jefe de proyecto y gerente. El rol comercial fue desempeñado directamente por el consultor y el jefe de proyecto bajo petición expresa de la Dirección de ONDO debido a las malas experiencias acumuladas en proyectos anteriores con los perfiles Ventas clásicos.

1.4.2 Relación de actividades y cronograma

Tarea	Duración (días)	Esfuerzo (horas)	Fecha Inicio	Fecha Fin
Planificación y Organización del Proyecto	5	20	28/02/2013	05/03/2013
Incidio TFC DW	0	0	28/02/2013	28/02/2013
Recopilación Información del TFC	1	4	28/02/2013	01/03/2013
Estudio Material Académico Introductorio Ámbito DW	2	8	01/03/2013	03/03/2013
Revisión Estado del Arte Soluciones y Herramientas DW	1	4	03/03/2013	04/03/2013
Planificación tareas Proyecto	1	4	04/03/2013	05/03/2013
Análisis Preliminar	7	28	05/03/2013	12/03/2013
Análisis de requerimientos cliente	1	4	05/03/2013	06/03/2013
Análisis de datos de origen	2	8	06/03/2013	08/03/2013
Elaboración Propuesta Solución Tecnológica	2	8	08/03/2013	10/03/2013
Imprevistos Análisis	2	8	10/03/2013	12/03/2013
Entrega PAC1 (hito)	0	0	12/03/2013	12/03/2013
Análisis y Diseño	35	140	12/03/2013	16/04/2013
Revisión Material Académico Ámbito DW	8	32	12/03/2013	20/03/2013
Diseño Modelo Conceptual y de Datos	5	20	20/03/2013	25/03/2013
Diseño Procesos Extracción y Carga	7	28	25/03/2013	01/04/2013
Diseño Informes	5	20	01/04/2013	06/04/2013
Diseño Arquitectura Técnica y preparación de Entorno	3	12	06/04/2013	09/04/2013
Imprevistos Diseño	7	28	09/04/2013	16/04/2013
Entrega PAC2 (hito)			16/04/2013	16/04/2013
Implementación	43	172	16/04/2013	29/05/2013
Implementación BBDD	7	28	16/04/2013	23/04/2013
Implementación Procesos ETL	22	88	23/04/2013	15/05/2013
Implementación Informes	9	36	15/05/2013	24/05/2013
Imprevistos	5	20	24/05/2013	29/05/2013
Entrega PAC3 (hito)	0	0	29/05/2013	29/05/2013
Memoria y Presentación Final	19	76	29/05/2013	17/06/2013
Elaboración de la memoria	10	40	29/05/2013	08/06/2013
Elaboración de la presentación virtual	5	20	08/06/2013	13/06/2013
Imprevistos	4	16	13/06/2013	17/06/2013
Entrega memoria (hito)	0	0	17/06/2013	17/06/2013

Tabla 1 Relación de actividades y Cronograma

1.4.3 Fases e hitos destacables

El cliente, sujeto a las rígidas normativas de gestión de proyectos TIC en el sector público ha solicitado dividir el mismo en las siguientes fases:

Fase	Inicio Fase	Fin Fase	Hito
FASE_PAC1 Plan de trabajo y Requerimientos	13/03/2013	12/03/2013	HITO_PAC1
FASE_PAC2 Análisis y Diseño	13/03/2013	16/04/2013	HITO_PAC2
FASE_PAC3 Implementación	17/04/2013	29/05/2013	HITO_PAC3
FASE_PAC4 Entrega Final y Defensa	13/03/2013	17/06/2013	HITO_PAC4

Tabla 2 Fases e Hitos requeridos

1.4.4 Diagrama de Gantt

A modo de documento resumido de planificación del proyecto, se preparó un diagrama de Gantt con fases, hitos, tareas y dependencias.

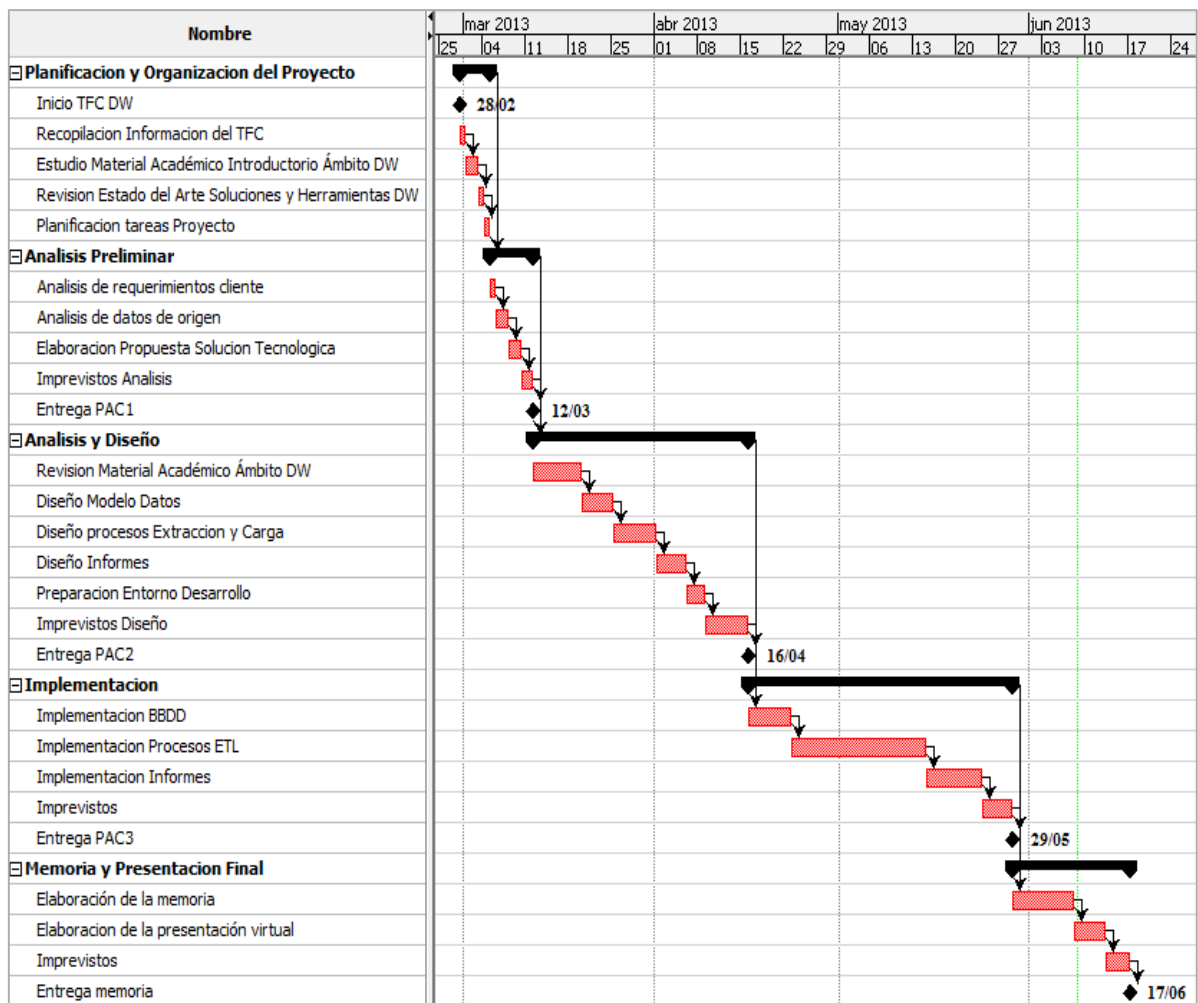


Ilustración 4 Diagrama de Gantt

1.4.5 Organización del proyecto

Se definió el organigrama del proyecto en correspondencia a los gremios de nivel estratégico, táctico y operativo de seguimiento, en los que se participó con los roles de gerente, consultor y jefe de proyecto respectivamente:



Ilustración 5 Organigrama Proyecto

1.4.6 Plan de comunicación

La comunicación es uno de los principales factores para la consecución exitosa de proyectos tecnológicos, y es especialmente crítica en el caso de proyectos para grandes clientes del sector público como es ONDO.

Con el fin de mantener informadas a las partes interesadas de los avances del proyecto y asegurar la intervención de las apropiadas en caso de necesidad, se definió al inicio del proyecto el siguiente plan de comunicación:

ID Actividad	Información	Receptores	Rangos de tiempo	Métodos
1	<i>Kick-Off</i> Proyecto	Dirección ONDO	Inicio	Presentación Ejecutiva
2	Presentación Informes Entregables Parciales	Comité Seguimiento ONDO	Cada Hito definido	Reunión Táctica
3	Estatus Proyecto, riesgos y acciones siguiente semana	Responsable Proyecto ONDO	Semanal	Email
4	Cierre Proyecto	Dirección ONDO	Finalización	Presentación Ejecutiva

Tabla 3 Plan de Comunicación

1.4.7 Gestión de riesgos

Durante la planificación del proyecto se consideraron posibles riesgos para la consecución de los objetivos del mismo. Para cada uno se fijaron los planes de respuesta específicos según probabilidad e impacto que se detallan continuación:

ID	Descripción del riesgo	Probabilidad	Impacto	Plan de respuesta al riesgo	
				Estrategia	Acción de respuesta
R1	Indisponibilidad de recursos intermitente	ALTA	ALTO	Mitigar	Extender jornada laboral a festivos
R2	Indisponibilidad de recursos persistente	MEDIA	ALTO	Aceptar	Replanificar y/o renegociar alcance con ONDO
R3	Fallo Entorno Desarrollo	BAJA	ALTO	Evitar	Desarrollo sobre máquina virtual redundada
R4	Cambio requerimientos Cliente	MEDIA	MEDIO	Transferir	ONDO asume costes y retrasos
R5	Entorno y Plataforma de Desarrollo desconocidos	ALTA	MEDIO	Evitar	Negociar con ONDO cambiar Entorno y Plataforma a Tecnologías ya conocidas

Tabla 4 Gestión de Riesgos

1.4.8 Planificación inicial vs final

Gracias a la experiencia acumulada en proyectos anteriores y al uso de herramientas ya conocidas se pudieron ajustar muy bien las cargas de trabajo previstas inicialmente y no ha habido grandes desviaciones con respecto al esfuerzo real requerido.

Los riesgos R1 y R5 ya anticipados con alta probabilidad se materializaron y el plan que se había articulado para ellos pudo ejecutarse de manera satisfactoria.

Cabe destacar sin embargo que contra todo pronóstico el riesgo R3, a priori de baja probabilidad, se materializó también a pocos días del fin del proyecto en forma de borrado accidental de la máquina virtual y puso a prueba la capacidad de contingencia que se había previsto, con 2 semanas-hombre de trabajo perdidas y llegando a poner en peligro la entrega final.

1.5 Productos obtenidos

A continuación se resumen los diferentes entregables generados durante las sucesivas fases del proyecto de acuerdo con los requerimientos del Cliente.

FASE	Entregable	Descripción
FASE_PAC_1	“Plan de Trabajo de Proyecto y Análisis Preliminar de Requerimientos”	Documento donde se describe la situación de partida, los requerimientos del Cliente y se detalla la planificación completa del proyecto. Integrado dentro de la Memoria.
FASE_PAC_2	“Análisis y Diseño de la Solución”	Documento con el análisis detallado de requisitos y diseños conceptual, técnico y físico de la Solución de Almacén de Datos, incluyendo descripción de los procesos ETL y los elementos asociados. Integrado dentro de la Memoria.
FASE_PAC_3	“Evidencias de Implementación”	Documento donde se describen los detalles de la implementación y muestras de los informes desarrollados. Integrado dentro de la Memoria.
FASE_PAC_4	“Producto”	Máquina Virtual con el producto desarrollado.
FASE_PAC_4	“Memoria”	Memoria del Proyecto.
FASE_PAC_4	“Presentación”	Presentación Ejecutiva Virtual del Proyecto realizado.

Tabla 5 Entregables del Proyecto

1.6 Breve descripción otros capítulos

■ ANALISIS

Se analizan los requerimientos del cliente, se elaboran los casos de uso y se detallan los requerimientos no funcionales de la solución. Adicionalmente se analizan a fondo los datos de partida que tenemos a disposición con el fin de conocer sus características y sus limitaciones.

■ DISEÑO

Se presenta el diseño de alto nivel de la solución identificando grandes módulos y sus relaciones para entrar luego en el diseño detallado de los mismos. Se presenta la arquitectura del software y del hardware de la solución, así como se justifica la selección de productos elegidos.

En este capítulo también se presenta el modelo conceptual diseñado, que satisface los requerimientos analizados y lo condicionantes de los datos de partida.

Para finalizar se explica el diseño general del sistema de ETL, su arquitectura modular y el funcionamiento de los procesos de extracción, transformación y carga que lo componen

■ INFORMES

En este último capítulo se introduce y justifica el entorno de Acceso a datos elegido y se presenta el modelo analítico con las medidas analíticas creadas y todos los informes desarrollados sobre el mismo.

2 ANÁLISIS

2.1 Análisis de requerimientos funcionales

Entender los factores clave que fundamentan el proceso de negocio del Cliente es crítico para obtener una comprensión sólida de sus requerimientos, y crucial para la traslación de estos requerimientos de negocio en consideraciones de diseño.

Para ello se desarrollaron una serie de reuniones y entrevistas en las dependencias de ONdO con los principales departamentos y contactos suministrados por el Comité de Seguimiento, y en donde se pone de manifiesto que la información mínima que necesitan en el almacén de datos para realizar su misión es la relativa los siguientes indicadores:

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

Esta información se debe poder consultar con una temporalidad a nivel de año y agregación por comarca, provincia, tipo de establecimiento y categoría, debiendo existir también informes predefinidos que la presenten.

2.2 Casos de uso

Se han identificado los siguientes casos de uso de la solución:

Perfil Usuario	Descripción Caso uso
Básico	Consume informes interactivos o estáticos existentes.
Analista	Consume informes interactivos o estáticos existentes, analiza datos a través de consultas <i>ad-hoc</i> y crea nuevos informes.
Administrador	Gestiona accesos de usuarios al sistema, ejecuta los procesos de ETL y consulta informes ejecución ETL.

Tabla 6 Perfiles de usuario

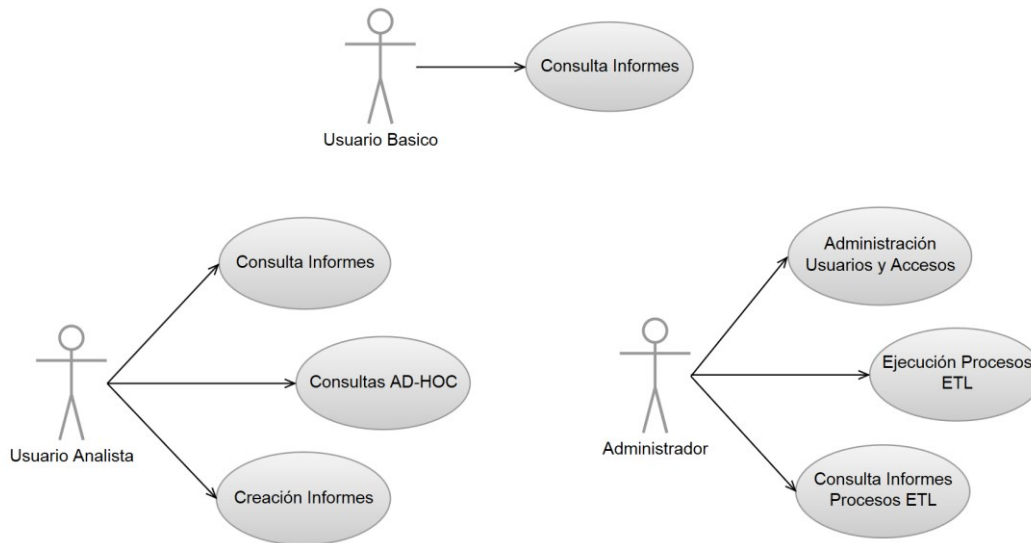


Ilustración 6 Diagramas Casos de Uso

2.3 Requerimientos no funcionales

Durante la ronda de entrevistas se recogieron toda una serie de requisitos que no son parte de las funciones a realizar por la aplicación pero sí son atributos de calidad que se usaran para juzgar la operación de la misma. Se acordó además de antemano la importancia relativa de cada uno para ONdO de manera que se pudiese priorizar adecuadamente en caso de tener que elegir entre alternativas de cualidades contrapuestas.

Requerimiento	Descripción	Ponderación
Rendimiento	Se espera que la solución permita acceder a la información y obtener respuestas inmediatamente o con un tiempo de espera muy corto.	9
Disponibilidad	La solución tiene que alcanzar la disponibilidad estándar de la Generalitat para sistemas no críticos. Así mismo debe permitir esquemas de Alta Disponibilidad en caso de cambio en la clasificación de criticidad en el futuro.	10
Seguridad	La solución debe permitir eventuales restricciones de seguridad según perfiles a nivel de atributo. Se apreciarán facilidades de restricción de seguridad a nivel de registro.	6
Accesibilidad	La solución debe permitir operar hasta cierto punto en condiciones de escasa o nula conectividad.	10
Usabilidad	Se desea una solución que sea ágil de usar y que no requiera más pasos de los necesarios para realizar una acción.	9
Estabilidad	La solución debe alcanzar la consistencia y fiabilidad requeridas para Aplicaciones de Productividad de Generalitat.	5
Coste y Mantenimiento	La solución debe tener un Coste Total de Propiedad competitivo con otras soluciones comparables en la Administración.	9
Interoperabilidad	La solución debe basarse única y exclusivamente en herramientas del Libro de Estándares TI de la Administración Pública	10
Escalabilidad	La solución tiene que permitir crecer en escala sin requerir replantear la arquitectura.	8
Concurrencia	Se requiere el acceso concurrente de hasta 100 usuarios, y permitir crecer en escala sin requerir replantear la arquitectura.	9
Interfaz	La interfaz debe ser familiar o de muy fácil aprendizaje para los usuarios tipo de ONdO	10

Tabla 7 Requerimientos no funcionales

2.4 Análisis datos de partida

El tercer y último punto tratado con los usuarios clave de ONdO fue el de las 3 fuentes de datos de partida puestas a disposición para el desarrollo de la solución.

- **EQUIPAMIENTOS**
- **POBLACION**
- **ESTABLECIMIENTOS**

Durante propias las entrevistas sin embargo ONdO expresó su deseo de no supeditar los objetivos del proyecto a las limitaciones de las mismas, y su disposición a la integración de otras fuentes que permitieran aumentar las capacidades de análisis y por tanto el valor aportado por la solución.

En esa línea, se realizó un análisis preliminar para detectar y valorar cuanto antes estas posibles limitaciones y presentar alternativas al Comité de Seguimiento, y cuyo resultado final se resume en la siguiente tabla:

ID	Origen	Descripción Limitación	Decisión
L1	POBLACIONES	No incluye las poblaciones de todos los municipios, sino sólo las de aquellos con población superior a 1.000 habitantes	Por la alta variabilidad de los datos censales de municipios pequeños ONdO confirma la idoneidad de usar únicamente los suministrados para los escenarios de análisis previstos.
L2	POBLACIONES	Sólo incluye el desglose por sexos para el año 2012.	En base al carácter aproximativo de los datos de población manifestado en L1 se acuerda también registrar datos sin distinción de sexo y utilizar el ratio medio mujeres-hombres 2006-20012 de IDESCAT.
L3	ESTABLECIMIENTOS	Contiene información a nivel de comarca, cuando en los otros está a nivel de municipio.	La solución debe encontrar la forma de resolver los 2 niveles de granularidad conservando el detalle a nivel de municipios donde se disponga.
L4	EQUIPAMIENTOS	Sólo contiene información de 2012 y carece de otro tipo de información suficiente para poder historificar.	Se acuerda no disponer de historificación de equipamientos pero sí diseñar la solución con la previsión de incorporar esta capacidad en el futuro.
LI	TODOS	No se ha suministrado información que relacione municipios a comarcas o a otras unidades de agregación territorial.	Se acepta la inclusión de una cuarta fuente de datos, MUNICIPIOS con información territorial proveniente de la entidad MUNICAT .

Una vez cerrado el conjunto de orígenes de datos de partida se efectuó un análisis detallado con el fin de conocer a fondo los mismos

2.4.1 Fuente Datos Equipamientos

Propietario: IDESCAT

Nombre: EQUIPAMIENTOS

Descripción: Relación de equipamientos públicos a fecha 31/12/2012.

Formato: Fichero Único

Nombre	Codificación	Rango	Calidad de datos
Equipaments.csv	Fichero texto codificado UTF-8 separado por comas	~ 31.000 registros	707 registros duplicados completos

Tabla 8 Origen de datos Equipamientos

Campos	Semántica	Tipo datos	Rango	Calidad de datos
nom	Nombre	texto	< 150 caracteres	Uso inconsistente mayúsculas y minúsculas
adreja	Dirección donde se ubica	texto	< 150 caracteres	Vía y numero separado por carácter ",". Si es un punto de interés la vía es "-"
municipi	Municipio	texto	< 50 caracteres	Consistente a nivel columna, con articulo al final separado por espacio
cp	Código postal	texto	< 50 caracteres	"-" si no tiene
comarca	Comarca	texto	< 50 caracteres	Consistente a nivel columna, con articulo al final separado por espacio
telefon	Teléfono	texto	< 50 caracteres	9, 3 o 0 dígitos numéricos separados inconsistentemente por carácter " "
longitud	Longitud donde se localiza	decimal	entre 0 y 4 y 15 decimales	Valores inconsistentes a nivel fila, valores mal codificados
latitud	Latitud donde se localiza	decimal	entre 40 y 43 con 15 decimales	Valores inconsistentes a nivel fila, valores mal codificados
categories	categorización del equipamiento	texto	< 500 caracteres	Categorías codificadas en el mismo campo con el carácter separador " "
location	Coordenadas	texto	< 500 caracteres	Codificación Sistema GIS. Redundante.

Tabla 9 Análisis origen de datos Equipamientos

2.4.2 Fuente Datos Población

Propietario: IDESCAT

Nombre: POBLACIÓN

Descripción: Relación con los habitantes por año y la superficie de cada municipio.

Formato: Fichero Único

Nombre	Tipo datos	Rango	Calidad de datos
Poblacio.csv	Fichero texto codificado UTF-8 separado por “;”	~ 450 registros	No incluye todos los municipios sino las de más de 1.000 habitantes. Hay registros que no corresponden a son municipios sino agregados.

Tabla 10 Origen de datos Población

Campo	Semántica	Tipo datos	Rango	Calidad de datos
municipi	Municipio	texto	< 100 caracteres	Consistente a nivel columna, con articulo al final entre caracteres "(" y ")"
Codi INE	Código INE	texto	< 100 caracteres	Consistente a nivel columna
Població 2012	Población 2012	entero	<10.000.0000	Carácter "." como separador de miles. Valor especial "n.d." si la cifra no está disponible.
Població 2011	Población 2011	entero	<10.000.0000	Carácter "." como separador de miles. Valor especial "n.d." si la cifra no está disponible.
Població 2010	Población 2010	entero	<10.000.0000	Carácter "." como separador de miles. Valor especial "n.d." si la cifra no está disponible.
Població 2009	Población 2009	entero	<10.000.0000	Carácter "." como separador de miles. Valor especial "n.d." si la cifra no está disponible.
Població 2008	Población 2008	entero	<10.000.0000	Carácter "." como separador de miles. Valor especial "n.d." si la cifra no está disponible.
Població 2007	Población 2007	entero	<10.000.0000	Carácter "." como separador de miles. Valor especial "n.d." si la cifra no está disponible.
Població 2006	Población 2006	entero	<10.000.0000	Carácter "." como separador de miles. Valor especial "n.d." si la cifra no está disponible.
Població 2012 homes	Población 2012 Hombres	entero	<10.000.0000	Carácter "." como separador de miles. Valor especial "n.d." si la cifra no está disponible.
Població 2012 dones	Población 2012 Mujeres	entero	<10.000.0000	Carácter "." como separador de miles. Valor especial "n.d." si la cifra no está disponible.
Extensió (km2)	Extensión Territorio	entero	<10.000.0000	Carácter "." como separador de miles. Valor especial "n.d." si la cifra no está disponible.

Tabla 11 Análisis Origen de datos Población

2.4.3 Fuente Datos Establecimientos

Propietario: FCAT

Nombre: ESTABLECIMIENTOS

Descripción: Relación con número de establecimientos, tipología, plazas ofrecidas y clasificación por comarcas.

Formato: Fichero por año

Nombre	Tipo datos	Rango	Calidad de datos
Establiments 20XX.txt	Fichero texto codificado UTF-8 separado por tabulador	~ 100 registros	Inconsistencia en el formato de los mismos. Se expusieron posibles alternativas en el proyecto al Comité de Seguimiento pero ONdO optó por exigir a la FCAT la entrega de los ficheros en un único formato estable.

Tabla 12 Origen de datos Establecimientos

Campo	Semántica	Tipo datos	Rango	Calidad de datos
comarcaambitprovincia	Unidad de agregación comarca, ámbito o provincia	texto	< 100 caracteres	carácter "." separador miles
establiments hotelers	establecimientos tipo hotel	entero	<100.000	carácter "." separador miles
campings	establecimientos tipo camping	entero	<100.000	carácter "." separador miles
turisme rural	establecimientos tipo casa rural	entero	<100.000	carácter "." separador miles
total establiments	establecimientos	entero	<100.000	carácter "." separador miles
places establiments hotelers	plazas tipo hotel	entero	<100.000	carácter "." separador miles
places campings	plazas tipo camping	entero	<100.000	carácter "." separador miles
places turisme rural	plazas tipo casa rural	entero	<100.000	carácter "." separador miles
places total	plazas	entero	<100.000	carácter "." separador miles
establiments estrelles d'or	establecimientos tipo hotel categoría estrellas oro	entero	<100.000	carácter "." separador miles
establiments estrelles d'argent	establecimientos tipo hotel categoría estrellas plata	entero	<100.000	carácter "." separador miles
total estrelles	establecimientos total categoría	entero	<100.000	carácter "." separador miles
places estrelles d'or	plazas tipo hotel categoría estrellas oro	entero	<100.000	carácter "." separador miles
places estrelles d'argent	plazas tipo hotel categoría estrellas plata	entero	<100.000	carácter "." separador miles
places total estrelles	plazas tipo hotel categoría	entero	<100.000	carácter "." separador miles

campings Lujo	establecimientos tipo camping categoría Lujo	entero	<100.000	carácter "." separador miles
campings 1a	establecimientos tipo camping categoría 1a	entero	<100.000	carácter "." separador miles
campings 2a	establecimientos tipo camping categoría 2a	entero	<100.000	carácter "." separador miles
campings 3a	establecimientos tipo camping categoría 3a	entero	<100.000	carácter "." separador miles
campings total	establecimientos tipo camping categoría	entero	<100.000	carácter "." separador miles
places campings Lujo	plazas tipo camping categoría Lujo	entero	<100.000	carácter "." separador miles
places campings 1a	plazas tipo camping categoría 1a	entero	<100.000	carácter "." separador miles
places campings 2a	plazas tipo camping categoría 2a	entero	<100.000	carácter "." separador miles
places campings 3a	plazas tipo camping categoría 3a	entero	<100.000	carácter "." separador miles
places campings categoría total	plazas tipo camping categoría	entero	<100.000	carácter "." separador miles
casa pueble compartida	establecimientos tipo casa rural categoría casa pueble compartida	entero	<100.000	carácter "." separador miles
casa pueble independent	establecimientos tipo casa rural casa pueble independiente	entero	<100.000	carácter "." separador miles
masia	establecimientos tipo casa rural categoría masia	entero	<100.000	carácter "." separador miles
masoveria	establecimientos tipo casa rural categoría masoveria	entero	<100.000	carácter "." separador miles
total turisme rural	establecimientos tipo casa rural categoría total	entero	<100.000	carácter "." separador miles
places casa pueble compartida	plazas tipo casa rural categoría casa pueble compartida	entero	<100.000	carácter "." separador miles
places casa pueble independent	plazas tipo casa rural categoría casa pueble independiente	entero	<100.000	carácter "." separador miles
places masia	plazas tipo casa rural categoría masia	entero	<100.000	carácter "." separador miles
places masoveria	plazas tipo casa rural categoría masoveria	entero	<100.000	carácter "." separador miles
places total	plazas tipo casa rural categoría total	entero	<100.000	carácter "." separador miles

Tabla 13 Análisis Origen de datos Establecimientos

2.4.4 Fuente Datos Municipios

Propietario: MUNICAT

Nombre: MUNICIPIOS

Descripción: Relación con los municipios, las comarcas, las zonas territoriales y las provincias de Catalunya.

Formato: Fichero Único

Nombre	Tipo datos	Rango	Calidad de datos
municipios.csv	Fichero texto codificado 1252-ANSI separado por “;”	~ 100 registros	Corresponde a un maestro de datos

Tabla 14 Origen de datos Municipios

Campo	Semántica	Tipo datos	Rango	Calidad de datos
Codi_10	Código territorial	texto	50	Consistente
Tipus_ens	Tipo ente	texto	50	Consistente
Delegacio_govern	Delegación Territorial	texto	50	Nombre corto, sin artículos, consistente
Comarca	Comarca	texto	50	Nombre corto, sin artículos, consistente
Provincia	Provincia	texto	50	Nombre corto, sin artículos, consistente
Nom_curt	Municipio	texto	50	Nombre corto, sin artículos, consistente
Capital	Municipio es capital	booleano	1	Consistente

Tabla 15 Análisis Origen de datos Municipios

3 DISEÑO

Una vez que la fase de análisis describió el problema, pasamos a presentar la fase de Diseño. En primer lugar se presenta un diseño de alto nivel o arquitectónico que define la estructura de la solución identificando grandes módulos y sus relaciones para entrar luego en el diseño detallado de los mismos.

3.1 Arquitectura general

La solución consta de los componentes comunes de una solución de almacén de datos. En una primera fase de extracción se leen e interpretan los datos de origen y se copian los datos necesarios para el almacén de datos en el Área de Staging.

Una vez allí se procede a aplicar transformaciones de limpieza de datos como resolver conflictos de dominio, tratar elementos que faltan y cambiar formatos, y de estructura de datos como conformar y desnormalizar datos. Esta fase también se encarga de asegurar la calidad de los datos, comprobando las condiciones de error e inconsistencias anticipadas durante el análisis.

Estas transformaciones son las precursoras de cargar los datos en la zona de presentación del almacén de datos en forma de tablas con modeladas dimensionalmente con esquemas en estrella.

Finalmente el acceso al área de presentación se realiza a través de diferentes herramientas de acceso o reportes configurados y desarrollados para los clientes finales.

3.2 Diseño de la arquitectura del Software

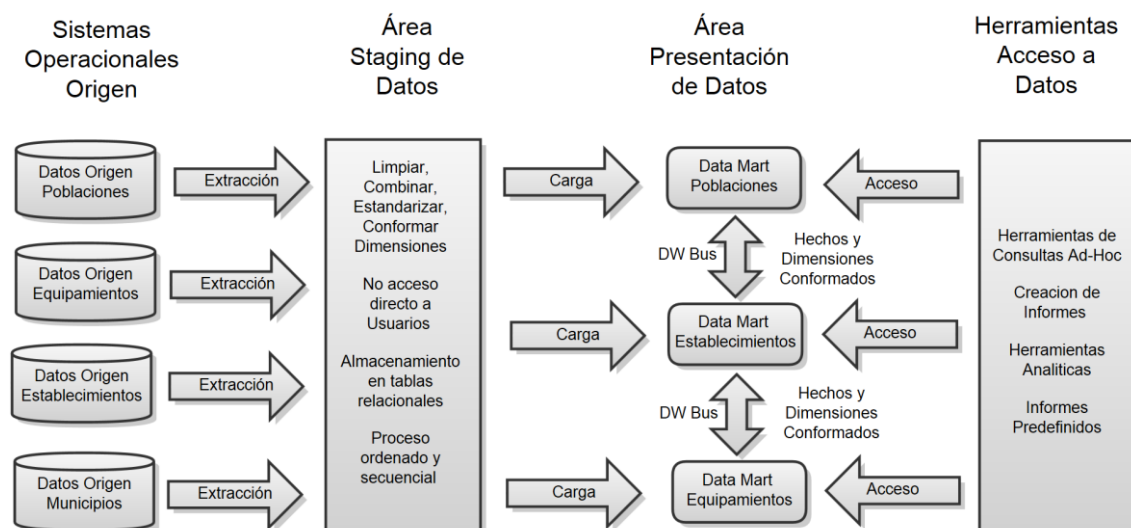


Ilustración 7 Diagrama de la Arquitectura del Software

3.3 Diagrama de la arquitectura del Hardware

En base a los requisitos planteados se propone una arquitectura sencilla con un único servidor de base de datos en las instalaciones de ONdO a los que los usuarios accedan desde sus equipos cliente a través de la red local. Adicionalmente se propone utilizar servicios en la nube para posibilitar el acceso de clientes externos cumpliendo la política de Seguridad IT de ONdO así como para aumentar la disponibilidad y la recuperabilidad global de la solución.

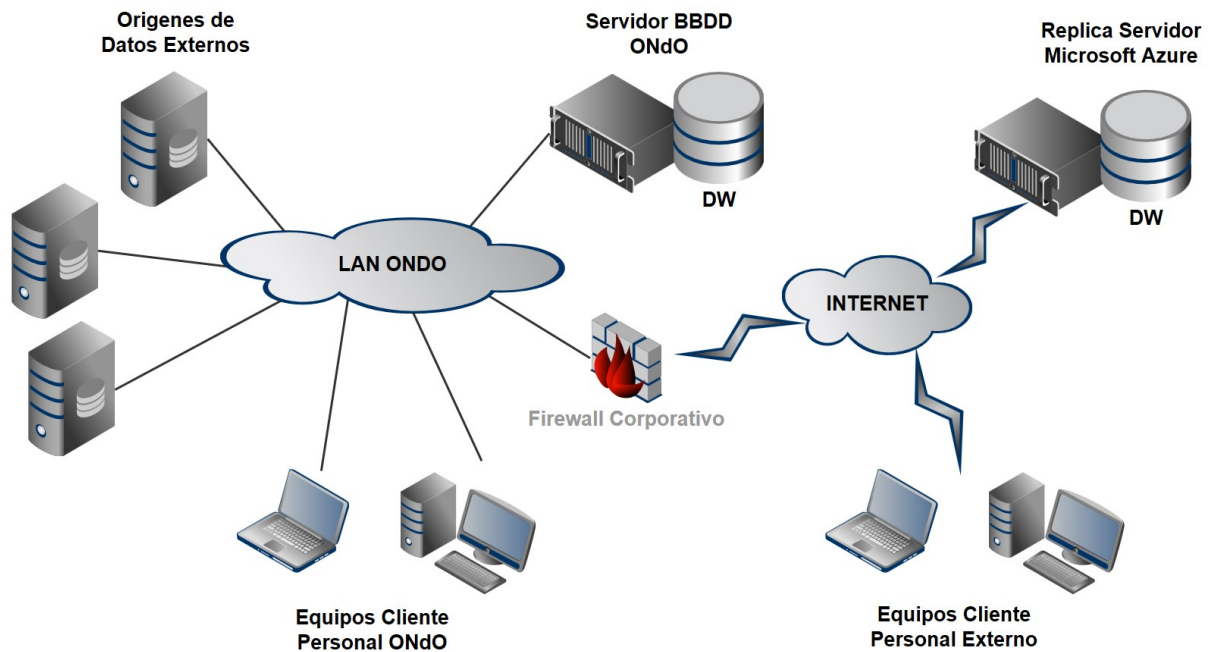


Ilustración 8 Arquitectura del Hardware

3.4 Selección de productos

En base a los objetivos generales del proyecto, a los requisitos no funcionales planteados y para evitar los riesgos característicos que se dan al utilizar una nueva plataforma, se propuso a la Dirección del Proyecto desarrollar la solución completamente sobre la plataforma de productos Microsoft.

Para la cultura de toma de decisiones en base a datos que ONdO quiere instaurar a todos los niveles de su organización es fundamental que los usuarios sean capaces no sólo de generar sus propios informes sino de integrar nuevas fuentes de datos con los de la solución y crear de manera efectiva nuevos escenarios de análisis sin requerir la intervención de un tercero. Es igualmente crítico además que todos los usuarios -incluidos los más reticentes- sean capaces de utilizar la solución sin pronunciadas curvas de aprendizaje.

En este sentido la Dirección del Proyecto aceptó entusiastamente la propuesta de integrar la solución dentro del proceso de trabajo estándar a través de la herramienta de análisis y planificación ya en uso, Microsoft Excel 2013.

3.5 Selección de productos de la solución

Plataforma Solución	Producto
Plataforma Servidor	Microsoft Windows Server 2012
Base de datos relacional Área Staging y Área Presentación	Microsoft SQL Server 2012
Extracción, Transformación y Carga	Microsoft Integration Services 2012
Plataforma Cloud Clientes Externos	Microsoft Azure
Consultas Ad-Hoc	Microsoft Excel PowerPivot 2013
Creación de Reportes	Microsoft Excel PowerView 2013
Acceso a Datos	Microsoft Excel PowerPivot 2013

Tabla 16 Selección de productos software de plataforma

3.6 Selección de productos entorno de desarrollo

Plataforma Desarrollo	Producto
Plataforma Virtualización	Virtual Box 4.2
Plataforma Servidor Desarrollo	Microsoft Windows Server 2012
Base de datos relacional Desarrollo	Microsoft SQL Server 2012
Extracción, Transformación y Carga Desarrollo	Microsoft Integration Services 2012
Entorno de Desarrollo Integrado	Microsoft Visual Studio 2012
Diseño y Creación BBDD	Microsoft SQL Server Management Studio
Acceso a Datos	Microsoft Excel PowerPivot 2013
Creación de Reportes Predefinidos	Microsoft Excel PowerView 2013

Tabla 17 Selección de productos software de desarrollo



Ilustración 9 Productos del stack Microsoft Business Intelligence

3.7 Modelo conceptual

Se analizaron en detalle los datos del proceso de negocio para realizar un modelo dimensional, identificando la tabla de hechos y su granularidad, sus dimensiones y atributos asociados y sus hechos numéricos.

3.7.1 Tablas de hechos y granularidad

A partir de los requerimientos de ONdO se identificaron 3 tablas de hechos correspondientes a 3 objetos de análisis diferenciados: *Poblaciones*, *Establecimientos* y *Servicios*.

Siguiendo la recomendación de Kimball de llegar al mayor grado de granularidad que posibilite el máximo número de escenarios de análisis mientras no sea inviable, se eligió el máximo nivel de granularidad posibilitado por los datos de origen.

Tablas de Hechos	Descripción y Granularidad
Poblaciones	Censo de Población en un municipio y un año.
Equipamientos	Inventario de servicios de un municipio y en una localización
Establecimientos	Registro de establecimientos y plazas en una comarca y un año

Tabla 18 Tablas de hechos y granularidades

3.7.2 Dimensiones de análisis

Se identificaron las siguientes dimensiones con información descriptiva:

Tablas de Hechos	Dimensiones
Poblaciones	Municipio, Año
Equipamientos	Localización, Municipio, Servicio
Establecimientos	Municipio, Tipo_Categoría, Año

Tabla 19 Tablas de Hechos y Dimensiones

Municipio y *Año* son dimensiones conformadas o referenciadas desde diversas tablas de hechos.

3.7.3 Descripción y Atributos de cada Dimensión

Cada dimensión consta de los atributos siguientes:

Dimensión	Descripción	Atributos
Año	Dimensión temporal	Año, Año_Número, Década
Municipio	Dimensión geográfica	Código_INE, Delegacion_Govern, Comarca, Provincia, Municipio, Es_Capital
Tipo_Categoria	Tipo y Categoría de establecimiento turístico	Tipo, Categoría
Localización	Localización y nombre de un equipamiento	Nombre, Dirección, Código Postal, Teléfono, Longitud, Latitud
Servicio	Categorización y nombre servicio disponible en un equipamiento	Categoría, Subcategoria1, Subcategoria2, Servicio

Tabla 20 Dimensiones y Atributos

Los atributos Categoría, Código_INE y Década de las dimensiones Servicio, Municipio y Año respectivamente no son estrictamente necesarios pero se mantienen para posibles futuras extensiones.

3.7.4 Jerarquías de Agregación

Los atributos de una dimensión pueden ser descriptores -que permiten hacer selecciones basadas en sus valores- definir niveles en una jerarquía de agregación. La Dimensión **Municipio** contiene 2 jerarquías ambas con Municipio como base, que puede agregarse en Comarca y Delegacion_Govern por una lado o Comarca y Provincia por otro. La dimensión **Año** tiene nivel de agregación Década. La Dimensión **Servicio** se agrega en Subcategoria2, Subcategoria 1 y Categoría y finalmente en la dimensión **Tipo_Categoria**, categoría se agrega en tipo.

Cabe destacar que la información descriptiva jerárquica se almacena redundantemente siguiendo un diseño en estrella en aras de facilidad de uso y rendimiento al hacer consultas. La alternativa de representar esta jerarquía con tablas separadas de manera normalizada daría como resultado una estructura que recibe el nombre de copo de nieve.

Las tablas de dimensiones son típicamente órdenes de magnitud más reducidas que las tablas de hechos, y optimizar el espacio a base de normalizar no tiene prácticamente impacto en el tamaño global de la base de datos. Con las tecnologías actuales de optimización de consultas ya no es tan crítico mantener un esquema de estrella por rendimiento, pero sigue siendo un modelo más simple y más fácil de mantener.

3.7.5 Medidas

Para cada tabla de hechos se escogieron las medidas relacionadas a continuación:

Tablas de Hechos	Medidas	Significado
Poblaciones	Población, Extensión	Numero Población y número kilómetros de extensión.
Equipamientos	Equipamiento	Dispone de servicio, valor numérico 1.
Establecimientos	Establecimientos, Plazas	Número de establecimientos y de plazas turísticas.

Tabla 21 Tablas de hechos y Medidas

La medida Equipamiento simplemente expresa que cierta localización en cierto municipio dispone de cierto servicio, y toma siempre el valor del natural 1 para posibilitar el cálculo eficiente de recuentos.

3.7.6 Restricciones de integridad

Para las 3 tablas de hechos y respectivas dimensiones se cumplen las restricciones siguientes:

- **Restricción de entidad de la base:** El conjunto de niveles que forman la base son funcionalmente independientes.
- **Restricción de unicidad:** todas las asociaciones de los niveles con la celda tienen multiplicidad 1.
- **Restricciones de agregación:** todas las medidas son aditivas a lo largo de todas las dimensiones.

En el caso de equipamientos, Localización y Municipio no cumplen la primera restricción pero al ser esta última una dimensión conformada se optó por simplicidad mantener la referencia directa desde la tabla de hechos.

3.7.7 Viabilidad

Tomando cotas superiores para el número de registros y tamaños por registro estimadas a partir los datos de origen, se determina un tamaño inicial del almacén de datos del orden de <100 Mb. Con la previsión de crecimiento lineal anticipada por ONDO se trataría de un diseño perfectamente viable.

Cabe destacar que es atípico que sea una dimensión la que consuma la mayor parte del espacio. Esto se debe al mayor tamaño relativo de la información de texto y sobre todo al comparativamente reducido número de hechos.

Concepto	Descripción	Número Registros	Tamaño Registro (bytes)	Tamaño Total (bytes)
Dimensión	Año	50	50	2500
Dimensión	Municipio	100	100	10.000
Dimensión	Tipo_Categoria	50	50	2.500
Dimensión	Localización	50.000	500	25.000.000
Dimensión	Servicio	500	200	100.000
Hecho	Poblaciones	5.000	10	5.000
Hecho	Equipamientos	50.000	10	500.000
Hecho	Establecimientos	500	50	25.000

Tabla 22 Estimación cotas superiores volumen de datos requerido

3.8 Diagrama E-R del diseño

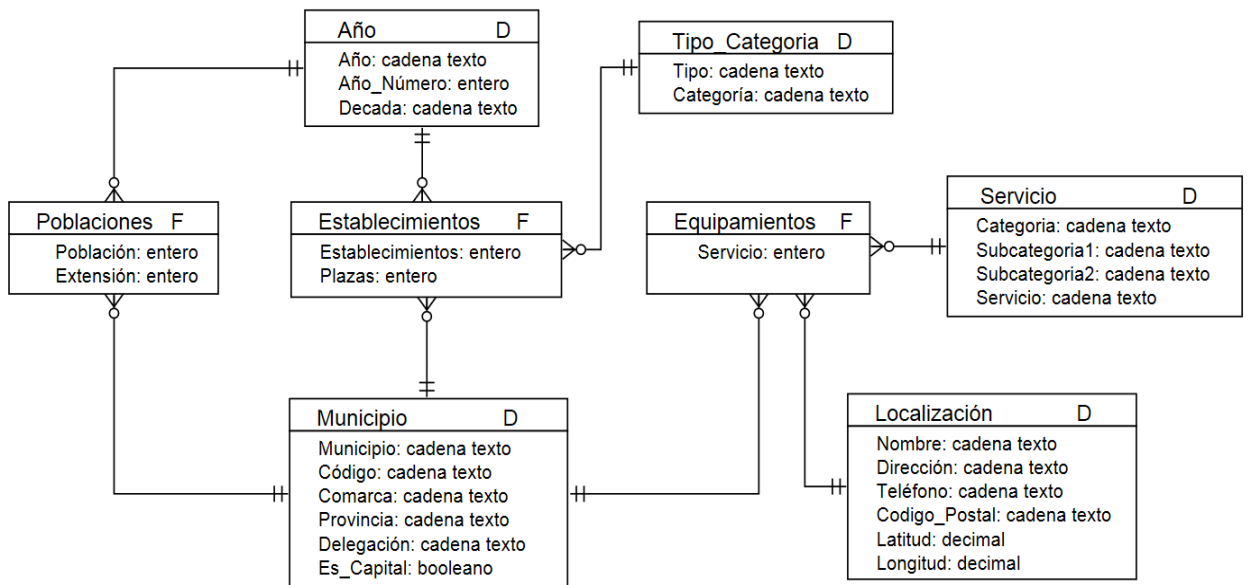


Ilustración 10 Diagrama Entidad Relación

3.9 Diseño Lógico

A partir del diagrama conceptual y siguiendo un diseño basado en un modelo de estrella con claves naturales reemplazadas por claves sustitutas se obtuvo el siguiente diseño lógico.

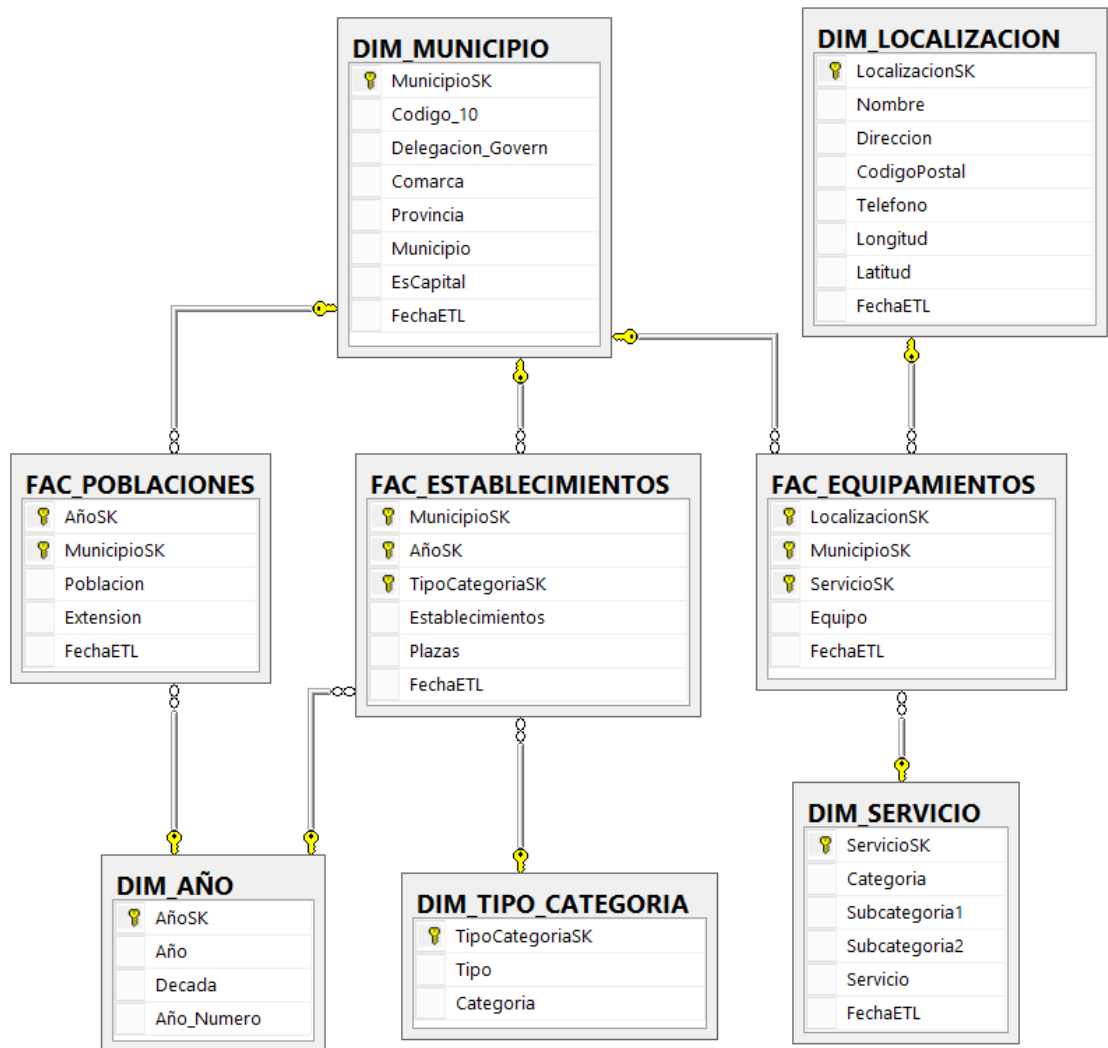


Ilustración 11 Diagrama Diseño Lógico

3.10 Diseño Físico

Finalmente, el modelo físico diseñado fué el siguiente. La nomenclatura *PK* corresponde a índice de clave primaria, e *IX* a índice compuesto. Este último aparece en las tablas correspondientes a dimensiones y tiene como objetivo agilizar el procesado de las mismas.

DIM_AÑO		
Columna	Tipo Datos	Índices
AñoSK	int	PK
Año	nvarchar(50)	IX
Decada	nvarchar(50)	
Año_Numero	int	

DIM_TIPO_CATEGORIA		
Columna	Tipo Datos	Índices
TipoCategoriaSK	int	PK
Tipo	nvarchar(50)	IX
Categoria	nvarchar(50)	IX

DIM_MUNICIPIO		
Columna	Tipo Datos	Índices
MunicipioSK	int	PK
Codigo_10	nvarchar(50)	
Delegacion_Govern	nvarchar(50)	
Comarca	nvarchar(50)	IX
Provincia	nvarchar(50)	
Municipio	nvarchar(50)	IX
EsCapital	bit	

DIM_SERVICIO		
Columna	Tipo Datos	Índices
ServicioSK	int	PK
Categoria	nvarchar(100)	IX
Subcategoria1	nvarchar(100)	IX
Subcategoria2	nvarchar(100)	IX
Servicio	nvarchar(100)	IX
FechaETL	datetime	

DIM_LOCALIZACION		
Columna	Tipo Datos	Índices
LocalizacionSK	int	PK
Nombre	nvarchar(150)	IX
Direccion	nvarchar(150)	IX
CodigoPostal	nvarchar(50)	IX
Telefono	nvarchar(50)	
Longitud	numeric(18, 15)	IX
Latitud	numeric(18, 15)	IX
FechaETL	datetime	

FAC_POBLACIONES		
Columna	Tipo Datos	Índices
AñoSK	int	PK
MunicipioSK	int	PK
Poblacion	int	
Extension	int	
FechaETL	datetime	

FAC_EQUIPAMINENTOS		
Columna	Tipo Datos	Índices
LocalizacionSK	int	PK
MunicipioSK	int	PK
ServicioSK	int	PK
Equipo	tinyint	
FechaETL	datetime	

FAC_ESTABLECIMIENTOS		
Columna	Tipo Datos	Índices
MunicipioSK	int	PK
AñoSK	int	PK
TipoCategoriaSK	int	PK
Establecimientos	int	
Plazas	int	
FechaETL	datetime	

3.11 Diseño procesos ETL

En el Sistema ETL es donde se extraen los datos en crudo de los sistemas de origen operacionales y se transforman en información valiosa para el negocio. Los Procesos ETL son por tanto la parte más importante y la más crítica de un Almacén de Datos. Típicamente dos terceras partes del riesgo y el esfuerzo de un proyecto de Almacenes de Datos se atribuyen a esta fase.

El Sistema ETL del proyecto desarró en base a una arquitectura modular. Consiste en una serie de procesos independientes para cada origen de datos y para cada fase de extracción, transformación y carga, y un proceso de control que se encarga de ejecutarlos, monitorizarlos y asegurar la secuenciación requerida por las dependencias entre los mismos.

Un sistema de ETL bien diseñado aspira alcanzar un alto rendimiento y alta calidad de los datos. Para el primer objetivo el diseño propuesto realiza la ejecución de ciertas tareas en paralelo, pero no aplica esquemas de paralización más profunda sobre todo el proceso o técnicas más avanzadas como particionamiento de los datos que corresponderían a proyectos de otra envergadura.

En cuanto a la de Calidad de los datos, los procesos fueron diseñados para detectar todas las condiciones de error en los formatos y los rangos de los datos origen así como algunas inconsistencias semánticas básicas, como que las coordenadas de un equipamiento pertenezcan al ámbito territorial de Catalunya, o que las sumas parciales de establecimientos y plazas por tipo coincidan con el valor total.

No se realizan por el contrario comprobaciones más profundas como determinar si un conjunto de valores suministrado configura una dirección válida, si prefijos de numeración telefónica y ámbitos territoriales corresponden, o si un dato de población cambia mucho más de lo que es demográficamente plausible de un periodo de tiempo a otro.

3.11.1 Diseño zona o área de staging

La zona de Staging diseñada para este proyecto presenta una estructura común para todos los procesos, consistente en dos tablas intermedias y 2 tablas de errores para que cada proceso guarde su output refinado o desvíe los registros erróneos con su código descriptivo de error.

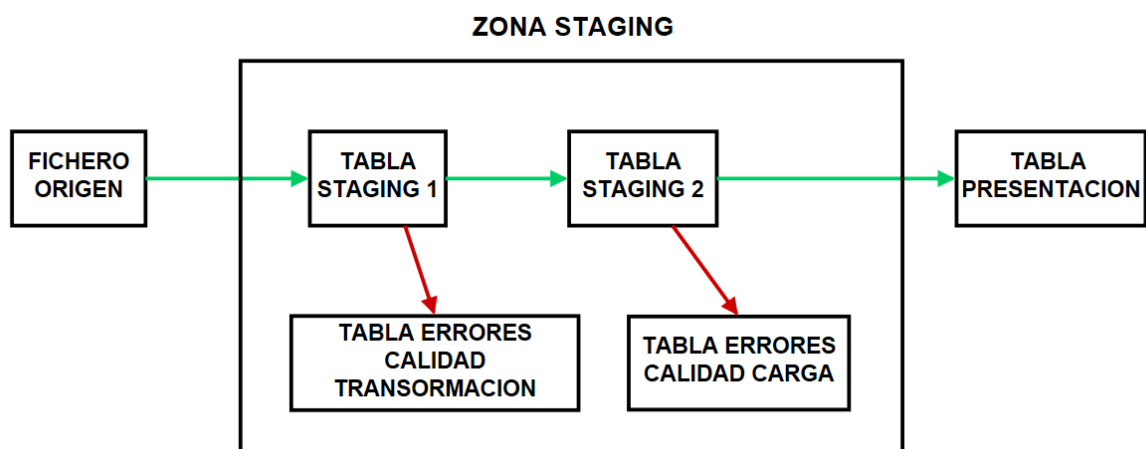


Ilustración 13 Tablas Zona Staging

3.11.1 Diseño Registro

Dada la importancia central de la calidad de datos dentro del proceso de ETL se hace indispensable no sólo detectar situaciones anómalas sino también registrarlas detalladamente. De esta manera el Administrador del Almacén de Datos puede conocer en todo momento exactamente qué datos se han cargado en el almacén y cuáles no. Junto con la causa del problema anotada a nivel de registro en las tablas de errores conforma una herramienta básica para diagnosticar problemas y corregir los datos en los sistemas de origen.

El subsistema de registro o log de la solución escribe en una tabla dedicada, a través de una rutina encapsulada llamada *EntradaLog* y con parámetros de entrada el mensaje y las variables de número de registros de entrada, salida y error. Esta única rutina es reutilizada a lo largo de toda la solución.



Ilustración 14 Rutina Escritura Log

SQLQuery1.sql - uoc_server.master (UOC_SERVER\Administrator (52))*

```
SELECT * FROM [DW_ONDO].[dbo].[ETL_REGISTRO]
```

100 %

IdLog	NombrePaquete	Descripcion	Inicio	Fin	RegistrosEntrada	RegistrosSalida	RegistrosError	
1	1	ETL_CONTROL_ONDO	Inicio Recarga Completa DW ONdO	16/06/2013 23:08:01	2013-06-16 23:08:02.180	0	0	0
2	2	ETL_CONTROL_ONDO	Inicio Fase Extracción	16/06/2013 23:08:01	2013-06-16 23:08:02.203	0	0	0
3	3	STG_MUNICIPIOS	Carga Tabla Staging 1 Fichero Municipios	16/06/2013 23:08:02	2013-06-16 23:08:03.790	947	947	0
4	4	STG_POBLACION	Carga Tabla Staging 1 Fichero poblacions	16/06/2013 23:08:04	2013-06-16 23:08:07.897	468	468	0
5	5	STG_ESTABLECIMIENTOS	Carga Tabla Staging 1 Fichero establiments 2006	16/06/2013 23:08:04	2013-06-16 23:08:08.010	41	41	0
6	6	STG_ESTABLECIMIENTOS	Carga Tabla Staging 1 Fichero establiments 2007	16/06/2013 23:08:04	2013-06-16 23:08:08.363	53	53	0
7	7	STG_ESTABLECIMIENTOS	Carga Tabla Staging 1 Fichero establiments 2008	16/06/2013 23:08:08	2013-06-16 23:08:08.593	53	53	0
8	8	STG_ESTABLECIMIENTOS	Carga Tabla Staging 1 Fichero establiments 2009	16/06/2013 23:08:08	2013-06-16 23:08:08.793	53	53	0
9	9	STG_ESTABLECIMIENTOS	Carga Tabla Staging 1 Fichero establiments 2010	16/06/2013 23:08:08	2013-06-16 23:08:08.990	53	53	0
10	10	STG_ESTABLECIMIENTOS	Carga Tabla Staging 1 Fichero establiments 2011	16/06/2013 23:08:08	2013-06-16 23:08:09.210	54	54	0
11	11	STG_ESTABLECIMIENTOS	Carga Tabla Staging 1 Fichero establiments 2012	16/06/2013 23:08:09	2013-06-16 23:08:09.413	54	54	0
12	12	STG_EQUIPAMIENTOS	Carga Tabla Staging 1 Equipamientos	16/06/2013 23:08:04	2013-06-16 23:08:10.120	31771	31771	0
13	13	ETL_CONTROL_ONDO	Inicio Fase Transformación	16/06/2013 23:08:01	2013-06-16 23:08:10.200	0	0	0
14	14	STG_ESTABLECIMIENTOS_2	Carga Tabla Staging 2 Establecimientos	16/06/2013 23:08:11	2013-06-16 23:08:16.607	361	2770	0
15	15	STG_POBLACION_2	Carga Tabla Staging 2 Poblacion	16/06/2013 23:08:10	2013-06-16 23:08:16.890	468	3269	1
16	16	STG_EQUIPAMIENTOS_2	Carga Tabla Staging 2 Equipamientos	16/06/2013 23:08:10	2013-06-16 23:08:24.000	31771	35226	771
17	17	ETL_CONTROL_ONDO	Inicio Fase Carga	16/06/2013 23:08:01	2013-06-16 23:08:24.120	0	0	0
18	18	DIM_TIPO_CATEGORIA	Carga Dimension Tipo_Categoria	16/06/2013 23:08:24	2013-06-16 23:08:25.143	10	10	0
19	19	DIM_AÑO	Carga Dimension Año	16/06/2013 23:08:24	2013-06-16 23:08:25.177	11	11	0
20	20	DIM_MUNICIPIO	Carga Dimension Municipio	16/06/2013 23:08:24	2013-06-16 23:08:25.863	947	947	0
21	21	FAC_POBLACIONES	Carga Tabla Hechos Poblaciones	16/06/2013 23:08:26	2013-06-16 23:08:31.047	3269	3269	0
22	22	FAC_ESTABLECIMIENTOS	Carga Table Hechos Establecimientos	16/06/2013 23:08:26	2013-06-16 23:08:31.190	2770	2730	40
23	23	FAC_EQUIPAMIENTOS	Carga Tabla Hechos Equipamientos	16/06/2013 23:08:26	2013-06-16 23:08:44.147	35226	35226	0
24	24	ETL_CONTROL_ONDO	Fin Recarga Completa DW ONdO	16/06/2013 23:08:01	2013-06-16 23:08:44.320	0	0	0

Query executed successfully. | uoc_server (11.0 SP1) | UOC_SERVER\Administrat... | master | 00:00:00 | 24 rows

Ilustración 15 Registro log ejecución

3.11.2 Proceso principal

Según explicado en la introducción, el proceso principal se encarga de controlar la secuencia precisa en que se deben ejecutar los procesos específicos de ETL. Realiza la carga completa del almacén de datos a partir de los datos de origen y registrando las diferentes fases realizadas así como el resultado de cada proceso.

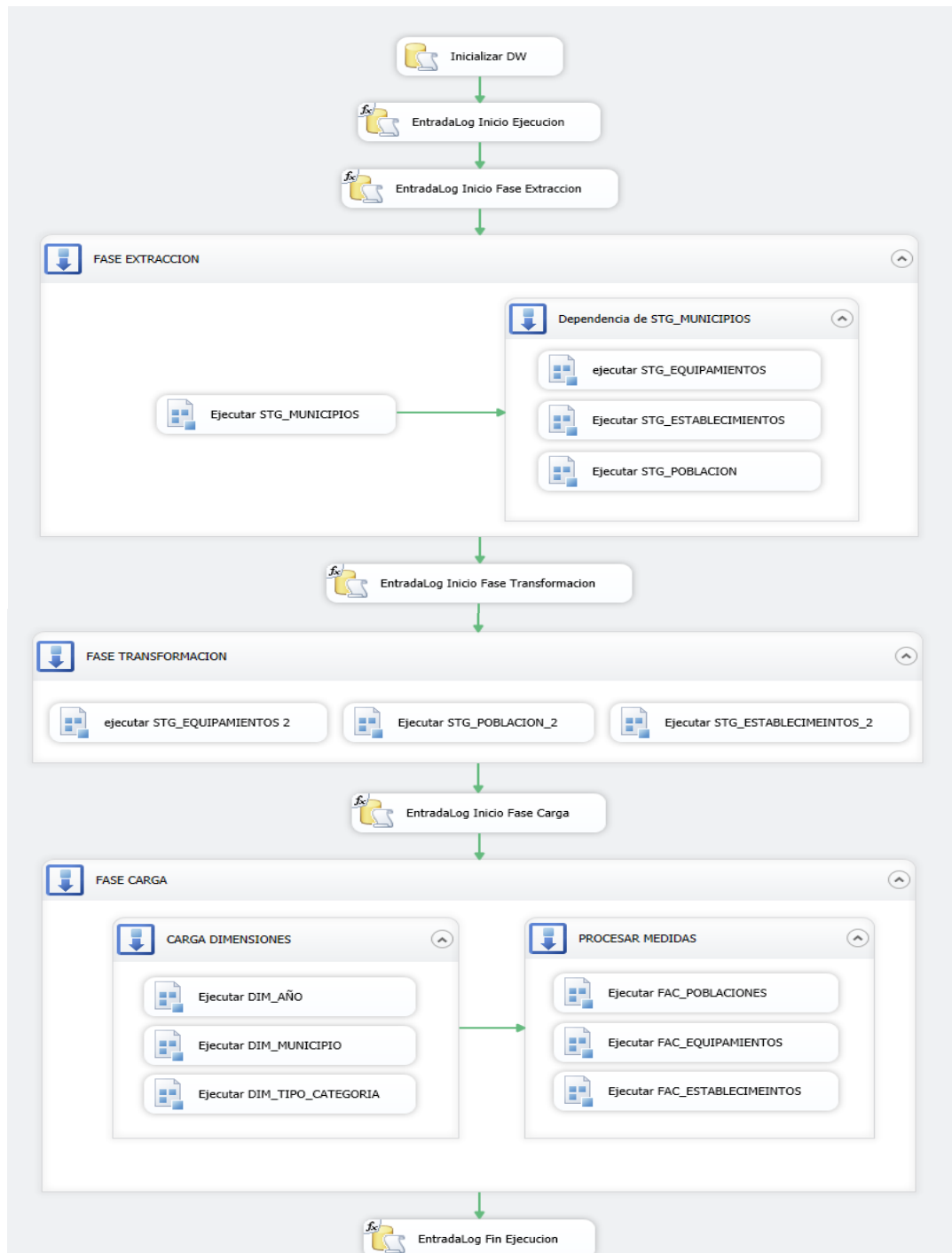


Ilustración 16 Proceso ETL de control

3.11.3 Fase Extracción

Los procesos correspondientes a la fase de extracción se centran puramente en extraer los datos de origen y guardarlos en tablas de la zona de staging, comprobando únicamente que se es capaz de leer los datos en el formato adecuado. Todos procesos de extracción desarrollados siguen exactamente el mismo patrón. En el ejemplo a continuación la única peculiaridad es que el flujo de control itera sobre todos los ficheros existentes de entrada en cierto directorio. Se trata el caso 2006 de manera especial por tener un formato de entrada ligeramente distinto a los demás. En todos los demás procesos, el flujo de control simplemente inicia el flujo de datos y escribe en el log al acabar.

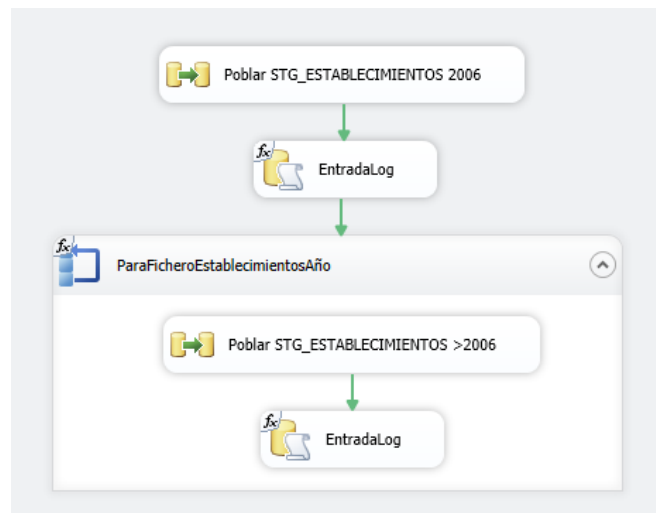


Ilustración 17 Ejemplo Proceso Extracción Equipamientos – Flujo Control

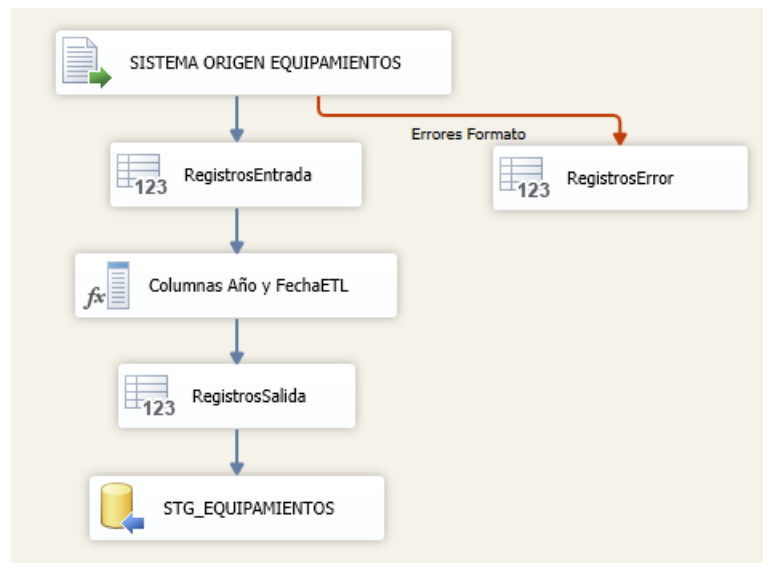


Ilustración 18 Ejemplo Proceso Extracción Equipamientos – Flujo Datos

3.11.4 Fase Transformación

En la fase de transformación los procesos diseñados toman la información guardada en las tablas staging 1 y aplican las transformaciones necesarias para alinear la estructura de los datos de entrada con la requerida por el diseño lógico y físico planteado. Es donde se realizan además las operaciones de limpieza de datos y comprobación de condiciones de error con el fin de obtener unos datos de salida con los formatos especificados y con el menor número de inconsistencias posibles. Los registros con errores se guardan en tablas específicas, y el resultado neto se guarda en tablas staging 2.

En el ejemplo, correspondiente al proceso de transformación de Equipamientos se corrigen los datos de los municipios con una búsqueda aproximada, se comprueban las condiciones de registros repetidos, errores de formato en las coordenadas y consistencia de las mismas, para finalmente aplicar una desnormalización sobre las columnas correspondientes al servicio.

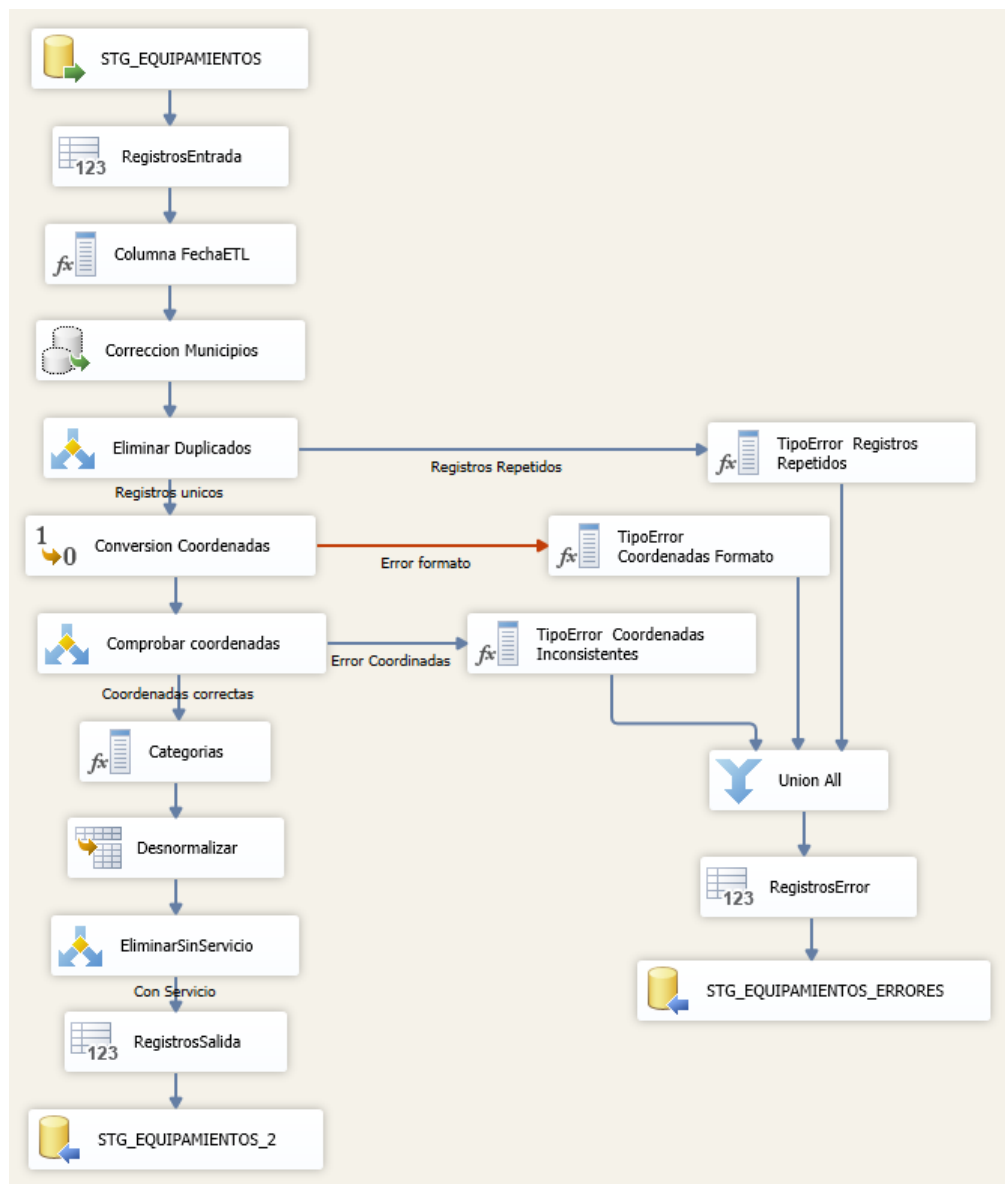


Ilustración 19 Ejemplo Proceso Transformación Equipamientos

3.11.4.1 Fase Carga

En la fase de carga se procesan los datos de tablas staging 2 para cargarlos en el modelo de estrella. Esta fase comprende la carga de las dimensiones y el procesado de las tablas de medidas, cambiando claves naturales por claves sustitutas por medio de búsquedas en las tablas de dimensiones. Los miembros desconocidos se infieren o se señalan como condiciones de error dependiendo del caso.

En el ejemplo a continuación, correspondiente al proceso de carga de las medidas de Equipamientos, los la dimensión Localización y la dimensión Servicio se infieren, mientras que miembros no hallados en la dimensión Año o Municipio se consideran una condición de error y se tratan como tal.

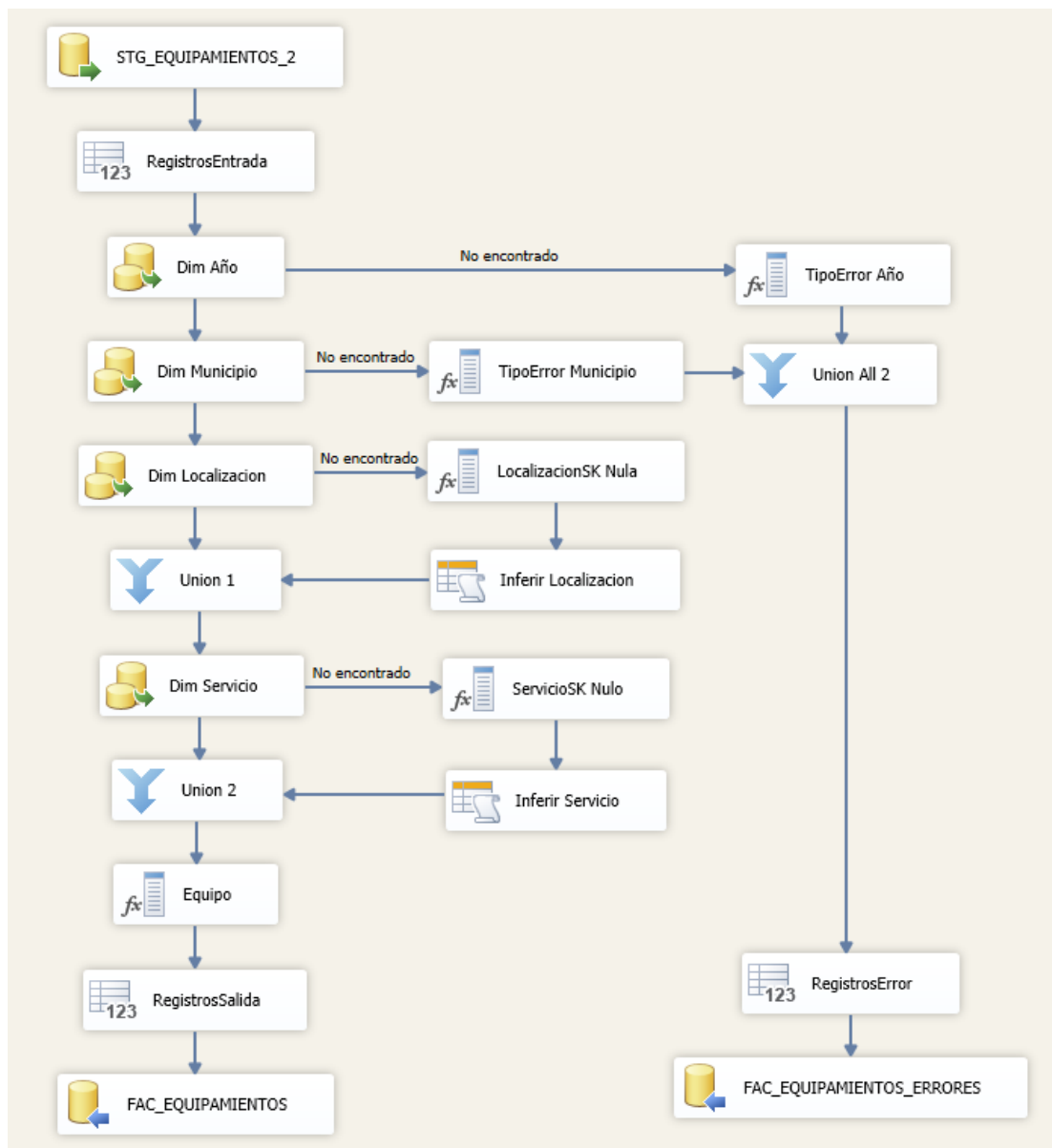
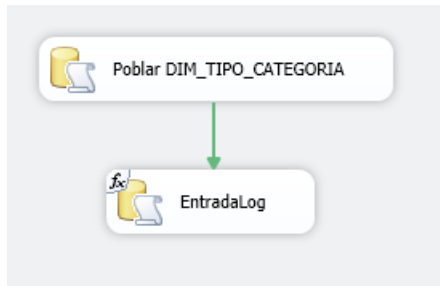


Ilustración 20 Ejemplo fase carga - procesado medidas equipamientos

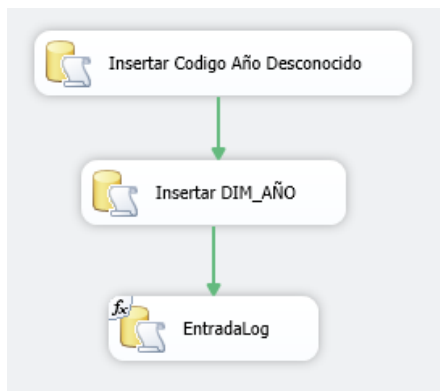
El procesado de las dimensiones, no requiriendo tratar cambios en el tiempo en sus atributos se convierte en una simple carga de los miembros individuales como es en el caso de Año y Tipo_Categoría o de la tabla de staging como en el de Municipios.



```
USE [DW_ONDO]
GO
```

```
INSERT INTO [DW_ONDO].[dbo].[DIM_TIPO_CATEGORIA] (Tipo, Categoría)
VALUES
('Hotel', 'Estrellas_Plata'),
('Camping', 'Lujo'),
('Hotel', 'Estrellas_Oro'),
('Camping', '3A'),
('Casa_Rural', 'Casa_Pueblo Compartida'),
('Casa_Rural', 'Casa_Pueblo Independiente'),
('Casa_Rural', 'Masia'),
('Camping', '2A'),
('Casa_Rural', 'Masoveria'),
('Camping', '1A')
```

Ilustración 22 Ejemplo fase carga - carga dimensión TipoCategoría



```
USE [DW_ONDO]
GO
```

```
INSERT INTO [DW_ONDO].[dbo].[DIM_AÑO] ([Año] , [Decada] , [Año_Numero])
VALUES
('2006', '00', 2006),
('2007', '00', 2007),
('2008', '00', 2008),
('2009', '00', 2009),
('2010', '10', 2010),
('2011', '10', 2011),
('2012', '10', 2012),
('2013', '10', 2013),
('2014', '10', 2014),
('2015', '10', 2015)
```

Ilustración 23 Ejemplo fase carga - carga dimensión Año

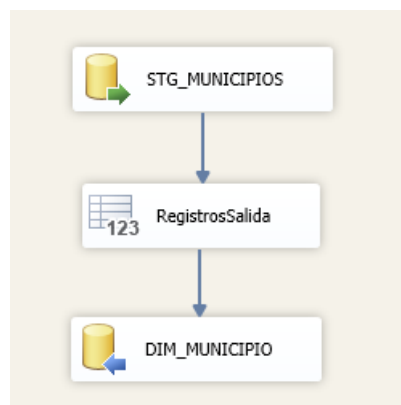


Ilustración 24 Ejemplo fase carga - carga dimensión Municipio

3.12 Diseño Informes

3.12.1 Sobre la herramienta

Para la funcionalidad de Acceso a Datos del Área de Presentación se ha utilizado la única herramienta de Microsoft Excel 2013 PowerPivot en las diversas facetas requeridas, y tanto para el desarrollo como para consumo o uso por el cliente final. Esta unicidad y simplicidad es uno de los puntos excepcionalmente poderosos de la solución escogida:

- El usuario dispone potencialmente del mismo poder expresivo y la capacidad de crear valor a partir del Almacén de Datos que el equipo desarrollador.
- El equipo desarrollador dispone de herramientas con facilidad de uso, sencillez y productividad al nivel de herramientas de usuario final.
- Las barreras para la interacción entre unos son más bajas que nunca, propiciando un entorno propenso a la colaboración y creación de nuevas oportunidades de valor para el negocio.

Concretamente sobre Microsoft Excel 2013 PowerPivot se ha creado el modelo analítico o modelo semántico en terminología Microsoft en base a los datos del Área de Presentación sobre el que luego se basan todos los informes interactivos y las consultas ad-hoc.

PowerPivot es una instancia local del motor de analítico en-memoria de Microsoft Análisis Services 2012 dentro de Microsoft Excel que carga datos en formato comprimido y los hace disponibles para elementos de visualización de hojas de trabajo. Este motor además interpreta y calcula medidas analíticas creadas en un lenguaje propio llamado DAX (*Data Analysis Expressions*).

Soporta la carga y la combinación de datos de una gran variedad de fuentes incluyendo bases de datos relacionales, orígenes multidimensionales, servicios de datos en la nube, otros ficheros Excel, ficheros de texto y datos en formato tabular de la Web, permitiendo al usuario crear nuevos escenarios analíticos sin requerir de un tercero.

PowerPivot permite crear datos portables y reusables, que permanecen dentro del libro de trabajo. Los usuarios no necesitan gestionar conexiones externas, ya que en el mismo fichero .xlsx contiene los datos embebidos. Esto permite asimismo que una vez que los datos están cargados el usuario pueda trabajar en modo desconectado sin requerir más conectividad a ningún servidor.

Como herramienta de desarrollo, los modelos diseñados pueden ser cargados en un servidor externo Microsoft Análisis Services 2012, lo que sumado a las capacidades ya mencionadas lo convierte en una gran herramienta de prototipado rápido y Desarrollo Ágil.

3.12.2 Modelo Analítico Acceso a Datos

A continuación se presenta el modelo analítico creado, con sólo los atributos relevantes visibles y jerarquías de agregación y medidas calculadas creadas.

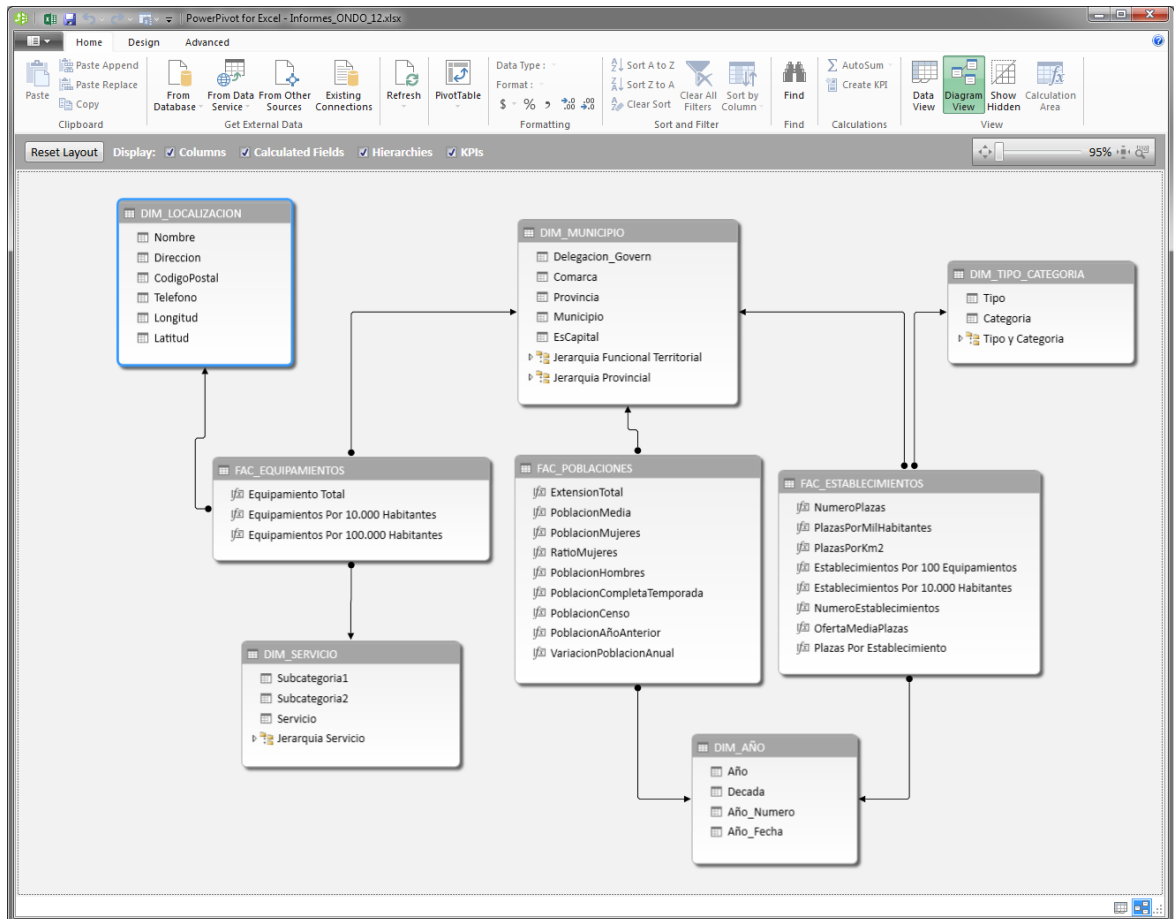


Ilustración 25 Modelo analítico de acceso a datos

3.12.1 Medidas creadas

Para cumplir los requerimientos de información y de análisis del cliente se crearon las diferentes medidas en DAX:

3.12.1.1 Medidas relativas a la población

- Población en base a la población censal original

PoblacionCenso := SUM([Poblacion])

- Población en base a la media de la población censal de un año y el siguiente usado como medida base de población a petición del cliente. Sólo es válida en un contexto analítico en que el año está definido.

```
PoblacionMedia:= if (countrows (values ('DIM_AÑO' [Año_Número])) = 1;
CALCULATE (SUM ([Poblacion]) /countrows (values (FAC_POBLACIONES [AñoNumero]));
'DIM_AÑO' [Año_Número]=VALUES ('DIM_AÑO' [Año_Número]) ||
'DIM_AÑO' [Año_Número]=VALUES ('DIM_AÑO' [Año_Número]) +1;ALL ('DIM_AÑO' ) ) ;
BLANK ())
```

- Población del año anterior. Sólo es válida en un contexto analítico en que el año está definido.

```
PoblacionAñoAnterior:=if (countrows (values ('DIM_AÑO' [Año_Número]))=1;
CALCULATE ([PoblacionMedia]; 'DIM_AÑO' [Año_Número]=VALUES ('DIM_AÑO' [Año_Número]) -
1;ALL ('DIM_AÑO' ) );BLANK ())
```

- Variación interanual de población

```
VariacionPoblacionAnual:=IF ([PoblacionAñoAnterior]<>0; [PoblacionMedia] / [Poblacion
AñoAnterior]-1;0)
```

- Constante Ratio Medio Mujeres – Hombres Catalunya 2006-2012

```
RatioMujeres:=0,504355832
```

- Población mujeres

```
PoblacionMujeres:=[PoblacionMedia]*[RatioMujeres]
```

- Población Hombres

```
PoblacionHombres:=[PoblacionMedia]*(1-[RatioMujeres])
```

- Medida de población y plazas total

```
PoblacionCompletaTemporada:=[PoblacionMedia]+[NumeroPlazas]
```

3.12.1.2 Medidas relativas a establecimientos y plazas turísticas

- Numero de plazas total

```
NumeroPlazas:=SUM ([Plazas])
```

- Numero establecimientos

```
NumeroEstablecimientos:=SUM ([Establecimientos])
```

- Medida con la oferta media de plazas a través de todos los años registrados

```
OfertaMediaPlazas:=if(countrows(values(FAC_ESTABLECIMIENTOS[AñoNumero])) >0 ;  
CALCULATE(SUM([Plazas])/countrows(values(FAC_ESTABLECIMIENTOS[AñoNumero]));ALL('D  
IM_AÑO' ) );BLANK())
```

- Indicador Plazas por Habitantes

```
PlazasPorMilHabitantes:=IF([PoblacionMedia]>0;[NumeroPlazas]/[PoblacionMedia]*100  
0;BLANK())
```

- Indicador Plazas por kilómetro cuadrado

```
PlazasPorKm2:=[NumeroPlazas]/[ExtensionTotal]
```

- Indicador Plazas por establecimiento

```
Plazas Por Establecimiento:=[NumeroPlazas]/[NumeroEstablecimientos]
```

- Indicador Plazas por Habitantes

```
Establecimientos Por 10.000 Habitantes:= IF ( [ PoblacionMedia]>0;  
[NumeroEstablecimientos] / [PoblacionMedia]*10000;BLANK())
```

3.12.1.3 Medidas relativas al equipamiento

- Equipamiento total

```
Equipamiento Total:=SUM([Equipo])
```

- Indicador de equipamientos por habitantes

```
Equipamientos Por 10.000 Habitantes:=IF([PoblacionMedia]>0; [Equipamiento  
Total]/[PoblacionMedia]*10000;BLANK())
```

```
Equipamientos Por 100.000 Habitantes:=IF([PoblacionMedia]>0; [Equipamiento  
Total]/[PoblacionMedia]*100000;BLANK())
```

- Indicador de establecimientos por equipamientos

```
Establecimientos Por 100 Equipamientos:=IF([Equipamiento Total]>0;  
[NumeroEstablecimientos]/ [Equipamiento Total]*100;BLANK())
```

4 INFORMES

4.1 Informe de indicadores principales – usuarios analistas

Informe de indicadores principales – usuarios analistas

INDICADORES POBLACIÓN, ESTABLECIMIENTOS Y SERVICIOS POR AÑO, TIPO Y ÁMBITO TERRITORIAL

Año: 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015

Establecimientos: Camping, Casa_Ru..., Hotel

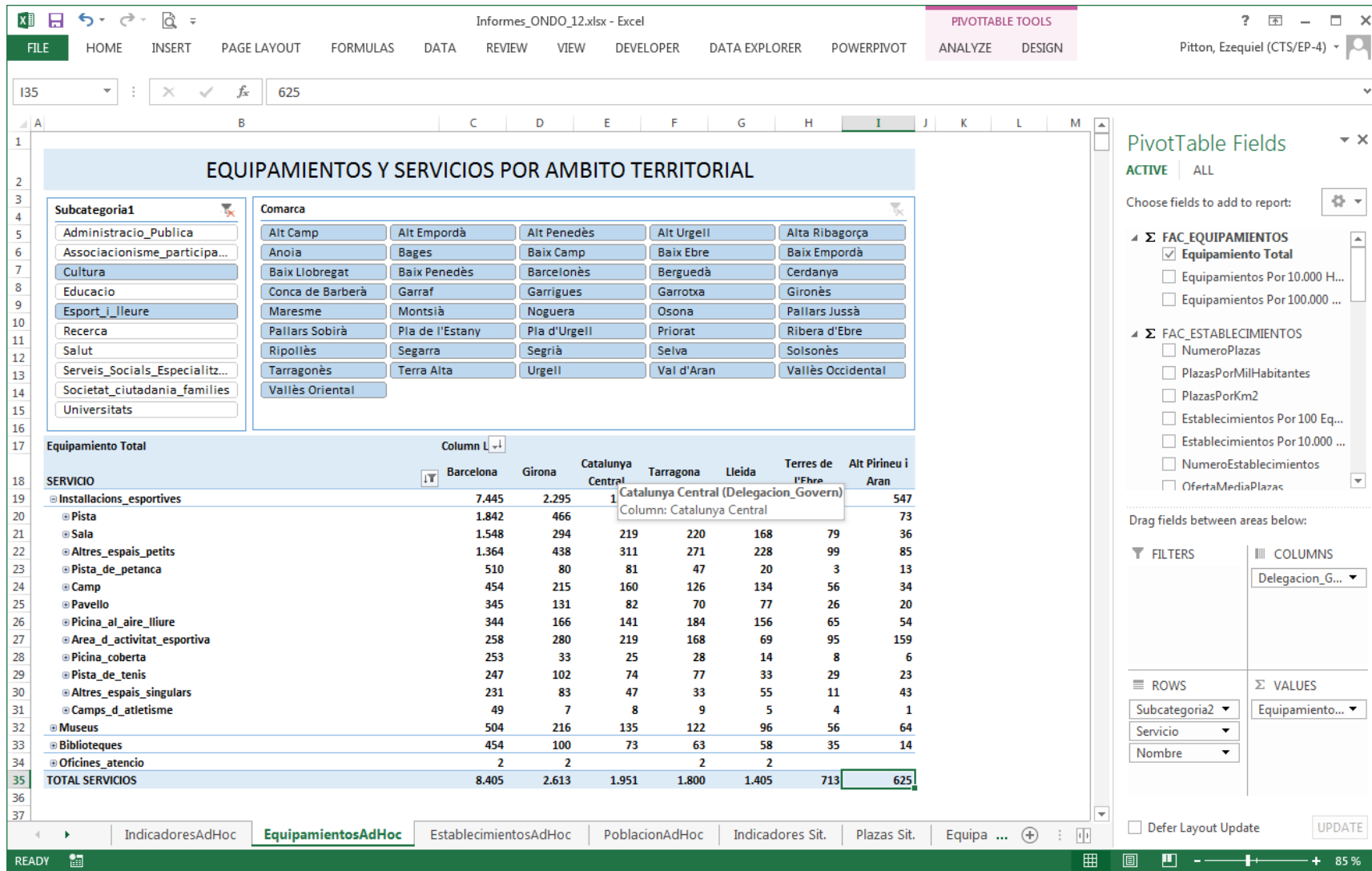
Categoría Establecimiento: 1A, 2A, 3A, Casa_Pueblo_Compartida, Casa_Pueblo_Independi..., Estrellas_Oro, Estrellas_Plata, Lujo, Masia, Masoveria

Tipo de Equipamiento: Administracio_Publica, Asociacionisme_particip..., Cultura, Educacio, Esport_i_lleure, Recerca, Salut, Serveis_Socials_Especialit..., Societat_ciudadania_famil..., Universitats

Ámbito Geográfico	Población Media	Extensión Total Km2	Establecimientos	Establecimientos Por 10.000 Habitantes	Establecimientos Por 100 Equipamientos	Plazas	Plazas Por Establecimiento	Oferta Media Plazas Anual	Plazas Por Mil Habitantes	Plazas Por Km2	Equipamientos Por 10.000 Habitantes
Barcelona	4.992.480	2.946	1.174	2	6	150.887	129	145.040	30	51,2	37
Alt Penedès	100.313	549	81	8	12	1.296	16	1.271	13	2,4	68
Baix Llobregat	796.062	486	90	1	3	12.045	134	11.903	15	24,8	37
Barcelonès	2.251.315	144	549	2	0	62.007	117	59.186	28	444,4	32
Garraf	143.862	185	73	Establecimientos Por 10.000 Habitantes				12.433	87	67,9	38
Canyelles	4.150	14		Value: 2							51
Cubelles	13.853	13		Row: Barcelona - Barcelonès							29
Olivella	3.373	39		Column: Establecimientos Por 10.000 Habitantes							65
Sant Pere de Ribes	28.376	41									32
Sitges	27.899	44									49
Vilanova i la Geltrú	66.211	34	73	11	32	12.554	172	12.433	190	369,2	35
Maresme	428.134	394	222	5	12	49.592	223	51.245	116	125,9	45
Vallès Occidental	881.777	546	44	0	1	4.773	108	4.510	5	8,7	38
Vallès Oriental	391.019	642	115	3	6	6.630	58	6.374	17	10,3	46
Girona	686.383	2.931	1.563	23	33	194.274	124	184.246	283	66,3	68
Tarragona	588.481	1.598	540	9	16	113.523	210	109.095	193	71,0	57
Catalunya Central	480.472	2.049	700	15	20	19.958	29	19.639	42	9,7	74
Lleida	320.593	2.296	250	8	9	7.291	29	7.257	23	3,2	89
Terres de l'Ebre	176.919	2.402	212	12	16	12.035	57	11.696	68	5,0	75
Alt Pirineu i Aran	56.873	1.864	774	136	74	36.615	47	35.531	644	19,6	183
TOTAL	7.302.199	16.086	5.213	7	15	534.583	103	512.504	73	33,2	48

IndicadoresAdHoc | EquipamientosAdHoc | EstablecimientosAdHoc | PoblacionAdHoc | Indicadores Sit. | Plazas Sit. | Equipa ...

4.2 Informe equipamientos y servicios – usuarios analistas



EQUIPAMIENTOS Y SERVICIOS POR AMBITO TERRITORIAL

Subcategoria1

- Administracio_Publica
- Associacionisme_participa...
- Cultura
- Educacio
- Esport_i_lluire
- Recerca
- Salut
- Serveis_Socials_Especialitz...
- Societat_ciudadania_families
- Universitats

Comarca

- Alt Camp
- Alt Empordà
- Alt Penedès
- Alt Urgell
- Alta Ribagorça
- Anoia
- Bages
- Baix Camp
- Baix Ebre
- Baix Empordà
- Baix Llobregat
- Baix Penedès
- Barcelonès
- Berguedà
- Cerdanya
- Conca de Barberà
- Garraf
- Garrigues
- Garrotxa
- Gironès
- Maresme
- Montsià
- Noguera
- Osona
- Pallars Jussà
- Pallars Sobirà
- Pla de l'Estany
- Pla d'Urgell
- Priorat
- Ribera d'Ebre
- Ripollès
- Segarra
- Segrià
- Selva
- Solsonès
- Tarragonès
- Terra Alta
- Urgell
- Val d'Aran
- Vallès Occidental
- Vallès Oriental

Equipamiento Total	Barcelona	Girona	Catalunya Central	Tarragona	Lleida	Terres de l'Ebre	Alt Pirineu i Aran
SERVICIO							
● Instal·lacions esportives	7.445	2.295	Catalunya Central (Delegacion_Govern)				547
● Pista	1.842	466	Column: Catalunya Central				73
● Sala	1.548	294	219	220	168	79	36
● Altres espais petits	1.364	438	311	271	228	99	85
● Pista_de_petanca	510	80	81	47	20	3	13
● Camp	454	215	160	126	134	56	34
● Pavello	345	131	82	70	77	26	20
● Picina_al_aire_lluire	344	166	141	184	156	65	54
● Area_d_activitat_esportiva	258	280	219	168	69	95	159
● Picina_coberta	253	33	25	28	14	8	6
● Pista_de_tenis	247	102	74	77	33	29	23
● Altres espais singulars	231	83	47	33	55	11	43
● Camps_d_atletisme	49	7	8	9	5	4	1
● Museus	504	216	135	122	96	56	64
● Biblioteques	454	100	73	63	58	35	14
● Oficines_atencio	2	2		2	2		
TOTAL SERVICIOS	8.405	2.613	1.951	1.800	1.405	713	625

PivotTable Fields

ACTIVE | ALL

Choose fields to add to report:

- Σ FAC_EQUIPAMIENTOS
 - Equipamiento Total
 - Equipamientos Por 10.000 H...
 - Equipamientos Por 100.000 ...
- Σ FAC_ESTABLECIMIENTOS
 - NumeroPlazas
 - PlazasPorMilHabitantes
 - PlazasPorKm2
 - Establecimientos Por 100 Eq...
 - Establecimientos Por 10.000 ...
 - NumeroEstablecimientos
 - OfertaMediaPlazas

Drag fields between areas below:

FILTERS

- Subcategoria1
- Comarca

COLUMNS

- Delegacion_G...

ROWS

- Subcategoria2
- Servicio
- Nombre

VALUES

- Equipamiento...

Defer Layout Update UPDATE

4.3 Informe establecimientos – usuarios analistas

The screenshot shows an Excel spreadsheet with a PivotTable titled "ESTABLECIMIENTOS POR TIPO, AÑO Y ÁMBITO". The PivotTable is set to show data by "AMBITO" (rows) and "AÑO" (columns). The data is summarized in the table below:

AMBITO	2006	2007	2008	2009	2010	2011	2012
Barcelona	891	1.050	1.097	1.136	1.174	1.209	1.215
Barcelonès	464	476	505	536	549	568	579
Maresme	228	221	218	212	222	224	211
Vallès Oriental	87	91	102	107	115	116	115
Garraf	74	74	76	75	73	74	73
Vallès Occidental	38	37	39	41	44	50	52
Alt Penedès		68	72	78	81	86	93
Baix Llobregat		83	85	87	90	91	92
Girona	783	1.450	1.494	1.539	1.563	1.527	1.616
Alt Pirineu i Aran	473	735	750	772	774	776	781
Catalunya Central	440	621	650	675	700	721	723
Tarragona	276	505	472	477	540	536	539
Lleida	226	234	230	239	250	319	252
Terres de l'Ebre	126	197	199	203	212	222	220
TOTAL ESTABLECIMIENTOS	3.215	4.792	4.892	5.041	5.213	5.310	5.346

The PivotTable Fields task pane on the right shows the following configuration:

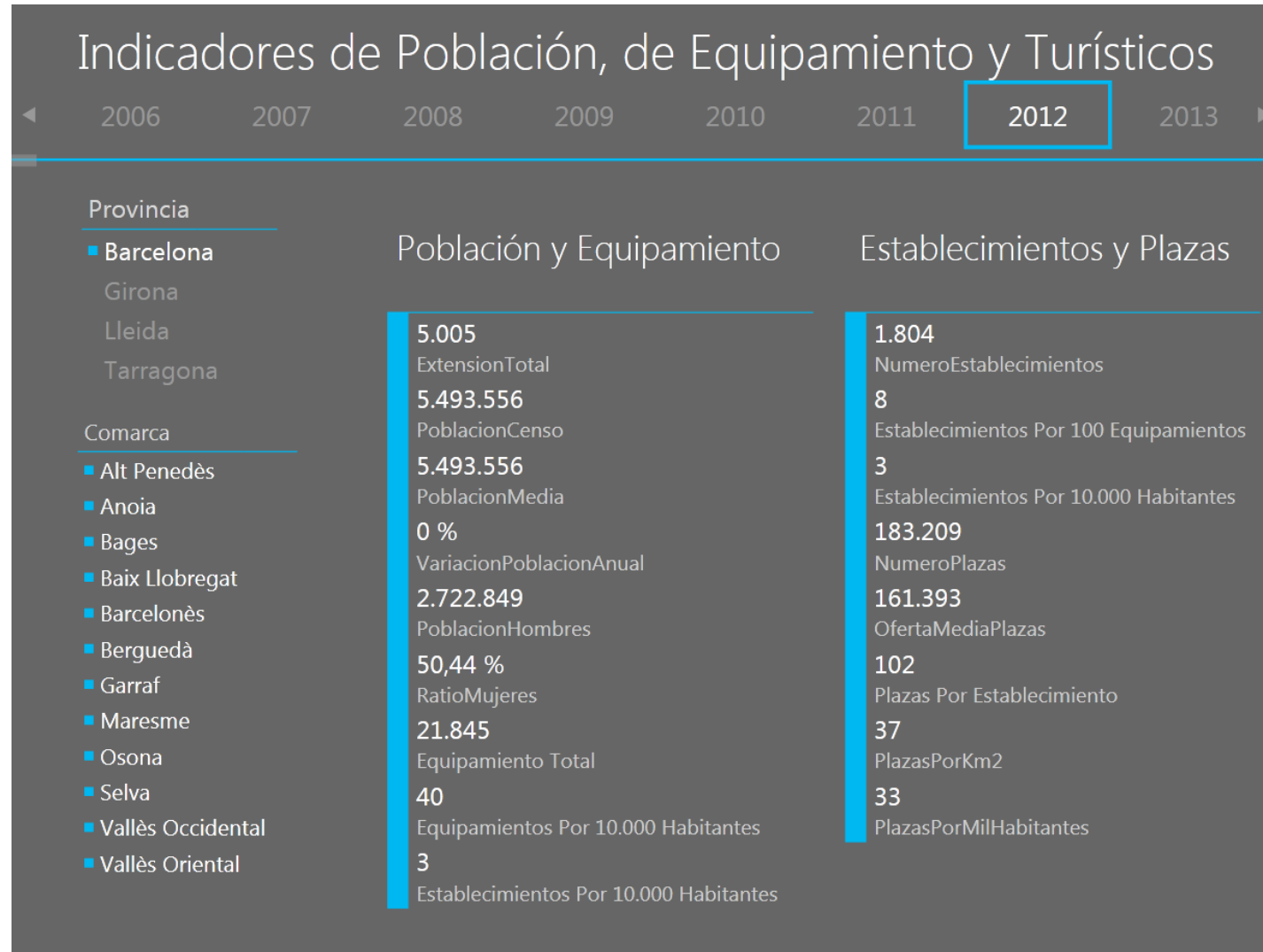
- ACTIVE** | ALL
- Choose fields to add to report:
 - FAC_EQUIPAMIENTOS
 - Equipamiento Total
 - Equipamientos Por 10.000 H...
 - Equipamientos Por 100.000 ...
 - FAC_ESTABLECIMIENTOS
 - NumeroPlazas
 - PlazasPorMilHabitantes
 - PlazasPorKm2
 - Establecimientos Por 100 Eq...
- Drag fields between areas below:
 - FILTERS**: (empty)
 - COLUMNS**: Año
 - ROWS**: Delegacion..., Comarca
 - VALUES**: NUMERO
- Defer Layout Update UPDATE

4.4 Informe población – usuarios analistas

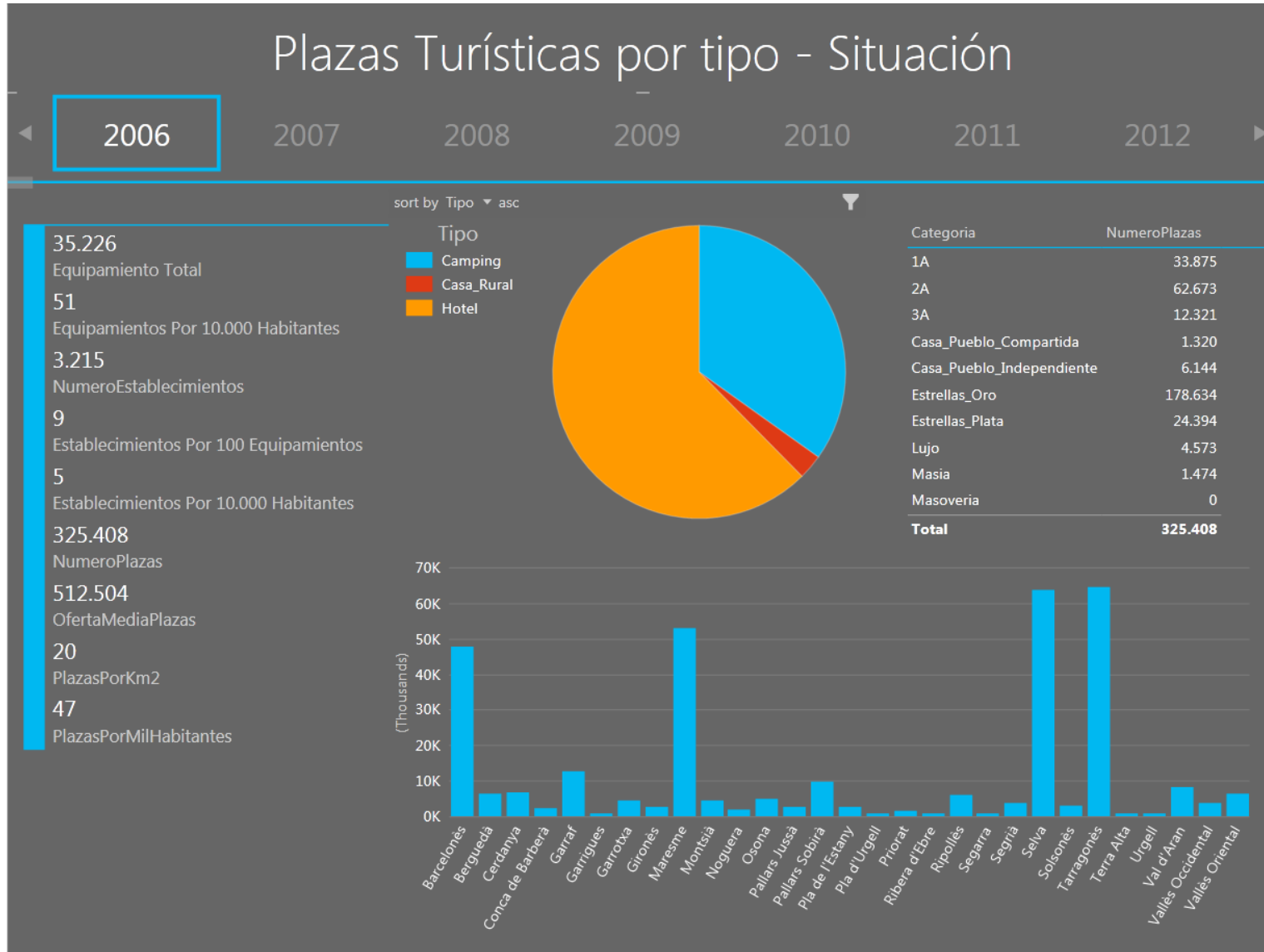
POBLACION POR AÑO Y AMBITO TERRITORIAL

PoblacionMedia	Column La	2006	2007	2008	2009	2010	2011	2012	Tendencia
Alt Pirineu i Aran		52.507	53.893	55.722	56.880	56.873	56.886	57.048	
Barcelona		4.797.127	4.839.830	4.883.084	4.950.615	4.992.480	5.010.794	5.018.821	
Alt Penedès		91.494	93.986	96.271	98.947	100.313	100.881	101.374	
Baix Llobregat		762.891	769.742	776.633	787.702	796.062	801.087	803.705	
Barcelonès		2.221.247	2.219.786	2.224.118	2.243.589	2.251.315	2.248.655	2.246.280	
Badalona		220.037	218.861	215.765	217.438	219.217	219.336	219.786	
Barcelona		1.599.339	1.600.356	1.605.509	1.618.723	1.620.437	1.617.393	1.615.448	
Hospitalet de Llobregat		250.517	249.999	252.815	255.410	257.840	257.354	256.065	
Sant Adrià de Besòs		32.763	32.660	32.979	33.492	33.933	34.131	34.157	
Santa Coloma de Gramene		118.593	117.911	117.051	118.527	119.889	120.442	120.824	
Garraf		130.523	134.723	138.370	141.739	143.862	145.563	146.469	
Maresme		403.509	411.212	416.740	422.906	428.134	432.298	434.253	
Vallès Occidental		825.011	840.134	853.266	869.719	881.777	888.453	891.318	
Vallès Oriental		362.454	370.249,5	377.687	386.014	391.019	393.858	395.422	
Catalunya Central		446.302	455.968	465.555	475.285	480.472	482.613	483.573	
Girona		616.701	636.484	656.954	676.324	686.383	690.484	692.071	
Lleida		291.457	296.968	304.577	314.348	320.593	323.604	324.783	
Tarragona		517.460	539.375	562.075	580.190	588.481	589.189	587.767	
Terres de l'Ebre		162.196	165.863	171.463	175.580	176.919	177.456	177.536	
Grand Total		6.883.748	6.988.380	7.099.429	7.229.221	7.302.199	7.331.025	7.341.599	

4.5 Informe indicadores principales



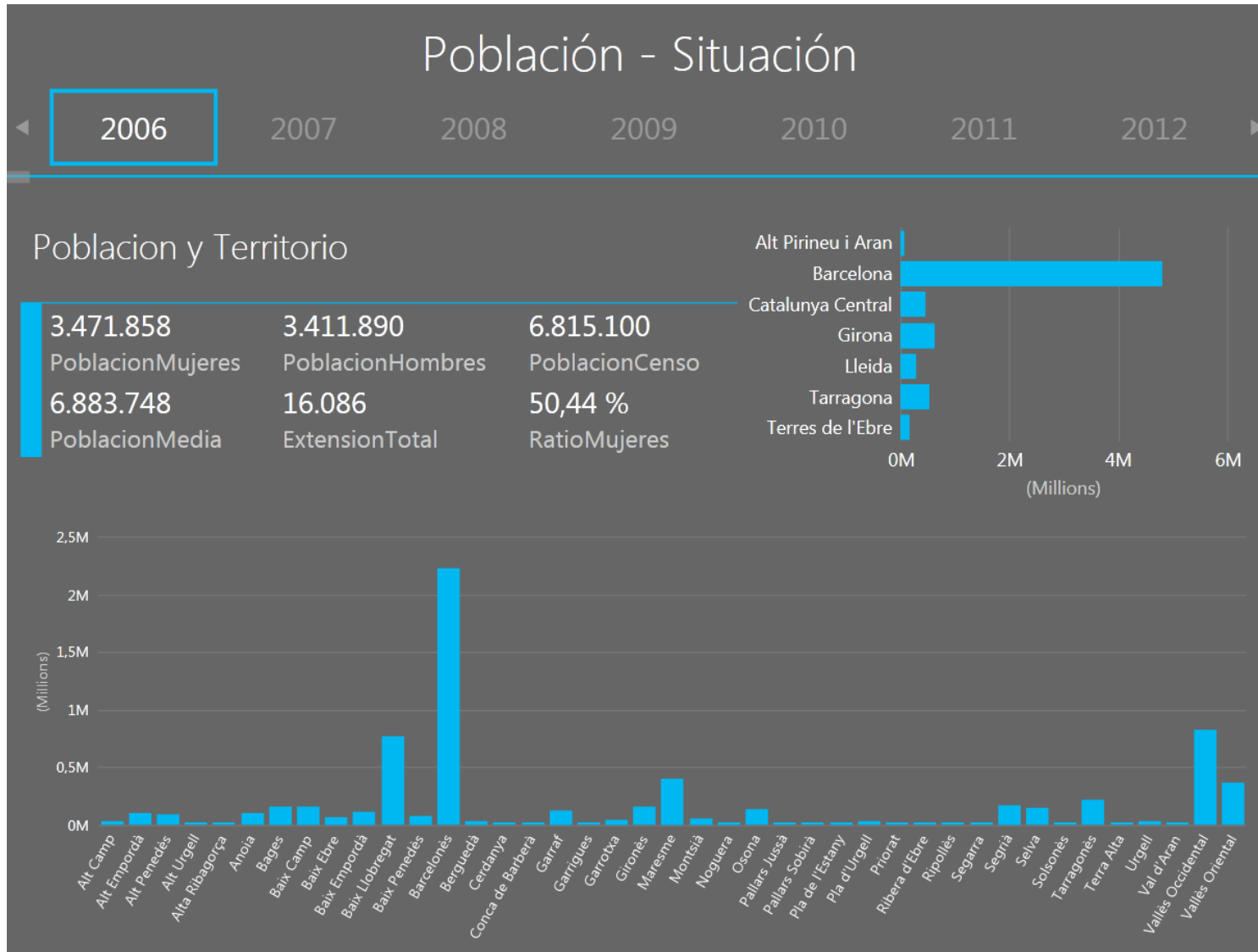
4.6 Informe Plazas turísticas por tipo



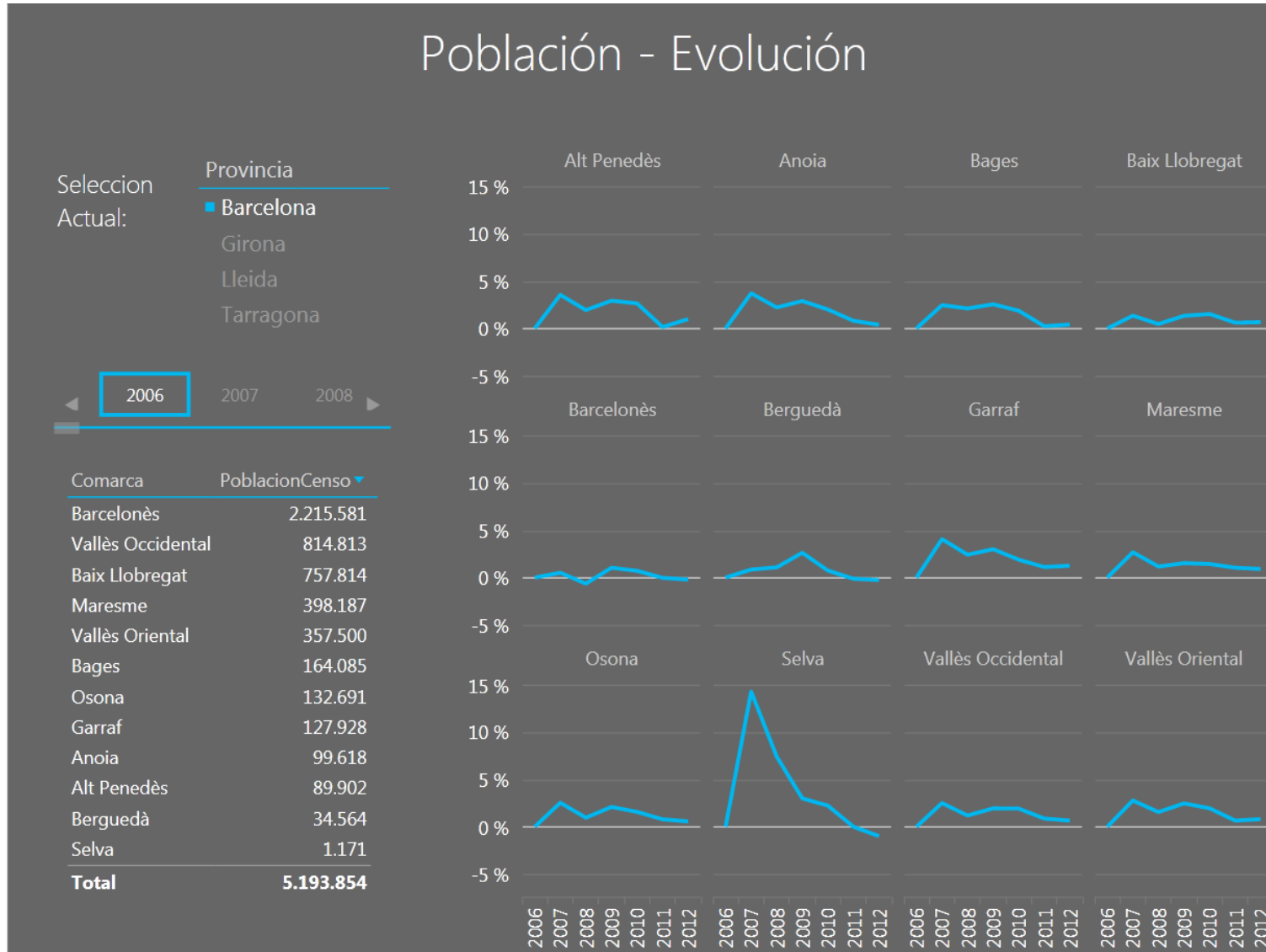
4.7 Informe comparativa equipamientos por comarca



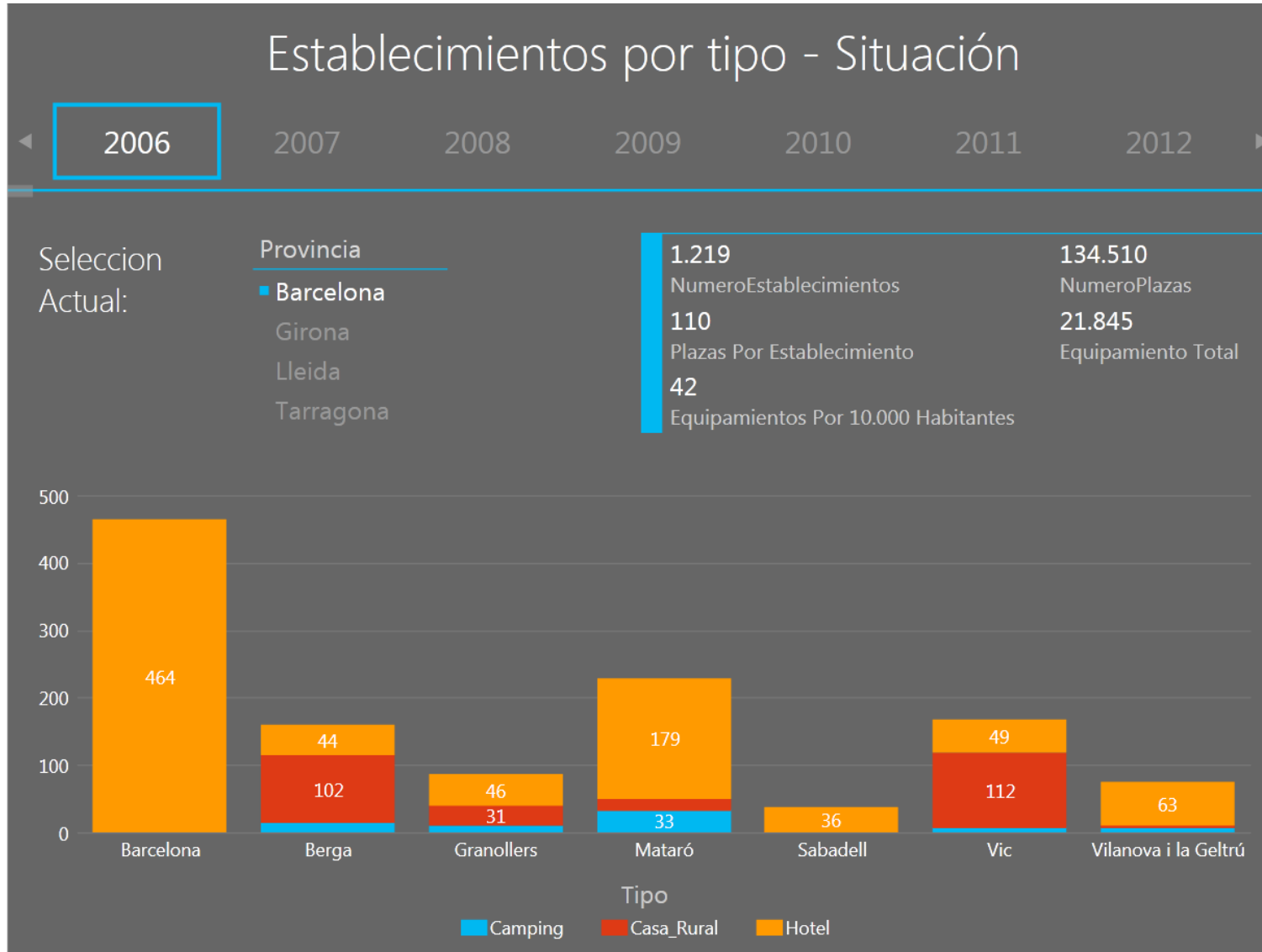
4.8 Informe comparativa de población por comarca



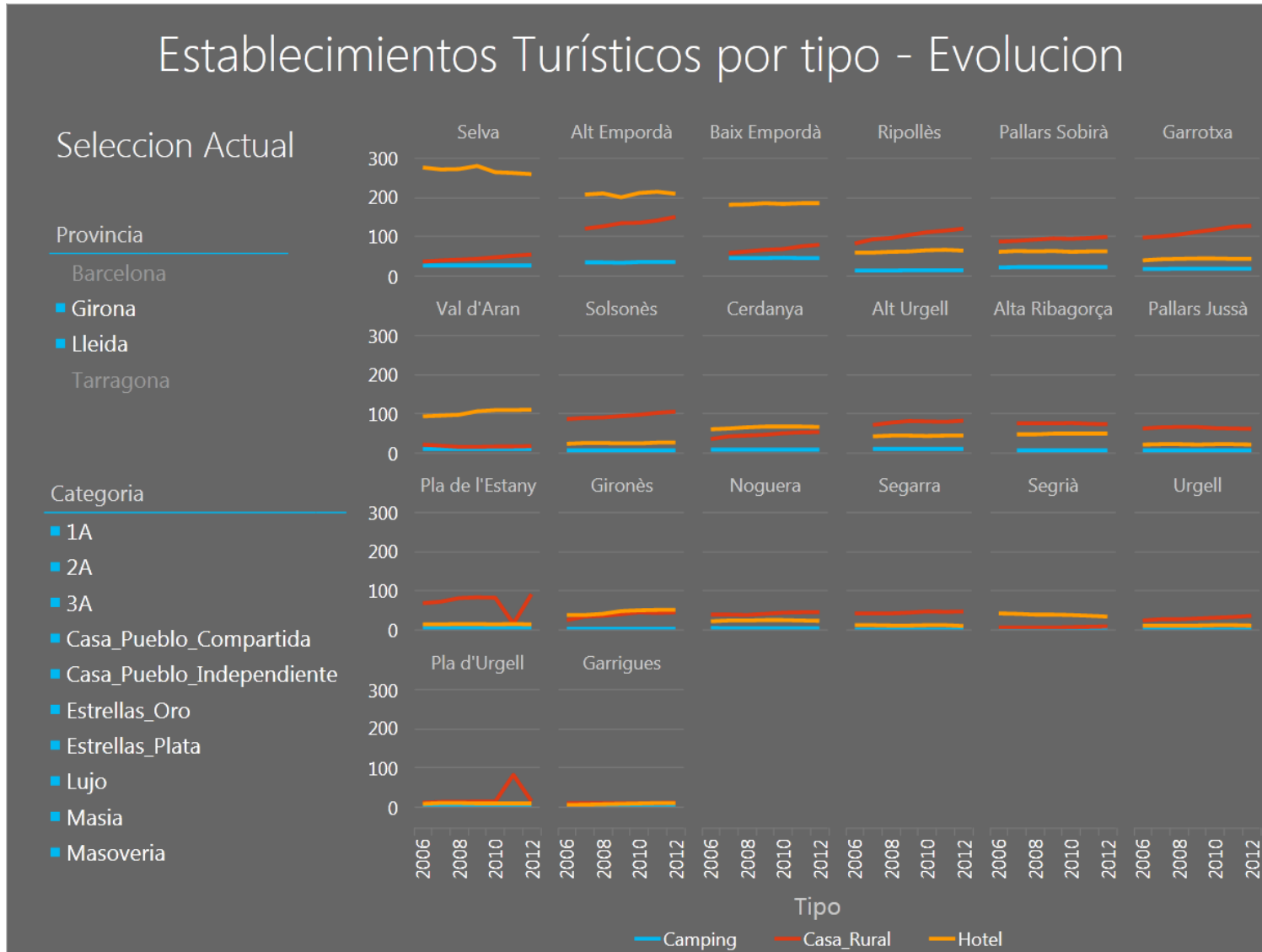
4.9 Informe evolución comparativa de población por comarca



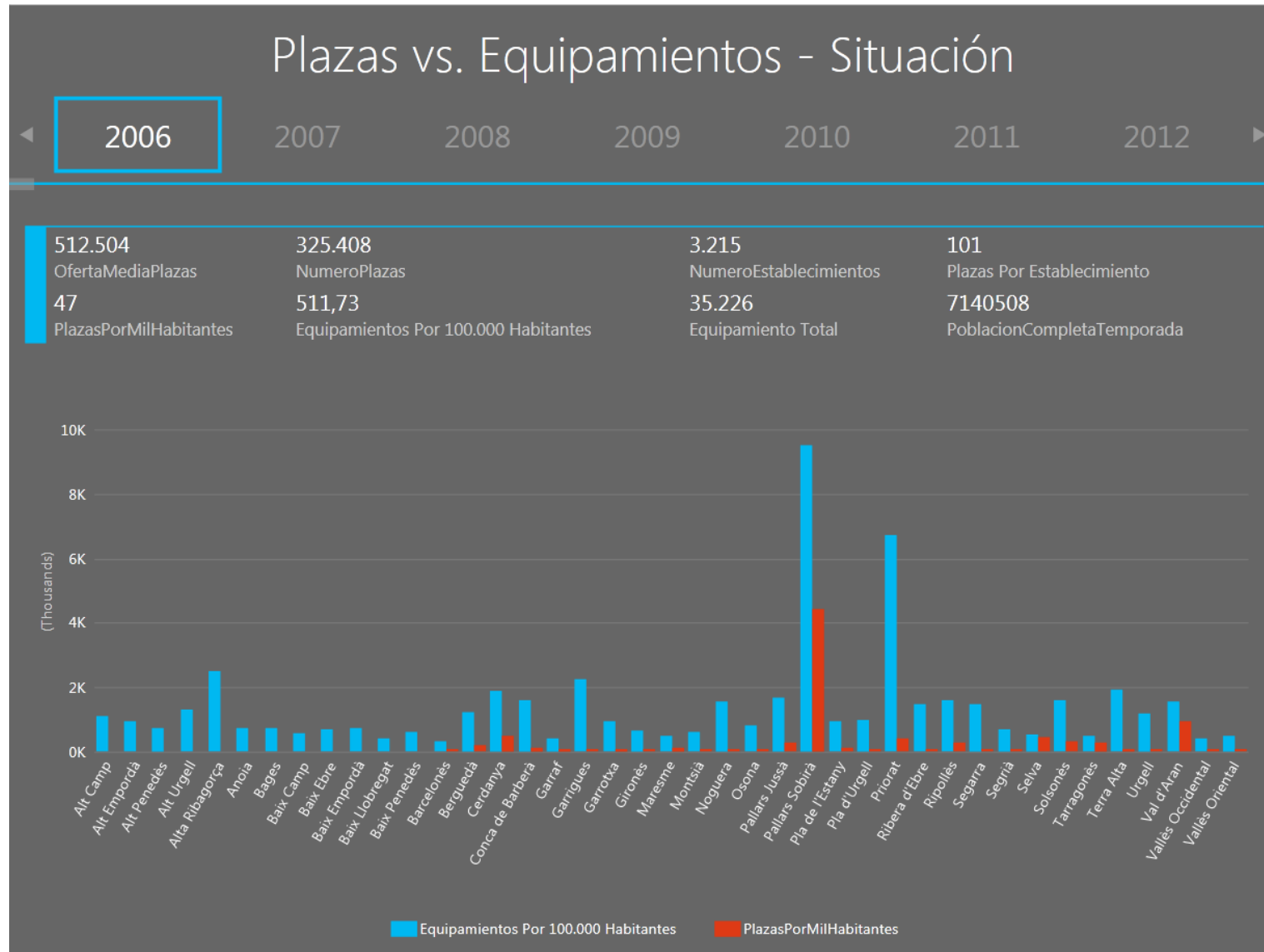
4.10 Informe comparativa número y tipo de establecimientos por comarca



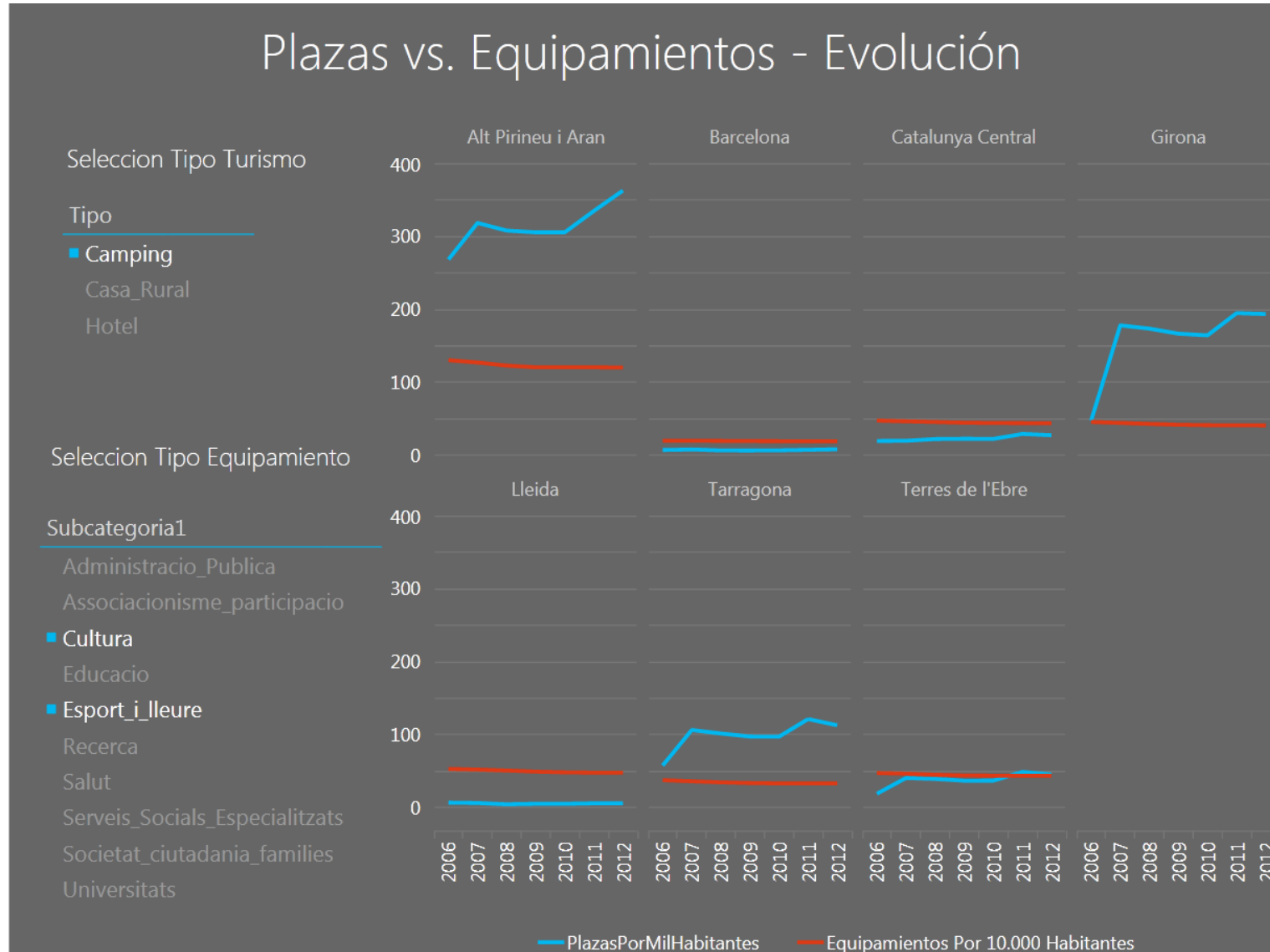
4.11 Informe evolución comparativa número y tipo de establecimientos por comarca



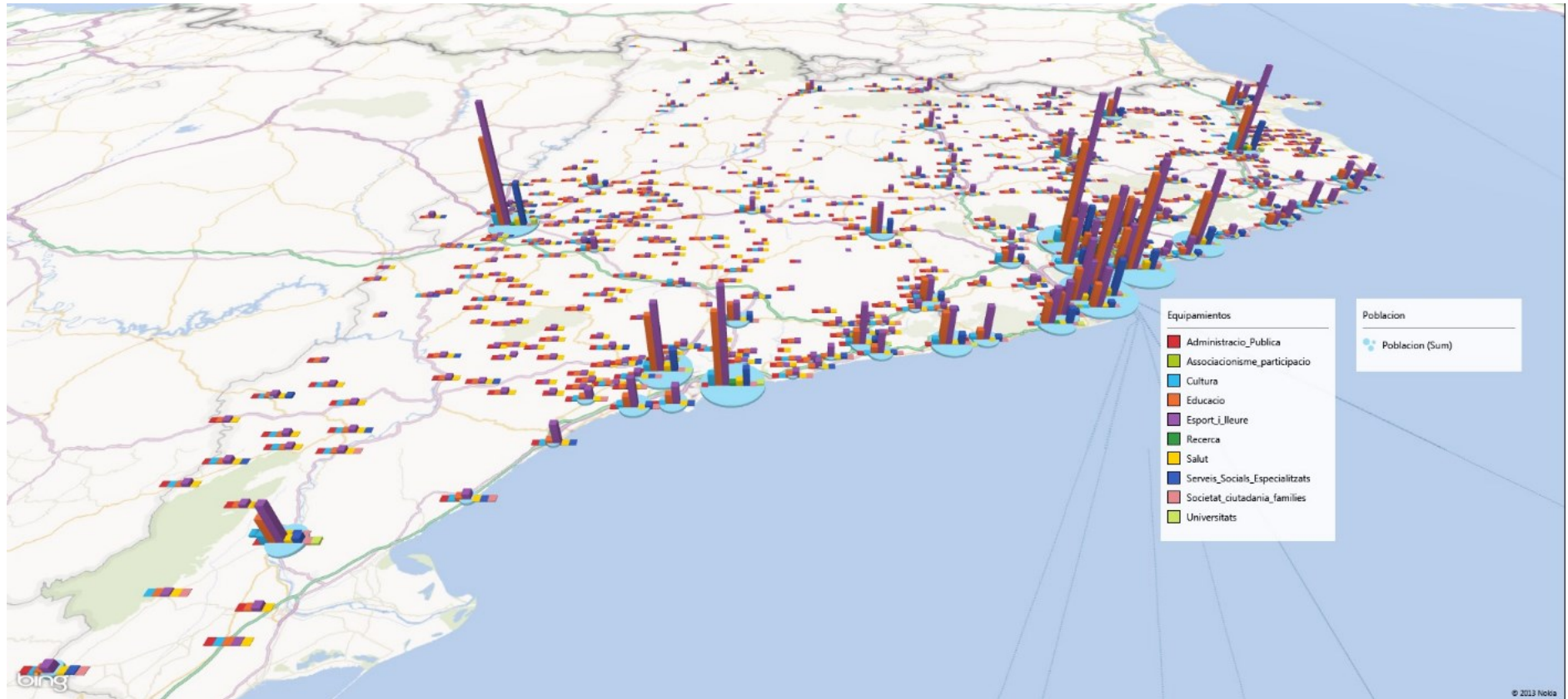
4.12 Informe comparativa plazas y equipamientos relativos a la población por comarca



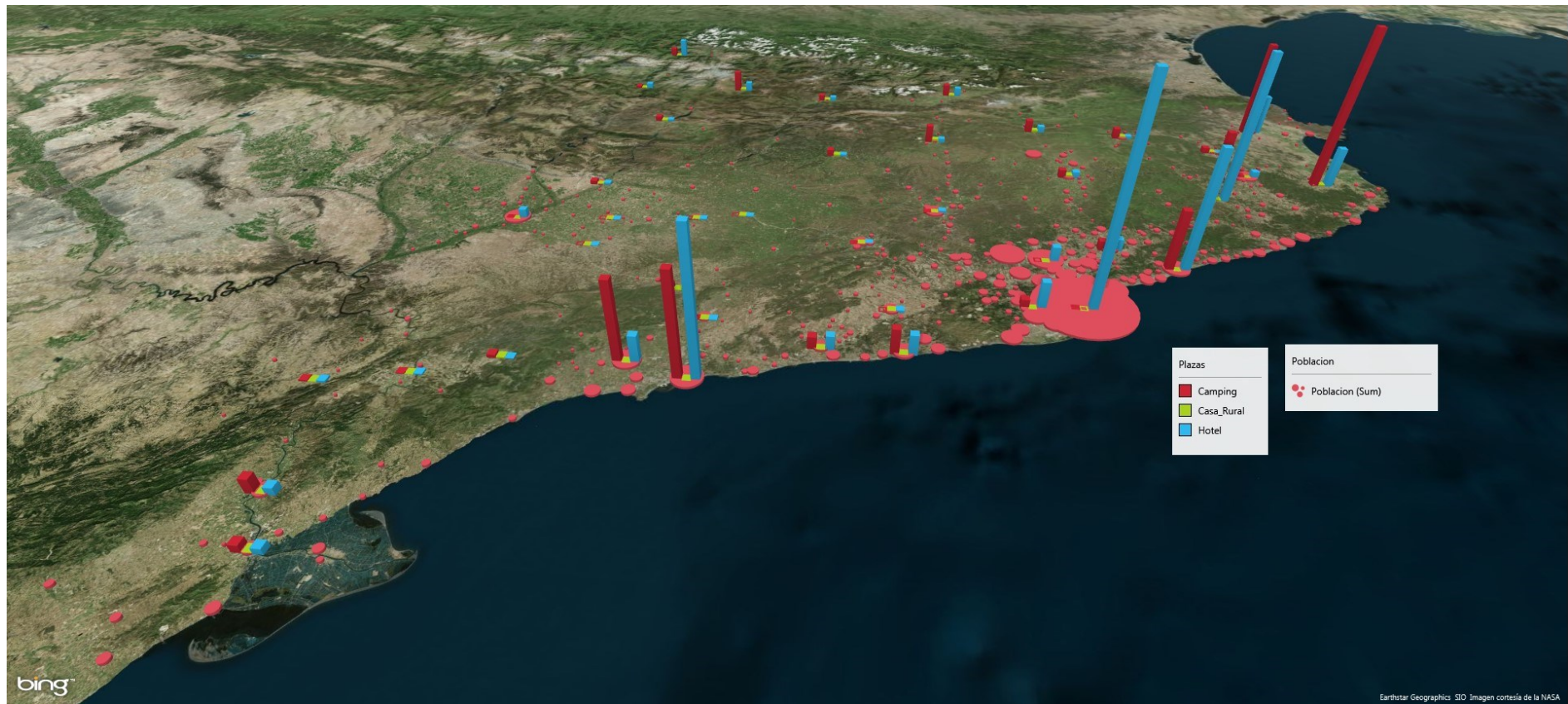
4.13 Informe evolución comparativa plazas y equipamientos relativos a la población



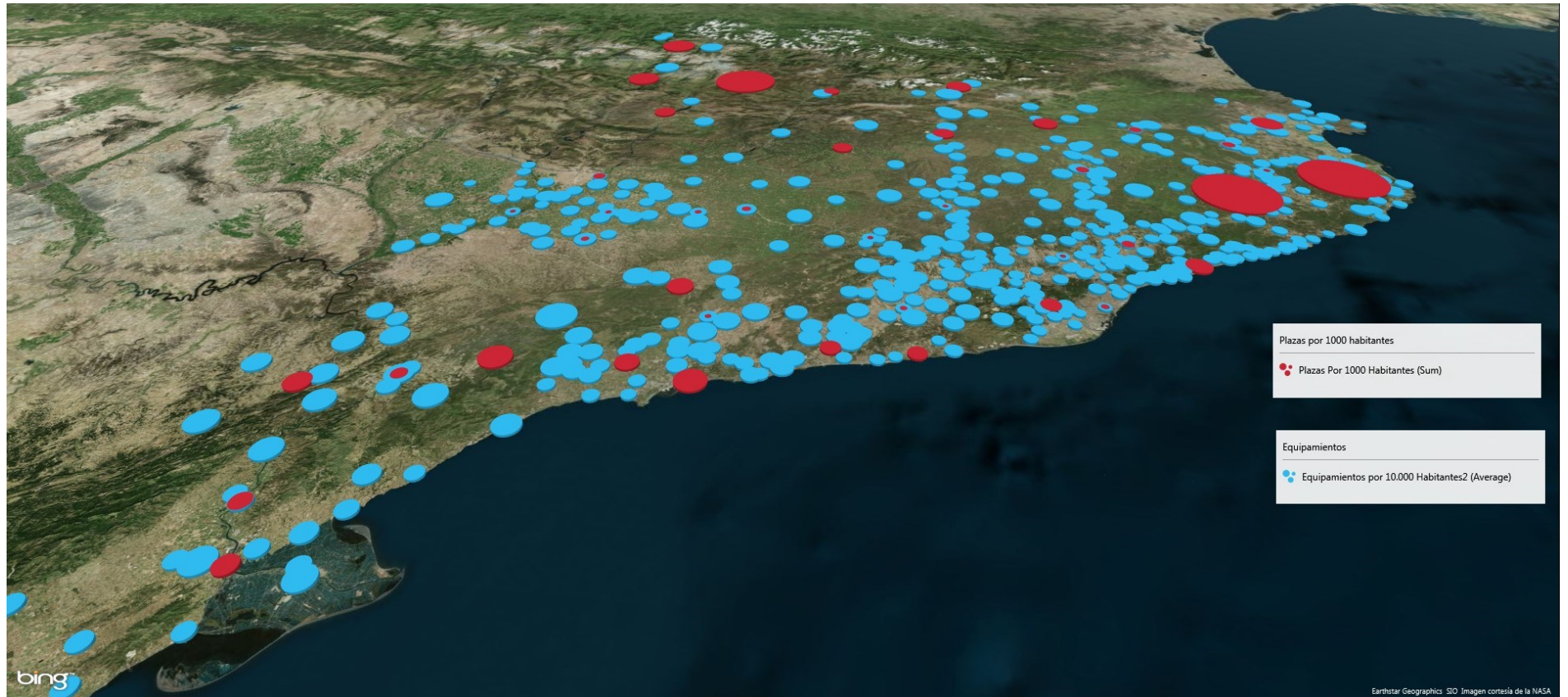
4.14 Informe geoespacial Población y Equipamientos por tipo



4.15 Informe geoespacial de Plazas Turísticas y Población en valores absolutos



4.16 Informe geoespacial de plazas turísticas y equipamientos relativos a la población



5 Líneas de evolución futura

Como parte del programa inicial de ONdO, si el proyecto es exitoso se planteará con toda seguridad la expansión del almacén de datos. Las nuevas peticiones se evaluarían y se presentarían a la Dirección de ONdO para su priorización, y empezaría el ciclo de nuevo, construyendo sobre la base establecida y enfocándonos en los nuevos requerimientos.

En la vertiente tecnológica, para promover el uso de la solución se planteará a ONdO la creación un espacio donde todos los usuarios puedan publicar y consultar de manera ordenada los informes y los análisis que produzcan, y que puedan ser reutilizados por toda la organización, así como un portal donde todos los usuarios puedan publicar y consumir sets de datos para integrar en sus escenarios de análisis. Este entorno colaborativo estaría basado en tecnología *Microsoft Sharepoint*.

Finalmente, en línea con las tendencias más actuales se investigará para ONdO la plausibilidad de llegar a un acuerdo con los principales entes procesadores de pagos de Catalunya para crear un gran almacén de datos con información de pagos en los establecimientos, servicios e instalaciones de Catalunya con el fin de analizar qué tipo de servicios son los más solicitados, más consumidos y por tanto más atractivos para las diferentes clases de turismo.

6 Conclusiones

Durante el desarrollo de este trabajo he constatado una vez más la importancia caudal de los aspectos de gestión proyectos, como son la planificación, la gestión de los recursos y sobre todo la adecuada gestión de los riesgos (!). Para mantener y cultivar la confianza de los clientes es fundamental poder *entregar* (del inglés *deliver*) en alcance y plazos, y para ello es fundamental planificar y sobre todo saber gestionar los recursos disponibles para afrontar con éxito las certeras contingencias.

En cuanto a la parte más técnico del proyecto, he tenido la oportunidad de hacer un repaso de muchos de los conceptos adquiridos durante los estudios, como son el desarrollo modular, los sistemas gestores de bases de datos y conceptos de ingeniería del software en general, así como adquirir conocimientos nuevos y afianzar los ya adquiridos en el ámbito específico de los almacenes de datos. En este sentido considero los objetivos académicos plenamente alcanzados.

A modo de conclusión, a pesar del esfuerzo y los sacrificios, y un poco por ellos también, la valoración final que hago del trabajo realizado es positiva, y considero con alegría este proyecto como mi entrada definitiva en una de las áreas más interesantes y de más proyección dentro de la industria de las tecnologías de la información, y a la que espero seguir próximo en el futuro.

7 Bibliografía

“The Datawarehouse toolkit. The Complete Guide to Dimensional Modeling, Second Edition”

Ralph Kimball, Margy Ross, Wiley, 2002

“The Microsoft Data Warehouse Toolkit: With SQL Server 2008 R2 and the Microsoft Business Intelligence Toolset”

Joy Mundy, Warren Thornthwaite with Ralph Kimball, Wiley, 2008

“Practical Business Intelligence with SQL Server 2005”

By John C. Hancock, Roger Toren, Addison Wesley Professional, 2006