



TRABAJO DE FINAL DE CARRERA

CONSTRUCCIÓN Y EXPLOTACIÓN
DE UN ALMACEN DE DATOS
PARA EL ANÁLISIS DE INFORMACIÓN INMOBILIARIA

José Ángel Santiago

ETGI

CONSULTOR: José Angel Martín
Junio 2008

RESUMEN Y PALABRAS CLAVE

Resumen:

El presente proyecto aborda el diseño y la puesta en funcionamiento del un almacén de datos que sirva de herramienta de apoyo en la toma de decisiones de una empresa inmobiliaria.

Se parte de un conjunto de datos en formato texto procedentes de un proceso de extracción desde un conjunto de hojas de M.S. Excel.

En función del análisis previo de los datos se diseñan los modelos conceptual, lógico y físico que definen el cubo de información multidimensional (cubo OLAP) que ha de dar soporte al almacén de datos

Con posterioridad se crean las tablas y se procede a la transformación de los datos, su carga en la base de datos y la creación de los índices pertinentes.

Por ultimo se crea un área de negocio que sirve de base al diseño e implementación de una serie de consultas e informes que constituyen el resultado visible del proyecto.

Palabras clave:

Almacén de datos, Data warehouse, Gestión inmobiliaria, ETL, OLAP, Consultas, Oracle, Discoverer, SQL, PL/SQL.

ÍNDICE

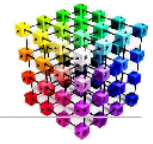
RESUMEN Y PALABRAS CLAVE.....	2
RESUMEN:	2
PALABRAS CLAVE:	2
1. INTRODUCCIÓN	6
1.1. JUSTIFICACIÓN.....	6
1.2. OBJETIVOS	7
1.3. OBJETIVOS DEL TFC	7
1.3.1. <i>Relación con los objetivos de la asignatura</i>	7
1.3.2. <i>Objetivos concretos del proyecto</i>	8
1.3.3. <i>Informes a realizar</i>	8
1.4. ENFOQUE Y METODOLOGÍA.....	9
1.5. PLANIFICACIÓN	10
1.5.1. <i>Fases del proyecto según el plan docente</i>	11
1.5.2. <i>Adaptación a fases del plan docente</i>	11
1.5.3. <i>Hitos y fechas claves</i>	12
1.5.4. <i>Planificación de tareas</i>	12
1.5.5. <i>Diagrama de Gantt</i>	13
1.5.6. <i>Incidencias y Riesgos</i>	16
1.5.7. <i>Desviación sobre la planificación prevista</i>	18
1.6. PRODUCTOS	18
1.7. OTROS CAPÍTULOS	19
2. ANÁLISIS.....	19
2.1. DATOS DE PARTIDA	19
2.2. MODELO CONCEPTUAL	20
2.2.1. <i>Identificar el Hecho</i>	21
2.2.2. <i>Encontrar la granularidad adecuada</i>	21
2.2.3. <i>Escoger las dimensiones a emplear en el análisis</i>	22
2.2.4. <i>Establecer los atributos de cada dimensión</i>	22
2.2.5. <i>Distinguir entre descriptores y jerarquías de agregación</i>	23
2.2.6. <i>Identificar las medidas a emplear</i>	23
2.2.7. <i>Definir las celdas</i>	23
2.2.8. <i>Explicitar las restricciones de integridad</i>	24
2.2.9. <i>Estudiar la viabilidad</i>	24
2.2.10. <i>Resultado</i>	25
3. DISEÑO	26
3.1. ARQUITECTURA DEL HARDWARE	26
3.2. ARQUITECTURA DEL SOFTWARE	26
3.3. DISEÑO LÓGICO.....	27
3.4. DISEÑO FÍSICO	27
3.4.1. <i>Ofertas</i>	28
3.4.2. <i>Inmuebles</i>	28
3.4.3. <i>Tiempo</i>	28
3.4.4. <i>Zona</i>	28
3.4.5. <i>Planificación de las consultas</i>	29
3.5. PROCESOS ETL.....	33
3.5.1. <i>Proceso de Precarga</i>	33
3.5.2. <i>Proceso de Transformación</i>	35
3.5.3. <i>Proceso de Carga</i>	38
3.5.4. <i>Creación de índices</i>	40
3.6. CREACIÓN DE LAS CONSULTAS	40
3.6.1. <i>Proceso de creación</i>	40

3.6.1.1.	<i>Creación del área de negocio</i>	40
3.6.1.2.	<i>Creación del libro de trabajo</i>	42
3.6.1.3.	<i>Creación de las consultas</i>	43
3.6.2.	<i>Consultas creadas</i>	43
3.6.2.1.	<i>Precios máximos, mínimos y medios por zona tipología y características</i>	44
3.6.2.2.	<i>Superficies máximas, mínimas y medias por zona tipología y características</i>	45
3.6.2.3.	<i>Tiempos de venta máximos mínimos y medios por zona tipología y características</i>	45
3.6.2.4.	<i>Inmuebles vendidos por zona tipología y características</i>	47
3.6.2.5.	<i>Inmuebles disponibles por zona tipología y características</i>	47
3.6.2.6.	<i>Inmuebles que bajan de precio en un mes con respecto al mes anterior</i>	48
3.6.2.7.	<i>Inmuebles que suben de precio en un mes con respecto al mes anterior</i>	50
3.6.2.8.	<i>Inmuebles que bajan de precio en un mes con respecto al precio inicial</i>	50
3.6.2.9.	<i>Inmuebles que suben de precio en un mes con respecto al precio inicial</i>	51
3.6.2.10.	<i>Inmuebles que cambian de precio en un mes con respecto al precio inicial</i>	52
3.6.2.11.	<i>Evolución anual de precios por zona y tipología de los inmuebles que cambian de precio</i>	53
3.7.	DETERMINACIÓN DEL INMUEBLE TIPO	53
3.7.1.	<i>Medidas de tendencia central</i>	54
3.7.2.	<i>proceso y resultados</i>	55
4.	CONCLUSIONES	56
5.	LINEAS DE EVOLUCIÓN FUTURA	57
6.	GLOSARIO	58
7.	BIBLIOGRAFÍA	59
7.1.	PUBLICACIONES:	59
7.2.	WEBS:	59
8.	ANEXOS	60
8.1.	RESTAURACIÓN DE LA BASE DE DATOS	60



Índice de Figuras

Figura 1. <i>Diagrama de Gantt (parte I)</i>	14
Figura 2. <i>Diagrama de Gantt (parte II)</i>	15
Figura 3 <i>Modelo conceptual</i>	25
Figura 4 <i>Arquitectura del hardware</i>	26
Figura 5 <i>Arquitectura del Software</i>	26
Figura 6 Esquema del diseño lógico	27
Figura 7 Esquema físico. Tabla ofertas.....	28
Figura 8 Esquema físico. Tabla inmuebles	28
Figura 9 Esquema físico. Tabla tiempo	28
Figura 10 Esquema físico. Tabla zona.....	28
Figura 11 <i>Ejecución del script TEMPORAL_UOC en SQLPLUS</i>	34
Figura 12 <i>Carga de los datos mediante sqldr</i>	34
Figura 13 <i>Ejecución del fichero de sentencias de correccion</i>	37
Figura 14 <i>Creación de las tablas finales</i>	38
Figura 15 <i>Ejecución del script de carga de datos</i>	39
Figura 16 <i>Creación del EUL</i>	41
Figura 17 <i>Datos, jerarquías y clases de elementos definidos en el área de trabajo</i>	42
Figura 18 <i>Consulta 1</i>	44
Figura 19 <i>Consulta 2</i>	45
Figura 20 <i>Consulta 3</i>	46
Figura 21 <i>Consulta 4</i>	47
Figura 22 <i>Consulta 5</i>	48
Figura 23 <i>Inserción de un campo calculado</i>	48
Figura 24 <i>Consulta 6</i>	49
Figura 25 <i>Inserción de una condición</i>	49
Figura 26 <i>Consulta 7</i>	50
Figura 27 <i>Consulta 9</i>	51
Figura 28 <i>Consulta 10</i>	52
Figura 29 <i>Inserción de una condición</i>	52
Figura 30 <i>Consulta 11</i>	53



1. INTRODUCCIÓN

1.1. JUSTIFICACIÓN

El presente proyecto nace como consecuencia de la petición realizada por la empresa inmobiliaria “Sostre per a tothom” que ha encargado el diseño y la realización de un almacén de datos, que sirva de ayuda en la toma de decisiones que permitan ajustar la oferta y la demanda teniendo como base el análisis de las zonas geográficas, las tipologías de los pisos ofertados, los precios de venta, etc.

Nadie pone en duda hoy en día la importancia que tiene la calidad de la información en los procesos de toma de decisiones. Las empresas que desean ser competitivas deben orientar sus decisiones en función de información concreta y fiable.

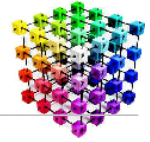
Es habitual que las empresas cuenten con diferentes bases de datos organizadas por departamentos o áreas de negocio, pero según aumenta el volumen de estos datos se hace cada vez más difícil extraer información global y resumida.

Se hace por tanto necesario unificar las diferentes fuentes de información en un único repositorio, al que se incorpore solamente información relevante, bien organizada, estructurada, integrada, y de fácil explotación. Esta unificación se consigue con la implantación de almacenes de datos.

El Almacén de Datos o Data Warehouse solicitado y que constituye el centro de este proyecto, puede definirse como una colección de datos orientado a temas, integrado, no volátil, de tiempo variante, que se emplea fundamentalmente para dar soporte al proceso de toma de decisiones.

Según la definición de Bill Inmon, un Data Warehouse debe poseer las siguientes características:

- **INTEGRADO:** Los datos que residen en el Data Warehouse deben estar integrados en una estructura consistente, lo que implica que las posibles inconsistencias existentes entre los diversos sistemas operacionales deberán ser eliminadas.
- **TEMÁTICO:** Los datos se han de organizar por temas para facilitar su acceso y permitir su entendimiento por parte de los usuarios finales, y sólo los datos necesarios para el proceso de generación del conocimiento del negocio en cuestión se integran desde el entorno operacional.
- **HISTÓRICO:** El almacén de datos se carga con los distintos valores que toma una variable en el tiempo, de forma que se generan históricos que son como fotografías de los diferentes momentos por los que ha se ha pasado. Esto es fundamental para poder realizar análisis sobre las tendencias.
- **NO VOLÁTIL:** El almacén de datos puede ser leído pero no modificado. Es decir, la incorporación de los últimos valores que pueden tomar las distintas variables contenidas



en él, no debe influir sobre la información que ya existía previamente.

En los siguientes apartados de este documento se describen las fases y los procesos seguidos en la elaboración del almacén de datos solicitado.

1.2. OBJETIVOS

En este apartado se describen los objetivos del proyecto entendido como trabajo final de carrera y en relación con los objetivos generales de la asignatura, así como los objetivos específicos del proyecto como tal.

1.3. Objetivos del TFC

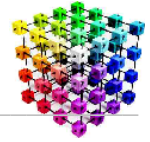
El objetivo fundamental de este TFC es la creación y puesta en funcionamiento de un almacén de datos, partiendo de la información procedente de bases de datos transaccionales. Este objetivo general del proyecto se puede descomponer en una serie de objetivos puntuales más concretos que son:

- Consolidar los conocimientos previamente adquiridos acerca de como elaborar un documento de especificaciones basado en los requerimientos de usuario.
- Obtener experiencia en la planificación y secuenciación de un proyecto de implantación de un almacén de datos.
- Adquirir conocimientos sobre el diseño de almacenes de datos.
- Obtener experiencia en la explotación de almacenes de datos.
- Entender la importancia del uso de lenguajes como PL/SQL en la implantación, gestión y explotación de un almacén de datos.
- Obtener experiencia en el uso de las herramientas que facilitan la explotación de los datos y la generación de informes.

1.3.1. Relación con los objetivos de la asignatura

Los objetivos de este TFC consisten en la aplicación a un caso particular los objetivos generales de la asignatura, que entre otros son:

- Familiarizarse con la estructura general de un Data Warehouse.
- Asimilar los aspectos teóricos del diseño de un Data Warehouse.
- Entrar en contacto con los lenguajes de consultas de bases de datos.



- Manejar y asimilar conceptos como dimensiones, atributos, hechos, desnormalización, agregación, etc.

1.3.2. Objetivos concretos del proyecto

El objetivo general del proyecto “Sostre per a tothom” consiste en la creación de un Almacén de Datos que dé soporte a la toma de decisiones de la empresa inmobiliaria. Este objetivo general se puede descomponer en objetivos más sencillos, prácticos y fácilmente abordables:

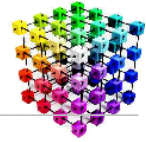
- Primer Objetivo: Crear un proceso automatizado de importación de los datos facilitados por “Sostre per a tothom” que se encontraran en formato texto, al sistema de gestión de bases de datos elegido Oracle v.10. Este proceso de importación debe seleccionar solamente la información de interés para las posibles consultas a fin de no sobrecargar el sistema y debe evitar los datos que sean claramente inconsistentes o erróneos. Para lograr este objetivo será necesario un análisis previo de los datos que ayudará a determinar su disposición y los posibles errores recurrentes que se puedan encontrar.
- Segundo Objetivo: Diseñar y poner en marcha un almacén de datos que aproveche la información de partida facilitada por “Sostre per a tothom” y facilite la generación de determinados informes que constituyan una buena herramienta de toma de decisiones.
- Tercer Objetivo: La aplicación desarrollada debe ser capaz de generar en un tiempo aceptable una serie de informes automatizados, que muestren determinados aspectos referentes a las propiedades en venta y a las ya vendidas tal como se detalla más adelante en otro apartado de este mismo documento. Además de estos informes preestablecidos la aplicación debe estar construida de manera tal que facilite la creación de consultas “*ad-hoc*”.

1.3.3. Informes a realizar.

La empresa ha solicitado una serie de informes automatizados que se han de implementar. Estos informes se pueden agrupar según su contenido de la siguiente manera:

- Inmuebles: vendidos, en oferta y totales, por zona, tipología, características,
- Precios: Evolución de precios (subidas, bajadas, máximos mínimos y medios) por zona, tipología y características
- Metros cuadrados: máximos, mínimos y medios por zona, tipología y características
- Tiempos de ventas: medios por zonas, tipología y características

Estos informes se pueden complementar con otros que se consideren interesantes durante el proceso de desarrollo y además la aplicación debe permitir y facilitar la generación de informes “*ad-hoc*”.



1.4. ENFOQUE Y METODOLOGÍA

La implantación de un Data Warehouse en cualquier ámbito en el que sea necesaria la toma de decisiones basadas en informes, se puede descomponer una serie de pasos o fases claramente diferenciadas, algunas de las cuales pueden variar según el autor que consultemos, pero otras, las más comunes, son ampliamente aceptadas.

En el desarrollo de este proyecto de final de carrera se ha seguido la metodología descrita en www.1keydata.com adaptándola a los requerimientos y la temporalidad impuestos por la signatura:

1. Análisis de requerimientos

Esta primera fase del proyecto consiste en recoger los requerimientos de los usuarios, que generalmente no poseen conocimientos técnicos y con ellos elaborar el documento de requisitos. De esta primera fase ha de salir también una planificación temporal y una aproximación al cubo OLAP que sirve de base a los informes.

2. Preparación del soporte físico

En esta fase se configura y se deja en funcionamiento todo el soporte físico del Data Warehouse. Esto incluye el hardware necesario y el sistema de gestión de bases de datos a emplear

3. Generación del modelo de datos

De los requerimientos de los usuarios se obtiene un modelo lógico que es trasladado a un modelo físico de datos. Una parte importante de este paso es la correcta identificación de las fuentes de datos.

4. ETL

La fase de ETL (Extraction, Transformation, Loading) es una de las fases más largas y comprende la extracción de los datos, su transformación y depuración y la carga de los mismos en la base de datos.

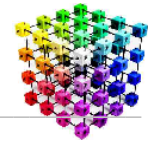
5. Diseño del cubo OLAP

En esta fase se define ya de forma más precisa el cubo OLAP que se deriva de los requerimientos de los usuarios y de los informes que se han de generar.

6. Desarrollo del Front End

Durante esta fase se define el Front End a emplear. Generalmente se suele optar por una herramienta comercial, aunque también puede ser diseñado a medida. En el proyecto que nos ocupa se ha optado por emplear Oracle Discoverer.

7. Ajuste del rendimiento



En esta fase se verifican y ajustan los rendimientos de los procesos que más tiempo y potencia de cálculo pueden consumir. Generalmente el análisis se centra en los procesos ETL, las consultas y los informes.

8. Aseguramiento de la calidad

En esta fase el equipo de desarrollo, junto con los representantes de los futuros usuarios verifica que el software desarrollado cumple las funciones deseadas.

9. Puesta en producción

En esta fase se pone en marcha el DW en condiciones de producción.

10. Mantenimiento

Incluye las tareas de mantenimiento posteriores a la puesta en producción.

11. Mejoras incrementales

Contempla los posibles añadidos no recogidos en el proyecto inicial.

Como se ha indicado anteriormente no todos los autores reconocen las mismas fases, y algunas de ellas no tienen sentido en un proyecto de final de carrera como este.

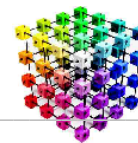
Es evidente que el mantenimiento en fase de producción y las mejoras incrementales están claramente fuera del alcance de este TFC. La fase de aseguramiento de la calidad o garantía de calidad se basa fundamentalmente el contacto y las pruebas con usuarios reales o representantes de un usuario tipo, y dado que este contacto no es posible dentro del alcance de este proyecto, por lo que esta fase tampoco ha sido llevada a la práctica.

Por lo que respecta a la fase de ajuste del rendimiento y las prestaciones, para evitar dilatar en exceso el proyecto, el contenido de esta fase se integrará en otras fases, fundamentalmente en la de ETL, y en la de creación de las consultas sobre el Front-End.

1.5. PLANIFICACIÓN

Para poder realizar adecuadamente la planificación del proyecto, es necesario conocer en primer lugar las tareas a realizar, distribuir las correctamente y adaptarlas a las fechas de desarrollo propuestas.

En este apartado tomando como base las fases típicas de la implantación de un Almacén de Datos comentadas en el punto anterior, se comparan y combinan estas con las fases propuestas en el programa de la asignatura y finalmente, se adaptan e integran las fases resultantes a las fechas claves previstas en el del plan docente. Por último se realiza una distribución de la carga de trabajo dentro de cada fase para lograr una planificación equilibrada.



1.5.1. Fases del proyecto según el plan docente

Del enunciado del proyecto y del plan docente de la signatura se deducen las siguientes fases:

- Plan de trabajo (cronograma) y análisis preliminar de requisitos.
- Análisis de requisitos y diseño conceptual y técnico
- Implementación:
 - Construcción del almacén de datos: bases de datos, cargas, etc.
 - Instalación de la herramienta de explotación de datos.
 - Construcción de los informes y análisis de la información.

1.5.2. Adaptación a fases del plan docente.

En este apartado se explica como se ajustan las fases mas comúnmente aceptadas de desarrollo de un DW al plan docente de la asignatura. Algunas fases no se han llevado a cabo, como ya se ha indicado en párrafos anteriores y otras se variaran y se reorganizarán para adaptarlas al plan de trabajo de la asignatura según el siguiente esquema.

• PAC I

La PAC I coincide plenamente con la fase Análisis de requerimientos (1) de un proceso estándar de implantación de un DW. En ella se generará un primer esbozo del análisis de requisitos y la planificación del trabajo. De esta fase se obtiene también un primer análisis de los datos y una aproximación somera de los elementos que definirán el análisis OLAP.

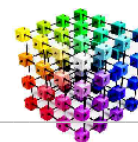
Dada la naturaleza de este TFC, a diferencia de la estructura aconsejada para un proyecto de implantación real, ha sido en esta primera fase cuando se ha realizado la preparación del hardware y la instalación del software necesario, trabajos estos que se corresponden con la fase (2) De esta manera se pretendió lograr lo antes posible la familiarización con las herramientas a emplear.

• PAC II

Según el programa de la asignatura en esta fase se incluyen un análisis de requisitos más completo y los diseños conceptual y técnico. Para completar los pasos típicos de un proyecto DW expuestos anteriormente, en esta fase se continuó el análisis y la depuración de los datos iniciado en la fase (1) y se realizó también la generación del modelo de datos (fase 3), que dará como resultado la obtención del modelo lógico y el modelo físico.

• PAC III

Según la planificación de la asignatura esta PAC es la de implementación, y esta constituida por tres apartados que son, la construcción del almacén de datos (bases de datos, cargas, etc.), la instalación de la herramienta de explotación de datos y la



generación de los informes. Esta fase se corresponden con el final de la fase ETL (4), y con la fase de puesta en producción (9).

1.5.3. Hitos y fechas claves

El plan docente de la asignatura recoge los siguientes hitos y fechas claves que han de ser tenidos en cuenta a la hora de elaborar la planificación del TFC:

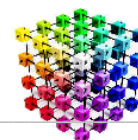
Evento	Fechas
Inicio	28/02/2008
PAC I	16/03/2008
PAC II	20/04/2008
PAC III	18/05/2008
Entrega Final	08/06/2008
Debate	18/06/2008 - 26/06/2008

Estas fechas han actuado como referencias inamovibles que han marcado la organización y el ritmo de la planificación y la distribución de tareas del TFC.

1.5.4. Planificación de tareas

La siguiente tabla muestra la distribución de tareas y la planificación del trabajo en fases temporales concretas tal como se concibió en las primeras fases del proyecto. Esta planificación resultó del compendio de las fases de diseño e implantación de un Data Warehouse combinadas con los hitos y fechas claves de la asignatura.

ID	Nombre de la tarea	Duración	Comienzo	Fin	Predec.
1	1. Arranque del proyecto	3 días	28-feb-08	1-mar-08	
2	1.1. Descarga y estudio del material de la asignatura: Plan docente, tablones de anuncios, foros, etc.	3 días	28-feb-08	1-mar-08	
3	1.2. Búsqueda preliminar de bibliografía	2 días	29-feb-08	1-mar-08	
4	2. Elaboración de un avance de la PAC1	14 días	1-mar-08	10-mar-08	
5	2.1. Descarga y lectura del enunciado del TFC	1 día	1-mar-08	1-mar-08	
6	2.2. Búsqueda de bibliografía y referencias	1 día	3-mar-08	3-mar-08	
7	2.3. Consulta de las dudas surgidas al leer el enunciado	1 día	4-mar-08	4-mar-08	5
8	2.4. Primeras consultas de la bibliografía	3 días	5-mar-08	7-mar-08	6
9	2.5. Elaboración del borrador de un análisis preliminar de los datos	5 días	6-mar-08	9-mar-08	
10	2.6. Elaboración del borrador del primer avance del documento de requisitos	5 días	6-mar-08	9-mar-08	
11	2.7. Elaboración del borrador del documento de planificación.	9 días	6-mar-08	10-mar-08	
12	2.8. Entrega de los borradores generados	0 días	9-mar-08	9-mar-08	9;10
13	2.9. Preparación del hardware e instalación del software recomendado	3 días	5-mar-08	7-mar-08	
14	2.10. Familiarización con el software recomendado	3 días	8-mar-08	9-mar-08	13
15	3. Elaboración de la PAC1	8 días	10-mar-08	16-mar-08	
16	3.1. Lectura de las correcciones a los borradores presentados	1 día	10-mar-08	10-mar-08	
17	3.2. Lectura amplia de la bibliografía disponible	7 días	11-mar-08	16-mar-08	6



ID	Nombre de la tarea	Duración	Comienzo	Fin	Predec.
18	3.3. Elaboración del documento de análisis preliminar de los datos	7 días	11-mar-08	16-mar-08	16
19	3.4. Elaboración del primer avance del documento de requisitos	7 días	11-mar-08	16-mar-08	
20	3.5. Elaboración del documento de planificación	7 días	11-mar-08	16-mar-08	
21	3.6. Entrega de los documentos definitivos de la PAC I	0 días	16-mar-08	16-mar-08	18;19;20
22	4. Elaboración de un avance de la PAC II	26 días	17-mar-08	13-abr-08	
23	4.1. Lectura detallada de los enunciados y directrices	1 día	17-mar-08	17-mar-08	
24	4.2. Redacción de capítulo de introducción del TFC	2 días	18-mar-08	19-mar-08	
25	4.3. Continuación del proceso de limpieza y depuración de los datos de origen	3 días	22-mar-08	24-mar-08	
26	4.4. Elaboración del borrador del documento de análisis de requisitos, basado en la primera aproximación generada en la PAC I	20 días	25-mar-08	12-abr-08	23
27	4.5. Elaboración del borrador del documento que recoja el modelo de datos lógico y su traslación al modelo de datos físico	20 días	25-mar-08	12-abr-08	23
28	4.6. Redacción de capítulo de objetivos del TFC	20 días	25-mar-08	12-abr-08	
29	4.7. Entrega de los borradores generados	0 días	13-abr-08	13-abr-08	26;27
30	5. Elaboración de la PAC II	6 días	15-abr-08	20-abr-08	
31	5.1. Lectura de las correcciones a los borradores.	1 día	15-abr-08	15-abr-08	
32	5.2. Redacción de los documentos definitivos de los modelos lógico y físico.	5 días	16-abr-08	19-abr-08	31
33	5.3. Entrega de los documentos definitivos de la PAC II	0 días	20-abr-08	20-abr-08	32
34	6. Elaboración de un avance de la PAC III	20 días	21-abr-08	11-may-08	
35	6.1. Lectura detallada de los enunciados y directrices	1 día	21-abr-08	21-abr-08	
36	6.2. Configuración de la base de datos	3 días	22-abr-08	24-abr-08	
37	6.3. Carga de datos	1 día	25-abr-08	25-abr-08	36
38	6.4. Elaboración del conjunto de informes	6 días	26-abr-08	2-may-08	37
39	6.5. Elaboración del borrador de la PAC III	8 días	3-may-08	10-may-08	38
40	6.6. Redacción del apartado Glosario del TFC	9 días	3-may-08	10-may-08	
41	6.7. Entrega borrador PAC III	0 días	11-may-08	11-may-08	39
42	7. Elaboración de la PAC III	6 días	13-may-08	18-may-08	
43	Lectura de las correcciones al borrador	1 día	13-may-08	13-may-08	
44	Depuración y testing de los informes	5 días	14-may-08	17-may-08	43
45	Elaboración del documento definitivo de la PAC III	5 días	14-may-08	17-may-08	
46	Entrega de los documentos definitivos de la PAC III	0 días	18-may-08	18-may-08	44;45
47	8. Elaboración definitiva de la memoria	21 días	19-may-08	8-jun-08	
48	8.1. Redacción de los capítulos restantes	17 días	19-may-08	4-jun-08	
49	8.2. Formato final del documento	3 días	5-jun-08	7-jun-08	48
50	8.3. Elaboración de la presentación	9 días	30-may-08	7-jun-08	
51	8.4 Entrega final (memoria y presentación)	0 días	8-jun-08	8-jun-08	49;50
52	9. Debate	8 días	18-jun-08	26-jun-08	
53	9.1. Intervención en el debate	8 días	18-jun-08	26-jun-08	

1.5.5. Diagrama de Gantt

La planificación indicada en la tabla anterior puede verse representada gráficamente en forma de diagrama de Gantt en las dos gráficas siguientes:

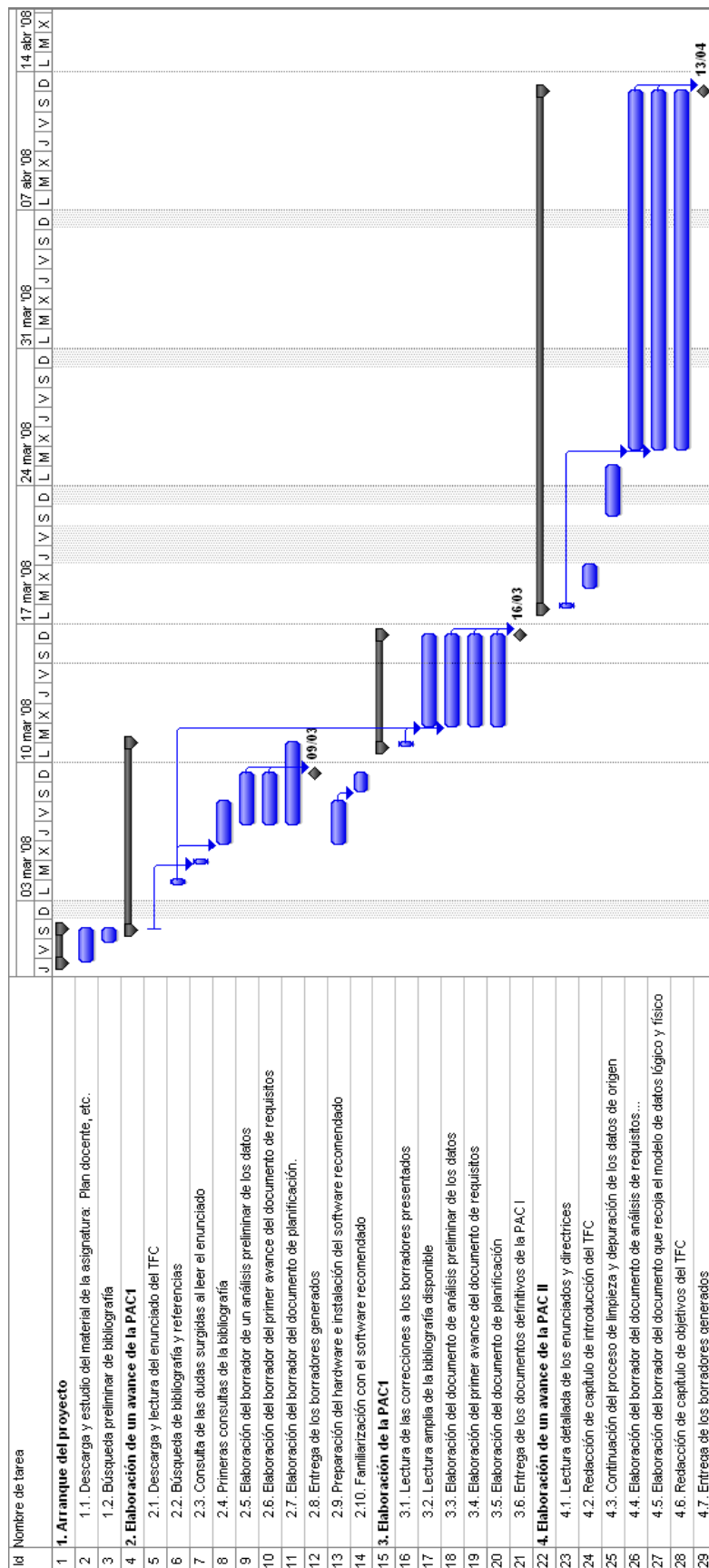
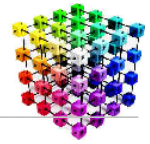


Figura 1. Diagrama de Gantt (parte I)

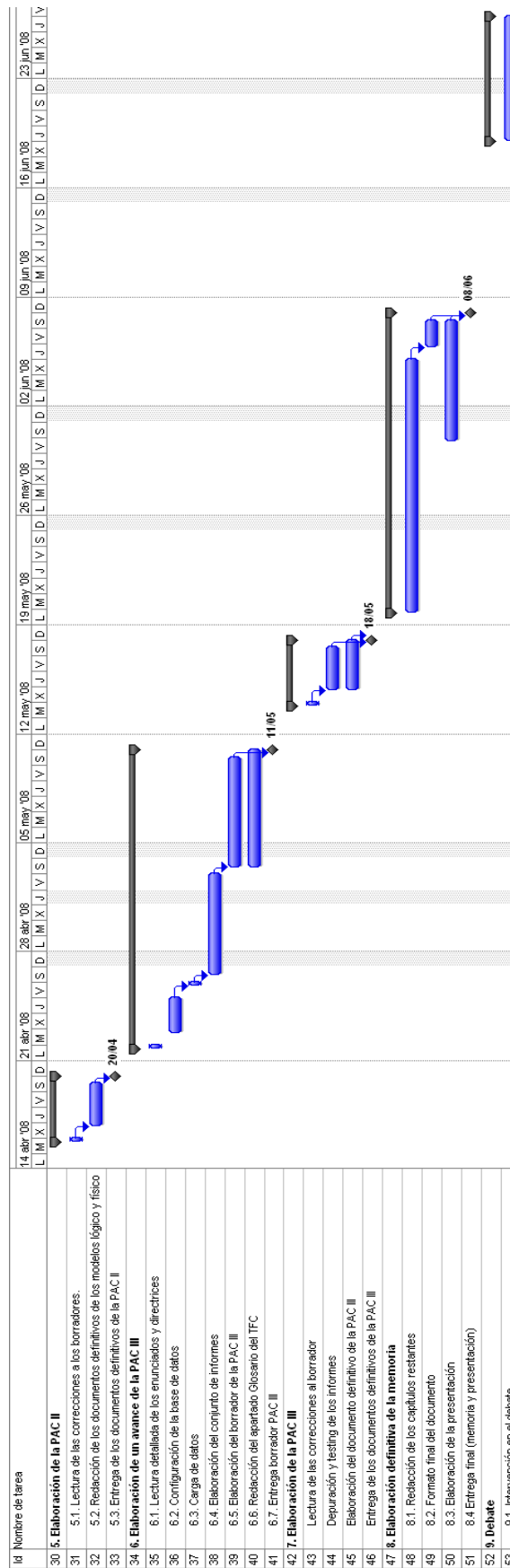
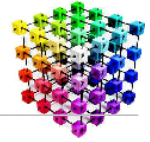
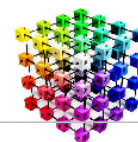


Figura 2. Diagrama de Gantt (parte II)



1.5.6. Incidencias y Riesgos

La siguiente tabla recoge los posibles riesgos e incidencias que fue posible prever en el momento de realizar la planificación del proyecto, así como las soluciones aplicables en cada caso concreto. A continuación se exponen de forma más detallada cada uno de los riesgos identificados en ese primer momento y la solución que se determinó para cada uno de ellos.

Incidencia	Solución
Problemas laborales	Días de asuntos propios, vacaciones
Problemas de salud	Prescindir de la entrega de borradores. Ajuste de la programación.
Estudios de Master	El TFC tendrá prioridad
Problemas técnicos	Disponibilidad de otros equipos. Copias de seguridad
Viajes no previstos	Disponibilidad de ordenador y conexión

Problemas por motivos laborales

Descripción:

Estos problemas pueden generarse por la coincidencia de las fechas de mayor carga de trabajo del TFC con situaciones excepcionales en el trabajo que exijan un mayor tiempo de presencia.

Solución:

En casos de desfases leves sobre el plan previsto, la redacción de los capítulos previstos para las primeras fases del TFC (introducción, objetivos, glosario...) pueden ser desplazadas y acomodadas en otras fechas, de forma que la redacción de estos capítulos actúe como un colchón que permita un seguimiento sin sobresaltos.

En caso de desfases graves se emplearán días de asuntos propios y días de vacaciones que permitan dedicar más horas al TFC

Problemas de Salud

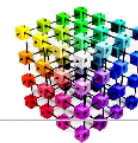
Descripción:

Esta es la contingencia más difícilmente solucionable y no existe solución satisfactoria para todos los posibles casos.

Solución:

La redacción de determinados capítulos se ha distribuido a lo largo de todo el TFC de manera que sean fácilmente movibles, adelantando trabajo cuando sea posible, y retrasándolo en caso de problemas. Llevar al día o incluso adelantada la redacción de estos capítulos puede suponer una medida preventiva ante este tipo de contingencias. Está previsto también prescindir de la entrega de borradores de las PAC si se presentan incidencias de este tipo.

Estudios de Master



Descripción:

En la actualidad estoy cursando un master de Gestión de la I+D+i y es posible que las fechas de entregas de dicho master coincidan con las del TFC.

Solución:

Una correcta planificación de las tareas del Master y el mantenimiento al día del trabajo tanto del Master como del TFC, adelantando trabajo sobre las fechas previstas cuando sea posible, pueden solventar el problema. En este sentido he planificado las asignaturas de la carrera para realizar el TFC en solitario y sólo tras haber superado todas las asignaturas.

En cualquier situación de conflicto el TFC será prioritario.

Problemas técnicos

Descripción:

En el transcurso del TFC se pueden presentar problemas técnicos con el equipo empleado para el alojamiento de DW y para la redacción de los trabajos y documentos.

Solución:

Se ha adelantado la fase de instalación del software y preparación del hardware a la PAC I para prever en la medida de lo posible este tipo de incidencias.

Si se producen problemas de este tipo, se dispone de un segundo ordenador que a pesar de ser un poco más lento, puede soportar sin problemas los requerimientos de este proyecto.

Para evitar pérdidas de datos o documentos se realizarán periódicamente copias de seguridad del trabajo que se alojarán en un servidor remoto.

Viajes no previstos

Descripción:

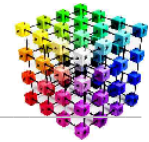
No se descarta la necesidad de realizar viajes no previstos por motivos laborales o personales

Solución:

Como en los casos descritos anteriormente se contempla la posibilidad de cambiar de fecha la redacción de determinados capítulos y el no entregar borradores de las PAC.

El proyecto se está llevando a cabo con un ordenador portátil con conexión inalámbrica, lo que facilitará continuar el trabajo a pesar de los viajes.

De forma general se ha previsto un tiempo de trabajo de dos horas diarias dejando libres los festivos y los domingos (excepto para el proceso de realizar las entregas). Esto unido al hecho de no cursar otras asignaturas permite disponer de un cierto margen para cumplir con éxito la planificación, aún el caso de que surjan problemas.



1.5.7. Desviación sobre la planificación prevista

Tal como se ha indicado en los dos apartados precedentes, tanto la planificación como la evaluación de riesgos expuestas anteriormente son las que se realizaron en la primera fase del proyecto.

Una vez alcanzados los hitos finales del proyecto, se puede constatar que la planificación ha resultado adecuada para los fines previstos ya que ha permitido abordar cada una de las fases en los tiempos estimados sin desviaciones significativas respecto al plan trazado.

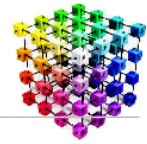
Durante el transcurso del proyecto se han producido ligeras modificaciones sobre el plan previsto. Estas modificaciones han afectado fundamentalmente a la fecha de entregas de borradores de las PAC que se han retrasado levemente sin que ello llegase a afectar en ningún momento a las entregas reales de las PAC. La no entrega de los borradores ya se contemplaba en la planificación como una solución a posibles contratiempos.

La única desviación sobre cronograma fijado que merece ser destacada tuvo lugar durante la PACIII. En esta fase y por un error en el cálculo del volumen real de trabajo que implicaba esta PAC no se realizó una entrega previa tal como estaba prevista. Esta desviación no afectó en ningún aspecto a la entrega definitiva de la PAC III

1.6. PRODUCTOS

Durante la elaboración de este proyecto de final de carrera y como resultado del mismo se han generado además de los documentos que han constituido las PAC I II y III, los siguientes documentos y archivos:

- Un fichero de exportación del esquema de Oracle Database 10g conteniendo los datos, tablas, e índices que permitirá restaurar la base de datos en otro equipo.
- Un fichero de exportación del área de negocio de Oracle Discoverer Administrator
- Un libro de trabajo de Oracle Discoverer Desktop conteniendo 11 consultas.
- Once informes en formato html correspondientes a las consultas.
- Scripts de creación de todas las tablas
- Índices aplicables a las tablas.
- Script de carga de datos
- Script de transformación y limpieza de datos de datos
- Un Fichero de control para la carga de datos con Oracle Loader.
- Documento de requisitos del proyecto y documento de análisis preliminar de datos (PAC I y II)
- Memoria final del proyecto.
- Presentación gráfica del proyecto.



1.7. OTROS CAPÍTULOS

En los siguientes capítulos se aborda el proceso de análisis, comenzando por el estudio de los datos de de partida y continuando con un análisis de los elementos que han de definir el cubo en el que se basará el almacén de datos, lo que constituye el modelo conceptual.

A continuación, y ya en el apartado de diseño, se estudian el modelo lógico y el modelo físico así como las tablas resultantes de este proceso. Se abordan seguidamente los procesos ETL y se diseñan las consultas requeridas.

2. ANÁLISIS

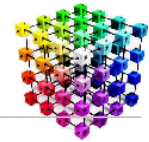
En los siguientes apartados se analizan los datos de origen que constituyen el punto de inicio del proyecto y a partir de la información obtenida del análisis de estos datos se realiza el diseño conceptual, que más adelante, junto con el diseño lógico y físico que nos llevará hasta la definición de las tablas que soportarán los datos.

2.1. DATOS DE PARTIDA

“Sostre per a tothom” proporciona los datos de los inmuebles en un fichero de texto resultado de la concatenación de diversas hojas de cálculo de MS Excel.

Los ficheros estan codificados en **UTF-8** (8-bit Unicode Transformation Format), con los campos delimitados por tabulaciones y consta de 93360 líneas organizadas según los siguientes campos:

- Codi: Identificador único de cada propiedad.
- Zona: Codificación del área geográfica. Se trata de un código alfanumérico que identifica completamente la zona geográfica en la que está situado el piso.
- Tipus: Tipología del inmueble. Este campo describe el tipo del inmueble. La tipología posible está compuesta por Áticos, Chalets, duplex, estudios y pisos. Estos son los cinco tipos de propiedades en los que la inmobiliaria clasifica sus productos.
- Planta: Planta del inmueble. Este campo recoge el número o la posición de la planta de la propiedad en el inmueble en el que esta situada.
- Dormitoris: Numero de dormitorios.
- Metres: Metros cuadrados del inmueble.
- Lista de precios: listado de precios por mes y año (una columna para cada precio y fecha) Cuando el precio aparece vacío para una determinada fecha es por que se ha realizado la venta.



Los campos `codi zona` y `tipus` almacenan valores de texto. `Codi` contiene un código identificador alfanumérico de la forma “VW0000000676364” compuesto por dos letras y trece números, que puede ser empleado como identificador único. El campo `planta` posee fundamentalmente valores numéricos aunque incluye alguna cadena de texto. Los campos `dormitorios` y `metres` son valores numéricos. Los campos referentes a los precios son también campos numéricos.

`Planta` abarca desde 1 hasta 99 y presenta valores nulos

`Dormitorios` abarca desde 0 hasta 99 y presenta valores nulos

`Metres` abarca desde 1 a 47700 y presenta valores nulos

`Tipus` presenta uno de los siguientes valores: Áticos, Chalets, duplex, estudios, pisos.

La ubicación geográfica de los inmuebles se especifica de forma textual mediante líneas de cabecera intercaladas en el fichero de datos y en forma de código mediante el campo `zona`. Esta clasificación incluye inmuebles de España y de Andorra y presenta cinco niveles de detalle partiendo generalmente de la provincia y llegando hasta niveles que permiten identificar un barrio o una zona concreta de una ciudad.

Estos niveles aparecen indicados por el orden de la línea (la primera línea de un conjunto indica un nivel más general de organización) y por la posición dentro de la cadena de caracteres. Los niveles de detalle se ordenan de mayor a menor de izquierda a derecha y se emplea como separador la cadena “:::”.

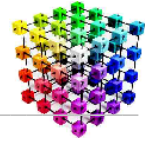
Teniendo en cuenta la organización que se ha descrito para el formato de las cabeceras podemos observar que la información está jerarquizada en 5 niveles y no cuatro como se indica en el documento de partida:

- Primer nivel: Este primer nivel de organización presenta 11 áreas o zonas diferentes que son alicante, andorra, barcelona, castellón, girona, ibiza-formentera, lleida, mallorca, menorca, tarragona y valencia,
- Segundo nivel: El segundo nivel de precisión presenta 94 descripciones diferentes
- Tercer nivel: presenta 269 cadenas de texto diferentes.
- Cuarto nivel: presenta 228 cadenas de texto diferentes.
- Quinto nivel: presenta 123 cadenas de texto diferentes.

2.2. MODELO CONCEPTUAL

En el apartado de metodología se estudiaron ya las fases que debe seguir un proyecto de almacén de datos. En este apartado se describe la secuencia de pasos seguida para abordar el diseño conceptual y se detallan los pasos seguidos para obtener la estrella que modelizará nuestro diseño.

Tal como se expuso en el apartado 5 de la metodología, para poder llevar a cabo un análisis de los datos basado en dimensiones, análisis multidimensional, se ha de definir un cubo



OLAP (*On-Line Analytical Processing*) y como primera fase de este proceso se siguen los pasos descritos a continuación

2.2.1. Identificar el Hecho.

Como primer paso de esta cadena debemos identificar el hecho que constituye el núcleo de la estrella que pretendemos diseñar. Para ello analizamos los procesos del nivel de negocio que estamos tratando, e intentamos buscar un proceso individual que en si mismo sea lo suficientemente significativo como para definir la actividad.

En nuestro caso, se trata de una agencia inmobiliaria, y los datos de que disponemos reflejan la venta de inmuebles en diferentes momentos de tiempo y aportan cierta información asociada a los inmuebles.

Según este primer análisis podríamos definir nuestro hecho como la venta de una propiedad inmobiliaria, pero un análisis más detallado de los datos, nos muestra que la gran mayoría de los mismos se refieren a la oferta de un inmueble y su evolución en el tiempo, donde la venta no es más que un registro puntual en esta serie de datos organizados temporalmente.

Por tanto resulta mucho mas adecuado establecer como hecho la oferta de la propiedad inmobiliaria, y entender la venta como un caso especial de la oferta que se diferenciará de esta, mediante algún método que se determinará más adelante.

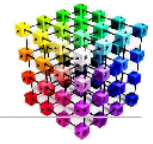
2.2.2. Encontrar la granularidad adecuada.

En esta fase del proceso tratamos de fijar cual ha de ser el grado de detalle de las celdas que compondrán nuestro cubo.

En una primera aproximación podemos tender a establecer una granularidad que nos permita el nivel de detalle más preciso que los datos puedan ofrecernos y de hecho esto es un principio a tener en cuenta en un buen diseño. Pero esta misma elección puede ser debida en muchas ocasiones a la falta de previsión sobre los informes que el sistema debe ofrecer y las consultas necesarias para obtenerlos. Un exceso de detalle puede provocar que el sistema tenga que manejar una cantidad de datos que bien sea excesiva desde un principio, o bien crezca de una manera que provoque que las consultas se vuelvan desasido lentas y poco operativas.

Es necesario por tanto buscar un equilibrio entre nivel de detalle y carga del sistema. En el caso que nos ocupa, la oferta de propiedades esta registrada por periodos mensuales y se registran ofertas o ventas atómicas. Según un primer análisis de las consultas requeridas vemos que pueden ser llevadas a cabo con la granularidad original de los datos tal como son facilitados por la empresa.

Teniendo en cuenta esta aproximación, la granularidad con la que se nos facilitan los datos, esto es, un apunte por propiedad en oferta y por mes parece una granularidad que no supondrá una carga excesiva del sistema y que permitirá el nivel de detalle suficiente para resolver todas las consultas prefijadas y la gran mayoría de las que podamos construir “ad-



hoc”.

Deberemos comprobar si los pasos posteriores de este análisis confirman la idoneidad de esta decisión o si es necesario modificarla a la vista de resultados como el correspondiente al tamaño total del cubo multidimensional.

2.2.3. Escoger las dimensiones a emplear en el análisis.

En un primer momento analizamos el hecho en busca de aquellos otros datos que ayudan a definir, o que sitúan en su contexto al hecho, de forma que estos sean los primeros candidatos a constituir las dimensiones.

Posteriormente intentamos determinar si se nos ha quedado fuera del análisis algún candidato que debería ser tenido en cuenta, y esto podemos hacerlo estudiando si las diferentes instancias que se produzcan de nuestro hecho están perfectamente definidas y delimitadas con las dimensiones escogidas, y finalmente tratamos de determinar si alguna dimensión se podría representar como una combinación de algunas de la otras, si es así, estaremos ante una dimensión fuertemente correlacionada con otra u otras y podemos prescindir de ella para nuestros propósitos.

En este caso las dimensiones que contextualizar el hecho son el “Tiempo” (mes en que se produce la oferta o la venta), la “Zona” geográfica en que se localiza el inmueble ofertado y las características que conforman el inmueble en si, que podemos agruparlas en la dimensión “Inmueble”

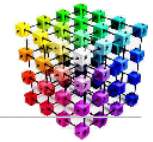
2.2.4. Establecer los atributos de cada dimensión.

Una vez encontradas las dimensiones que formarán parte de nuestro análisis, pasamos a averiguar cuales son los atributos de esas dimensiones y a establecer cuales de ellos serán relevantes para nuestro modelo.

Comenzando por la dimensión “Tiempo”, debemos tener en cuenta que los atributos continuos no son adecuados para nuestros propósitos y que los datos nos vienen dados con una periodicidad mensual. Según esto el atributo que caracterice la dimensión tiempo debería ser la fecha expresada como mes o como año según el nivel de agregación que escojamos.

En cuanto a la dimensión “Zona” partimos tal como se indicó en el análisis previo, de que contamos con cinco niveles de precisión en la definición de la zona, pero todos ellos no son más que una descripción textual de la zona en la que se encuentra el inmueble, por tanto, podemos establecer un atributo llamado zona que representa el nombre o la denominación del nivel de zona que estemos tratando.

Por último para la dimensión “Inmueble” podemos encontrar como atributos: un código identificativo del inmueble, el número de metros cuadrados “metros”, el número de habitaciones disponibles “habitaciones”, y la planta en que se encuentra situado el inmueble “planta”. Este último atributo contendrá valores alfanuméricos dentro de un rango de valores discretos prefijados.



2.2.5. Distinguir entre descriptores y jerarquías de agregación

En esta fase del proceso hemos de establecer que atributos de las dimensiones actuarán como descriptores, permitiéndonos hacer selecciones basadas sus valores y los que nos permitirán hacer agrupaciones y conformar jerarquías.

Teniendo esto en cuenta podemos encontrar que en la dimensión “Tiempo” podemos establecer una jerarquía de agregación en la que la base sean los meses y estos se agrupen en años.

De igual forma, las “Zonas geográficas” presentan también una jerarquía claramente marcada. Estas zonas establecen una agregación que va desde lo que se ha denominado nivel 5, que es la descripción del barrio o zona urbana en que está situado el inmueble hasta lo que hemos denominado nivel 1 que corresponde, en el caso de los inmuebles españoles a la comunidad autónoma.

Para facilitar algunas de las consultas que se solicitan se establece una nivel de jerarquía adicional constituido por el “país”.

2.2.6. Identificar las medidas a emplear.

En el caso que nos ocupa, podríamos plantearnos interpretar el precio del inmueble como un atributo del propio inmueble, pero hemos de tener en cuenta que el precio es una característica que previsiblemente variará con cada instancia del hecho, de forma que la aproximación correcta consistiría en considerar el precio como una medida propia de la oferta. Si podemos considerar en cambio como atributo del inmueble su precio inicial.

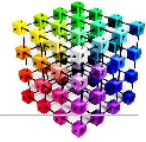
Por otro lado, según los requerimientos del proyecto, hemos de poder realizar consultas que distingan entre los inmuebles ofertados y los vendidos. Como en el caso del precio, este valor cambiará en algún momento, por lo que debe incluirse en el hecho en forma de un atributo que refleja la condición de vendido o no.

Se exige también la posibilidad de diferenciar aquellos inmuebles que han bajado de precio. Este requerimiento resulta bastante impreciso, ya que no se especifica si se trata de una bajada entre dos meses concretos, una bajada con respecto al precio inicial, o una bajada considerada entre el primer y último mes, independientemente de los cambios intermedios. Para resolver la mayor parte de las posibles interpretaciones se opta por incluir también la diferencia con el precio del mes anterior (apreciación) en el hecho.

Por último es necesario añadir un nuevo atributo que nos permita controlar cuanto tiempo lleva cada inmueble en venta, y siguiendo el mismo razonamiento introduciremos un atributo que haga referencia a la antigüedad de la oferta.

2.2.7. Definir las celdas.

En este paso definimos cuales de las posibles celdas que resultan del análisis, pueden ser consideradas interesantes para nuestro modelo y por tanto han de ser almacenadas y cuales de



ellas se pueden considerar derivadas de otras y por tanto no será necesario almacenarlas.

En nuestro caso solo tenemos una celda que refleje el hecho analizado. Usaremos una celda que representará una oferta unitaria de un inmueble.

2.2.8. Explicitar las restricciones de integridad.

En esta fase definimos las bases y establecemos las restricciones que se consideren necesarias. La base nos indicará que dimensiones son realmente necesarias para definir de forma única una celda. En nuestro caso hemos de tener presente que de las tres dimensiones que habíamos definido anteriormente Tiempo, Inmueble y Zona, la dimensión zona queda determinada por la dimensión inmueble, es decir un inmueble estará asociado a la misma zona durante toda su existencia, por lo que será suficiente el tiempo y el inmueble para determinar una celda individual.

Es a la base definida de esta forma a la que debemos aplicarle las reglas de integridad

2.2.9. Estudiar la viabilidad.

Una vez que tenemos perfilado el diseño conceptual debemos realizar los cálculos que nos permitan establecer, aunque sea de una manera aproximada, cuanto ocuparán los datos de nuestro modelo una vez almacenados en el sistema, ya que esto determinará si nuestro modelo es aplicable o no en la realidad.

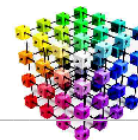
Partimos de unos datos iniciales que contienen algo más de 92000 inmuebles que están o han estado en oferta. Contamos también con datos referidos a 25 meses, con lo que en principio tenemos algo más de 2300000 instancias del hecho tratado, cada una de ellas con los atributos correspondientes.

Es necesario destacar aquí, que aunque la zona constituye una dimensión ya que conforma un punto de vista del modelo, cada inmueble solo puede estar en una zona, por lo que no multiplicamos por esta dimensión o multiplicamos por 1 con lo que el resultado no varía.

Ahora debemos tener en cuenta que hemos incluido en cada celda el precio, la antigüedad de la oferta, el valor que indica si la venta se ha realizado o no y la apreciación del inmueble. Sobreestimando el número de bits necesarios podemos suponer que se emplearán 8 bits para el booleano, 32 para cada uno de los restantes atributos y 24 para cada uno de los tres índices que apuntan a las dimensiones. En total y por sobreestimación tendremos $8 + 96 + 72 = 176$ bits = 21 bytes

Según esto ya podemos hacer una estimación más o menos aproximada del total: $2300000 * 21 = 48.300.000$ o lo que es lo mismo $48.300.000 / 1024 = 47167,9$. Podemos por tanto concluir que de forma aproximada emplearemos unos 48 Mbytes.

Es necesario tener en cuenta en este cálculo, por un lado, que no todos los pisos se ofertan durante todo el tiempo, por lo que el número de instancias disminuye y por otro lado que es de prever que mensualmente se recibirán nuevos datos por lo que la carga del sistema



aumentará.

En resumen el volumen de datos resultante esta claramente dentro de los rangos que pueden ser fácilmente manejados por cualquier pequeño servidor o incluso por cualquier PC relativamente nuevo y tampoco ha de constituir ningún problema para el motor de bases de datos a emplear.

2.2.10. Resultado

Según los pasos descritos anteriormente, el modelo conceptual resultante sería el siguiente:

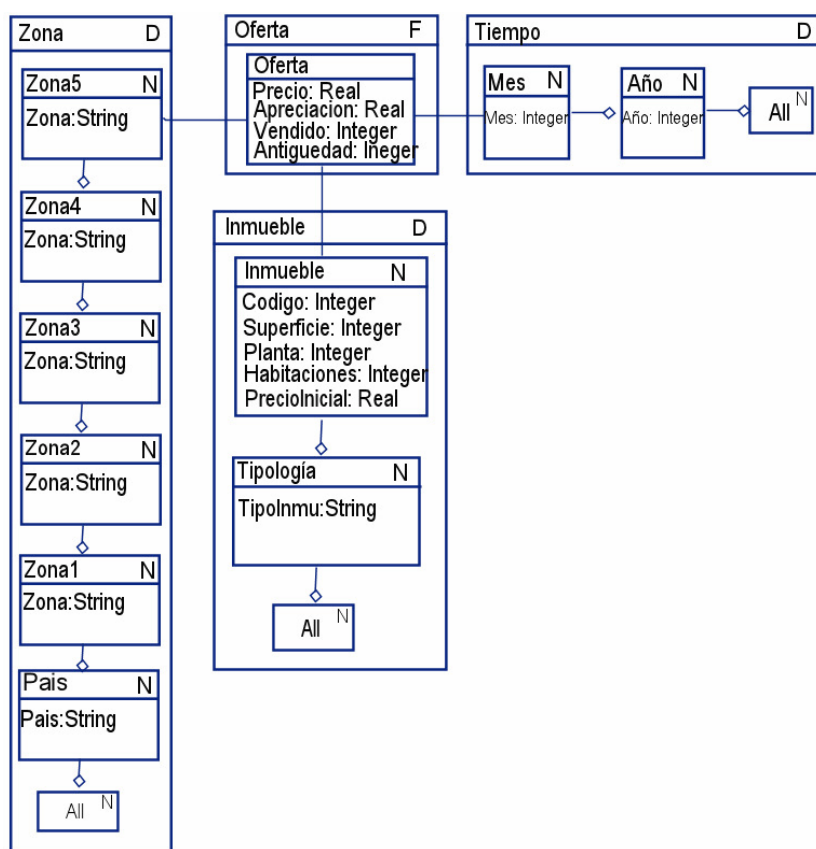
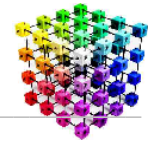


Figura 3 Modelo conceptual



3. DISEÑO

3.1. ARQUITECTURA DEL HARDWARE

En la siguiente imagen se muestra un esquema de la arquitectura del hardware propuesta para el diseño y la explotación de este proyecto.

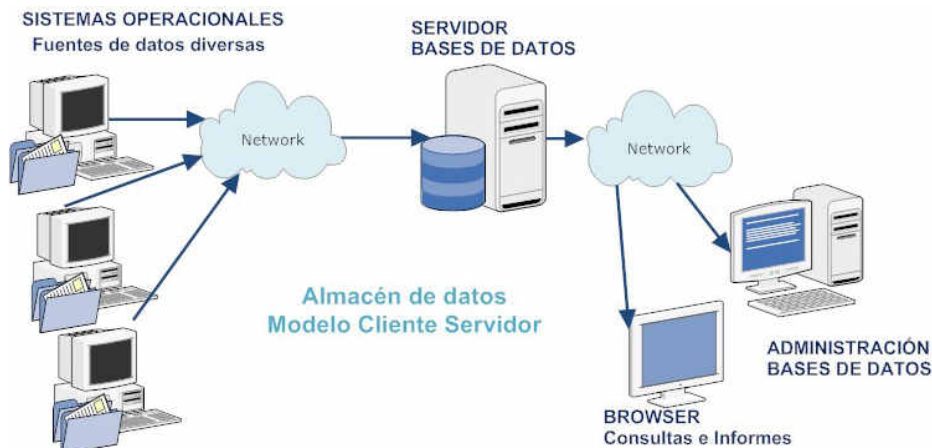


Figura 4 Arquitectura del hardware

Los diferentes sistemas operaciones que posea la empresa nutrirán de datos de diversos orígenes al servidor de bases de datos y los clientes accederán al servidor para realizar las consultas. Desde una estación de administración se realizan los procesos de carga y mantenimiento del almacén de datos.

3.2. ARQUITECTURA DEL SOFTWARE

En la siguiente imagen se muestra un esquema de la arquitectura del software usado y propuesto para el diseño y la explotación de este proyecto.

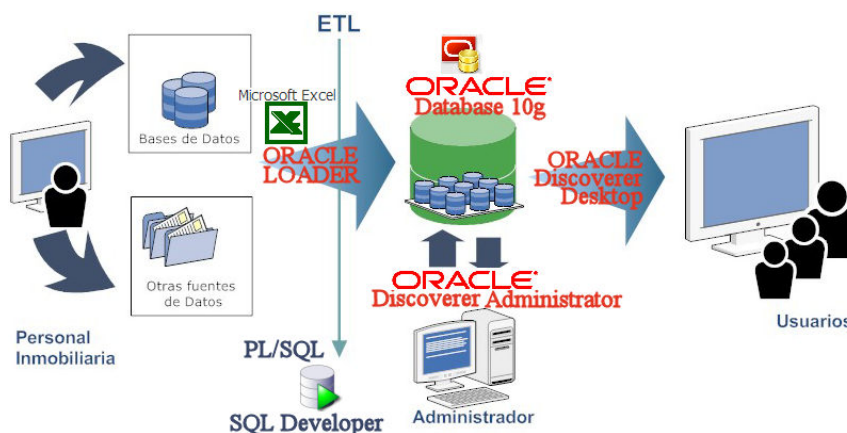
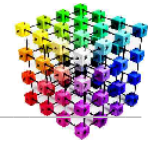


Figura 5 Arquitectura del Software

Según la arquitectura propuesta se parte de ficheros exportados desde MS Excel, que son pre-cargados con Oracle Loader, transformados y cargados definitivamente haciendo uso de



PL/SQL mediante SQL Developer y alojados en tablas de Oracle Database.

Las consultas se diseñan con O. Discoverer Administrador y Desktop y se visualizan con Oracle Discoverer Desktop.

3.3. DISEÑO LÓGICO

Tras completar el modelo conceptual, ha de abordarse el diseño lógico. Partimos del análisis de la estrella definida en los pasos anteriores y para abordar a partir de ella el diseño lógico debemos crear una tabla para el hecho y una por cada una de las dimensiones que componen nuestro modelo.

Una vez definida la tabla del hecho y las de las dimensiones se ha de proceder a ligar las entre sí añadiendo las claves primarias de las dimensiones a la tabla del hecho en forma de claves foráneas.

Para evitar la carga que el empleo de claves foráneas muy pesadas supondría en el sistema, no se emplean claves muy descriptivas, sino que se sustituyen por claves mas sencillas suelen coincidir con el número identificador de la fila.

Aplicando la metodología descrita se puede obtener un modelo lógico como el siguiente:

Oferta (idTiempo, idInmueble, idZona, precio, apreciacion, antigüedad, vendido)

Tiempo (RowId, mesAño, año)

Inmueble(RowId, código, superficie, planta, habitaciones, tipoInmueble, precioInicial)

Zona(RowId, zona1, zona2, zona3, zona4, zona5, país)

Como se explicó en apartados anteriores la clave foránea referida a la tabla Zona no se incorpora como clave primaria del hecho (Oferta) ya que no es necesaria para determinar una celda concreta, ya que un inmueble siempre estará asociado a la misma zona.

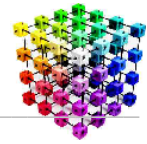


Figura 6 Esquema del diseño lógico

3.4. DISEÑO FÍSICO

La estructura obtenida a partir del diseño lógico se transforma ahora en tablas que puedan ser gestionadas por un motor de bases de datos.

Las tablas resultantes son las siguientes:



3.4.1. Ofertas

Field Name	Field Type	Size	Precision	Not Null	Default
idTiempo	INTEGER	11	0	<input checked="" type="checkbox"/>	0
idInmueble	INTEGER	11	0	<input checked="" type="checkbox"/>	0
idZona	INTEGER	5	0	<input checked="" type="checkbox"/>	0
precio	FLOAT	9	3	<input type="checkbox"/>	Null
vendido	TINYINT	4	0	<input type="checkbox"/>	Null
antiguedad	INTEGER	11	0	<input type="checkbox"/>	Null
apreciacion	FLOAT	9	3	<input type="checkbox"/>	Null

Figura 7 Esquema físico. Tabla ofertas

Los campos idTiempo e idInmueble constituyen la clave primaria de la tabla Ofertas y por tanto no pueden ser nulos. El campo idZona tampoco puede ser nulo ya que como los dos anteriores es una clave foránea de las tablas de dimensiones.

El precio y el precio inicial se han definido como real, por que aunque en los datos iniciales todos sus valores son enteros, no podemos descartar que en el futuro contengan valores reales.

3.4.2. Inmuebles

Field Name	Field Type	Size	Precision	Not Null	Default
codigo	INTEGER	11	0	<input type="checkbox"/>	Null
habitaciones	INTEGER	6	0	<input type="checkbox"/>	Null
id	INTEGER	11	0	<input checked="" type="checkbox"/>	Null
planta	VARCHAR	10	0	<input type="checkbox"/>	Null
precioinicial	FLOAT	9	3	<input type="checkbox"/>	Null
superficie	FLOAT	5	3	<input type="checkbox"/>	Null
tipologia	VARCHAR	10	0	<input type="checkbox"/>	Null

Figura 8 Esquema físico. Tabla inmuebles

El identificador de de fila (id) cosntituye la clave principal de la tabla Inmueble

3.4.3. Tiempo

Field Name	Field Type	Size	Preci...	Not ...	Default
mes	TINYINT	4	0	<input type="checkbox"/>	Null
id	SMALLINT	6	0	<input type="checkbox"/>	Null
año	INTEGER	6	0	<input type="checkbox"/>	Null

Figura 9 Esquema físico. Tabla tiempo

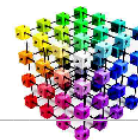
El identificador de de fila (id) cosntituye la clave principal de la tabla Tiempo

3.4.4. Zona

Field Name	Field Type	Size	Precision	Not Null	Default
id	INTEGER	11	0	<input checked="" type="checkbox"/>	0
zona5	VARCHAR	50	0	<input type="checkbox"/>	Null
zona4	VARCHAR	50	0	<input type="checkbox"/>	Null
zona3	VARCHAR	50	0	<input type="checkbox"/>	Null
zona2	VARCHAR	50	0	<input type="checkbox"/>	Null
zona1	VARCHAR	50	0	<input type="checkbox"/>	Null
pais	VARCHAR	50	0	<input type="checkbox"/>	Null

Figura 10 Esquema físico. Tabla zona

El identificador de de fila (id) cosntituye la clave principal de la tabla Zona



3.4.5. Planificación de las consultas

En esta fase del proyecto, antes de implementar las consultas mediante el front-end elegido se planteó la conveniencia de comprobar si las consultas requeridas en el documento de especificaciones se pueden realizar con los modelos y las tablas propuestas.

En este apartado se prueban las consultas que el documento de requisitos especifica de forma concreta para determinar si con la estructura y las tablas planificadas dichas consultas serían posibles. Si en estas pruebas se detectase la imposibilidad de realizar determinadas consultas, sería conveniente replantear el modelo físico e incluso los modelos conceptual y lógico.

Los datos sobre los que se han realizado estas pruebas no son reales y han sido creados únicamente con el fin de poder verificar la idoneidad del diseño sobre un conjunto limitado y manejable de elementos. Lo nombres y la disposición de los campos son también diferentes de los empleados en la configuración definitiva de las tablas, por este motivo los resultados de estas consultas no tienen más valor que el de demostrar que la extracción de la información es posible con el modelo propuesto.

Número de inmuebles por zona, tipología y características

Este primer requerimiento referente a las consultas puede interpretarse de dos formas distintas:

- 1) En primer lugar podemos obtener el número de inmuebles agrupados por zona, tipología y características.

#Número de inmuebles agrupados por zona

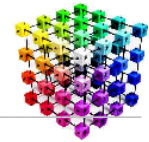
```
SELECT Zona.zona5, COUNT( Distinct( Inmuebles.codigo)) AS NumeroInmuebles
FROM (Ofertas, Tiempo,Inmuebles, Zona)
WHERE (Ofertas.idTiempo = Tiempo.id
AND Ofertas.idInmueble = Inmuebles.id
AND Ofertas.idZona = Zona.id)
GROUP BY Zona.`zona5`;
```

zona	NumeroInmuebles
Ciudad Jardín	1
Entrepinares	1
Playa	2

Numero de inmuebles por número de habitaciones

```
SELECT Inmuebles.`habitaciones`, COUNT( Distinct(Inmuebles.codigo)) AS NumeroInmuebles
FROM (Ofertas, Tiempo,Inmuebles, Zona)
WHERE (Ofertas.idTiempo = Tiempo.id
AND Ofertas.idInmueble = Inmuebles.id
AND Ofertas.idZona = Zona.id)
GROUP BY Inmuebles.`habitaciones`;
```

habitaciones	NumeroInmuebles
3	1
4	1
6	1
9	1



De igual forma se pueden realizar las consultas para las tipologías.

2) En segundo lugar se puede interpretar que lo que se pretende es que el usuario, mediante la interfaz, seleccione una zona, un inmueble o una tipología:

```
SELECT COUNT(Distinct(Inmuebles.codigo) AS Inmuebles
FROM (Ofertas, Tiempo, Inmuebles, Zona)
WHERE (Ofertas.idTiempo = Tiempo.id
AND Ofertas.idInmueble = Inmuebles.id
AND Ofertas.idZona = Zona.id
AND Zona.zona5 = 'Playa');
```

Inmuebles
2

Listado de inmuebles que han bajado de precio

Esta consulta admite diferentes interpretaciones. El siguiente ejemplo muestra como se podría encontrar aquellos inmuebles que han bajado de precio desde su puesta en venta.

#Listado de inmuebles que han bajado de precio

```
SELECT Distinct(Inmuebles.codigo) AS InmueblesBajada
FROM (Ofertas, Tiempo, Inmuebles, Zona)
WHERE (Ofertas.idTiempo = Tiempo.id
AND Ofertas.idInmueble = Inmuebles.id
AND Ofertas.idZona = Zona.id
AND Ofertas.precio < Inmuebles.precioInicial)
ORDER BY Inmuebles.codigo;
```

InmueblesBajada
3

Evolución de precios por zona, tipología y características

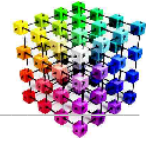
#Evolución de precios para una zona concreta

```
SELECT Ofertas.idInmueble, Tiempo.mes, Tiempo.año, Ofertas.precio
FROM (Ofertas, Tiempo, Inmuebles, Zona)
WHERE (Ofertas.idTiempo = Tiempo.id
AND Ofertas.idInmueble = Inmuebles.id
AND Ofertas.idZona = Zona.id
AND Zona.zona5 = 'playa')
ORDER BY Ofertas.idInmueble, Tiempo.id;
```

idInmueble	mes	año	precio
Click here to select visible columns			
		2008	110,000
2	2	2008	120,000
2	3	2008	130,000
3	1	2008	140,000
3	2	2008	150,000
3	3	2008	100,000
3	4	2008	4,000,000

De igual forma puede realizarse la consulta para tipología o características

Precios máximos, mínimos y medios por zona, tipología y características



#Máximo precio por zonas

```
SELECT MAX(Ofertas.`precio`) AS MaxPrecio,Zona.`zona`
FROM (Ofertas, Tiempo,Inmuebles, Zona)
WHERE (Ofertas.idTiempo = Tiempo.id
AND Ofertas.idInmueble = Inmuebles.id
AND Ofertas.idZona = Zona.id)
GROUP BY Zona.`zona`
ORDER BY Zona.`zona`;
```

MaxPrecio	zona
260,000	Ciudad Jardin
300,000	Entrepinares
4.000,000	Playa

#Mínimo precio por Características (numero de habitaciones)

```
SELECT MIN(Ofertas.`precio`) AS MinPrecio,Inmuebles.`habitaciones`
FROM (Ofertas, Tiempo,Inmuebles, Zona)
WHERE (Ofertas.idTiempo = Tiempo.id
AND Ofertas.idInmueble = Inmuebles.id
AND Ofertas.idZona = Zona.id)
GROUP BY Inmuebles.`habitaciones`
ORDER BY Inmuebles.`habitaciones`;
```

MinPrecio	habitaciones
110,000	3
230,000	4
100,000	6
300,000	9

#Precio Promedio por Tipología

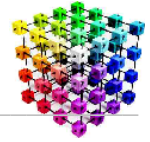
```
SELECT AVG(Ofertas.`precio`) AS PecioMedio,Inmuebles.`tipología`
FROM (Ofertas, Tiempo,Inmuebles, Zona)
WHERE (Ofertas.idTiempo = Tiempo.id
AND Ofertas.idInmueble = Inmuebles.id
AND Ofertas.idZona = Zona.id)
GROUP BY Inmuebles.`tipología`;
```

PecioMedio	tipología
1.097,500	atico
120,000	casa
300,000	chalet
245,000	piso

Metros cuadrados máximos, mínimos y medios por zona, tipología y características

#Metros cuadrados promedio por características (numero de habitaciones)

```
SELECT AVG(Inmuebles.`superficie`) AS MetrosMedios,Inmuebles.`tipología`
FROM (Ofertas, Tiempo,Inmuebles, Zona)
WHERE (Ofertas.idTiempo = Tiempo.id
AND Ofertas.idInmueble = Inmuebles.id
AND Ofertas.idZona = Zona.id)
GROUP BY Inmuebles.`tipología`;
```



MetrosMedios	tipología
35,000	atiko
24,000	casa
56,000	chalet
30,000	piso

#Metros Mínimos por Características (numero de habitaciones)

```
SELECT MIN(Inmuebles.`superficie`) AS MinPrecio, Inmuebles.`habitaciones`
FROM (Ofertas, Tiempo, Inmuebles, Zona)
WHERE (Ofertas.idTiempo = Tiempo.id AND Ofertas.idInmueble = Inmuebles.id AND Ofertas.idZona =
Zona.id)
GROUP BY Inmuebles.`habitaciones`;
```

MinPrecio	habitaciones
24	3
30	4
35	6
56	9

Tiempo medio de venta por zona, tipología y características

#Tiempos medios de venta por zonas

```
SELECT AVG(Ofertas.`antiguedad`) AS TiempoMedio, Zona.zona5
FROM (Ofertas, Tiempo, Inmuebles, Zona)
WHERE (Ofertas.idTiempo = Tiempo.id
AND Ofertas.idInmueble = Inmuebles.id
AND Ofertas.idZona = Zona.id
AND vendido = 1)
GROUP BY Zona.zona5
Order by Zona.zona5;
```

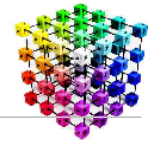
NumeroInmu	zona
1	Ciudad Jardin
1	Entrepinares
1	Playa

Al tratarse de tiempos de ventas constamos solo las ofertas con vendido=1

Número de inmuebles vendidos y existentes por zona, tipología y características

#Número de inmuebles vendidos por zona

```
SELECT COUNT (Distinct(Inmuebles.codigo)) AS NumeroInmuebles , Zona.zona5
FROM (Ofertas, Tiempo, Inmuebles, Zona)
WHERE (Ofertas.idTiempo = Tiempo.id AND Ofertas.idInmueble = Inmuebles.id
AND Ofertas.idZona = Zona.id AND Ofertas.vendido = 1)
GROUP BY Zona.zona5;
```

NumeroInmu	zona
1	Ciudad Jardin
1	Entrepinares
1	Playa

En este caso la diferenciación entre las consultas para vendidos y existentes se realiza mediante el atributo vendido

De igual forma son posibles las consultas por tipología y características.

Tras las pruebas realizadas no se ha encontrado ningún error de diseño que imposibilite una consulta, por lo que no se considera necesario realizar cambios en los modelos conceptual, lógico o físico.

3.5. PROCESOS ETL

En los apartados anteriores se han descrito el modelo conceptual, el lógico y el físico que debían de dar soporte al proyecto.

En los siguientes apartados se realiza una descripción de tallada de los procesos de creación de tablas, carga de datos, como precarga y carga, y transformación de los mismos. Se analizan en detalle los scripts y las sentencias empleados y se describe el proceso de creación y la interpretación de las consultas solicitadas en los requerimientos del proyecto.

Se incluye además un apartado dedicado a la restauración de la base de datos y el esquema de usuario a partir de los ficheros de exportación.

3.5.1. Proceso de Precarga.

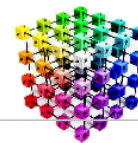
En un primer planteamiento del proceso de carga, se optó por realizar esta actividad leyendo el fichero de datos original desde un script de PL/SQL que sería también el encargado de rellenar las diferentes tablas.

Tras algunos ensayos siguiendo esta metodología se optó por cambiar la forma de realizar esta primera fase del proceso en aras de simplificar el código resultante y facilitar su mantenimiento y evolución.

Se decidió optimizar el proceso realizando en un primer momento una carga del fichero original a una tabla de Oracle empleando para ello la herramienta de Oracle SQLLoader y empleando esta primera tabla de datos en bruto como una tabla temporal desde la que se cargarán los datos a las diferentes tablas definitivas mediante un script de PL/SQL.

Este proceso de “precarga” permitió realizar los procesos de corrección y transformación de los datos sobre la tabla temporal antes de que estos se cargasen en las tablas definitivas.

Para realizar este proceso, en primer lugar se debe crear la tabla que acogerá temporalmente los datos provenientes del fichero de texto. Este paso se ha automatizado con la creación de



un Script denominado CREA_TEMPORAL_UOC.pls¹ que será el encargado de generar la tabla denominada TEMPORAL_UOC así como la secuencia y el trigger asociados a la misma.

```
C:\WINDOWS\system32\cmd.exe - sqlplus TFC1@localhost/uoc
D:\oracle>sqlplus TFC1@localhost/uoc
SQL*Plus: Release 10.1.0.4.2 - Production on SBb May 17 17:26:13 2008
Copyright (c) 1982, 2005, Oracle. All rights reserved.

Conectado a:
Oracle Database 10g Express Edition Release 10.2.0.1.0 - Production
SQL> @D:\oracle\final\CREA_TEMPORAL_UOC.PLS
Procedimiento PL/SQL terminado correctamente.

Disparador creado.
SQL>
```

Figura 11 Ejecución del script TEMPORAL_UOC en SQLPLUS

Una vez ejecutado el script en SQLPLUS y creada la tabla, se realiza el proceso de carga mediante SQLLoader. Este proceso requiere, por un lado, del fichero con los datos de origen, que en este caso se ha denominado catalogo2.dat¹, y por otro de un fichero de control, catalogo2.ctl, que contiene las instrucciones necesarias para la correcta carga de los mismos.

El fichero de control contiene, además del nombre de la tabla de destino, una relación de los campos de destino en la tabla temporal junto con la descripción del tipo de los mismos. Figura también en este fichero de control la identificación del carácter que actúa como delimitador de campo, que en este caso es el tabulador, así como otras instrucciones que controlan el proceso como son el número de filas a ignorar, el juego de caracteres de los datos y otros.

Desde una consola de DOS se invoca al Loader de la siguiente manera

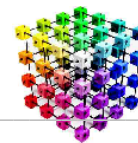
```
sqlldr userid=TFC1@localhost/uoc control=catalogo2.ctl data=catalogo2.dat log=catalogo2.log
```

```
C:\WINDOWS\system32\cmd.exe
D:\oracle>CD FINAL
D:\oracle\final>sqlldr userid=TFC1@localhost/uoc control=catalogo2.ctl data=catalogo2.dat log=catalogo2.log
SQL*Loader: Release 10.2.0.1.0 - Production on SBb May 17 17:41:50 2008
Copyright (c) 1982, 2005, Oracle. All rights reserved.

Carga terminada - recuento de registros lógicos 93360.
D:\oracle\final>
```

Figura 12 Carga de los datos mediante sqlldr

¹ Los scripts mencionados se encuentran en el directorio scripts de la entrega



Donde *userid* identifica al usuario, la conexión y el password, *control* hace referencia al fichero de control, *data* identifica el fichero de datos y *log* establece un fichero de incidencias.

Tras la ejecución del *sqlldr*, que toma del orden de 40 segundos en un Intel Centrino M a 1.6 GHz, los datos son cargados en la tabla *TEMPORAL_UOC*. En total se cargan 93360 registros.

Con los registros ya cargados en la tabla temporal se da paso al proceso de limpieza y adecuación de los mismos.

3.5.2. Proceso de Transformación.

Desde un primer momento, la empresa ha advertido de posibles errores e inconsistencias en los datos, hecho este que el análisis preliminar de los mismos se encargó de poner de manifiesto.

Debido a la existencia de estos errores en los datos y a que el formato de los mismos no es en algunos casos el más adecuado, se hace necesario un proceso de transformación previo a la carga de los mismos en las tablas definitivas. Esta limpieza de los datos evitará en la medida de lo posible ofrecer información errónea en las consultas, sobre todo en aquellas que impliquen cálculos o el empleo de valores máximos, mínimos o medios.

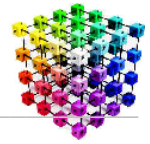
La limpieza y corrección de los datos posibilitará además calcular de forma adecuada el “inmueble tipo” por país tal como se solicita en las especificaciones.

Ya en informes previos se analizaron algunas inconsistencias que debían ser solventadas en este momento. Aprovechando que los datos ya están cargados en la base de datos se emplean sentencias SQL para la transformación de los mismos lo que facilita y automatiza el proceso.

Las correcciones efectuadas son las siguientes:

Inconsistencias entre la información almacenada en tipología y la almacenada en planta:

En los casos e inconsistencia entre la información proporcionada por la tipología y la proporcionada por el campo planta, se decide adoptar como válida la información de la tipología, ya que esta incluye menos valores y esto implica que también sean menores las posibilidades de error. Esta suposición se ve reforzada por la presencia de gran cantidad de valores indeterminados en el campo planta, marcados como "--" lo que indica que la información de este campo es menos precisa.



- Combinación de bj (planta baja) y ático. Los inmuebles que pertenecen a la tipología de ático y que tienen el campo planta fijado a bj son actualizados fijando el campo planta a NULL. Las sentencias empleadas para la consulta y la actualización son las siguientes:

```
SELECT count(distinct(CODIGO))
FROM TEMPORAL_UOC
WHERE TIPUS = 'áticos'
AND PLANTA = 'bj';
--Result 26;

UPDATE TEMPORAL_UOC
SET PLANTA = NULL
WHERE TIPUS = 'áticos'
AND PLANTA = 'bj';
--Query OK, 26 rows affected (250 ms)
```

- Combinaciones de chalet y sótano:

```
SELECT count(distinct(CODIGO))
FROM TEMPORAL_UOC
WHERE TIPUS = 'chalets'
AND PLANTA = 'st';
--Result 1;

UPDATE TEMPORAL_UOC
SET PLANTA = NULL
WHERE TIPUS = 'chalets'
AND PLANTA = 'st';
--Query OK, 1 rows affected (160 ms)
```

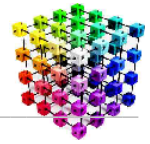
- Poner a NULL los valores de la superficie de los inmuebles con menos de 10 metros cuadrados, ya que la tipología no incluye aparcamientos ni trasteros.

```
SELECT count(distinct(CODIGO))
FROM TEMPORAL_UOC
WHERE METRES < 10;
-- Result 18;

UPDATE TEMPORAL_UOC
SET METRES = NULL
WHERE METRES < 10;
--Query OK, 18 rows affected (90 ms)
```

- Poner a NULL los valores de la superficie de los pisos, áticos y estudios con mas de 1000 metros.

```
SELECT count(distinct(CODIGO))
FROM TEMPORAL_UOC
WHERE METRES >1000
AND ((TIPUS = 'pisos')
OR (TIPUS = 'áticos')
OR (TIPUS = 'estudios'));
-- Result 10;
```



```
UPDATE TEMPORAL_UOC
SET METRES = NULL
WHERE METRES >1000
AND ((TIPUS = 'pisos')
OR (TIPUS = 'áticos')
OR (TIPUS = 'estudios'));
--Query OK, 10 rows affected (300 ms)
```

- Poner a NULL los valores de los dormitorios de los inmuebles con más de 20 dormitorios.

```
SELECT count(distinct(CODIGO))
FROM TEMPORAL_UOC
WHERE DORMITORIS >20;
-- Result 3
```

```
UPDATE TEMPORAL_UOC
SET DORMITORIS = NULL
WHERE DORMITORIS >20;
--Query OK, 3 rows affected (200 ms)
```

```
C:\WINDOWS\system32\cmd.exe - sqlplus TFC1@localhost/uoc @transf...
D:\oracle\final>sqlplus TFC1@localhost/uoc @transformacion.sql
SQL*Plus: Release 10.1.0.4.2 - Production on 5 May 17 20:02:17 2008
Copyright (c) 1982, 2005, Oracle. All rights reserved.

Conectado a:
Oracle Database 10g Express Edition Release 10.2.0.1.0 - Production

COUNT(DISTINCT(CODIGO))
-----
26
26 filas actualizadas.

COUNT(DISTINCT(CODIGO))
-----
1
1 fila actualizada.

COUNT(DISTINCT(CODIGO))
-----
18
18 filas actualizadas.

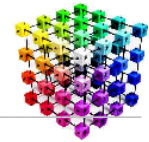
COUNT(DISTINCT(CODIGO))
-----
10
10 filas actualizadas.

COUNT(DISTINCT(CODIGO))
-----
3
3 filas actualizadas.

SQL>
```

Figura 13 Ejecución del fichero de sentencias de corrección².

² Los scripts mencionados se encuentran en el directorio scripts de la entrega



Además de estas correcciones que resultan evidentes a primera vista, podrían haberse realizado otras muchas que también son fácilmente detectables, pero para estos casos hubiese sido necesario contar con el asesoramiento de un representante de la inmobiliaria que pudiese confirmar la idoneidad de la transformación. Así por ejemplo se ha optado por no cambiar la información de planta en los chalet que presentan algún valor numérico, por desconocer si la tipología puede incluir algún tipo de excepción no especificada.

3.5.3. Proceso de Carga.

Una vez finalizada la limpieza y transformación de los datos se procede a la carga de los mismos desde la tabla temporal “TEMPORAL_UOC” a las tablas definitivas siguiendo los siguientes pasos:

En primer lugar han de crearse las tablas definitivas. Tal como se indicaba en la definición del modelo físico del proceso OLAP , las tablas definitivas serán cuatro:

- TIEMPO
- ZONAS
- INMUEBLES
- OFERTAS

La creación de las tablas definitivas se ha automatizado mediante un script PL/SQL llamado CREA_TODAS_TABLAS.pls³ que se encarga de construir las tablas, las secuencias y los triggers asociados a cada una.

```
C:\WINDOWS\system32\cmd.exe - sqlplus TFC1@localhost/uoc ...
D:\oracle\final>sqlplus TFC1@localhost/uoc @CREA_TODAS_TABLAS.PLS
SQL*Plus: Release 10.1.0.4.2 - Production on 5b May 17 20:17:35 2008
Copyright (c) 1982, 2005, Oracle. All rights reserved.

Conectado a:
Oracle Database 10g Express Edition Release 10.2.0.1.0 - Production

Procedimiento PL/SQL terminado correctamente.

Disparador creado.

Procedimiento PL/SQL terminado correctamente.

Disparador creado.

Procedimiento PL/SQL terminado correctamente.

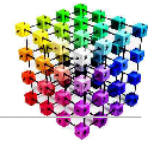
Disparador creado.

Procedimiento PL/SQL terminado correctamente.

SQL>
```

Figura 14 Creación de las tablas finales

³ Los scripts mencionados se encuentran en el directorio scripts de la entrega



Una vez creadas las tablas definitivas se ha de proceder a la carga de los datos desde la tabla temporal a cada una de las cuatro tablas finales. De este proceso se encarga un script denominado InsertarDatos.pls. Este script contiene un procedimiento anónimo principal y cuatro funciones asociadas

De forma resumida el proceso realizado por el script consiste en

- Leer mediante un cursor los nombres de las columnas de la tabla temporal que comienzan por “Preu”
- Trocear estas cadenas y almacenar los valores de meses y años en la tabla TIEMPO
- Leer mediante un cursor cada registro de la tabla temporal
- Trocear el valor del campo registro siempre que comience con # y rellenar con los valores la tabla ZONAS.
- Rellenar la tabla INMUEBLES.
- Rellenar la tabla OFERTAS en función de los valores de la tabla temporal y de las claves primarias de las otras tres tablas.

El proceso de carga genera:

- 25 registros en la tabla TIEMPO
- 613 registros en la tabla ZONAS
- 92555 registros en la tabla INMUEBLES
- 191388 registros en la tabla OFERTAS

El script de carga realiza la inserción de los datos en un tiempo aproximado de 2,19 minutos sobre un Intel Centrino M a 1.8GHz

```
C:\WINDOWS\system32\cmd.exe - sqlplus TFC1@localhost/uoc @Insert...
D:\oracle\final>sqlplus TFC1@localhost/uoc @InsertarDatos.pls
SQL*Plus: Release 10.1.0.4.2 - Production on 5b May 17 20:19:52 2008
Copyright (c) 1982, 2005, Oracle. All rights reserved.

Conectado a:
Oracle Database 10g Express Edition Release 10.2.0.1.0 - Production

Tipo creado.

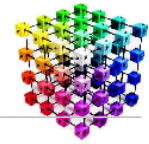
Función creada.

Función creada.

Función creada.

Procedimiento PL/SQL terminado correctamente.
SQL>
```

Figura 15 Ejecución del script de carga de datos



Una vez finalizado el proceso de carga la tabla temporal puede ser borrada mediante a sentencia

```
DROP TABLE TEMPORAL_UOC;
```

Aunque en este caso se ha mantenido para facilitar la comprensión del proceso.

3.5.4. Creación de índices

Una vez cargados los datos en las tablas definitivas es el momento de crear los índices que nos permitirán acelerar las consultas.

Se ha optado por crear estos índices una vez insertados los datos y no tras la creación de las tablas para no penalizar el proceso de carga, ya que la actualización de los índices por parte de Oracle supone un tiempo de proceso que puede ralentizar de forma innecesaria el proceso de carga.

El proceso de creación de los índices ha constado de dos fases claramente separadas, una primera en que se han definido como índices las claves primarias y foráneas de todas las tablas y otra segunda fase en que se han definidos índices más específicos para las consultas a generar.

Se han estudiado en una primera aproximación los campos que constituyen las uniones de tablas “join” y se han creado índices específicos.

En segundo lugar se han tenido en cuenta los campos que retornan las consultas y el orden de los mismos, para definir índices encadenados sobre los grupos más frecuentes.

En todos los casos, se ha estudiado el plan de la consulta mostrado por Oracle Discoverer para comprobar si el índice propuesto constituía o no una mejora en el tiempo empleado por la consulta y si en consecuencia debía ser mantenido o eliminado.

3.6. CREACIÓN DE LAS CONSULTAS

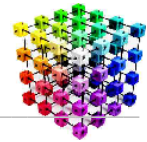
En los siguientes apartados se describe el proceso de creación de las consultas y se explica de forma detallada cada una de las consultas creadas.

3.6.1. Proceso de creación

Tal como se recogía en el documento de especificaciones las consultas y los informes finales se han realizado empleando Oracle Discoverer.

El área de negocios necesaria para alimentar las consultas se ha definido con Discoverer Administrator mientras que los libros de trabajo se han realizado con Discoverer Desktop.

3.6.1.1. Creación del área de negocio



Mediante Oracle Discoverer Administrator se ha creado un único área de negocio denominada AreaInmobiliaria que contiene los datos, las jerarquías y las clases de elementos necesarios para la realización de las consultas.

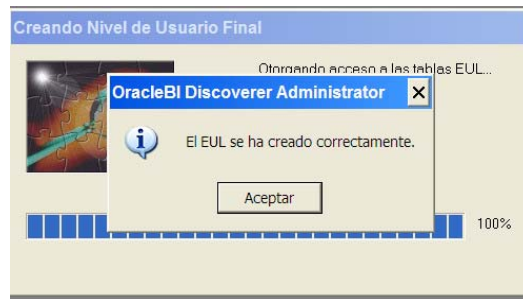


Figura 16 Creación del EUL

Antes de trabajar propiamente con un área de negocio es necesario dar de alta un EUL (End User Layer) para el usuario que realizará la tarea. Una vez creado el EUL y almacenada la información correspondiente en la base de datos ya es posible comenzar a trabajar creando un nuevo área de negocio dentro de la que se han de seguir los siguientes pasos:

Como primer paso se han asignados nombres más claros e identificativos en los elementos que pudiesen resultar poco claros.

En segundo lugar se han creado todas las relaciones (uniones) necesarias entre todos los elementos de las carpetas.

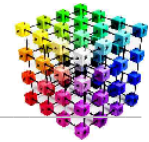
Se ha creado posteriormente una nueva carpeta denominada análisis en la que incluir todos aquellos elementos y sus relaciones que se considerasen necesarios para la generación de las consultas.

Para obtener una relación ordenada de los meses del año se ha creado una carpeta personalizada OrdenMeses en la que se ha incluido la siguiente sentencia SQL:

```
select 'GENER' Mes, 1 ORDEN_MESES from dual
union
select 'FEBRER' Mes, 2 ORDEN_MESES from dual
union
select 'MARÇ' Mes, 3 ORDEN_MESES from dual
union
select 'ABRIL' Mes, 4 ORDEN_MESES from dual
union
. . . (continúa)
select 'DESEMBRE' Mes, 12 ORDEN_MESES from dual.
```

La sentencia anteriormente creada se ha usado al insertar una nueva clase de elemento que permite la ordenación de los meses. (MesesOrd).

Se ha añadido una clase de elemento para cada elemento que se prevé que intervendrá en las consultas.



Datos	Clases de Elementos	Jerarquías

Figura 17 Datos, jerarquías y clases de elementos definidos en el área de trabajo.

Además de las carpetas de datos, las ordenaciones y las clases de elementos es necesario definir también las jerarquías que han de relacionar los diferentes elementos. En este caso se ha establecido una jerarquía que relaciona el país del inmueble con los diferentes niveles de detalle del área geográfica.

Se ha establecido además una segunda jerarquía que relaciona los años con los meses contenidos el elemento anteriormente creado que representa los meses ordenados.

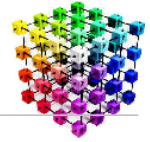
Como último paso se han otorgado los permisos de acceso al área de negocio recién creada.

3.6.1.2. Creación del libro de trabajo.

Una vez creada el área de negocio con Discoverer Administrator podemos empezar a crear consultas sobre ella desde el Discoverer Desktop.

Comenzamos por identificarnos como un usuario con permisos sobre el área de trabajo y abrimos el área anteriormente creada.

Sobre esta área creamos un nuevo libro de trabajo que es la forma que tiene Discoverer de organiza las consultas. Cada libro puede contener múltiples hojas y cada una de ellas corresponde a una consulta que refleja a modo de tabla o de matriz los resultados de la misma.



Para cumplir con los requerimientos del proyecto se ha creado un libro de trabajo denominado LibroConsultasInmobiliaria que contiene las siguientes hojas:

- Precios
- Superficie
- Tiempos de Venta
- Inmuebles Vendidos
- Inmuebles Disponibles
- Baja Mensual
- Subida Mensual
- Bajada Absoluta
- Subida Absoluta
- Cambio de Precio
- Evolución Anual

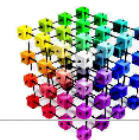
3.6.1.3. Creación de las consultas.

Tal como se ha indicado anteriormente, cada hoja del libro de trabajo alberga una consulta que responde a cada una de las necesidades de información planteada en documento de requisitos. Se han creado 11 consultas que cubren los requerimientos y añaden información extra, no solicitada, que se ha considerado de interés para la inmobiliaria. Cada hoja representa una información específica y las diferentes hojas han sido ordenados de acuerdo al tema que tratan. La estructuración de los datos en el área de trabajo, facilita la creación de consultas “ad-hoc” que no fueron previstas en un primer momento.

Los informes relativos a las consultas en formato html se entregaron en la PAC III

3.6.2. Consultas creadas.

A continuación se describe cada hoja del libro de trabajo y la consulta que alberga, así como las condiciones, cálculos, porcentajes totales y formatos condicionales empleados en cada una de ellas.



3.6.2.1. Precios máximos, mínimos y medios por zona tipología y características

PRECIOS MÁXIMOS, MÍNIMOS Y MEDIOS Por Zona Tipología y Características									
Elementos de Página: Superficie: <Todo> Tipología: <Todo> Habitaciones: <Todo> Planta: <Todo>									
	2006			2007			2008		
	Precio Mínimo	Precio Medio	Precio Máximo	Precio Mínimo	Precio Medio	Precio Máximo	Precio Mínimo	Precio Medio	Precio Máximo
Andorra	€135.000	€493.926	€3.700.000	€70.000	€491.906	€5.048.501	€135.000	€509.779	€5.048.501
andorra	€135.000	€493.926	€3.700.000	€70.000	€491.906	€5.048.501	€135.000	€509.779	€5.048.501
encamp - canillo - ordino	€135.000	€394.664	€860.000	€70.000	€350.180	€1.630.600	€135.000	€362.261	€1.630.600
massana - andorra - escaldes	€210.000	€524.718	€1.914.442	€159.000	€617.208	€5.048.501	€159.000	€664.344	€5.048.501
	€139.000	€514.920	€3.700.000	€150.000	€530.379	€3.700.000	€185.000	€549.533	€3.700.000
España	€1.000	€370.525	€16.000.000	€460	€365.339	€25.000.000	€460	€361.403	€8.500.000
alcante	€1.000	€273.429	€5.250.000	€1.000	€269.484	€12.661.095	€1.000	€266.405	€4.000.000
alt vinalopó	€60.101	€232.505	€666.000	€48.000	€212.755	€875.000	€48.000	€241.254	€875.000
baix segura	€36.000	€164.370	€1.250.000	€33.000	€163.621	€4.500.000	€33.000	€164.006	€1.800.000
baix vinalopó	€75.000	€221.045	€850.000	€65.000	€217.225	€1.200.000	€65.000	€216.134	€830.000
comtat	€41.500	€267.858	€900.000	€41.500	€212.799	€680.000	€50.000	€199.793	€680.000
izalacanti	€41.640	€295.472	€5.250.000	€41.640	€300.499	€3.000.000	€43.450	€294.558	€3.000.000
izalcoya	€20.000	€253.689	€3.000.000	€20.000	€213.256	€1.320.000	€39.000	€214.568	€1.320.000
marina alta	€3.000	€339.854	€4.000.000	€1.600	€340.118	€12.661.095	€3.000	€337.054	€4.000.000
marina baixa	€1.000	€337.235	€3.726.275	€1.000	€338.953	€3.726.275	€1.000	€332.179	€2.900.000
vinalopó mitjà	€29.990	€217.772	€1.200.000	€36.500	€231.514	€1.200.000	€62.000	€240.029	€1.200.000
àrea de eida	€29.990	€209.178	€1.200.000	€49.000	€218.137	€1.200.000	€62.000	€215.598	€1.200.000
àrea de novelda	€36.500	€232.141	€901.518	€36.500	€250.353	€901.518	€72.000	€271.167	€901.518
barcelona	€50.000	€442.965	€9.500.000	€500	€431.972	€9.500.000	€66.000	€430.643	€8.500.000
castellón	€9.000	€259.040	€7.362.398	€800	€258.843	€7.362.398	€1.800	€257.300	€5.500.000
girona	€45.000	€463.181	€16.000.000	€2.400	€440.320	€16.000.000	€49.000	€419.672	€6.000.000
ibiza-formentera	€12.000	€465.950	€4.808.097	€18.000	€472.599	€4.808.097	€18.000	€456.004	€4.808.097
leida	€23.000	€292.604	€5.000.000	€33.000	€276.827	€5.000.000	€33.000	€261.160	€5.000.000
mallorca	€55.000	€391.314	€5.000.000	€55.000	€410.742	€25.000.000	€58.217	€417.266	€7.655.000
menorca	€110.000	€368.534	€3.790.000	€110.000	€394.510	€3.780.000	€110.000	€385.236	€3.780.000
tarragona	€9.000	€302.175	€3.875.000	€460	€294.278	€3.000.000	€460	€283.754	€3.000.000
valencia	€24.041	€298.780	€3.500.000	€18.030	€299.446	€4.267.185	€18.030	€297.811	€4.267.185

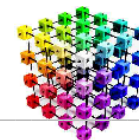
Figura 18 Consulta 1

Esta hoja del libro de trabajo refleja los precios máximos, mínimos y medios de los inmuebles organizados en forma de matriz. En la parte izquierda es posible seleccionar el área geográfica con mayor o menor detalle desde país hasta el nivel más bajo (zona5).

La primera barra de elementos permite seleccionar valores para tipología, superficie, habitaciones y planta (por defecto muestran todos los valores).

La segunda barra superior, en color azul más oscuro, permite seleccionar la escala temporal. Por defecto se encuentra en años, pero cada año permite mostrar un mayor detalle cambiando el nivel a meses.

Para la confección de esta consulta no ha sido necesario insertar ninguna condición ni realizar ningún cálculo.



3.6.2.2. Superficies máximas, mínimas y medias por zona tipología y características

SUPERFICIES MÁXIMAS, MÍNIMAS Y MEDIAS Por Zona Tipología y Características (Metros cuadrados)										
Elementos de Página: Tipología: <Todo> Habitaciones: <Todo> Planta: <Todo>										
	2006			2007			2008			
	Superficie	MIN	Superficie MAX	Superficie	MIN	Superficie MAX	Superficie	MIN	Superficie MAX	
Andorra	32,00	1080,00	118,71	25,00	2180,00	116,09	30,00	2180,00	120,26	
andorra	32,00	1080,00	118,71	25,00	2180,00	116,09	30,00	2180,00	120,26	
encamp - canillo - ordino	32,00	306,00	101,06	25,00	344,00	85,06	30,00	344,00	85,95	
massana - andorra - escaldes	35,00	340,00	117,34	35,00	2180,00	141,08	35,00	2180,00	156,19	
	32,00	1080,00	123,93	38,00	1080,00	126,85	38,00	1080,00	129,56	
España	11,00	47700,00	122,82	11,00	47700,00	121,91	14,00	47700,00	121,45	
alcante	14,00	2000,00	121,75	14,00	1600,00	121,14	14,00	1880,00	120,01	
barcelona	11,00	47700,00	115,61	11,00	47700,00	115,48	20,00	47700,00	115,64	
castellón	26,00	1100,00	110,69	26,00	1277,00	110,09	26,00	1277,00	110,87	
girona	18,00	2427,00	157,02	18,00	2427,00	152,09	18,00	2427,00	148,06	
ibiza-formentera	30,00	1500,00	135,26	28,00	1800,00	132,58	28,00	1500,00	126,30	
leida	22,00	800,00	146,71	22,00	1000,00	140,03	31,00	1000,00	137,20	
malorca	20,00	2400,00	145,86	20,00	2400,00	146,14	24,00	2400,00	147,43	
menorca	30,00	1000,00	126,14	38,00	1000,00	131,19	38,00	1000,00	128,58	
tarragona	20,00	2400,00	124,86	20,00	2400,00	120,38	20,00	2400,00	117,23	
alt camp	38,00	1200,00	152,22	38,00	1200,00	149,81	38,00	1200,00	162,97	
baix camp	35,00	1000,00	125,75	30,00	1000,00	120,04	30,00	800,00	117,47	
baix ebre	24,00	385,00	121,53	24,00	385,00	118,34	30,00	385,00	114,58	
baix penedès	30,00	1505,00	138,45	26,00	1505,00	131,56	30,00	1505,00	128,51	
conca de barberà	25,00	2400,00	303,30	47,00	2400,00	281,77	39,00	2400,00	274,00	
montsià	36,00	510,00	97,05	30,00	450,00	89,91	30,00	450,00	84,75	
priorat	50,00	750,00	178,36	63,00	750,00	199,53	63,00	750,00	192,06	
ribera d'ebre	42,00	390,00	153,37	42,00	238,00	135,64	42,00	231,00	127,58	
tarragonès	20,00	900,00	113,01	20,00	900,00	110,86	20,00	900,00	108,76	
área de catllar	46,00	150,00	95,27	46,00	360,00	109,54	46,00	360,00	107,19	
área de la canonja	40,00	270,00	122,88	40,00	270,00	122,67	40,00	270,00	131,71	
área de pobla de montemès	55,00	300,00	151,77	50,00	300,00	143,68	55,00	300,00	138,38	
área de roda de barà	60,00	280,00	137,05	33,00	325,00	123,01	33,00	325,00	108,02	

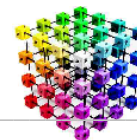
Figura 19 Consulta 2

Esta hoja refleja las superficies, en metros cuadrados, máximas, mínimas y medias de los inmuebles organizados en forma de matriz. En la parte izquierda es posible seleccionar el área geográfica con mayor o menor detalle desde país hasta el nivel más bajo (zona5), mientras que en la parte superior podemos fijar las características y el periodo temporal de la consulta.

Al igual que en el caso anterior, para la confección de esta consulta no ha sido necesario insertar ninguna condición ni realizar ningún cálculo.

3.6.2.3. Tiempos de venta máximos mínimos y medios por zona tipología y características

Al igual que en las consultas anteriores, esta hoja refleja los tiempos de venta, en meses, máximos, mínimos y medios de los inmuebles organizados en forma de matriz. En la parte izquierda es posible seleccionar el área geográfica con mayor o menor detalle, mientras que en la parte superior podemos fijar las características y el periodo temporal de la consulta.



TIEMPO DE VENTA MÁXIMO, MÍNIMO Y MEDIO Por Zona Tipología y Características (Meses)						
Elementos de Página: Tipología: <Todo> Habitaciones: <Todo> Planta: <Todo> Superficie: <Todo>						
	2006			2007		
	Antigüedad Máxima	Antigüedad Mínima	Antigüedad Media	Antigüedad Máxima	Antigüedad Mínima	Antigüedad Media
Andorra	8	1	5	24	1	11
andorra	8	1	5	24	1	11
encamp - canillo - ordino	5	1	3	20	1	7
massana - andorra - escaldés				20	2	10
	8	4	6	24	13	19
España	12	1	6	24	1	12
alicante	12	1	6	24	1	13
barcelona	12	1	6	24	1	12
castelón	12	1	6	24	1	12
girona	12	1	6	24	1	13
ibiza-formentera	12	1	6	24	1	12
leida	12	1	6	23	1	11
malorca	12	1	6	24	1	11
menorca	12	1	6	24	1	12
tarragona	12	1	6	24	1	12
alt camp	12	1	7	22	5	13
baix camp	12	1	6	24	1	12
baix ebre	12	1	7	24	2	13
baix penedès	12	1	6	24	1	12
conca de barberà	10	4	6	24	5	11
montsià	12	1	7	24	1	10
priorat	7	3	5	24	3	17
ribera d'ebre	12	3	6	17	6	10
tarragonès	12	1	6	24	1	13
área de catlar				19	1	7
área de la canonja	10	10	10	14	9	12
área de pobla de monteròs				15	1	8

Figura 20 Consulta 3

Para la realización de esta consulta se ha aplicado la condición de que el inmueble esté vendido, ya que solo de estos inmuebles puede saberse el tiempo de venta.

Editar Condición

Nombre: Vendido = 'SI'

Descripción: SoloVendidos

Ubicación: LibroConsultasInmobiliaria

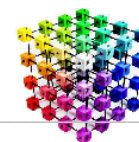
Fórmula: Introduzca el nombre de un elemento o selecciónelo de la lista desplegable.

Elemento	Condición	Valor(es)
Vendido	=	'SI'

Coincidir Mayúsculas/Minúsculas

Aceptar Cancelar Ayuda

No ha sido en cambio necesario incluir ningún campo calculado ya que la antigüedad es un valor que se almacena en las ofertas



3.6.2.4. Inmuebles vendidos por zona tipología y características

INMUEBLES VENDIDOS Por Zona Tipología y Características													
Elementos de Página: Tipología: <Todo> Habitaciones: <Todo> Planta: <Todo> Año: <Todo>													
	Inmuebles Vendidos												
	GENE	FEBRER	MARÇ	ABRIL	MAY	JUNY	JULIOL	AGOST	SETEMBRE	OCTUBRE	NOVEMBRE	DESEMBRE	Suma
Andorra	1	1		1	2	5	7	7	2	2	4	2	34
andorra	1	1		1	2	5	7	7	2	2	4	2	34
encamp - canillo - ordino	1			1	2	1	1	3	1		1	1	12
massana - andorra - escaldes		1				2	1	1		1			6
					2	5	3	1	1	3	1		16
España	2410	2904	3908	4587	3367	3302	3854	4585	3181	3479	4343	3304	43224
alicante	285	273	487	597	395	345	524	467	332	356	532	368	4961
barcelona	1266	1517	1664	1990	1562	1540	1632	2180	1484	1632	2120	1594	20181
castellón	73	78	132	116	95	146	100	240	94	149	156	113	1492
grona	109	134	122	178	135	147	166	204	127	146	196	189	1853
alt empordà	19	20	15	17	21	19	22	94	22	24	33	34	340
bax empordà	24	24	33	78	34	41	25	33	34	31	62	57	476
gronès	15	23	26	22	14	28	34	22	22	33	34	20	293
àrea de celrà	1		2	1			2		1		4	1	12
àrea de grona	9	12	13	11	7	17	12	16	12	16	15	11	151
àrea de lagostera	1	4	7	3	5	2	9	1		6	3	3	44
àrea de quart		1				1	2			1	2		7
àrea de salt	3	2	3	2	1	3	5	3	5	5	2	4	38
àrea de sant gregori	1	4	1	5	1	5	4	2	4	5	8	1	41
la cerdanya	5	8	8	8	6	5	3	2	1	9	8	9	72
la garroba	2	2		3	1	3	2	3	2	3	3		24
la selva	38	55	38	45	58	45	77	47	45	42	53	65	608
pa de Zestany	1			2		2	1	1	1			2	10
rpòles	5	2	2	3	1	4	2	2		4	3	2	30
l'iza-formentera	11	8	14	13	17	6	13	9	15	13	15	9	143
leida	29	35	17	18	27	43	16	27	30	28	37	31	338

Figura 21 Consulta 4

Esta hoja muestra en forma de matriz los inmuebles vendidos, esto es aquellos que tienen almacenado SI en el campo vendido de la tabla ofertas. En este caso para mostrar de forma más evidente la evolución temporal se ha optado por situar el año junto con el resto de las características y tipologías en la primera barra superior y situar los meses como cabecera de columna, aunque como ya se comentó anteriormente el nivel de detalle es fácilmente modificable por el usuario.

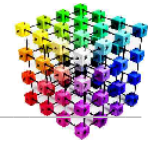
Al igual que en el caso anterior esta consulta solo es posible si se establece la condición de “vendido = SI”

Se ha añadido a la derecha un campo suma que totaliza los valores de cada fila.

3.6.2.5. Inmuebles disponibles por zona tipología y características

Esta consulta muestra en forma de tabla un listado de aquellos inmuebles no vendidos. Para facilitar la comprensión de la información se ha situado el selector de país en la barra superior junto al resto de las características y la información de zonas se ha incluido en el cuerpo de la tabla para hacer evidente la agrupación de los valores por zonas.

La información central de la consulta, el número de inmuebles disponibles se sitúa en la primera fila.



INMUEBLES DISPONIBLES Por Zona Tipología y Características					
Elementos de Página: Año: 2006 Mes: GENER Tipología: chalets País: España					
	Zona1	Zona2	Zona3	Zona4	Inmuebles Disponibles
1	girona	gironès	área de salt		3
2	girona	gironès	área de quart		1
3	girona	gironès	área de celrà		2
4	girona	gironès	área de girona	example	1
5	girona	gironès	área de girona	montjuic	1
6	girona	gironès	área de girona	fontajau - taiala	1
7	girona	gironès	área de girona	palau - montilivi	5
8	girona	gironès	área de girona	sta. eugenia - mas xirgu	1
9	girona	gironès	área de lagostera		21
10	girona	gironès	área de sant gregori		31
11	girona	la selva	área de blanès		13
12	girona	la selva	área de vidreres		10
13	girona	la selva	área de tossa de mar		20
14	girona	la selva	área de lloret de mar		152
15	girona	la selva	área de caldes de malavela		6
16	girona	la selva	área de arbúcies - hostalric		3
17	girona	la selva	área de santa coloma de farners		178
18	girona	ripollès	área de ripoll		3
19	girona	ripollès	área de ribes de freser		12
20	girona	la cerdanya	área de puigcerdà		24
21	girona	la cerdanya	área de belver de cerdanya		5
22	girona	la garrotxa			10
23	girona	alt empordà	área de figueres		2
24	girona	alt empordà	área de llançà		12
25	girona	alt empordà	área de cadaqués		32
26	girona	alt empordà	área de la jonquera		48
27	girona	alt empordà	área de castelló d'empúries		56
28	girona	baix empordà	área de país		49

Figura 22 Consulta 5

Para la realización de la consulta ha sido necesario imponer la condición “vendido = NO”, agrupar los valores por zona y aplicar el calculo siguiente para contar los valores diferentes:

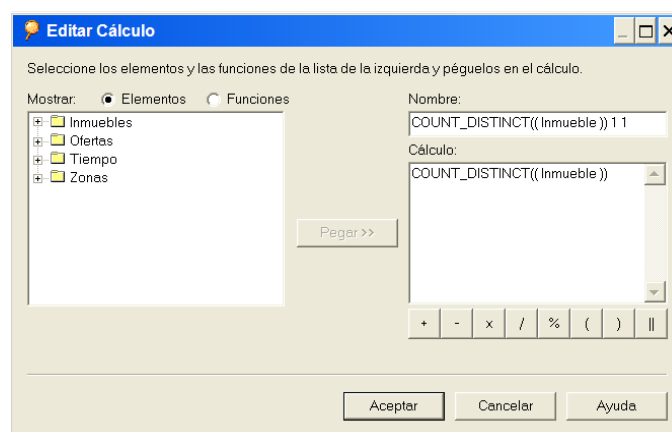
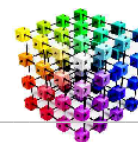


Figura 23 Inserción de un campo calculado

Es evidente que una consulta como esta solo tiene sentido si se aplica para un mes concreto

3.6.2.6. Inmuebles que bajan de precio en un mes con respecto al mes anterior

El almacenamiento en la tabla OFERTAS de la diferencia de precio de cada inmueble con el precio del mes anterior permite obtener una consulta como esta, en la que se muestra este valor solo para los inmuebles que han bajado de precio en la columna depreciación.



INMUEBLES QUE BAJAN DE PRECIO (En un mes con respecto al anterior) Por Zona Tipología y Características									
Elementos de Página: Tipología: pisos Anio: 2007 Mes: JULIOL									
	Pais	Zona1	Zona2	Codigo	Habitaciones	Planta	Superficie	Depreciación	
> 1	España	tarragona	baix camp	VW0000000660072	2	bj	85,00	-€6.010	
> 2	España	tarragona	baix camp	VW0000000728260	2	bj	95,00	-€15.000	
> 3	España	malorca	levante	VP0000000948095	1	3	55,00	-€16.000	
> 4	España	leida	segarra	VW0000000800012	4	2	100,00	-€14.700	
> 5	España	girona	gironès	VW0000000418806	3	2	75,00	-€5.000	
> 6	España	girona	alt empordà	VW0000000817777	3	5	75,00	-€11.000	
> 7	España	alicante	l'Alacantí	VC0000000532732	4	1	100,00	-€29.000	
> 8	España	alicante	l'Alacantí	VP0000000751751	3	5	110,00	-€30.000	
> 9	España	alicante	l'Alacantí	VW0000000716692	3	6	110,00	-€6.400	
> 10	España	alicante	vinalopó mitjà	VW0000000665777	3	2	114,00	-€3.400	
> 11	España	alicante	marina alta	VW0000000458188	3	bj	142,00	-€5.000	
> 12	España	alicante	baix segura	VW0000000606831	1	2	46,00	-€4.000	
> 13	España	ibiza-formentera	área de eivissa	VW0000000657193	3	1	98,00	-€10.000	
> 14	España	barcelona	barcelonès	VW0000000726864	4	2	85,00	-€5.985	
> 15	España	barcelona	barcelonès	VW0000000619486	3	2	65,00	-€7.500	
> 16	España	barcelona	barcelonès	VP0000000900200	3	2	100,00	-€72.000	
> 17	España	barcelona	barcelonès	VP0000000987496	2	4	65,00	-€45.000	
> 18	España	barcelona	barcelonès	VW0000000670523	2	2	50,00	-€13.294	
> 19	España	barcelona	barcelonès	VW0000000693464	1	bj	42,00	-€6.000	
> 20	España	barcelona	barcelonès	VW0000000597447	4	3	115,00	-€12.057	
> 21	España	barcelona	barcelonès	VW0000000542053	3	1	75,00	-€5.000	
> 22	España	barcelona	barcelonès	VP0000000838644	3	2	60,00	-€18.031	
> 23	España	barcelona	barcelonès	VW0000000628063	2	1	65,00	-€5.000	
> 24	España	barcelona	barcelonès	VW0000000515915	3	5	90,00	-€11.250	
> 25	España	barcelona	barcelonès	VW0000000636236	4	1	110,00	-€30.000	

Figura 24 Consulta 6

Como en el caso anterior la consulta ha sido preparada para que muestre la información por meses.

Las cabeceras de columnas referentes al país y las zonas geográficas pueden ser modificadas por el usuario en cuanto al nivel de detalle que muestran.

Se ha empleado la condición “*apreciación < 0*”, para mostrar solo los inmuebles que han bajado de precio.

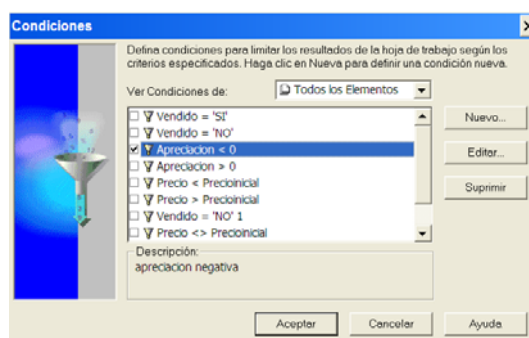
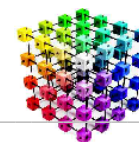


Figura 25 Inserción de una condición



3.6.2.7. Inmuebles que suben de precio en un mes con respecto al mes anterior

Esta consulta esta construida de forma similar a la anterior pero para mostrar solo los inmuebles que han subido de precio se emplea la condición “*apreciación < 0*”.

INMUEBLES QUE SUBEN DE PRECIO (En un mes con respecto al anterior) Por Zona Tipología y Características								
Elementos de Página: Tipología: pisos Año: 2007 Mes: JULIOL								
	Pais	Zona1	Zona2	Codigo	Habitaciones	Planta	Superficie	Apreciación
> 1	España	girona	la cerdanya	VP0000000895206	2	3	45,00	€3.180
> 2	España	alicante	izalacantí	VW0000000760616	3	1	108,00	€1.742
> 3	España	barcelona	baix llobregat nord	VW0000000698829	3	bj	250,00	€4.749
> 4	España	valencia	valencia	VP0000000859979	3	1	95,00	€1
> 5	España	valencia	valencia	VP0000000893114	3	1	80,00	€50
> 6	España	valencia	valencia	VP0000000956581	4	7	170,00	€55.000
> 7	España	valencia	el camp de túria	VP0000000868193	4	3	140,00	€12.000
> 8	España	alicante	izalacantí	VP0000000628503	3	5	104,00	€4.000
> 9	España	alicante	marina alta	VW0000000668009	2	1	85,00	€3.000
> 10	España	alicante	marina alta	VP0000000825649	2	bj	82,00	€60.000
> 11	España	alicante	marina alta	VW0000000452236	3	3	90,00	€100.000
> 12	España	barcelona	barcelonès	VP0000000998602	0	1	70,00	€12.028
> 13	España	barcelona	garraf	VW0000000645236	3	1	94,00	€15.000
> 14	España	barcelona	barcelonès	VP0000000955831	3	5	50,00	€12.000
> 15	España	barcelona	valès occidental	VP0000000961752	3	4	74,00	€1.204
> 16	España	barcelona	valès occidental	VW0000000518737	4	9	105,00	€3.000
> 17	España	valencia	valencia	VP0000000889871	1	1	117,00	€4.000
> 18	España	valencia	valencia	VC0000000789964	3	1	60,00	€500
> 19	España	valencia	valencia	VW0000000659925	4	8	151,00	€30.000
> 20	España	castelón	plana baixa	VC0000000601954	1	3	42,00	€10.000
> 21	España	girona	baix empordà	VW0000000804380	3	1	90,00	€29.545
> 22	España	alicante	baix segura	VW0000000585486	2	1	76,00	€1.000
> 23	España	barcelona	barcelonès	VW0000000627648	3	5	101,00	€3.000
> 24	España	barcelona	barcelonès	VP0000000885099	2	2	80,00	€166.096
> 25	España	barcelona	valès oriental	VW0000000678060	2	2	95,00	€6.010

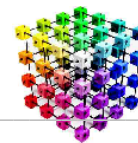
Figura 26 Consulta 7

3.6.2.8. Inmuebles que bajan de precio en un mes con respecto al precio inicial

Por razones de eficiencia el precio inicial del inmueble se guarda en la tabla INMUEBLE, mientras que el precio mensual se almacena en la tabla OFERTAS. Esta consulta recupera ambos precios y muestra en forma de tabla, para un mes seleccionable por el usuario el precio inicial de venta de cada inmueble, su precio actual y un campo calculado con la diferencia entre ambos.

El nivel de detalle geográfico es seleccionable por el usuario final.

La consulta emplea la condición “*precio < precioInicial*” y el campo calculado diferencia.



INMUEBLES QUE BAJAN DE PRECIO (Diferencia con el precio inicial) Por Zona Tipología y Características							
Elementos de Página: Tipología: pisos Anio: 2007 Mes: JULIOL Pais: España Zona1: Alicante							
	Codigo	Habitaciones	Planta	Superficie	PrecioInicial	Precio Actual	Diferencia
1	VW0000000670388	3	3	90,00	105000,00	95000,00	-10000,00
2	VW0000000714968	1	3	60,00	97000,00	95000,00	-2000,00
3	VP0000000824201	2	3	68,00	215000,00	206000,00	-9000,00
4	VP0000000827141	3	2	110,00	284000,00	263000,00	-21000,00
5	VC0000000652420	3	1	95,00	180000,00	160000,00	-20000,00
6	VC0000000778296	4	4	140,00	126213,00	124000,00	-2213,00
7	VP0000000860990	3	5	110,00	310000,00	300000,00	-10000,00
8	VP0000000880149	3	3	90,00	370000,00	350000,00	-20000,00
9	VP0000000974182	1	bj	66,00	166000,00	163000,00	-3000,00
10	VP0000000892585	3	8	110,00	291000,00	280000,00	-11000,00
11	VC0000000532732	4	1	100,00	330000,00	270000,00	-60000,00
12	VC0000000583215	2	1	70,00	250000,00	245000,00	-5000,00
13	VC0000000702728	2	2	49,00	228000,00	216364,00	-11636,00
14	VP0000000555741	3	bj	110,00	235000,00	210000,00	-25000,00
15	VW0000000715359	4	3	130,00	128000,00	120000,00	-8000,00
16	VP0000000793702	3	3	80,00	146000,00	141000,00	-5000,00
17	VP0000000821958	2	1	85,00	216364,00	201339,00	-15025,00
18	VW0000000196620	2	en	103,00	93000,00	90000,00	-3000,00
19	VP0000000822222	2	8	85,00	210354,00	189318,00	-21036,00
20	VP0000000822290	2	2	65,00	189318,00	176698,00	-12620,00
21	VC0000000459434	3	5	100,00	190000,00	179900,00	-10100,00
22	VC0000000561727	3	5	105,00	240405,00	225000,00	-15405,00
23	VP0000000765097	3	2	100,00	195329,00	186312,00	-9017,00
24	VP0000000797650	3	2	120,00	441745,00	407000,00	-34745,00
25	VW0000000756849	2	5	82,00	280000,00	275000,00	-5000,00
26	VW0000000776460	3	3	120,00	375000,00	360000,00	-15000,00
27	VC0000000585006	2	8	78,00	199000,00	193000,00	-6000,00

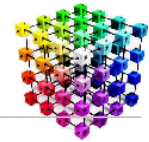
Consulta 1 Consulta 8

3.6.2.9. Inmuebles que suben de precio en un mes con respecto al precio inicial

INMUEBLES QUE SUBEN DE PRECIO (Diferencia con el precio inicial) Por Zona Tipología y Características							
Elementos de Página: Tipología: chalets Anio: 2007 Mes: JULIOL Pais: España Zona1: Alicante							
	Codigo	Habitaciones	Planta	Superficie	PrecioInicial	Precio Actual	Diferencia
1	VW0000000681053	4	--	120,00	36000,00	225.000,00	189000,00
2	VW0000000681068	5	--	132,00	41500,00	45.000,00	3500,00
3	VW0000000591019	4	--	357,00	432000,00	448.000,00	16000,00
4	VW0000000549589	5	--	260,00	275000,00	289.000,00	14000,00
5	VW0000000276758	5	--	377,00	700000,00	760.000,00	60000,00
6	VW0000000455933	3	--	120,00	225380,00	228.000,00	2620,00
7	VW0000000704480	4	--	320,00	420000,00	450.000,00	30000,00
8	VP0000000709729	5	--	250,00	420375,00	461.750,00	41375,00
9	VP0000000864925	6	--	450,00	812000,00	818.000,00	6000,00
10	VC0000000652413	4	--	200,00	270455,00	290.000,00	19545,00
11	VW0000000618812	4	--	300,00	374000,00	399.000,00	25000,00
12	VC0000000356789	3	--	220,00	332000,00	365.000,00	33000,00
13	VW0000000146582	5	--	370,00	2200000,00	2.500.000,00	300000,00
14	VW0000000401149	4	--	320,00	689000,00	690.000,00	1000,00
15	VW0000000692721	2	--	165,00	159000,00	169.000,00	10000,00
16	VW0000000692593	5	--	350,00	541000,00	571.000,00	30000,00
17	VW0000000709874	3	--	140,00	175000,00	180.000,00	5000,00
18	VW0000000607968	4	--	296,00	1200000,00	1.250.000,00	50000,00
19	VW0000000175837	3	--	97,00	245000,00	250.000,00	5000,00
20	VW0000000261240	3	--	135,00	164000,00	175.000,00	11000,00
21	VW0000000174071	5	--	620,00	550000,00	580.000,00	30000,00
22	VW0000000276399	3	--	146,00	187000,00	192.000,00	5000,00
23	VP0000000906710	3	--	120,00	240000,00	265.000,00	25000,00
24	VW0000000475074	2	--	67,00	210000,00	260.000,00	50000,00
25	VW0000000516218	5	--	380,00	530000,00	540.000,00	10000,00
26	VW0000000448827	4	--	140,00	225000,00	226.000,00	1000,00
27	VP0000000837861	4	--	154,00	390000,00	420.000,00	30000,00

Figura 27 Consulta 9

Esta consulta es similar a la anterior solo que incluyendo la condición “*precio > precioInicial*”.



3.6.2.10. Inmuebles que cambian de precio en un mes con respecto al precio inicial

INMUEBLES QUE CAMBIAN DE PRECIO (Diferencia con el precio inicial) Por Zona Tipología y Características									
Elementos de Página: Tipología: chalets Anio: 2007 Mes: JULIOL Pais: España Zona1: alcante									
	Zona2	Codigo	Habitaciones	Planta	Superficie	PrecioInicial	Precio Actual	Diferencia	
>	102	baix segura	VC0000000550858	3	--	75,00	126000,00	119500,00	-6500,00
>	103	baix segura	VC0000000756192	3	bj	250,00	350000,00	299000,00	-51000,00
>	104	baix segura	VP0000000602830	2	--	85,00	127000,00	125000,00	-2000,00
>	105	baix segura	VW0000000607968	4	--	296,00	1200000,00	1250000,00	50000,00
>	106	baix segura	VC0000000391301	3	--	80,00	145000,00	135000,00	-10000,00
>	107	baix segura	VP0000000867385	1	--	60,00	92697,00	74157,00	-18540,00
>	108	baix segura	VW0000000175837	3	--	97,00	245000,00	250000,00	5000,00
>	109	baix segura	VW0000000248226	3	--	320,00	500000,00	450000,00	-50000,00
>	110	baix segura	VW0000000261240	3	--	135,00	164000,00	175000,00	11000,00
>	111	baix segura	VP0000000323543	2	--	70,00	117000,00	110000,00	-7000,00
>	112	baix segura	VW0000000357390	2	--	65,00	120000,00	105000,00	-15000,00
>	113	baix segura	VW0000000174071	5	--	620,00	550000,00	580000,00	30000,00
>	114	baix segura	VW0000000276399	3	--	146,00	187000,00	192000,00	5000,00
>	115	baix segura	VP0000000626272	2	--	80,00	159000,00	155000,00	-4000,00
>	116	baix segura	VC0000000647911	3	--	100,00	190000,00	179000,00	-11000,00
>	117	baix segura	VP0000000621985	3	--	90,00	229000,00	215000,00	-14000,00
>	118	baix segura	VP0000000741819	3	--	70,00	171000,00	150000,00	-21000,00
>	119	baix segura	VP0000000906710	3	--	120,00	240000,00	265000,00	25000,00
>	120	baix segura	VP0000000971452	2	--	65,00	100000,00	99000,00	-1000,00
>	121	baix segura	VW0000000245816	2	--	70,00	130000,00	115000,00	-15000,00
>	122	baix segura	VW0000000319581	2	--	52,00	110000,00	108000,00	-2000,00
>	123	baix segura	VW0000000426533	1	--	36,00	74000,00	72000,00	-2000,00
>	124	baix segura	VW0000000475074	2	--	67,00	210000,00	260000,00	50000,00
>	125	baix segura	VW0000000312016	2	--	67,00	120202,00	114200,00	-6002,00
>	126	baix segura	VW0000000615020	2	--	50,00	100000,00	95000,00	-5000,00
>	127	baix segura	VW0000000643754	1	--	36,00	81140,00	72125,00	-9015,00
>	128	baix segura	VW0000000673211	2	--	78,00	106000,00	96000,00	-10000,00

Figura 28 Consulta 10

Esta consulta es similar a las dos anteriores, solo que en ella se muestran a la vez las subidas y las bajadas y se emplea el formato de excepciones para marcar de diferente color cada caso.

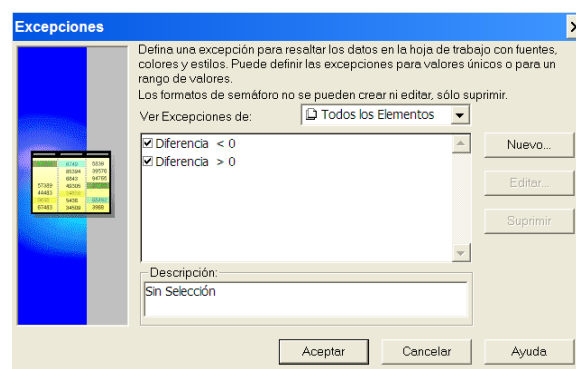
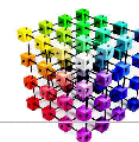


Figura 29 Inserción de una condición

Esta consulta se emplea la condición “precio <> precioInicial”.



3.6.2.11. Evolución anual de precios por zona y tipología de los inmuebles que cambian de precio

EVOLUCIÓN DE PRECIOS Por Zona Tipología y Características (Variación mensual)											
Elementos de Página: Tipología: estudios ▾ Habitaciones: <Todo> ▾ Planta: <Todo> ▾ País: España ▾ Año: 2006 ▾											
	> FEBRER	> MARÇ	> ABRIL	> MAIG	> JUNY	> JULIOL	> AGOST	> SETEMBRE	> OCTUBRE	> NOVENBRE	> DESEMBRE
	Apreciación	Apreciación	Apreciación	Apreciación	Apreciación	Apreciación	Apreciación	Apreciación	Apreciación	Apreciación	Apreciación
VC0000000317562				€5.833							
VC0000000390863							-€9.000				
VC0000000438918								€6.000			
VC0000000634356								-€8.000			
VC0000000651134				-€60.101					-€18.030		
VP0000000476961						-€8.717					
VP0000000499593				€9.181							
VP0000000540075							-€12.964				
VP0000000621193			-€37.000								
VP0000000625834					-€5.000						
VP0000000648250			-€4.500								
VP0000000671566		-€12.020									
VP0000000692086											-€10.000
VP0000000703783		€6.316									
VP0000000707823						-€7.811					
VP0000000715297		€59.580									
VP0000000739780		-€6.010									
VP0000000751546					-€736	€1.583					
VP0000000760995			-€11.000							-€17.700	
VP0000000762350					€4.000	€2.920					
VP0000000763588				-€1.000							
VP0000000774400						-€2.590					
VP0000000775418								-€1.568			

Figura 30 Consulta 11

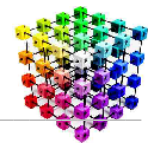
Esta consulta muestra una columna por cada mes del año seleccionado y en ella aparecen solamente aquellos inmuebles que han cambiado de precio con respecto al mes anterior, “apreciación $\neq 0$ ”. En la parte Izquierda puede consultarse el código del inmueble y la opción de formatear excepciones permite ver en colores diferentes las subidas y las bajadas de precio.

3.7. DETERMINACIÓN DEL INMUEBLE TIPO

Uno de los requerimientos impuestos por el documento de requisitos es la determinación del “inmueble tipo” español y andorrano tomando como base para esta estimación los datos de los inmuebles ofertados por la empresa inmobiliaria.

Podemos interpretar este requerimiento de diversas maneras y cada una de ellas llevará asociada un procedimiento de cálculo diferente.

Veamos en primer lugar que medidas podemos emplear para seleccionar posteriormente la más adecuada.



3.7.1. Medidas de tendencia central

Se entiende por medida de tendencia central un único valor que resume un conjunto de datos y señala el centro de los valores de una serie.

No hay una sola medida de tendencia central, se consideran hasta 5: la media aritmética, media ponderada, la mediana, la moda y la media geométrica. Veamos las que pueden sernos útiles:

La Media Aritmética

La medida de tendencia central más ampliamente usada es la media aritmética, usualmente abreviada como media.

Es la medida de tendencia central más obvia que se puede elegir, consiste en el valor obtenido sumando las observaciones y dividiendo esta suma por el número de observaciones que hay en el grupo. La media resume en un valor las características de una variable teniendo en cuenta a todos los casos.

Desventajas de la media aritmética

Si alguno de los valores es extremadamente grande o extremadamente pequeño, la media no es el promedio apropiado para representar la serie de datos. La media se ve afectada de modo notable por valores extraordinariamente grandes o pequeños, por lo que en nuestro caso la presencia de inmuebles con características extremas, como sería el caso de precios muy altos o muy bajos desvirtuaría el resultado obtenido.

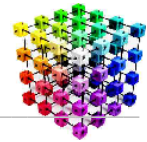
La Mediana

La Mediana es el punto medio de los valores después de ordenarlos de menor a mayor, o de mayor a menor. Se tiene entonces que 50% de las observaciones se encuentran por encima de la mediana y 50% por debajo de ella.

Desventajas de la mediana

Esta parece una medida más ajustada para nuestro propósito, pero si os encontramos ante un conjunto en el que priman los extremos, supongamos por ejemplo que solo se venden inmuebles muy grandes y muy pequeños, podríamos sacar la conclusión de que el inmueble representativo es aquel que presenta un tamaño intermedio aunque no exista ninguno en el catálogo.

La Moda



La moda es el valor de la observación que aparece con más frecuencia.

Puede determinarse para todos los niveles de datos: nominal, ordinal, de intervalo y de razón y no se ve afectada por valores muy altos o muy bajos.

Por todo lo visto hasta ahora podemos concluir que la moda es la tendencia de medida central que mejor se ajusta a nuestro propósito.

3.7.2. proceso y resultados

La moda se obtiene calculando la frecuencia de aparición de cada elemento, ordenando esta, y seleccionado el valor más alto.

Esta operación podría llevarse a cabo empleando un script de PL/SQL que realice la tarea, pero el SQL de Oracle ya dispone entre sus funciones estadísticas de una específica para el cálculo de la moda.

La función `STATS_MODE` facilita el cálculo de la moda y se ha empleado para obtener el valor más frecuente de tipología, superficie, número de dormitorios, planta y precio sobre la tabla de inmuebles.

Moda de todos los inmuebles sin diferenciar el país:

```
SELECT STATS_MODE(tipologia)      from INMUEBLES ;
SELECT STATS_MODE(superficie)     from INMUEBLES ;
SELECT STATS_MODE(habitaciones)   from INMUEBLES ;
SELECT STATS_MODE(precioinicial)  from INMUEBLES ;
SELECT STATS_MODE(planta)         from INMUEBLES ;
```

Resultados:

Tipología	Superficie	Dormitorios	Planta	Precio Inicial
piso	90 m ²	3	-- * 1	240000

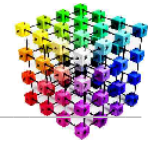
* La planta más frecuente en el listado de inmuebles corresponde a la indefinida (--) por eso se ha obtenido el segundo valor más frecuente mediante la siguiente sentencia SQL:

```
SELECT STATS_MODE(planta)
FROM(
    select planta from INMUEBLES where planta<>'--';
```

Obteniendose el valor 1 para la moda.

El resto de valores modales por país se han calculado con sentencias SQL que siguen el siguiente patrón:

```
SELECT
STATS_MODE( INMUEBLES.TIPOLOGIA) as ModaTipologia
```



```

FROM INMUEBLES
WHERE INMUEBLES.ID IN(
  SELECT
  DISTINCT (OFERTAS.INMUEBLE)
  FROM OFERTAS,ZONAS
  WHERE
  ZONAS.PAIS = 'Andorra' AND
  OFERTAS.ZONA =ZONAS.ID
)
    
```

Cambiando en cada caso el elemento sobre el que se realiza la consulta y el país al que corresponde el inmueble.

Del conjunto de sentencias se obtienen los siguientes resultados:

Pais	Tipología	Superficie	Dormitorios	Planta	Precio Inicial
España	piso	90 m ²	3	-- * 1	240000
Andorra	piso	100 m ²	3	1	210000

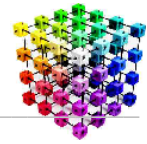
* Se toma el segundo valor más frecuente como ya se ha explicado

Podemos concluir que el inmueble tipo es un piso de tres dormitorios situado en la primera planta en ambos países, que en España es de 90 metros cuadrados mientras que en Andorra es de 100 m² y que cuesta 240.000€ en España y 210.000 en Andorra

4. CONCLUSIONES

Tras la exposición del proceso seguido y de los resultados obtenidos en cada fase del mismo se pueden extraer las siguientes conclusiones.

- Se han realizado un análisis previo de los datos que ha servido de punto de partida para la realización de las siguientes fases.
- Se han diseñado y llevado a la práctica los modelos conceptual, lógico y físico que definen el cubo multidimensional que ha dado soporte al almacén de datos.
- Se han generado procesos automatizados de creación de tablas, transformación y carga de datos.
- Se han definido índices que aceleran las consultas sobre la base de datos.
- Se ha creado un área de negocio en Discoverer Administrador y un libro de trabajo en Discoverer Desktop que permiten la generación y visualización de las consultas preestablecidas y “ad-hoc”.
- Se han generado 11 consultas que incluyen todas aquellas solicitadas en los requisitos



además de otras relacionadas.

- Se ha determinado el inmueble tipo, tanto de forma conjunta como por país.
- Se puede afirmar que se han cumplido los requisitos del proyecto

5. LINEAS DE EVOLUCIÓN FUTURA

Los requerimientos impuestos por la empresa inmobiliaria para la realización de este proyecto pueden ser considerados como básicos. Se trata de una primera solución al problema de la toma de decisiones basadas en datos estructurados, pero que es claramente mejorable mediante un proyecto de mayor alcance y duración.

El proyecto y los productos de él obtenidos son claramente mejorables atendiendo a los siguientes aspectos:

- Estandarización de la entrada de datos.

El proceso de análisis y transformación de datos ha permitido evidenciar incoherencias y errores en los mismos. La información obtenida de este proceso debería emplearse en la mejora del proceso de entrada de datos de forma que dichas incoherencia sean evitadas o reducidas.

- Aceleración del proceso de carga.

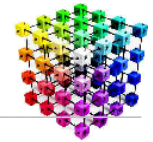
El proceso de carga de datos desde la tabla temporal a las tablas definitivas, ha consumido cerca de dos minutos ejecutándose en un equipo relativamente antiguo y poco potente, pero aún así, debe tenerse en cuenta que el volumen de datos tratados es relativamente pequeño en comparación común lo que suele ser habitual en este tipo de proyectos por lo que el proceso puede y debe ser optimizado para poder manejar un volumen de datos mucho más amplio en un tiempo aceptable.

- Tolerancia a fallos.

El script de carga definitiva de datos contempla diversos tipos de excepciones que pueden producirse durante su ejecución. Estas posibles excepciones han sido verificadas con el conjunto de datos facilitados, pero sería recomendable aumentar la tolerancia a fallos realizando ensayos con conjuntos de datos más amplios.

- Aceleración de las consultas.

Se ha conseguido reducir significativamente el tiempo empleado por las consultas mediante el empleo de índices sobre las tablas. Como línea de actuación futura sería recomendable mejorar más aún estos tiempos mediante el empleo de resúmenes una vez que la base de datos esté definitivamente instalada en el servidor del cliente.



- Consultas vía Web

Con las herramientas disponibles para la realización de este proyecto las consultas han de ser visualizadas empleando Oracle Discoverer Desktop que ha de estar instalado en la parte cliente. Con el uso de herramientas más avanzadas como Oracle Discoverer Plus las consultas podrían realizarse desde un navegador instalado en cualquier equipo.

6. GLOSARIO

Ad-Hoc (Consulta): Una consulta que no ha sido diseñada a priori en el momento de determinar las consultas. Suele construirse de manera dinámica con herramientas diseñadas al efecto.

Agregación: La agregación de datos es el proceso mediante el cual se redefinen los mismos mediante sumariación basada en algunas reglas o criterios.

Almacén de datos: Un almacén de datos es un conjunto de datos orientado a una temática concreta, integrado, variable en el tiempo y no volátil que da soporte a la toma de decisiones.

Atributo: Un atributo representa un tipo básico de información en una dimensión. Por ejemplo año es un atributo en la dimensión tiempo .

Cubo OLAP: También llamado cubo multidimensional o hipercubo . Se compone de hechos numéricos llamados medidas que se clasifican por dimensiones. El cubo de metadatos se suele crear a partir de un esquema en estrella o copo de nieve, esquema de las tablas en una base de datos relacional. Las medidas se obtienen de los registros de una tabla de hechos y las dimensiones se derivan de la dimensión de los cuadros.

Data Warehouse: Almacén de datos.

Dimensión: Una estructura que representa una de las caras de un cubo. Cada dimensión representa una categoría diferente, como región, tiempo, inmueble, etc., son las perspectivas de análisis de las variables (forman parte de las tablas de dimensiones).

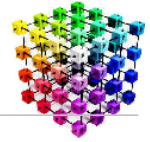
Estrella: Una de las posibles formas de representar un modelo dimensional. En un esquema de estrella, cada dimensión esta representada por una única tabla de dimensión.

ETL: Denominación genérica del proceso de extracción, transformación y carga (Extraction, Transformation, and Loading).

Jerarquía: Cada una de las dimensiones de un cubo OLAP puede resumirse mediante una jerarquía que define una estructura de navegación en función de un mayor o menor detalle sobre la dimensión estudiada.

Medida: Atributo numérico de una celda. Las medidas se obtienen de los registros de una tabla de hechos.

Metadato: Se les suele definir como "*datos acerca de los datos*", ya que se trata de datos que describen cuál es la estructura de los datos que se van a almacenar y cómo se relacionan entre ellos.



Modelo dimensional: Un tipo de modelo de datos empleado en el Almacén de Datos. En un modelo dimensional existen dos tipos de tablas de dimensiones y tablas de hechos.

OLAP: Acrónimo de On-Line Analytical Processing (procesamiento analítico en línea). Se trata de una solución utilizada en el campo de la llamada Inteligencia empresarial (o *Business Intelligence*) cuyo objetivo es agilizar la consulta de grandes cantidades de datos.

Tabla de dimensiones: Guarda los registros relativos a una dimensión concreta. No se almacenan hechos en las tablas de dimensiones.

Tabla de hechos: una de las tablas que constituyen el modelo dimensional. Una tabla de hechos típica contiene dos tipos de columnas, los hechos en sí y las claves relativas a las tablas de dimensiones.

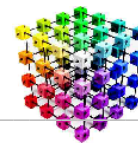
7. BIBLIOGRAFÍA

7.1. PUBLICACIONES:

- Building a Data Warehouse Using Oracle OLAP Tools. (1997) Oracle Technical Report Satish Mahajan
- Building the Data Warehouse, Fourth Edition (2005) W. H. Inmon
- Data Warehouse Database Design (2001) M. Lea Shaw.
- Getting Started with Data Warehouse and Business Intelligence (1999) Maria Sueli Almeida, Missao Ishikawa, Joerg Reinschmidt, Torsten Roeber
- Mastering Data Warehouse Design Relational And Dimensional Techniques (2003) Claudia Imhoff
- Oracle Data Warehouse Management. Secrets of Oracle Data Warehousing (2003) Mike Ault
- Warehouse Management Guide (2000) SAP AG

7.2. WEBS:

- <http://www.inmoncif.com>
- <http://www.rkimball.com>
- <http://www.dwreview.com>
- <http://www.1keydata.com/datawarehousing/datawarehouse.html>
- <http://www.intranetjournal.com/features/datawarehousing.html>



8. ANEXOS

8.1. RESTAURACIÓN DE LA BASE DE DATOS

Para permitir la portabilidad del proyecto el esquema de usuario incluyendo las tablas, las secuencias, los triggers los procedimientos y funciones almacenadas, índices, así como el EUL han sido exportados usando la aplicación exp de Oracle.

```
D:\oracle>exp file=d:\esquema.dmp log=d:\esquema.log owner=TFC1
```

- El usuario de la base de datos y del EUL es: **TFC1**
- La clave asociada a dicho usuario es: **UOC**
- El fichero de exportación es: *esquema.dmp*
- El fichero de exportación de la base de datos se entrega comprimido en formato RAR como *esquema_dmp.rar* en la carpeta BaseDatos
- El libro de trabajo se ha guardado como fichero de exportación de Discoverer en “**\Discoverer\LibroConsultasInmobiliaria.DIS**”
- El EUL completo del usuario ha sido exportado desde Discoverer Administrator como fichero de exportación conteniendo el área de negocio “**AreaInmobiliaria**” y el libro “**LibroConsultasInmobiliaria**”.

Procedimiento de restauración.

Un usuario con los suficientes privilegios en la base de datos debe ejecutar desde una consola el siguiente comando

```
imp <user>/<password> FILE=esquema.dmp
```

Donde user es el login del usuario y password su clave. El fichero *esquema.dmp* se encuentra comprimido en formato RAR como *esquema_dmp.rar* en la carpeta BaseDatos

Si la aplicación imp no se encuentra en el path de la máquina, deberá indicarse la ruta completa hasta ella.

Después de este paso la base de datos debe estar cargada y el usuario puede, si lo desea, cargar el EUL conteniendo el área y el libro de trabajo: Para ello debe buscar el libro de trabajo *AreaInmobiliariaEUL.eex* en el directorio Discoverer de la entrega:

“*\discoverer\AreaInmobiliariaEUL.eex*” y cargarlo mediante la opción importar del menu Archivo de Oracle Discoverer Administrator.

Pude también abrir el libro de trabajo “*LibroConsultasInmobiliaria.DIS*” desde el Discoverer Desktop con la opción abrir libro de trabajo y seleccionar desde mi computadora.