TRABAJO FIN DE CARRERA

Construcción y explotación de un almacén de datos para el análisis de información inmobiliaria.

José Manuel Gómez Diego
ETIS

Consultor: Pere Juanola Juanola

10 Enero 2012

Universitat Oberta de Catalunya

Resumen, palabras clave y área

Resumen

El presente trabajo aborda el diseño y puesta en funcionamiento de un almacén de datos a petición de una empresa inmobiliaria que desea contar con un sistema de información que le permita automatizar la recogida y explotación de la información inmobiliaria con la que cuentan, además de la realización de una serie de informes que le ayuden a la toma de decisiones de negocio y de esta forma obtener una ventaja competitiva respecto a otras empresas del sector.

Se parte de un fichero de texto aportado por la empresa y generado a partir de un conjunto de hojas de cálculo Microsoft Excel que contienen la información inmobiliaria.

Tras llevar a cabo un análisis previo de los datos se han obtenido los modelos conceptual, lógico y físico que definen el cubo de información multidimensional (cubo OLAP) sobre el que se soportará el almacén de datos.

A continuación se han creado los objetos y estructuras de la base de datos en la que se llevará a cabo la carga de los datos y la transformación de los mismos.

Finalmente, se define un área de negocio que es la base para el diseño de una serie de consultas e informes que constituyen el resultado final del trabajo.

Palabras clave

Almacén de datos, OLAP, ETL, análisis multidimensional, inmobiliaria,Data Warehouse.

Área del Trabajo de Fin de Carrera

Magatzems de dades

Índice de contenidos

Resumen ---------- 2
Palabras clave ---------- 2
Área del Trabajo Fin de Carrera ---------- 2

1 Capítulo 1 – Introducción ---------- 6

1.1 Justificación del TFC y contexto en el que se desarrollará:

punto de partida y aportación del TFC. ----------6

1.2 Objetivos del TFC ----------7

1.2.1 Objetivos del TFC como asignatura ----------7

1.2.2 Objetivos propios del TFC ----------7

1.3 Enfoque y método seguido ----------7

1.4 Planificación del proyecto ---------- 9

1.4.1 Planificación según plan docente ---------- 9

1.4.2 Hitos y fechas clave. ----------10

1.4.3 Planificación detallada de tareas. ---------- 11

1.4.4 Diagrama de Gantt. ---------- 12

1.4.5 Incidencias y riesgos. Plan de contingencia ---------- 14

1.4.6 Valoración y coste del proyecto. ---------- 15

1.5 Productos obtenidos ---------- 15

1.6 Breve descripción de los restantes capítulos de la memoria. ---------- 16

1.6.1 Capítulo 2 – Análisis ---------- 16

1.6.2 Capítulo 3 - Diseño ---------- 16

2 Capítulo 2 – Análisis ---------- 17

2.1 Diagrama de casos de uso ----------17

2.2 Diagrama del modelo conceptual ----------18

2.2.1 Identificar el hecho ----------18

2.2.2 Encontrar la granularidad adecuada ---------- 19

2.2.3 Escoger las dimensiones a emplear en el análisis ---------- 19

2.2.4 Establecer los atributos de cada dimensión ----------19

2.2.5 Distinguir entre descriptores y jerarquías de agregación ----------20

2.2.6 Decidir cuáles son las medidas de interés ----------21

2.2.7 Definir las celdas. ---------- 21

2.2.8 Expliciar las restricciones de integridad ---------- 21

2.2.9 Estudiar la viabilidad ----------22

3 Capítulo 3 – Diseño ---------- 23

3.1 Diagrama de la arquitectura software ---------- 23

3.2 Diagrama de la arquitectura hardware ----------24

3.3 Diseño de las base de datos y diagrama del modelo físico ---------- 25

3.3.1 Diseño lógico ---------- 25

3.3.2 Diseño físico ----------26

3.3.3 Proceso ETL ---------- 27

3.4 Diseño y descripción de los informes creados. ---------- 27

3.4.1 La creación del área de negocio ---------- 27

3.4.2 La creación del libro de trabajo. ---------- 28

3.4.3 Descripción de los informes creados. ---------- 29

3.4.3.1 Número de inmuebles por zona, tipología y características. ---------- 29

3.4.3.2 Listado de inmuebles que han bajado de precio -------- 30

3.4.3.3 Evolución del precio de inmuebles que han bajado de precio respecto a la oferta inicial y final. ---------- 31

3.4.3.4 Precios máximos, mínimos y medios por zona, tipología y características. ---------- 32

3.4.3.5 Metros cuadrados máximos, mínimos y medios por zona, tipología y características. ---------- 32

3.4.3.6 Tiempos medios de venta por zona, tipología y características. ---------- 33

3.4.3.7 Número de inmuebles vendidos y existentes por zona, tipología y característicias. ---------- 34

3.4.3.8 Modelo de inmueble español y andorrano. ---------- 34

4 Capítulo 4 – Conclusiones ---------- 35

5 Capítulo 5 - Líneas de evolución futuras. ---------- 36

Glosario ---------- 36

Bibliografía consultada ---------- 39

Anexos ---------- 39

Anexo A: "Acceso a la base de datos" ---------- 39

Anexo B: "Detalle y automatización del proceso ETL" ---------- 40

Anexo C: "Acceso a Oracle Discoverer Administrator" ---------- 49

Anexo D: "Acceso a Oracle Discoverer Desktop" ---------- 50

Anexo E: "Entorno de desarrollo" ---------- 50

E.1 Hardware ---------- 50

E.2 Software ----------51

Índice de figuras y tablas

Fig. 1 : Etapas del ciclo de vida en cascada y documentos obtenidos ---------- 8
Fig. 2 : Fechas de inicio y de entrega de las tareas planteadas. ---------- 11
Fig. 3 : Diagrama Gantt (parte 1 de 2) ---------- 13
Fig. 4 : Diagrama Gantt (parte 2 de 2) ---------- 13
Fig. 5 : Diagrama de casos de uso ---------- 18
Fig. 6 : Diagrama del modelo conceptual ---------- 23
Fig. 7 : Arquitectura software ---------- 24
Fig. 8 : Arquitectura hardware ---------- 25
Fig. 9 : Diagrama del modelo lógico ---------- 26
Fig. 10 : Esquema físico para la tabla inmuebles. ---------- 26
Fig. 11 : Esquema físico para la tabla ofertas ---------- 26
Fig. 12 : Esquema físico para la tabla Zonas ---------- 27
Fig. 13 : Esquema físico para la tabla Tiempo ---------- 27
Fig. 14 : Imagen editada del espacio de trabajo con los contenidos referidos ---------- 28
Fig. 15 : Imagen de informe Número de Inmuebles ---------- 29
Fig. 16 : Imagen de informe Listado de Inmuebles bajaron precio respecto al mes anterior ---------- 30
Fig. 17 : Imagen de informe Listado de Inmuebles bajaron precio ofertas inicial y final -31
Fig. 18 : Imagen de informe Evolución de precio de inmuebles ---------- 31
Fig. 19 : Imagen de informe de precios máximos, mínimos y medios ----------32
Fig. 20 : Imagen de informe de superficies máximas, mínimas y medias ----------33
Fig. 21 :Imagen de informe Tiempo medio de venta ----------34
Fig. 22 : Imagen de informe Número de inmuebles vendidos/disponbiles ---------- 34
Fig. 23 : Imagen de informe Modelo de inmueble por precio y tiempo ---------- 34
Fig. 24 : Imagen de informe Modelo de inmueble por superficie y número de dormitorios -- 35
Fig. 25 : Detalle del usuario y propiedades de conexión ---------- 40
Fig. 26 : Esquema uoc_usr ---------- 40
Fig. 27 : Imagen del resultado de los procesos de importación y procesado de datos -- 41
Fig. 28 :Imagen del contenido de los ficheros de carga y control con detalle de línea a editar 42
Fig. 29 : Imagen del contenido del fichero de log ---------- 42
Fig. 30 : Imagen del contenido del fichero Carga.bad ---------- 43
Fig. 31 : Imagen de la tabla TEMP_DATOS ---------- 44
Fig. 32 : Imagen de la tabla TEMP_INTERVALZONAS ---------- 45
Fig. 33 : Imagen del área de influencia de una jerarquía ---------- 45
Fig. 34 : Imagen de la tabla TEMP_INTERVALZONAS correspondiente ---------- 45
Fig. 35 : Imagen de las partes del procesado de datos ---------- 47
Fig. 36 : Imagen de ejemplo de zonas ---------- 47
Fig. 37 : Imagen proceso de extracción de ofertas de inmuebles ---------- 48
Fig. 38 : Imagen de ejemplo ofertas para un inmueble ---------- 49
Fig. 39 : Imagen de la ventana de acceso de usuario ---------- 49
Fig. 40 :Imágenes del acceso a a Oracle D. Desktop, asignación de privilegios y apertura de libro de trabajo ---------- 50

1. Capítulo 1 – Introducción

1.1 Justificación del TFC y contexto en el que se desarrollará: punto de partida y aportación del TFC

Actualmente las empresas u organizaciones generan gran cantidad de información de negocio de distinta naturaleza (contable, productiva, etc.) que queda registrada en las bases de datos departamentales cuyo volumen aumenta a lo largo del tiempo.

Es indiscutible la importancia que tiene la información del negocio para cualquier empresa, y que dicha información es un activo de carácter estratégico, en el cual las organizaciones fundamentan y justifican sus tomas de decisiones.

Con los sistemas tradicionales de almacenamiento de datos, se hace cada vez más complicada la extracción de información relevante, sea global o detallada, relativa a ciertos aspectos del negocio y comprometen la correcta toma de decisiones.

Por tanto, el disponer de estos datos no significa necesariamente disponer de información.

Para que los datos aporten información con el objetivo de mejorar la competitividad y ayudar a la toma de decisiones estratégicas para la empresa, es imprescindible que tengan una estructura organizada, integrada, lógica, dinámica y de fácil explotación y eso se consigue con los almacenes de datos.

Por otro lado, el acceso a la información no ha de estar condicionado por lo conocimientos informáticos de los distintos analistas por lo que se hace imprescindible disponer de algún tipo de herramienta que facilite dicha consulta, de manera que a partir del almacén de datos corporativo, se acostumbra a diseñar pequeños almacenes de datos departamentales que se diseñan bajo el modelo multidimensional, de manera que se puedan utilizar herramientas OLAP para consultarlos.

En este trabajo se llevará a cabo el diseño, implementación y explotación de un almacén de datos para la empresa “Sostre per a tothom” que le permita realizar diferentes análisis comparativos para ajustar su oferta y demanda y de esta forma obtener una ventaja competitiva respecto a otras empresas del sector.

Para llevarlo a cabo , se parte de un fichero con la información departamental relevante para el caso de estudio que se encuentra en distintas hojas de cálculo y proporcionada en un fichero de texto. A partir de estos datos y de las necesidades planteadas por la empresa, se procederá a la creación de un almacén de datos y a la especificación de un área de negocio que permitirá la consulta eficiente de la información por parte de los analisitas y/o usuarios autorizados a través de informes.

1.2 Objetivos del TFC

A continuación se describen los objetivos a alcanzar tanto de la asignatura como del proyecto en sí.

1.2.1 Objetivos del TFC como asignatura

El Trabajo Fin de Carrera tiene como principal objetivo mostrar el aprendizaje que se ha logrado con los estudios de Ingeniería Técnica en Informática.

Este objetivo se concreta:

- Analizar un problema complejo de tipo práctico transformándolo en un proyecto informático.
- Planificar y estructurar el desarrollo del proyecto mediante la elaboración de un plan de trabajo aplicando una metodología adecuada.
- Trabajar a fondo los aspectos formales del desarrollo de proyectos.
- Sintetizar una solución viable y realista el problema propuesto.
- Elaborar una memoria del proyecto según una estructura prefijada.
- Elaborar una presentación del desarrollo y de resultados finales del proyecto.

1.2.2 Objetivos propios del TFC.

El objetivo principal del proyecto es adquirir experiencia en el diseño, construcción y explotación de un almacén de datos a partir de la información disponible en una base de datos transaccional.

Dicho objetivo se alcanza a través de los siguientes puntos:

- Familiarización con la arquitectura general de un Data Warehouse (DW).
- Conocer los aspectos teóricos del diseño de un DW
- Ponerse en contacto con lenguajes de consultas de base de datos.
- Analizar y comparar las herramientas del mercado que trabajan con estos modelos y realizar los análisis dirigidos a usuarios finales del DW.

1.3 Enfoque y método seguido.

De manera general el enfoque y el método seguido para la realización de este trabajo se ajusta al del ciclo de vida clásico o ciclo de vida en cascada, de manera que en cada etapa se han ido obteniendo documentos que son la base

de partida para la siguiente etapa, con la particularidad de que al tratarse de un TFC, no hay fase de mantenimiento.

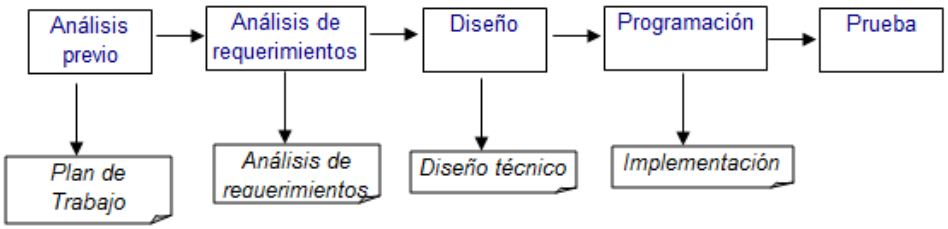

Fig. 1 : Etapas del ciclo de vida en cascada y documentos obtenidos

Ahora bien, al tratarse de un proyecto de almacén de datos las fases que se han ido cubriendo se encuentran mejor referenciadas en las fases típicas de las que consta un proyecto de este tipo, que se pueden descomponer en las siguientes (si bien existen algunas diferencias entre distintos autores) según la metodología descrita en www.1keydata.com y que de forma resumida presento a continuación:

Recopilación de requerimientos.

Se recogen los requerimientos del cliente y se elabora el documento de requisios así como una planificación temporal una aproximación al cubo OLAP que sirve de base a los informes.

Instalación del entorno físico.

Instalación y configuración de servidores y bases de datos.
En general de las herramientas informáticas necesarias.

Generación del modelo de datos.

Del documento de requerimientos se realiza un modelización de datos a tres niveles: conceptual, lógico y físico. Los dos primeros se pueden unificar en un único proceso en el que se identifican las entidades, sus claves primarias, atributos y relaciones entre ellas. La modelización física se convierten las entidades en tablas , los atributos en columnas y las relaciones en claves ajenas.

ETL.

Es la fase de extracción, transformación y carga de los datos proporcionados, para su puesta a punto y carga de los mismos en la base de datos.

Diseño del cubo OLAP.

De forma más precisa se define el cubo OLAP a partir del conocimiento de los requerimientos, donde se identifican los elementos necesarios para realizar las consultas que cumplan los requerimientos del cliente.

Desarrollo de las herramientas de consulta para el usuario.

En nuestro caso son herramientas comerciales ya diseñadas, pero que en cualquier caso han de ser amigables y adaptables a la empresa cliente. Suelen utilizarse herramientas que generan informes a través de la red , de manera que no sea necesario instalarlas en cada ordenador del cliente. En este trabajo se utiliza Oracle Discoverer.

Ajuste de rendimiento.
Se comprueban y ajustan los tiempos de los procesos tanto ETL como del procesamiento de consultas y generación de informes para que sean óptimos. Esto se puede conseguir con un diseño de carga adecuado y la utilización de índices, entre otros.

Garantía de calidad.
Las realiza el cliente, para comprobar el buen funcionamiento del almacén de datos.

Entrada en funcionamiento del producto.
Suele ir asociada a un curso de preparación del cliente.

Mantenimiento del producto.
Incluye tareas de mantenimiento posteriores a la puesta en producción.

Mejoras incrementales.
Contempla posibles ampliaciones no recogidas en los requerimientos iniciales.

Claramente algunas de las fases no tienen cabida en este TFC, como el mantenimiento del producto o las mejoras incrementales, así como la de garantía de calidad, para la que se necesitaría de un contacto con un usuario real que no será posible.

1.4 Planificación del proyecto

A continuación se hace un estudio de cada una de las tareas que se han de llevar a cabo y de los tiempos de los que se dispondrá para hacerlo.

Para llevar a cabo la planificación es necesario conocer las tareas a realizar, el tiempo que llevará realizarlas y adaptarlas a las fechas de los envíos que están programados según el plan docente de la asignatura y que se describen a continuación:

1.4.1 Planificación según plan docente.

INICIO DEL PROYECTO:

Descarga y lectura del material del TFC.

PAC1 : Análisis preliminar de requerimientos y Plan de Trabajo:

Elaboración del Plan de Trabajo, donde se indicará la planificación estimada de las diferentes tareas a realizar así como Análisis Preliminar de Requisitos (no detallado) con la enumeración y breve descripción de los elementos de análisis identificados (dimensiones, atributos, indicadores, etc.) disponibles para los usuarios, así como el número de informes aproximado a implementar y sus

contenidos. También se realizará un análisis de las fuentes de datos operacionales proporcionadas.

Esta fase coincide con la fase de Análisis de Requerimientos de un proceso de implantación de un DW.

Debido a la particularidad de este TFC, la instalación del soporte físico se ha llevado a cabo a lo largo de diferentes etapas. En este caso particular, se instaló aplicaciones como OpenProj, MS Excel y Word.

PAC2: Análisis de requerimientos y diseño conceptual y técnico:

Elaboración de un documento con el análisis detallado de los requerimientos basado en el Análisis Preliminar realizado. También se enviará un documento de diseño con la descripción del modelo dimensional que dará respuesta a las necesidades de los usuarios y el diseño de los procedimientos de extracción de datos a alto nivel (procesos, pseudocódigo, etc.)

PAC3: Implementación

Esta fase constará de las siguientes tareas:

- Construcción del almacén de datos: base de datos, cargas , etc.
- Instalación de la herramienta de explotación de los datos.
- Construcción de los informes y análisis de la información.

Las herramientas necesarias ya se encuentran instaladas en la máquina virtual aportada por el cliente.

MEMORIA y PRESENTACIÓN VIRTUAL.

Entrega de la memoria final del proyecto.

Elaboración de la documentación de la memoria final así como de una presentación virtual del proyecto.

y finalmente DEBATE.

1.4.2 Hitos y fechas clave.

Las fechas de entrega de cada fase serán las siguientes:

FASE	TAREA	INICIO	FIN
PAC1	Análisis preliminar de requerimientos y Plan de Trabajo	22/9/2011	05/10/2011
PAC2	Análisis de requerimientos y diseño conceptual y técnico	06/10/2011	09/11/2011
PAC3	Implementación	10/11/2011	21/12/2011
MEMORIA	Entrega de la memoria final del proyecto	22/12/2011	10/01/2012
PRESENTACION	Presentación virtual del proyecto de unos 20 minutos	22/12/2011	10/01/2012
DEBATE		23/01/2012	26/01/2012

Fig 2 : Fechas de inicio y de entrega de las tareas planteadas.

1.4.3 Planificación detallada de tareas.

Las tareas descritas anteriormente se desglosan y muestran en la siguiente tabla, donde se indican sus duraciones, fechas de inicio, fin y precedencia. Esta planificación está basada en una dedicación aproximada de 2 horas/día.

Id	Nombre	Duración	Inicio	Fin	Precedencia
1	**INICIO DEL PROYECTO**	1	21-sep-11	21-sep-11	
2	Descarga y lectura material TFC	1	21-sep-11	21-sep-11	
3	**AVANCE BORRADOR PAC1**	**8**	**22-sep-11**	**29-sep-11**	
4	Descarga y lectura enunciado PAC1	1	22-sep-11	22-sep-11	2
5	Estudio material docente	2	23-sep-11	24-sep-11	4
6	Estudio de otras fuentes de información	1	25-sep-11	25-sep-11	5
7	Resolución dudas enunciado	1	26-sep-11	26-sep-11	6
8	Instalación : OpenProj , Excel y Word	1	27-sep-11	27-sep-11	7
9	Análisis de datos fuente del cliente	1	27-sep-11	27-sep-11	7
10	Redacción del borrador PAC1	2	28-sep-11	29-sep-11	9
11	Envío borrador PAC1	0	29-sep-11	29-sep-11	10
12	**REDACCIÓN DEFINITIVA PAC1**	**6**	**30-sep-11**	**5-oct-11**	
13	Revisión correcciones del borrador	1	30-sep-11	30-sep-11	11
14	Redacción del Análisis Preliminar	2	1-oct-11	2-oct-11	13
15	Redacción del Plan de Trabajo	3	3-oct-11	5-oct-11	14
16	Envío PAC1	0	5-oct-11	5-oct-11	15
17	**AVANCE BORRADOR PAC2**	**28**	**6-oct-11**	**2-nov-11**	
18	Descarga y lectura enunciado PAC2	1	6-oct-11	6-oct-11	16
19	Instalación herramientas Case : ArgoUML	1	6-oct-11	6-oct-11	16
20	Redacción Análisis Requerimientos (detalle)	2	7-oct-11	8-oct-11	19
21	Análisis datos fuente (depuración)	3	9-oct-11	11-oct-11	20
22	Elaboración del modelo conceptual	5	12-oct-11	16-oct-11	21
23	Elaboración del modelo lógico	5	12-oct-11	16-oct-11	21
24	Descarga y lectura solución PAC1	1	17-oct-11	17-oct-11	16
25	Revisión Análisis Requerimientos (detalle)	1	18-oct-11	18-oct-11	23
26	Revisión modelo conceptual	7	19-oct-11	25-oct-11	25
27	Revisión modelo lógico	7	19-oct-11	25-oct-11	25
28	Pseudocódigo procs extracción datos	4	26-oct-11	29-oct-11	27
29	Redacción borrador PAC2	4	30-oct-11	2-nov-11	28
30	Envio borrador PAC2	0	2-nov-11	2-nov-11	29
31	**REDACCIÓN DEFINITIVA PAC2**	**7**	**3-nov-11**	**9-nov-11**	
32	Revisicón correcciones del borrador	4	3-nov-11	6-nov-11	30

33	Redacción definitiva PAC2	3	7-nov-11	9-nov-11	32
34	Envío PAC2	0	9-nov-11	9-nov-11	33
35	**AVANCE BORRADOR PAC3**	**35**	**10-nov-11**	**14-dic-11**	
36	Descarga y lectura del enunciado PAC3	1	10-nov-11	10-nov-11	34
37	Construcción almacén de datos: bd,descargas,etc	9	11-nov-11	19-nov-11	36
38	Instalación aplicación explotación de datos	1	20-nov-11	20-nov-11	37
39	Descarga y lectura solución PAC2	1	21-nov-11	21-nov-11	34
40	Revisión bd, procesos carga, etc	2	22-nov-11	23-nov-11	38
41	Diseño de informes solicitados	9	24-nov-11	2-dic-11	40
42	Pruebas de informes solicitados	4	3-dic-11	6-dic-11	41
43	Análisis de información solicitada	5	7-dic-11	11-dic-11	42
44	Redacción borrador PAC3	3	12-dic-11	14-dic-11	43
45	Envió borrador PAC3	0	14-dic-11	14-dic-11	44
46	**REDACCIÓN DEFINITIVA PAC3**	**7**	**15-dic-11**	**21-dic-11**	
47	Revisicón correcciones del borrador	4	15-dic-11	18-dic-11	45
48	Redacción definitiva PAC3	3	19-dic-11	21-dic-11	47
49	Envío PAC3	0	21-dic-11	21-dic-11	48
50	**REDACCIÓN MEMORIA**	**20**	**22-dic-11**	**10-ene-12**	
51	Redacción de la memoria	20	22-dic-11	10-ene-12	49
52	Redacción de la presentación	20	22-dic-11	10-ene-12	49
53	Descarga y lectura solución PAC3	1	2-ene-12	2-ene-12	49
54	Envío memoria y presentación	0	10-ene-12	10-ene-12	52
55	**DEBATE**	0	23-ene-12	26-ene-12	54

1.4.4 Diagrama de Gantt.

El diagrama de Gantt donde se muestran los tiempos de dedicación previstos para las distintas tareas del proyecto se muestra a continuación:

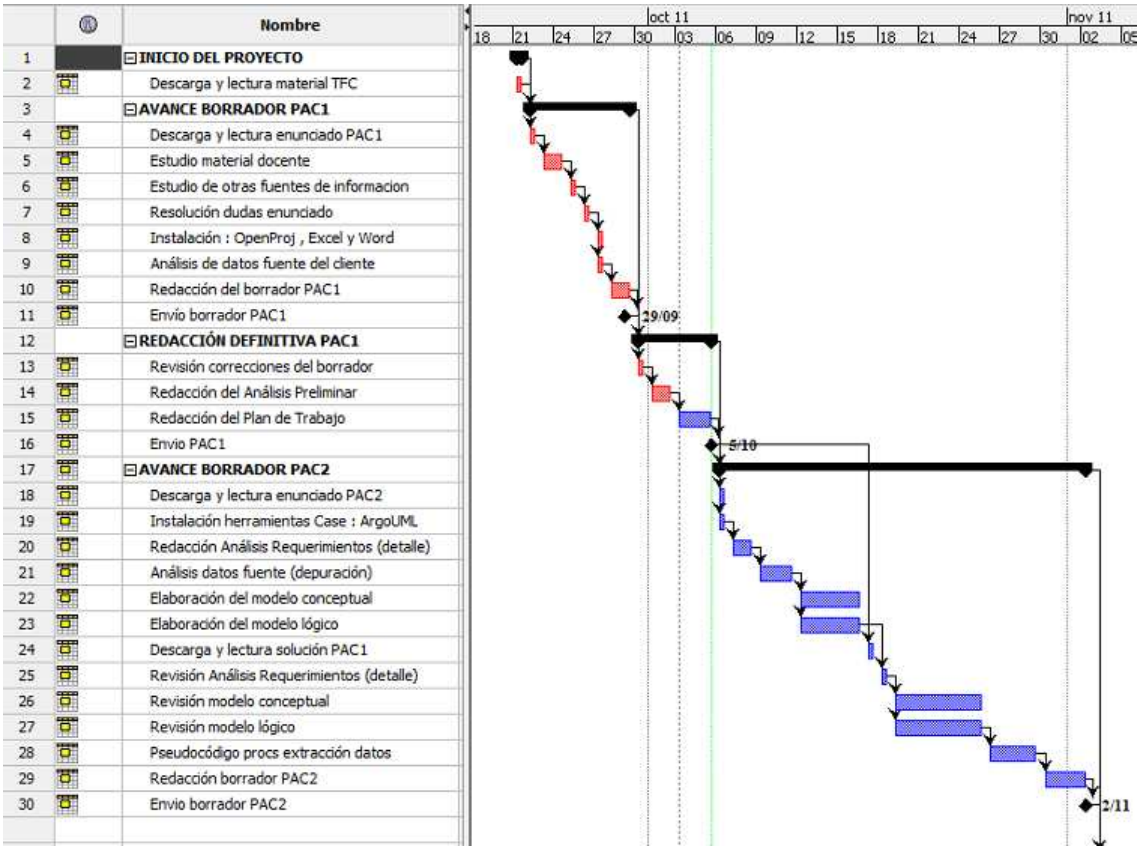

Fig. 3 : Diagrama Gantt (parte 1 de 2)

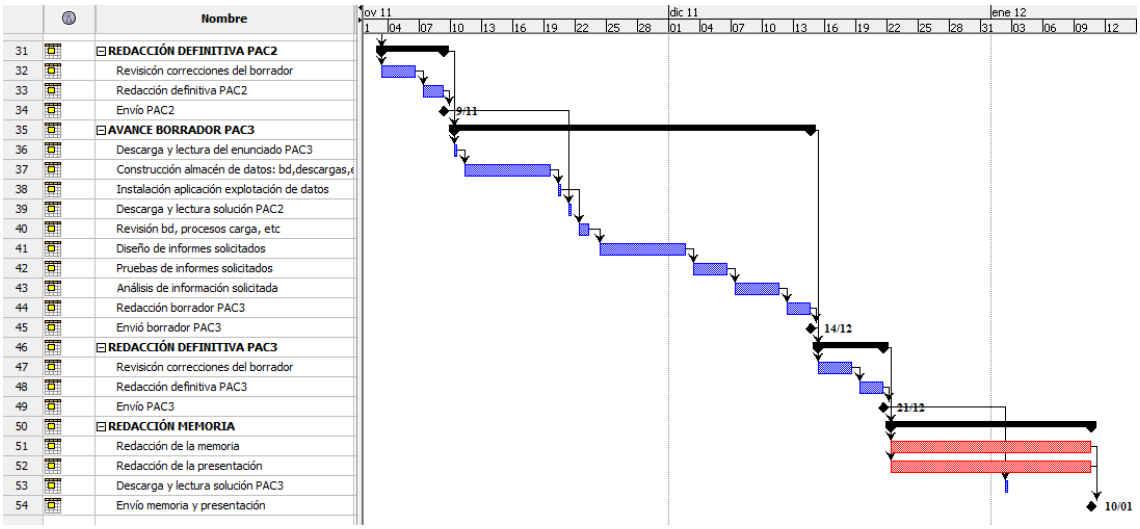

Fig. 4 : Diagrama Gantt (parte 2 de 2)

1.4.5 Incidencias y riesgos. Plan de contingencia

A continuación se determinan las posibles incidencias y riesgos que pueden surgir a lo largo del desarrollo del proyecto así como del correspondiente plan de contigencias.

- Desplazamientos por motivos laborales

Actualmente me encuentro trabajando, y aunque no es lo habitual en ocasiones he de llevar a cabo visitas a clientes lo que supone entre otras cosas que la jornada de trabajo en ocasiones se alargue, y tiene como consecuencia una reducción del tiempo dedicado al proyecto.

En estos casos se trataría de compensar el tiempo no dedicado al proyecto, con la lectura de documentación relacionada al mismo durante los desplazamientos.

- Avería del punto de trabajo

Aunque se trata de imprevistos poco probables, su efecto daría lugar a retrasos significativos, pues podría perderse toda la información y desarrollo del mismo.

Para minimizar este efecto , se llevará a cabo copias de seguridad diarias de los avances que se vayan realizando.

- Compromisos

Actualmente no hay ninguno programado. En caso de que surgiese alguno, se trataría de recuperar el posible retraso que se haya podido producir aumentando (durante un período de tiempo) las horas diarias dedicadas.

- Motivos de salud

Se trata de una contingencia impredecible que tendrá en mayor o menor medida un efecto sobre la planificación propuesta. Para estar prevenidos frente a estas indisposiciones se trataría de llevar la planificación al día.

- Retraso en publicaciones de enunciados y soluciones PAC y en el envío de materiales.

Por tratarse de causas ajenas que provocarían el retraso en ciertas tareas planificadas, se solicitaría al tutor el mismo retraso en las entregas correspondientes.

En cualquier caso, los retrasos que se pudieran producir tratarían de solventar con dedicación máxima, para lo que sería posible contar con días libres en el trabajo.

1.4.6 Valoración y coste del proyecto.

Desde el 21 de Septiembre hasta el 10 de Enero , que damos como fechas para la realización del proyecto, transcurren 112 días. Puesto que la planificación del mismo se ha hecho teniendo en cuenta una dedicación diaria, incluyendo festivos , de dos horas , el número total estimado de horas es de 112*2= 224 horas.

Puesto que el salario de un analista de BD es de unos 50 € / hora, el coste será de 11.200 €. + 18% IVA = 13.216 €.

1.5 Productos obtenidos

A lo largo del proyecto se han elaborado los siguientes documentos y productos.

En primer lugar se describen los documentos obtenidos:

- **Plan de Trabajo**: Descripción general del proyecto y sus objetivos así como la planificación temporal de las tareas , los recursos hardware y software a utilizar y las posibles incidencias y sus soluciones previstas cuando pudieran surgir.

- **Análisis preliminar de requerimientos**: Descripción y análisis preliminar de los datos proporcionados por el cliente así como la identificación de los elementos para el análisis multidimensional y de los informes solicitados.

- **Análisis de requerimientos**: Se pasan a definitivos los análisis realizados en el documento anterior (tras su revisión en su caso).

- **Diseño técnico**: Contiene el análisis conceptual, lógico y físico del modelo multidimensional para el almacén de datos así como la descripción a alto nivel de los procesos ETL y la descripción de la arquitectura del sistema.

- **Implementación**: Construcción de la base de datos y estructuras para dar soporte a los procesos ETL y almacén de datos, así como la creación del área de negocio y de los informes solicitados. Se describe con detalle la creación de usuarios.

- **Memoria final**: Documento actual que recoge la descripción detallada del proyecto, sus objetivos, planificación, metodología y resultados obtenidos así como las conclusiones.

- **Presentación**: Presentación multimedia del proyecto para dar una visión ejecutiva de la tarea desarrollada.

Los siguientes productos no son documentos:

- **01_ObjetosCargayAlmacen.sql** y **00_Elimina_ObjetosCargayAlmacen.sql** son ficheros PL/SQL de creación de la base de datos (tablas, indices, secuencias, triggers, procedimientos almacenados, etc.) y su correspondiente de eliminación.

- **CargaDatos.ctl** : fichero de control de carga de los datos proporcionados por la empresa cliente.

- **CargaDatos.bat** : fichero batch que ejecuta los comandos de carga de los datos proporcionados por la empresa cliente y de transformación de los mismos junto con su correspondiente carga en las tablas del almacén de datos.

- **CargaLog.log** y **CargaBad.bad** , ficheros de log del proceso de carga y fichero de registros no cargados.

- **Informes cliente.DIS**, se trata del libro de trabajo (en local) de los informes desarrollados, aunque también exite una versión cargada en la propia base de datos de dichos informes.

- Finalmente otro producto que se ha obtenido de manera indirecta es el fichero comprimido **jgomezdie_mv.zip** correspondiente al disco duro de la máquina virtual proporcionada por la empresa cliente.

1.6 Breve descripción de los restantes capítulos de la memoria.

1.6.1 Capítulo 2 - Análisis

A partir del análisi de los datos proporcionados por la empresa cliente así como de los requisitos que plantea, se lleva a cabo el diseño conceptual del modelo multidimensional, donde se establecen el hecho y su granularidad, dimensiones , atributos, descriptores, jerarquías y medida adecuadas a la resolución según los requisitos planteados.
Como siguiente paso, se obtiene el modelo lógico y por útlimo el modelo físico que se verá en el siguiente capítulo.

1.6.2 Capítulo 3 - Diseño

Se lleva a cabo la implementación del modelo multidimensional diseñado, para lo que hay que crear las estructuras que lo soporten en la base de datos como tablas, índices, triggers, procedimientos almacendados, funciones etc.

Además también se crea el usuario que gestionará el área de negocio que será creada y definida para la obtención de los informes así como los informes requeridos por la empresa cliente.

2. Capítulo 2 – Análisis

A partir de los requisitos planteados por la empresa cliente y de los datos que proporciona se lleva a cabo el estudio de las diferentes interacciones con el sistema así como del modelo conceptual que del modelo multidimensional para el almacén de datos.

El análsis de los datos proporcionados también puede ser consultado en el anexo B "Detalle y automatización del proceso ETL" en el apartado 4.3.2 Descripción de los elementos utilizados y en el subapartado SetUpDatosyUmbrales.

2.1 Diagrama de casos de uso.

Se trata de un diagrama que representa la interacción del sistema a desarrollar con el exterior, al que modeliza a través de entidades conocidas como actores que no son más que el modo o papel en que una entidad externa (como un individuo o bien otro sistema externo independiente) llevan a cabo una acción que da lugar a la obtención de datos finales o intermedios que alimentarán otras acciones o bien que dan lugar a otras operaciones.

Para este trabajo se han planteado los siguientes actores :

- Analista

Es el actor encargado de llevar a cabo el análisis de la información contenida en el almacén de datos a través de las herramientas de explotación seleccionadas ,Oracle Discoverer Desktop para este proyecto.
Su labor es únicamente consultiva.

Dicho análisis lo puede llevar a cabo tras actualizar los informes mediante una nueva consulta, que incorporará al informe los últimos datos actualizados.

Además puede solicitar la creación de nuevos informes al administrador del área de negocio.

- Administrador de las áreas de negocio

Se encarga de la gestión de las áreas de negocio existentes, definiendo los elementos de los informes, cálculos e informes así como los usuarios que podrán tener acceso a cada una de las áreas de negocio.

- Usuario/esquema Base de datos

Es el usuario propietario del esquema sobre el que se ha desarrollado el almacén de datos en la base de datos. Llevará a cabo la gestión de

usuarios de la base de datos y de sus privilegios, los cambios que se pudieran llevar a cabo en el esquema original y de realizar la carga de los datos.

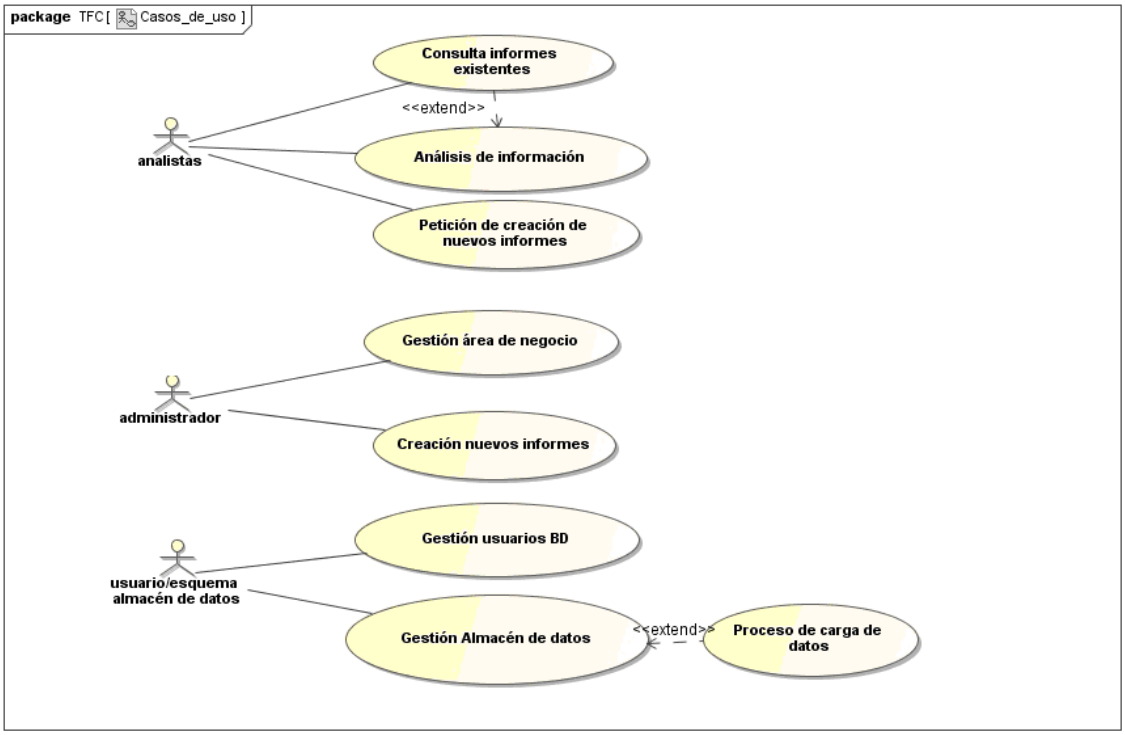

Fig. 5 : Diagrama de casos de uso.

2.2 Diagrama del modelo conceptual

En el diseño del modelo conceptual se desarrollarán con detalle, a través de una metodología las dimensiones, atributos, medidas y hechos de nuestro modelo así como el estudio de la viabilidad de la implementación.
Dicha metodología se define en una serie de nueve pasos que permitirá obtener el modelo en estrella en que se basará el diseño.

2.2.1 Identificar el hecho

Se trata de identificar el hecho que constituye el núcleo de la estrella que se pretende diseñar. Para llevarlo a cabo se analizan los procesos de negocio a tratar e intentamos buscar uno de ellos que en sí mismo sea los suficiente significativo como para definir la actividad.

Para este proyecto en particular y a partir de los datos de los que se dispone que reflejan la venta/oferta de inmuebles en diferentes zonas y momentos y según sus características (superficie, tipo, dormitorios, planta), el hecho que se puede identificar es la oferta de un inmueble en un lugar y fecha.

2.2.2 Encontrar la granularidad adecuada

El gránulo es el elemento último que se quiere analizar , y por tanto fijar el grado de detalle de las celdas que compondrán nuestro cubo es una tarea muy importante, ya que tiene un impacto directo en el tamaño de la base de datos , así como en la eficiencia de los procedimientos de extracción de la información de análisis.

Por tanto un exceso de detalle, (una granularidad pequeña) provocaría que el sistema tuviese que manejar una cantidad excesiva de datos a procesar lo que ralentizaría las consultas y dispararía el tamaño de la base de datos y un defecto de detalle (una granularidad grande) provocaría la pérdida de información y comprometería el objetivo final del proyecto.

Es pues necesario encontrar un equilibrio, y escoger la granularidad más pequeña posible

Para este caso, y a partir de los datos suministrados que representan ofertas de inmuebles por zona y tiempo, parece ser adecuada al propósito general del proyecto.

2.2.3 Escoger las dimensiones a emplear en el análisis

A partir de un análisis del hecho que se ha escogido se ponen de relevancia ciertos datos que sitúan al hecho dentro de un contexto, desde el que se puede observar desde distintos puntos de vista.

De forma inmediata se observa que las dimensiones son la zona geográfica donde se ubica el inmueble así como el momento en que se produce el hecho, por tanto, Zona y Tiempo son las dimensiones primeras que se obtienen.

Tras este primer análisis hay que preguntarse si existe algún otro candidato a tener en cuenta, algún otro punto de vista, que defina y delimite completamente las instancias del hecho oferta que se trata.

En este caso las características que definen al inmueble en sí, pueden agruparse en la dimensión Inmueble.

2.2.4 Establecer los atributos de cada dimensión

Se trata de escoger cualquier atributo que forme parte de las dimensiones que pueda ser útil para seleccionar, agrupar o simplemente para poner de cabecera de un informe que sean relevantes para nuestro modelo.

Para la dimensión “Tiempo” puesto que los datos de esta dimensión corresponden a períodos mensuales expresados como mesAño, parece adecuado un atributo fecha expresado como mes o año según el nivel de agregación que se escoja.

Como para esta dimensión, pese a ser una dimensión de un dominio de valores contínuo , sólo se aportan valores discretos , no es necesario clasificar/discretizar dichos datos.

Para la dimensión “Zona”, a partir de los datos aportados, se observa información de 4 ó 5 niveles como se indicó en el Análisis de requerimientos, sin embargo una inspección más exhaustiva de los datos, detalla hasta 5 niveles más el país. Cada zona viene identificada por un código textual, aunque también se puede obtener a partir de una jerarquía de cadenas de texto que representan los distintos niveles de agregación: país, provincia, comarca, población, distrito y barrio.

Para la dimensión “Inmueble” los atributos son código identificador (alfanumérico), metros cuadrados y dormitorios (numéricos) y la planta (alfanumérico).

2.2.5 Distinguir entre descriptores y jerarquías de agregación

De entre los atributos de una dimensión se pueden distinguir dos tipos de uso que se pueden dar:

- Utilizarlos para agrupar
- Utilizarlos para seleccionar

Una opción u otra depende mucho más de lo que los usuarios quieren que de los datos en sí mismos.

En este punto, por tanto , se establecen cuáles atributos actuarán como descriptores que permitirán hacer selecciones basadas en sus valores y cuáles otros permitirán hacer agrupaciones y conformar jerarquías. Estas jerarquías hay que consensuarlas con el usuario.

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

Para la dimensión “Zona”, también se puede establecer una jerarquía de agregación según niveles, así ésta comprendería desde el nivel 5 que es el barrio al nivel 5 de provincia/comunidad autónoma. También habría que añadir el nivel de país.

Para la dimensión “Inmueble”, se puede establecer una jerarquía de agregación según la tipología del inmueble.

2.2.6 Decidir cuáles son las medidas de interés

Las medidas son los atributos de una celda. La mayoría de las medidas son de tipo numérico, aunque algunas son booleanas y hasta textuales.

Para este proyecto se ha considerado el precio, ya que el hecho es la oferta y por tanto viene descrita por dicho valor.

Además, según los requerimientos del cliente, se han de poder realizar consultas que distingan entre los inmuebles que están en venta de los que no lo están.

Se podría pensar que dicha característica del inmueble deba de aparecer como atributo del mismo, sin embargo hay razones que llevan a considerarlo como una medida de la oferta, ya que por un lado los datos aportados los inmuebles vendidos aparecen sin precio en dicha fecha y por otro se trata de un dato variable con el tiempo es decir variará con cada instancia del hecho.

También se requiere que se puedan realizar estudios sobre aquellos pisos que han bajado de precio, y sobre el tiempo que llevan en venta.

Respecto al primer requerimiento (ya que no se indica en los requisitos, definición alguna sobre como considerar una bajada de precio) sobre las distintas alternativas que puede haber : bajada respecto de un mes al siguiente, respecto entre dos meses cualesquiera o entre el precio inicial y el actual , se opta por considerar como bajada de precio la diferencia entre el precio inicial y el actual , para ello he incluido el atributo precio inicial en el hecho y la diferencia entre el precio en un mes y el anterior (cuando éstos sean diferentes) para lo que incluyo el atributo revaloración.

En el segundo requerimiento se ha determinado incluir un atributo en el hecho que se corresponda con los meses que una oferta lleva en venta.

2.2.7 Definir las celdas

De las celdas resultantes de este análisis habrá que determinar cuáles son relevantes para el modelo y por tanto ser almacenadas, y además habrá que tener en cuenta las distintas granularidades que dichas celdas puedan tener y determinar también cuáles de ellas se pueden obtener derivadas de otras.

Para este proyecto se obtiene una única celda que refleja el hecho analizado, y se trata de la oferta del inmueble.

2.2.8 Explicitar las restricciones de integridad

Como último paso del modelo conceptual habrá que expresar las restricciones de integridad correspondientes a las bases, restricciones de agregación y de transitividad.

La base indicará qué dimensiones definen unívocamente una celda.

Para las dimensiones de Zona, Tiempo e Inmueble, hay que decir que puesto que un inmueble siempre estará en la misma zona, será suficiente el Tiempo e Inmueble para definir de forma única a una celda.

También para lograr que la agregación sea completa , se ha llevado a cabo en la dimensión Zona las modificaciones necesarias para que no aparezcan instancias de niveles por debajo de otro nivel, vacías (como por ejemplo para los inmuebles de Andorra que presentan los niveles por debajo de comarca vacíos).

2.2.9 Estudiar la viabilidad

Una vez terminado el diseño conceptual, hay que comprobar que la estrella resultante es realmente implementable o no, y por eso se ha de estimar el espacio que ocuparán todos los datos una vez almacenados en el sistema.

Para llevarlo a cabo se realiza una estimación a partir de las instancias correspondientes al hecho.

El número de registros en la tabla de hecho es aproximadamente de 92.555 ,el número de meses es de 25 , por tanto el número de registros en esta primera carga será de : 92.555 x 25 = 2.313.875.

En la tabla del hecho , se cuenta con 3 identificadores de las dimensiones que ocuparán 4 bytes cada uno, más 1 atributo de dos caracteres (2 bytes) , más 4 atributos numéricos con 4 bytes cada uno. En total (12 + 2 + 16) 30 bytes por registro, así que el total será 69.416.250 bytes ó 69,42 Megabytes, cantidad que se puede manejar con facilidad por cualquier SGBD.

El crecimiento mensual del tamaño estimado será de:

- 2 Megabytes , que son lo que ocuparán las ofertas de los 92.555 inmuebles al mes siguiente.

- Si se supone un crecimiento de un 1% mensual del stock de pisos ofertados (925 nuevos pisos) se obtiene (925*30) 27.750 bytes ó 27 Kb

Finalmente el crecimiento mensual será de aproximadamente (2,02 / 69,42)*100 ~ 3%

Tras estas consideraciones el modelo conceptual resultante es el siguiente:

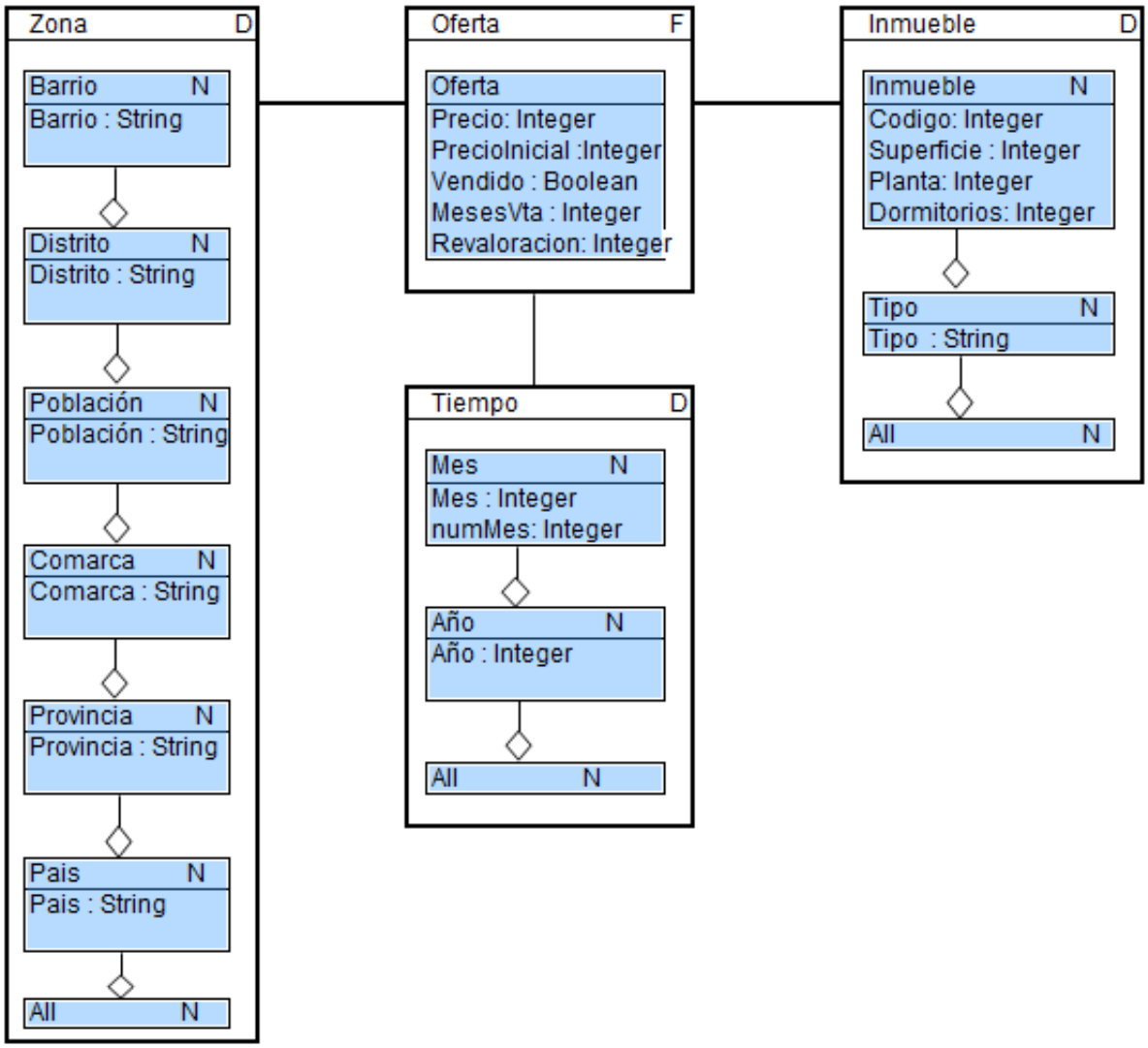

Fig. 6 : Diagrama del modelo conceptual.

3. Capítulo 3 - Diseño

3.1 Diagrama de la arquitectura software

El sistema se compone de los siguientes elementos software:

- Hojas de cálculo , Microsoft Excel: Mediante sus funcionalidades se puede tranformar la información a ficheros .txt ó .csv (y viceversa) listos para la carga.

- SqlLoader, como aplicación para la carga de datos en la base de datos, y cuya arquitectura se muestra en la figura siguiente.

- Un SGBD Oracle 10g XE, sobre el que se implementará el DataMart, y que se gestiona mediante un navegador web, y además se ha utilizado la herramienta Oracle SQL Developer para facilitar la programación.

- Mediante Oracle Discoverer Administrator, el administrador creará, mantendrá y administrará datos de la capa de usuario final y definirá cómo los usuarios interactúan con los datos.

- Mediante Oracle Discoverer Desktop, los usuarios finales ó analisitas ejecutarán las consultas ad hoc y generarán informes.

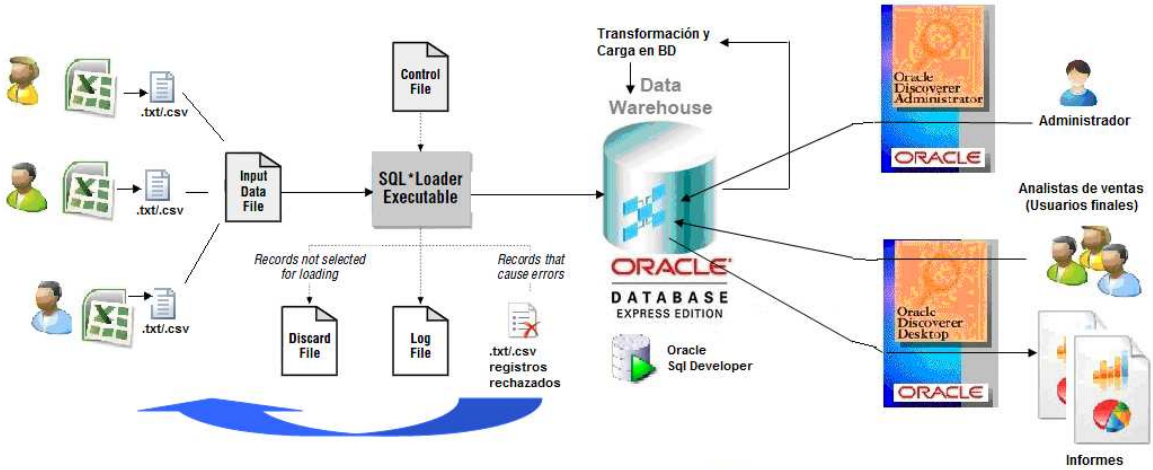

Fig. 7 : Arquitectura software.

3.2 Diagrama de la arquitectura hardware

El sistema se compone de los siguientes elementos hardware:

- Los sistemas operacionales de la empresa cliente desde los que se genera los datos, y que nutrirán al el almacén de datos.
- Los sistemas cliente que accederán al servidor para realizar las consultas.

- La estación de administración desde donde se lanzan los procesos de carga y mantenimiento del almacén de datos.

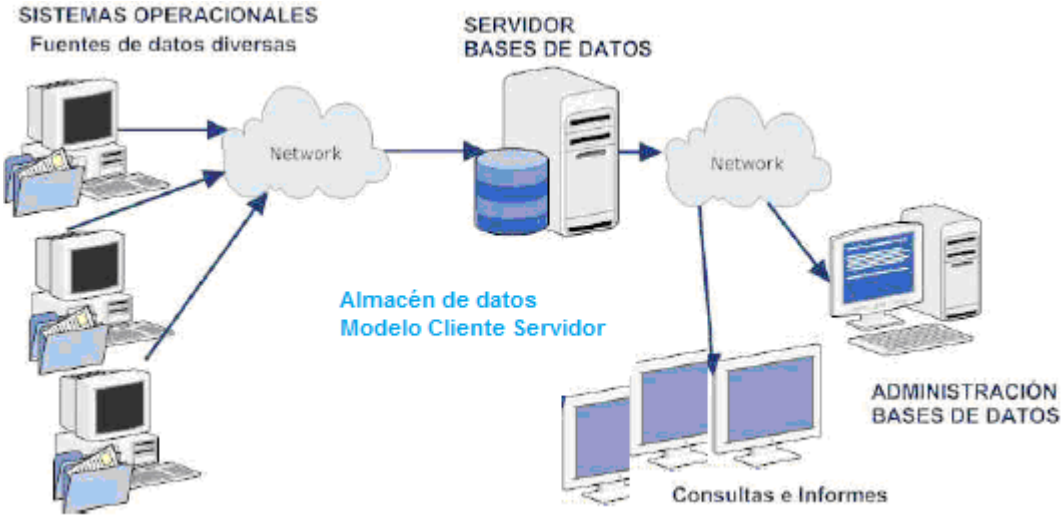

Fig. 8 : Arquitectura hardware.

3.3 Diseño de la base de datos y diagrama del modelo físico

3.3.1 Diseño lógico

Una vez obtenido el modelo en estrella como resultado del análisis del modelo conceptual, el siguiente paso es obtener el modelo lógico, que es el resultado de la implementación de los hechos y dimensiones en tablas que para este proyecto son relacionales.

Para implementar la estrella se necesita una tabla para el hecho (en el que cada fila representa una celda) y una tabla más por cada dimensión, donde las jerarquías de agregación quedarán implícitas por los valores de los atributos de las tablas de dimensión, no siendo necesario establecerlas con tablas diferentes.

Las tabla de hechos estará relacionada con las de dimensión mediante claves foráneas. Dichas claves foráneas se han generado por el SGBD, ya que la utilización de dichas claves para crear las relaciones es más ágil que utilizar valores descriptivos , correspondientes a los valores de los atributos de las tablas de dimensiones que actuasen como claves primarias.

Así pues, el modelo lógico obtenido es el siguiente:

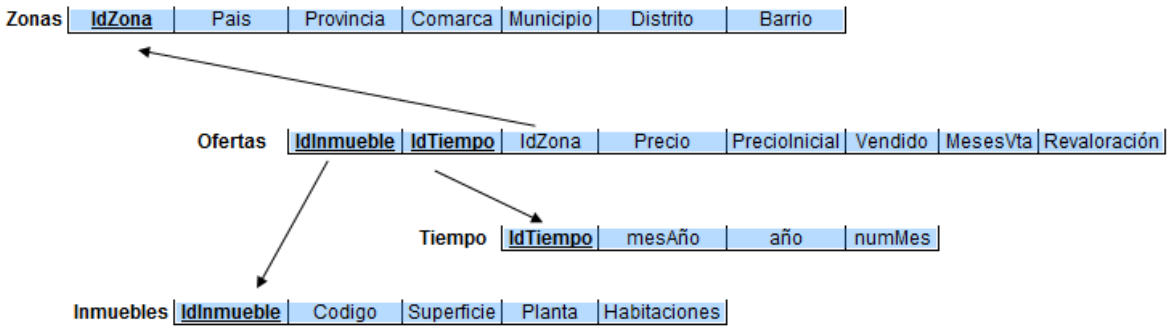

Fig. 9 : Diagrama del modelo lógico.

3.3.2 Diseño físico

A partir de las relaciones y tablas que compondrán el almacén de datos solamente queda hacer el diseño físico teniendo en cuenta obtener buenos tiempos de respuesta, para los que inicialmente se cuenta con las claves primarias y ajenas.

Si los tiempos de respuesta no fueran aceptables , en la fase de ajuste de rendimiento se pueden estudiar índices que mejoren las consultas requeridas.

- Inmuebles : Contiene las características atemporales de los inmuebles y su clave primaria es IDINMUEBLE.

Column Name	Data Type	Nullable	Data Default	COLUMN ID	Primary Key
IDINMUEBLE	NUMBER(15,0)	No		1	1
CODIGO	VARCHAR2(30 Bytes)	No		2	
SUPERFICIE	NUMBER(5,0)	Yes		3	
PLANTA	VARCHAR2(15 Bytes)	Yes		4	
DORMITORIOS	NUMBER(2,0)	Yes		5	
TIPO	VARCHAR2(30 Bytes)	Yes		6	

Fig. 10 : Esquema físico para la tabla inmuebles.

- Ofertas : La clave está constituida por IDINMUEBLE Y IDTIEMPO.

Column Name	Data Type	Nullable	Data Default	COLUMN ID	Primary Key
IDINMUEBLE	NUMBER(15,0)	No		1	1
IDTIEMPO	NUMBER(5,0)	No		2	2
IDZONA	NUMBER(5,0)	No		3	
PRECIO	NUMBER(15,0)	Yes		4	
PRECIOINICIAL	NUMBER(15,0)	Yes		5	
VENDIDO	VARCHAR2(2 Bytes)	Yes		6	
MESESVTA	NUMBER(3,0)	Yes		7	
REVALORACION	NUMBER(15,0)	Yes		8	

Fig. 11 : Esquema físico para la tabla ofertas.

- Zonas : Su clave primaria es IDZONA.

Column Name	Data Type	Nullable	Data Default	COLUMN ID	Primary Key
IDZONA	NUMBER(5,0)	No		1	1
PAIS	VARCHAR2(100 Bytes)	Yes		2	
PROVINCIA	VARCHAR2(100 Bytes)	Yes		3	
COMARCA	VARCHAR2(100 Bytes)	Yes		4	
MUNICIPIO	VARCHAR2(100 Bytes)	Yes		5	
DISTRITO	VARCHAR2(100 Bytes)	Yes		6	
BARRIO	VARCHAR2(100 Bytes)	Yes		7	

Fig. 12 : Esquema físico para la tabla Zonas.

- Tiempo : Su clave primaria es IDTIEMPO y se incluye un campo informativo que es nummes que ayudará a la ordenación de los meses.

Column Name	Data Type	Nullable	Data Default	COLUMN ID	Primary Key
IDTIEMPO	NUMBER(5,0)	No		1	1
ANYO	NUMBER	Yes		2	
MES	VARCHAR2(20 Bytes)	Yes		3	
NUMMES	NUMBER(2,0)	Yes		4	

Fig. 13 : Esquema físico para la tabla Tiempo.

En el anexo A “Acceso a la base de datos” se detalla cómo acceder al esquema y se muestran las estructuras creadas.

3.3.3 Proceso ETL

El proceso ETL, puede considerarse como una extensión previa del diseño de la base de datos. Sin dicho proceso no sería posible realizar un tratamiento correcto de los datos aportados por el cliente en la mayoría de los casos.

Esto supone la creación de tablas auxiliares ó temporales para el volcado y procesado de datos mediante procedimientos almacenados.

El detalle del proceso ETL diseñado para este proyecto puede encontrarse en el anexo B “Detalle y automatización del proceso ETL”

3.4 Diseño y descripción de los informes creados

3.4.1 La creación del área de negocio

Con la herramienta Oracle BI Discoverer Administrator se ha creado el área de negocio AN_Ventas que contiene los datos, las jerarquías y las clases de elementos necesarios para la realización y ejecución de informes.

Previamente a la creación del área de negocio, se ha creado un usuario en la base de datos con los privilegios necesarios para que pueda llevar a cabo la creación de un EUL (End User Layer). Una vez creado el EUL y almacenada la información correspondiente ya es posible comenzar a trabajar creando el nuevo área de negocio a la que me he referido.

Para conocer con más detalle cómo crear el usuario ver el anexo C “Acceso a Oracle Discoverer Administration”.

En dicha área se han definido los cálculos y filtros necesarios para la elaboración de los informes así como dos clases de elementos, uno de ellos para la ordenación correcta de las cadenas de texto que representan los meses.

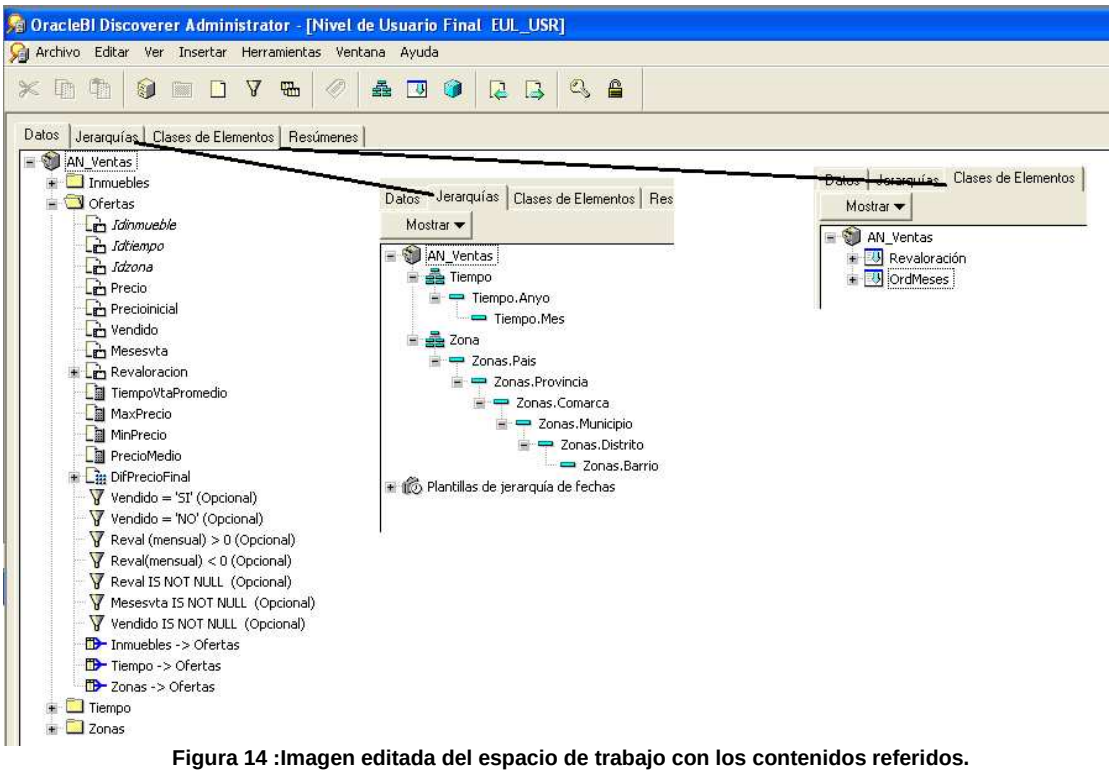

Figura 14 :Imagen editada del espacio de trabajo con los contenidos referidos.

3.4.2 La creación del libro de trabajo

Una vez creada el área de negocio con la herramienta Oracle BI Desktop se ha creado el libro de trabajo “Informes clientes.DIS” en la propia base de datos.
Para la creación de un libro de trabajo es necesario que exista un usuario al que el administrador del área de negocio otorgue los privilegios necesarios para ello.

Puede consultarse el anexo D “Acceso a Oracle Discoverer Desktop” para más detalle.

El libro de trabajo contiene 10 hojas cada una de las cuales contiene una tabla o una matriz que da respuesta a las consultas planteadas por la empresa cliente.

La información obtenida se puede mostrar a niveles generales para un gran volumen de datos o bien restringiendo las condiciones como puede ser para unos determinados tipos o características de inmuebles, o bien meses o años concretos, etc.

3.4.3 Descripción de los informes creados

El conjunto de informes solicitados por el cliente es el siguiente:

- Número de inmuebles por zona, tipología y características
- Listado de inmuebles que han bajado de precio
- Precios máximos, mínimos y medios por zona, tipología y características
- Metros cuadrados máximos, mínimos y medios por zona, tipología y características
- Tiempos medios de venta por zona, tipología y características
- Número de inmuebles vendidos y existentes por zona, tipología y características
- Modelo de inmueble español-andorrano

A continuación se pasa a describir el detalle de los informes junto con la captura de pantallas correspondientes.

3.4.3.1 Número de inmuebles por zona, tipología y características

El informe muestra la información por provincia, comarca, tipo y año.
Para ello se ha utilizado el filtro Vendido is not null, de manera que independientemente de si ha sido vendido o no, nos mostrará la información de los inmuebles.

Número de Inmuebles
por zona , tipología y características

Elementos de Página: Anyo: 2006

	Provincia	Comarca	Tipo	NumInmuebles
1	ALICANTE	ALT VINALOPÓ	chalets	11
2	ALICANTE	ALT VINALOPÓ	dúplex	3
3	ALICANTE	ALT VINALOPÓ	pisos	12
4	ALICANTE	BAIX SEGURA	áticos	106
5	ALICANTE	BAIX SEGURA	chalets	208
6	ALICANTE	BAIX SEGURA	dúplex	82
7	ALICANTE	BAIX SEGURA	estudios	53
8	ALICANTE	BAIX SEGURA	pisos	914
9	ALICANTE	BAIX VINALOPÓ	áticos	21
10	ALICANTE	BAIX VINALOPÓ	chalets	37
11	ALICANTE	BAIX VINALOPÓ	dúplex	36
12	ALICANTE	BAIX VINALOPÓ	estudios	2
13	ALICANTE	BAIX VINALOPÓ	pisos	269
14	ALICANTE	COMTAT	chalets	5
15	ALICANTE	COMTAT	dúplex	2
16	ALICANTE	COMTAT	estudios	1
17	ALICANTE	COMTAT	pisos	7
18	ALICANTE	LŽALACANTÍ	áticos	149
19	ALICANTE	LŽALACANTÍ	chalets	247
20	ALICANTE	LŽALACANTÍ	dúplex	41
21	ALICANTE	LŽALACANTÍ	estudios	16

Inmuebles disponibles | Precios máximos,mí

Fig. 15 : Imagen de informe Número de Inmuebles.

3.4.3.2 Listado de inmuebles que han bajado de precio

Se han diseñado dos informes, uno que indica aquellos que bajaron de precio en algún mes durante su período de venta, con respecto al mes anterior, y otro entre el precio de la oferta inicial y la final (independientemente de que se haya vendido o no).

Con respecto al mes anterior en un mes determinado

En este caso se utiliza el filtro "Reval(mensual)<0" que indica que son inmuebles que en algún momento bajaron de precio. Hay que observar que esto no condiciona el precio final de venta, ya que en otro momento pudiera haber subido de precio.

Listado de Inmuebles que han bajado de precio respecto al mes anterior.
por zona,tipología y características

Elementos de Página: Anyo: 2007 Mes: FEBRER

	Provincia	Comarca	Tipo	Codigo
1	ALICANTE	ALT VINALOPÓ	chalets	VW0000000454800
2	ALICANTE	ALT VINALOPÓ	pisos	VW0000000545756
3	ALICANTE	BAIX SEGURA	áticos	VW0000000364429
4	ALICANTE	BAIX SEGURA	áticos	VW0000000445194
5	ALICANTE	BAIX SEGURA	áticos	VW0000000530333
6	ALICANTE	BAIX SEGURA	áticos	VW0000000703182
7	ALICANTE	BAIX SEGURA	chalets	VP0000000906710
8	ALICANTE	BAIX SEGURA	chalets	VC0000000647911
9	ALICANTE	BAIX SEGURA	chalets	VW0000000323543
10	ALICANTE	BAIX SEGURA	chalets	VW0000000357390
11	ALICANTE	BAIX SEGURA	dúplex	VW0000000678988
12	ALICANTE	BAIX SEGURA	estudios	VW0000000545775
13	ALICANTE	BAIX SEGURA	pisos	VP0000000843899
14	ALICANTE	BAIX SEGURA	pisos	VC0000000633813
15	ALICANTE	BAIX SEGURA	pisos	VC0000000467486
16	ALICANTE	BAIX SEGURA	pisos	VW0000000730999
17	ALICANTE	BAIX SEGURA	pisos	VW0000000716131
18	ALICANTE	BAIX SEGURA	pisos	VP0000000895857
19	ALICANTE	BAIX SEGURA	pisos	VW0000000646387
20	ALICANTE	BAIX SEGURA	pisos	VC0000000757908

Inmuebles disponibles | Inmuebles bajaron mes | Precios máximos

Fig. 16 : Imagen de informe Listado de Inmuebles bajaron precio respecto al mes anterior.

Del precio final con respecto al precio inicial

En este caso se calcula una diferencia de precio de la oferta final y el de la inicial y si éste es menor que cero. En la oferta final del inmueble pudiera darse el caso de que se corresponde con la de venta o no, por eso se incluye el elemento Vendido para seleccionar cada uno de estos posibles casos.
En nuestro caso particular, se puede ver que en 2008 sólo hay una oferta por lo que no aparece como resultado en los informes.

Si un inmueble no se vendiera y no apareciesen más ofertas por él , porque por ejemplo causa baja del sistema, aparecería en este informe con el campo Vendido=No.

Listado de Inmuebles que han bajado de precio respecto las ofertas inicial y final
por zona,tipología y características

Elementos de Página: Vendido: SI | Anyo: 2007 | Mes: FEBRER

	Provincia	Comarca	Tipo	Codigo	DifPrecioFinal
1	ALICANTE	ALT VINALOPÓ	chalets	VC0000000503288	-41000
2	ALICANTE	BAIX SEGURA	áticos	VP0000000945773	-19000
3	ALICANTE	BAIX SEGURA	áticos	VP0000000946188	-12800
4	ALICANTE	BAIX SEGURA	áticos	VP0000000949623	-1500
5	ALICANTE	BAIX SEGURA	áticos	VP0000000949477	-4600
6	ALICANTE	BAIX SEGURA	áticos	VP0000000867611	-5818
7	ALICANTE	BAIX SEGURA	áticos	VP0000000945785	-19400
8	ALICANTE	BAIX SEGURA	chalets	VP0000000948490	-15000
9	ALICANTE	BAIX SEGURA	chalets	VW0000000747546	-20000
10	ALICANTE	BAIX SEGURA	pisos	VP0000000946305	-15000
11	ALICANTE	BAIX SEGURA	pisos	VW0000000750903	-19000
12	ALICANTE	BAIX SEGURA	pisos	VP0000000945777	-6200
13	ALICANTE	BAIX SEGURA	pisos	VW0000000751975	-20000
14	ALICANTE	BAIX SEGURA	pisos	VW0000000743822	-8000
15	ALICANTE	BAIX SEGURA	pisos	VP0000000948527	-3100
16	ALICANTE	BAIX SEGURA	pisos	VP0000000946184	-12800
17	ALICANTE	BAIX SEGURA	pisos	VP0000000945615	-7000
18	ALICANTE	BAIX SEGURA	pisos	VW0000000319108	-5000
19	ALICANTE	BAIX SEGURA	pisos	VP0000000946311	-3000

Inmuebles bajaron mes | **Inmuebles bajaron precio final** | Precios máximos,mínimos y medio

Fig. 17 : Imagen de informe Listado de Inmuebles bajaron precio ofertas inicial y final.

3.4.3.3 Evolución del precio medio de inmuebles que han bajado de precio respecto a la oferta inicial y final

Evolución precio medio de inmuebles que han bajado de precio respecto a su oferta inicial y final
por zona,tipología y características

Elementos de Página: Anyo: 2006 | Tipo: áticos | Vendido: SI

		PrecioMedio										
		FEBRER	MARÇ	ABRIL	MAIG	JUNY	JULIOL	AGOST	SETEMBRE	OCTUBRE	NOVEMBRE	DESEMBRE
ANDORRA	ANDORRA			267.712,5						625.000,0		
ESPAÑA	ALICANTE	189.319,0		73.200,0	200.000,0		159.000,0	184.000,0	227.650,0	166.770,5	168.000,0	206.841,0
	BARCELONA	391.393,5	280.139,0	337.270,3	380.840,6	402.647,9	326.944,0	389.456,8	403.430,0	420.406,2	391.173,7	396.335,4
	CASTELLÓN			380.000,0	165.000,0		276.000,0	341.503,0	283.000,0			245.000,0
	GIRONA			178.644,0			350.000,0		246.500,0	202.864,0	294.200,0	217.500,0
	IBIZA-FORMENTERA							215.000,0				
	LLEIDA	222.375,0						185.000,0	274.000,0	199.000,0		
	MALLORCA	360.000,0			221.000,0	331.000,0		215.000,0	259.666,7	269.166,7	201.000,0	229.000,0
	MENORCA										138.000,0	
	TARRAGONA				203.142,0		167.910,0	226.677,0	174.500,0		172.000,0	223.652,5
	VALENCIA				300.900,0	365.664,3	270.000,0	230.143,3	353.459,1	269.769,0	285.323,6	337.195,1

Fig. 18 : Imagen de informe Evolución de precio de inmuebles.

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

Tanto si el inmueble ha sido vendido como si no , aparecen los precios máximos , medios y mínimos por anyo , tipo y hasta el nivel del municipio.

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

Elementos de Página: Vendido: SI | Anyo: 2007 | Tipo: chalets

	Provincia	Comarca	Municipio	MinPrecio	PrecioMedio	MaxPrecio
1	ALICANTE	ALT VINALOPÓ	ÁREA DE BENEJAMA	399.000	399.000,0	399.000
2	ALICANTE	ALT VINALOPÓ	ÁREA DE VILLENA	180.304	276.384,0	342.000
3	ALICANTE	ALT VINALOPÓ	<MUNICIPIO>	319.000	319.000,0	319.000
4	ALICANTE	BAIX SEGURA	ÁREA DE GUARDAMAR DEL SEGURA	95.000	183.668,4	370.787
5	ALICANTE	BAIX SEGURA	ÁREA DE LA MATA	108.000	162.500,0	217.000
6	ALICANTE	BAIX SEGURA	ÁREA DE ORIHUELA	78.000	276.666,2	730.000
7	ALICANTE	BAIX SEGURA	ÁREA DE ORIHUELA COSTA	120.000	258.331,3	720.000
8	ALICANTE	BAIX SEGURA	ÁREA DE SAN MIGUEL DE SALINAS	95.000	254.636,7	840.000
9	ALICANTE	BAIX SEGURA	TORREVIEJA	72.125	174.319,7	450.000
10	ALICANTE	BAIX VINALOPÓ	CREVILLENTE	270.000	292.155,3	330.000
11	ALICANTE	BAIX VINALOPÓ	ELCHE/ELX	243.410	477.727,5	760.000
12	ALICANTE	BAIX VINALOPÓ	PARTIDAS DE ELCHE-NORTE	150.253	297.309,5	541.000
13	ALICANTE	BAIX VINALOPÓ	PARTIDAS DE ELCHE-SUR	104.000	104.000,0	104.000
14	ALICANTE	BAIX VINALOPÓ	SANTA POLA	180.000	268.168,8	360.000
15	ALICANTE	COMTAT	ÁREA DE BENIMASSOT	50.000	50.000,0	50.000
16	ALICANTE	COMTAT	ÁREA DE MURO DE ALCOY	200.000	200.000,0	200.000
17	ALICANTE	COMTAT	<MUNICIPIO>	73.000	216.583,3	330.000
18	ALICANTE	LŽALACANTÍ	AGOST	146.250	151.125,0	156.000
19	ALICANTE	LŽALACANTÍ	AIGÜES	252.000	252.000,0	252.000
20	ALICANTE	LŽALACANTÍ	ALICANTE / ALACANT	258.000	579.808,9	1.751.000

Inmuebles bajaron precio final | Precios máximos,mínimos y medios | Superficies máximas,mínimas y medias | Tiempo Ve

Fig. 19 : Imagen de informe de precios máximos, mínimos y medios.

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

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

Elementos de Página: Vendido: SI | Anyo: 2007 | Tipo: chalets

	Provincia	Comarca	Municipio	MinSup	SupMedia	MaxSup
1	ALICANTE	ALT VINALOPÓ	ÁREA DE BENEJAMA	200	200,0	200
2	ALICANTE	ALT VINALOPÓ	ÁREA DE VILLENA	120	206,7	280
3	ALICANTE	ALT VINALOPÓ	<MUNICIPIO>	263	263,0	263
4	ALICANTE	BAIX SEGURA	ÁREA DE GUARDAMAR DEL SEGURA	53	101,8	180
5	ALICANTE	BAIX SEGURA	ÁREA DE LA MATA	60	120,0	180
6	ALICANTE	BAIX SEGURA	ÁREA DE ORIHUELA	75	178,7	461
7	ALICANTE	BAIX SEGURA	ÁREA DE ORIHUELA COSTA	60	119,1	400
8	ALICANTE	BAIX SEGURA	ÁREA DE SAN MIGUEL DE SALINAS	50	165,1	530
9	ALICANTE	BAIX SEGURA	TORREVIEJA	35	90,9	250
10	ALICANTE	BAIX VINALOPÓ	CREVILLENTE	180	240,0	300
11	ALICANTE	BAIX VINALOPÓ	ELCHE/ELX	160	255,0	500
12	ALICANTE	BAIX VINALOPÓ	PARTIDAS DE ELCHE-NORTE	70	149,7	250
13	ALICANTE	BAIX VINALOPÓ	PARTIDAS DE ELCHE-SUR	66	66,0	66
14	ALICANTE	BAIX VINALOPÓ	SANTA POLA	75	97,4	145
15	ALICANTE	COMTAT	ÁREA DE BENIMASSOT	132	132,0	132
16	ALICANTE	COMTAT	ÁREA DE MURO DE ALCOY	160	160,0	160
17	ALICANTE	COMTAT	<MUNICIPIO>	100	178,3	430
18	ALICANTE	LŽALACANTÍ	AGOST	150	175,0	200
19	ALICANTE	LŽALACANTÍ	AIGÜES	120	120,0	120
20	ALICANTE	LŽALACANTÍ	ALICANTE / ALACANT	110	270,5	734
21	ALICANTE	LŽALACANTÍ	BUSOT	95	248,5	500

Inmuebles bajaron precio final | Precios máximos,mínimos y medios | Superficies máximas,mínimas y medias | Tiem

Fig. 20 : Imagen de informe de superficies máximas, mínimas y medias

3.4.3.6 Tiempos medios de venta por zona, tipología y características.

Tiempo medio de venta
por zona,tipología y características

Elementos de Página: País: ESPAÑA | Tipo: áticos

	Anyo	Provincia	Comarca	TiempoVtaPromedio
1	2006	ALICANTE	BAIX SEGURA	6,5
2	2006	ALICANTE	BAIX VINALOPÓ	6,8
3	2006	ALICANTE	LŽALACANTÍ	6,4
4	2006	ALICANTE	LŽALCOYA	4,5
5	2006	ALICANTE	MARINA ALTA	5,9
6	2006	ALICANTE	MARINA BAIXA	7,0
7	2006	ALICANTE	VINALOPÓ MITJÀ	5,3
8	2006	BARCELONA	ALT PENEDÈS	6,8
9	2006	BARCELONA	ANOIA	5,1
10	2006	BARCELONA	BAGES	5,5
11	2006	BARCELONA	BAIX LLOBREGAT NORD	7,0
12	2006	BARCELONA	BAIX LLOBREGAT SUD	6,1
13	2006	BARCELONA	BARCELONÈS	5,8
14	2006	BARCELONA	GARRAF	6,8
15	2006	BARCELONA	MARESME	5,9
16	2006	BARCELONA	OSONA	6,3
17	2006	BARCELONA	VALLÈS OCCIDENTAL	5,8
18	2006	BARCELONA	VALLÈS ORIENTAL	6,6
19	2006	CASTELLÓN	ALTO PALANCIA	8,7
20	2006	CASTELLÓN	BAIX MAESTRAT	6,8

Tiempo Venta | precio final | Precios máximos,mínimos y medios | Su

Fig. 21 :Imagen de informe Tiempo medio de venta.

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

Número de Inmuebles vendidos ó disponibles
por zona,tipología y características

Elementos de Página: Anyo: 2007 Vendido: SI

	Provincia	Comarca	Tipo	NumInmuebles
1	ALICANTE	ALT VINALOPÓ	chalets	8
2	ALICANTE	ALT VINALOPÓ	pisos	3
3	ALICANTE	BAIX SEGURA	áticos	70
4	ALICANTE	BAIX SEGURA	chalets	164
5	ALICANTE	BAIX SEGURA	dúplex	51
6	ALICANTE	BAIX SEGURA	estudios	37
7	ALICANTE	BAIX SEGURA	pisos	490
8	ALICANTE	BAIX VINALOPÓ	áticos	5
9	ALICANTE	BAIX VINALOPÓ	chalets	26
10	ALICANTE	BAIX VINALOPÓ	dúplex	3
11	ALICANTE	BAIX VINALOPÓ	estudios	2
12	ALICANTE	BAIX VINALOPÓ	pisos	91
13	ALICANTE	COMTAT	áticos	2
14	ALICANTE	COMTAT	chalets	8
15	ALICANTE	COMTAT	pisos	4
16	ALICANTE	LŽALACANTÍ	áticos	79
17	ALICANTE	LŽALACANTÍ	chalets	174
18	ALICANTE	LŽALACANTÍ	dúplex	21
19	ALICANTE	LŽALACANTÍ	estudios	2
20	ALICANTE	LŽALACANTÍ	pisos	585
21	ALICANTE	LŽALCOYA	áticos	2

Num Inmuebles ven/disp | Precios máximos,mínimos y medios | Super

Fig. 22 : Imagen de informe Número de inmuebles vendidos/disponbiles

3.4.3.8 Modelo de inmueble español y andorrano.

Se han generado dos informes para realizar una comparación bien por características físicas (superficie y número de dormitorios) o bien por las características de la venta (precio de salida precio de venta y tiempo de venta).

Modelo Inmueble por precio salida, precio de venta y tiempo de venta medio
por tipología

Elementos de Página:

	Tipo	Pais	PrecioInicialMedio	PrecioMedio	TiempoVtaPromedio
1	áticos	ANDORRA	508.771,3	504.172,9	13,0
2	áticos	ESPAÑA	363.610,2	360.244,7	12,0
3	chalets	ANDORRA	1.313.476,7	1.310.476,7	19,5
4	chalets	ESPAÑA	538.954,8	535.734,0	14,6
5	dúplex	ANDORRA	547.112,6	547.690,3	9,9
6	dúplex	ESPAÑA	375.414,1	372.497,3	12,7
7	estudios	ANDORRA	194.495,3	193.066,7	8,6
8	estudios	ESPAÑA	152.390,1	151.154,3	11,7
9	pisos	ANDORRA	362.363,7	360.349,4	10,2
10	pisos	ESPAÑA	295.598,9	292.950,5	11,3

Fig. 23 : Imagen de informe Modelo de inmueble por precio y tiempo.

El piso andorrano es más caro y tarda en venderse menos que el español, y la mayor diferencia de precios se establce en el precio de chalets.

Modelo Inmueble por dormitorios y superficie medias por tipología

Elementos de Página:

	Tipo	Pais	DormitoriosPromedio	SupMedia
1	áticos	ANDORRA	2,7	106,4
2	áticos	ESPAÑA	2,7	103,4
3	chalets	ANDORRA	3,5	324,3
4	chalets	ESPAÑA	3,8	214,5
5	dúplex	ANDORRA	2,9	126,2
6	dúplex	ESPAÑA	3,0	124,0
7	estudios	ANDORRA	0,0	41,6
8	estudios	ESPAÑA	0,0	39,3
9	pisos	ANDORRA	2,0	81,4
10	pisos	ESPAÑA	2,8	87,3

Fig. 24 : Imagen de informe Modelo de inmueble por superficie y número de dormitorios.

De los pisos se puede comprobrar que el piso español es de 2,8 dormitorios y una superficie de 87,3 y el modelo andorrano es de menor tamaño y número de habitaciones, probablemente de mayor superficie.

4. Capítulo 4 - Conclusiones

Una vez se llega a este punto, se hace necesario una reflexión para revisar lo que ha significado el desarrollo del proyecto y llevar a cabo una valoración.

Gran parte de mi trayectoria profesional está ligada al “reporting” pero desde la perspectiva del usuario final, sin la posibilidad de conocer qué hay por debajo de las aplicaciones finales, por lo que este proyecto ha servido para aumentar mi interés por este tipo de desarrollos.

Ha sido muy gratificante poder aplicar los conocimientos se han ido adquiriendo a lo largo de los años de estudios de la ETIS y de la experiencia laboral.

También ha sido muy enriquecedor la elaboración de un proyecto de inicio a fin en lugar de participar únicamente en alguna fase del desarrollo bien como analista o como programador o como interlocutor para la recogida de requerimientos del cliente, como suele ser lo habitual en el mundo profesional.

La realización de proyectos como éste ayuda a tener una visión global de todo lo que comporta la realización de un proyecto profesional.

5. Capítulo 5 - Líneas de evolución futuras

- Creación de una plantilla para la captura de datos que además contemple la normalización de los códigos y nombres correspondientes a las provincias, comarcas, poblaciones, distritos y barrios.

- La posibilidad de incorporar un mecanismo para establecer umbrales (como la máxima planta de un inmuebles o su superficie) sin que sea necesario que estén programados en triggers en la base de datos.

- La incorporación de otro tipo de información, como puede ser la forma de pago al realizar la compra, pudiera dar una idea de la posible financiación y ofertas como nuevas líneas de negocio.

- La creación de nuevos informes que indiquen los porcentajes de ventas conseguidas sobre el total ofertado por zonas, tipos y características para detectar donde llevar a cabo acciones de refuerzo o de incentivación. Para ello es necesario tener un conjunto de datos en una ventana temporal mayor que la que se ha aportado, para obtener datos relevantes.

- También se podría incluir un fichero de personal que ha llevado a cabo la venta para un programa de incentivos.

Glosario

Atributo: Cualquier detalle que sirva para calificar, identificar, clasificar, cuantificar o expresar el estado de una entidad.

Análisis multidimensional: El análisis multidimensional consiste en analizar los datos que hacen referencia a hechos desde la perspectiva de sus componentes o dimensiones, abarcando también los diferentes niveles que puedan tener las respectivas dimensiones.

Base de Datos: Es un sistema que almacena datos que están relacionados.

Bussines Intelligence (B.I.): Conjunto de estrategias y herramientas orientadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa.

Clave foranea: Una o más columnas de una tabla de una base de datos relacional que implementa una relación muchos a uno.

Clave primaria: Una o más columnas de una table de una base de datos relacional cuyos valores combinados entre sí son únicos dentro de la tabla.

Celda: Medida que hace referencia al mismo acontecimiento dentro de una misma estructura.

Celda : Conjunto de celdas del mismo hecho que están asociadas a instancias del mismo nivel para cada una de las dimensiones.

Cubo OLAP: Los cubos de información o cubos OLAP funcionan como los cubos de rompecabezas en los juegos, en el juego se trata de armar los colores y en el DWH se trata de organizar los datos por tablas o relaciones; los primeros (el juego) tienen 3 dimensiones, los cubos OLAP tienen un número indefinido de dimensiones, razón por la cual también reciben el nombre de hipercubos. Un cubo OLAP contendrá datos de una determinada variable que se desea analizar, proporcionando una vista lógica de los datos provistos por el sistema de información hacia el DWH, esta vista estará dispuesta según unas dimensiones y podrá contener información calculada. El análisis de los datos está basado en las dimensiones del hipercubo, por lo tanto, se trata de un análisis multidimensional.

Data Warehouse (DW ó DWH): Es la integración de datos consolidados, almacenados en un dispositivo de memoria no volátil, proveniente de múltiples y posiblemente diferentes fuentes de datos. Con el propósito del análisis y a partir de este tomar decisiones en función de mejorar la gestión del negocio. Contiene un conjunto de cubos de datos que permiten a través de técnicas de OLAP consolidar, ver y resumir los datos acorde a diferentes dimensiones de) estos. (Chaudhuri & Dayal, 1997)

Data Marts: Es un subconjunto del Data Warehouse, usado normalmente para el análisis parcial de los datos. Por ejemplo, el Data Mart de los datos del departamento ventas y el Data Mart de Inventarios. El objetivo de subdividir está dado por la complejidad computacional del análisis global de todas las dimensiones del Data Warehouse y por la necesidad de rapidez. (Microsoft Data Warehouse Training, 2000)

Descriptor: Atributo que se encuentra en un nivel. Contiene información no jerárquica y está definido sobre un dominio discreto.

Diagrama de casos: En el Lenguaje de Modelado Unificado (UML), un diagrama de casosde uso es una especie de diagrama de comportamiento.

Dimensiones: Las dimensiones de un cubo son atributos relativos a las variables, son las perspectivas de análisis de las variables (forman parte de la tabla de dimensiones). Son catálogos de información complementaria necesaria para la presentación de los datos a los usuarios, como por ejemplo: descripciones, nombres, zonas, rangos de tiempo, etc. Es decir, la información general complementaria a cada uno de los registros de la tabla de hechos.

Esquema o modelo en estrella: En las bases de datos usadas para data warehousing, un esquema en estrella es un modelo de datos que tiene una tabla de hechos (o tabla fact) que contiene los datos para el análisis, rodeada de las tablas de dimensiones. Este aspecto, de tabla de hechos (o central) más grande rodeada de radios o tablas más pequeñas es lo que asemeja a una estrella, dándole nombre a este tipo de construcciones.

EUL (End User Layer) : Enmascara la complejidad y la estructura del almacén de datos proporcionando a los usuarios finales una interfaz fácil de entender y de utilizar sin que sea necesario conocimientos específicos de bases de datos.

Oracle: Sistema de gestión de bases de datos relacional, desarrollado por Oracle Corporation.

Oracle Discoverer: Es una herramienta orientada a temáticas. Se utiliza para construir consultas ad hoc con datos de un área específica que son de interés para un cierto grupo de usuarios.

Oracle Discoverer Administrator: Este producto sirve para crear, mantener y administrar datos en la capa de usuario final (EUL) y para definir como los usuarios interactuan con los datos.

Oracle Discoverer Desktop: El usuario final (EUL) utiliza este componente para ejecutar consultas ad hoc y generar informes.

Proceso ETL (extracción, transformación y carga): Los procesos de extracción, transformación y carga (ETL) son importantes ya que son la forma en que los datos se guardan en un almacén de datos (o en cualquier base de datos). Implican la extracción (acción de obtener la información deseada a partir de los datos almacenados en fuentes externas), la transformación (cualquier operación realizada sobre los datos para que puedan ser cargados en el data warehouse o se puedan migrar de éste a otra base de datos) y la carga (almacenar los datos en la base de datos final, por ejemplo el almacén de datos objetivo normal).

Granularidad : se refiere a la especificidad a la que se define un nivel de detalle en una tabla, es decir, si hablamos de una jerarquía la granularidad empieza por la parte más alta de la jerarquía, siendo la granularidad mínima, el nivel más bajo. En Data Warehouse, no sólo existe granularidad para las tablas de hechos (Fact's), también existe granularidad en las dimensiones.

Hecho : Actividad de la empresa objeto de análisis.

Indicadores: También llamadas “indicadores de gestión”, son los datos que están siendo analizados. Forman parte de la tabla de hechos. Más formalmente, las variables representan algún aspecto cuantificable o medible de los objetos o eventos a analizar.

Jerarquía de agregación: Conjunto de relaciones entre las instancias de una dimensión que indica cómo se agrupan unas para formas las otras.

Medida: Atributo numérico de una Celda, normalmente aditivo.

Modelo conceptual: El modelo conceptual no es más que un esquema de los datos y de las asociaciones o interrelaciones que se han obtenido a partir del conocimiento de los requerimientos y especificaciones de lo que se quiere obtener al final.

OLAP (Online Analitical Proccesing): Son los sistemas que se usan para analizar los datos que las OLTP introducen en la Base de Datos. A diferencia de los primeros estos casi siempre usan el modelo multidimensional para organizar los datos en la Base de Datos ya que brindan mejores resultados a la hora del análisis de estos. (Microsoft Books Online, 2000)

PL/SQL: Lenguaje de programación que soporta todas las consultas y manipulación de datos que se utilizan en SQL, pero que incluye características como el manejo de variables, las estructuras modulares , las de control de flujo y el control de excepciones.

SQL*Loader: Utilidad de importación de datos que posibilita la carga automática de datos externos (residentes en ficheros) en tablas de las bases de datos. La

información puede cargarse en una o varías tablas previamente creadas y que pueden contener datos previos. Los nuevos datos podrán sustituir a los que ya existían o bien añadirse como nuevos registros.

Trigger : También conocido como disparador, es un procedimiento almacenado que se activa cuando se dan ciertas condiciones, llevando a cabo acciones que pueden modificar los datos insertados y/o borrados de una tabla.

Bibliografía consultada

Publicaciones

The Data Warehouse Toolkit. (2nd Edition). Ralph Kimball and Margy Ross
Building the Data Warehouse (3rd Edition). W.H. Inmon.
Magatzems de dades i models multidimensionals. Gavídia y Serra. UOC

Manuales y guias

Oracle Database SQL Reference 10g R2 2005
Oracle Database PL/SQL User's Guide and Reference 10g R2 2005
Oracle Bussines Intelligence Discoverer Administration Guide 10g Release 2 2005
Oracle Bussines Intelligence Discoverer Desktop User's Guide 10g Release 2 2005

Web

http://www.oracle.com/index.html
http://www.1keydata.comOracle SQL*Loader: The Definitive Guide. Gennikc y Mishra (2001)

Anexos

Anexo A: "Acceso a la base de datos"

Para la creación de las estructuras necesarias (tablas, índice, triggers, secuencias, procedimientos, etc.), se ha creado el usuario "uoc_usr" con contraseña "1234" y tablespace es USERS y tablespace temporal TEMP,a partir del usuario SYSTEM y contraseña "uoc" aportado.

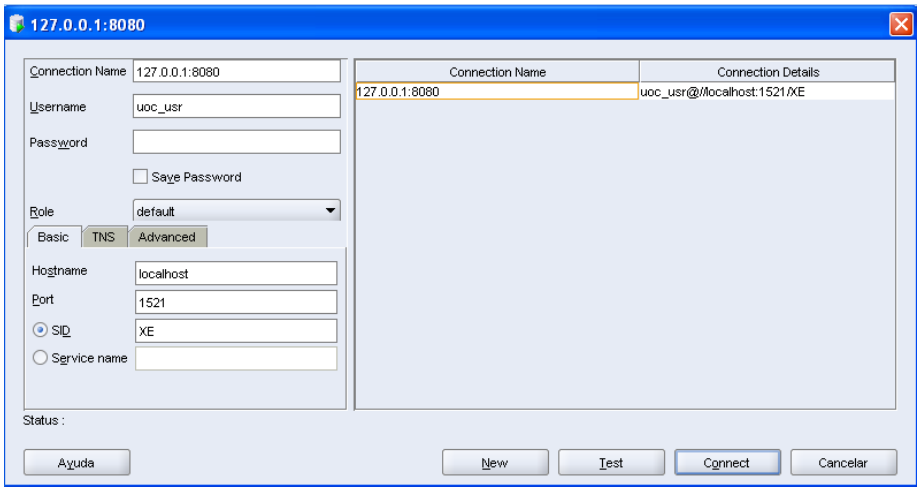

Fig. 25 : Detalle del usuario y propiedades de conexión.

El esquema asociado al usuario uoc_usr contiene los siguientes objetos.:

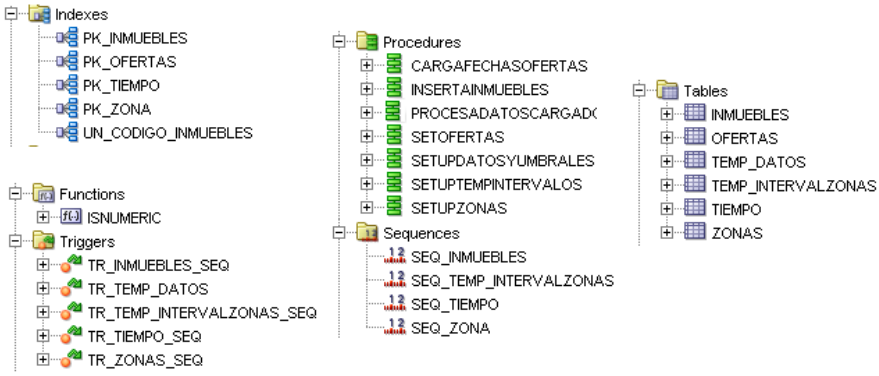

Fig. 26 : Esquema uoc_usr.

Anexo B : "Detalle y automatización del proceso ETL"

El proceso de carga ó de importación de los datos puede dividirse en las siguientes partes:

4.0 Creación de las estructuras sobre en la base de datos para la gestión del almacén de datos.

La creación de las tablas, secuencias, índices, triggers, funciones, claves y procedimientos almacenados se supone realizada en una fase previa, y que lógicamente solamente se lleva a cabo una única vez.

4.1 Transformación de las hojas de cálculo con los contenidos de las ofertas a formato csv ó txt, desde la propia herramienta Microsoft Excel.

Un usuario que desee llevar a cabo este proceso, generará a partir de su fichero Excel un fichero .csv ó .txt a partir del mismo en la carpeta especificada y además tendrá que **editar** el fichero **CargaDatos.ctl**, en

concreto la entrada INFILE para indicar el nombre del fichero .csv ó .txt generado que desea cargar.
(En la imagen inferior hay un detalle de la edición)

4.2 Carga del contenido del fichero .csv ó .txt en la base datos.

Concretamente la carga se lleva a cabo utilizando la herraienta SqlLoader que inserta los datos sobre la tabla TEMP_DATOS. Durante dicha carga se lleva a cabo un primer filtrado de los datos a través de un trigger sobre dicha tabla.

Los registros que no se hayan cargado se almacenarán en el archivo Carga.bad.

El proceso de carga se ha automatizado de la siguiente manera:

Se ha creado el archivo **CargaDatos.bat** que contiene el comando para que SqlLoader realice la misma sobre la tabla indicada, a través de un fichero de control, CargaDatos.ctl. Tras esta carga se ejecutará los procedimientos para la tranformación y carga de los datos en el almacén de datos que se detallarán en el siguiente punto.

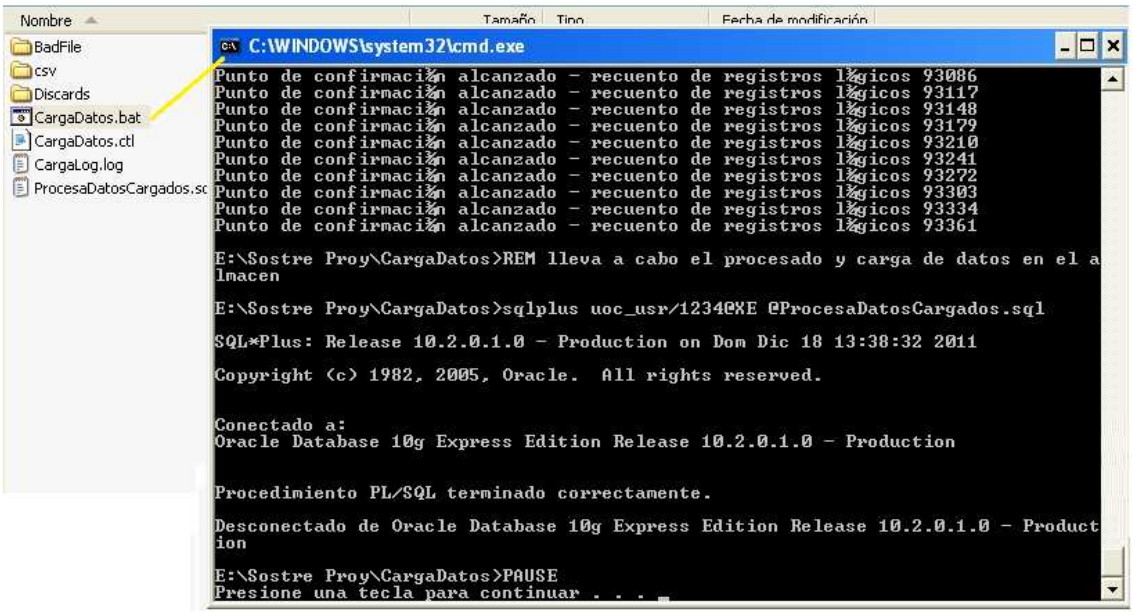

Fig. 27 : Imagen del resultado de los procesos de importación y procesado de datos.

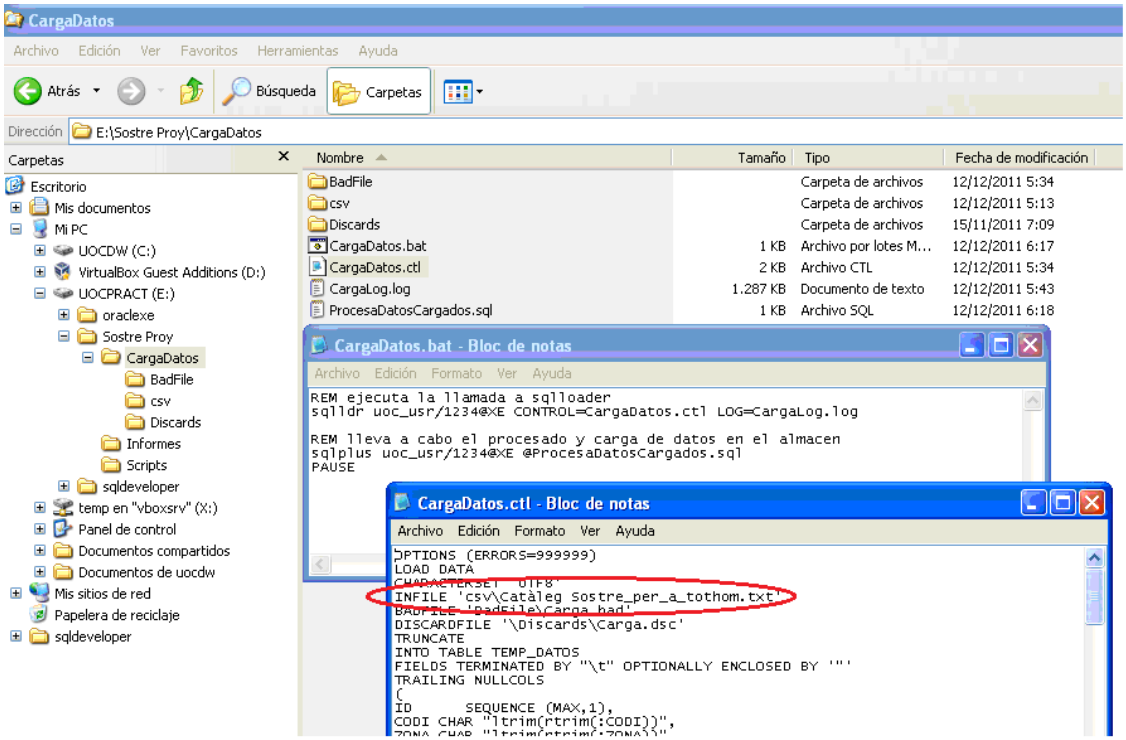

Fig. 28 :Imagen del contenido de los ficheros de carga y control con detalle de línea a editar.

El proceso de carga puede generar varios ficheros de salida, y uno de ellos siempre será el fichero de log, **Carga.log**, en donde se indica las incidencias encontradas durante el proceso de carga, así como un resumen final de los registros cargados, rechazados, etc.

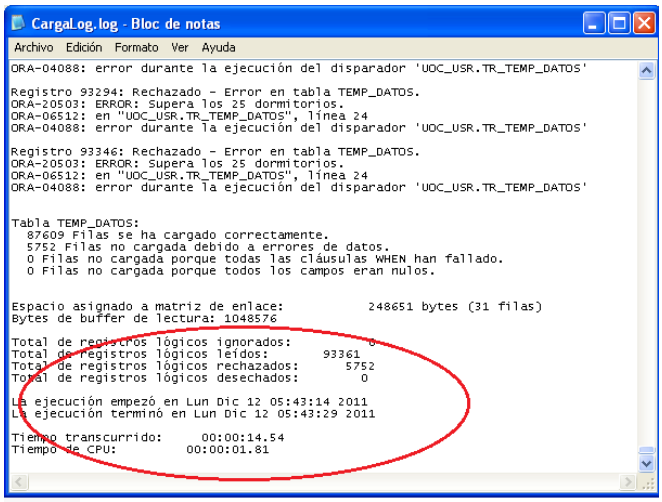

Fig. 29 : Imagen del contenido del fichero de log.

Otro de los ficheros que se generarán en caso de que haya registros rechazados, será el fichero **/BadFile/Carga.bad**. Como he comentado, aquí se encontrarán aquellos que han sido rechazados por el trigger programado sobre la tabla TEMP_DATOS, y que detallaré más adelante.

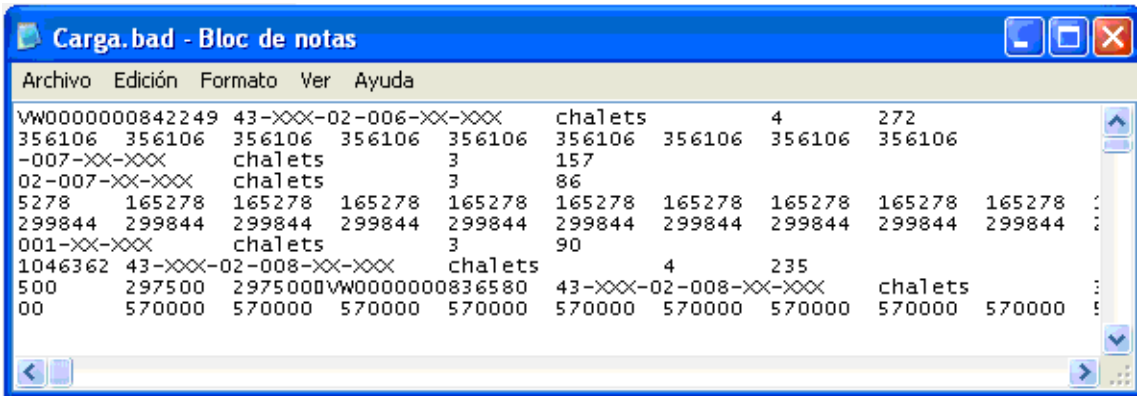

Fig. 30 : Imagen del contenido del fichero Carga.bad.

4.3 Procesado de los datos cargados e inserción en el almacén de datos.

Este proceso se lleva a cabo conjuntamente y a continuación del anterior, pues ha sido programado en el fichero CargaDatos.bat.

Concretamente, se trata de un comando sobre sqlplus, que ejecutará el contenido del fichero **ProcesaDatosCargados.sql**, que es una llamada al procedimiento ProcesaDatosCargados.

4.3.1 Otros detalles a tener en cuenta:

- Un fichero .csv ó .txt se puede volver a repetir el proceso.

- Cada fichero .csv ó .txt puede incluir hasta 25 ofertas por cada inmueble.

- No importa el orden temporal de las ofertas en cada fichero .csv ó .txt, y tampoco el orden en que se importen los ficheros .csv ó .txt es decir, pueden importarse ofertas del año 2011 antes que las ofertas del año 2010.

- El proceso reordenará siempre la temporalidad de las ofertas importadas junto con las existentes ya en el almacén de datos y calculará los campos de cada oferta.

- El sistema también está preparado para que un mismo inmueble pueda ser vendido en períodos diferentes.

4.3.2 Descripción de los elementos utilizados:

a) Disparador (trigger) sobre la tabla de carga TEMP_DATOS

Con la finalidad de facilitar al cliente aquellos registros con errores en su contenido y que éstos puedan ser corregidos para una inserción posterior, se ha diseñado un trigger que controla la inserción de los datos.

Dicho trigger podrá ser modificado para controlar algún otro tipo de errores.

Los registros rechazados se almacenarán en el fichero Carga.bad y las causas por las que un registro no se insertará en TEMP_DATOS son:

- Inmuebles que no sean chalets en los que no se haya definido la planta.
- Inmuebles que no sean estudios en los que no se haya definido el número de dormitorios.
- Inmuebles áticos en plantas 'bj','en','pp','ss','st'.
- Inmuebles con más de 25 dormitorios.
- Inmuebles que no sean chalets que superen la planta 60.

El fichero Carga.bad está en formato de texto con los campos delimitados por tabulación, y podrá ser editado fácilmente tras ser recuperado con la herramienta Excel.

b) Tablas utilizadas para el procesado de datos

Se trata de dos tablas. TEMP_DATOS sobre la que se lleva a cabo directamente la importación desde SqlLoader y TEMP_INTERVALZONAS que se rellenará durante el procesado. Ambas se vacian de datos en la importación y procesado respectivamente.

TEMP_DATOS es un reflejo (exceptuando los registros rechazados) del contenido del fichero .csv ó .txt y además contiene campos adicionales que facilitan el posterior procesado, como son :

- Un campo **ID** para identificar a cada uno de los registros.

- Un campo **IDZONA** y un campo **IDINMUEBLE** que se actualizarán durante el procesado y que son identificadores de la zona y del inmueble que le corresponde a cada registro de inmueble en las tablas del almacén de datos ZONAS e INMUEBLES.

Column Name	Data Type	Nullable	Data Default	COLUMN ID	Primary Key
ID	NUMBER	Yes		1	
CODI	VARCHAR2(255 Bytes)	Yes		2	
ZONA	VARCHAR2(30 Bytes)	Yes		3	
TIPUS	VARCHAR2(30 Bytes)	Yes		4	
PLANTA	VARCHAR2(15 Bytes)	Yes		5	
DORMITORIS	VARCHAR2(15 Bytes)	Yes		6	
METRES	VARCHAR2(15 Bytes)	Yes		7	
CAMPO1	VARCHAR2(30 Bytes)	Yes		8	
CAMPO2	VARCHAR2(30 Bytes)	Yes		9	
CAMPO3	VARCHAR2(30 Bytes)	Yes		10	
CAMPO4	VARCHAR2(30 Bytes)	Yes		11	
CAMPO5	VARCHAR2(30 Bytes)	Yes		12	
CAMPO20	VARCHAR2(30 Bytes)	Yes		27	
CAMPO21	VARCHAR2(30 Bytes)	Yes		28	
CAMPO22	VARCHAR2(30 Bytes)	Yes		29	
CAMPO23	VARCHAR2(30 Bytes)	Yes		30	
CAMPO24	VARCHAR2(30 Bytes)	Yes		31	
CAMPO25	VARCHAR2(30 Bytes)	Yes		32	
IDZONA	NUMBER(5,0)	Yes		33	
IDINMUEBLE	NUMBER(15,0)	Yes		34	

Fig. 31 : Imagen de la tabla TEMP_DATOS.

TEMP_INTERVALZONAS, se borra en cada proceso de carga. Es una tabla de control que me permite tener controlado el nivel de influencia de una determinada zona. Para ello identifica la jeraquía final de la zona, es decir, el registro anterior al siguiente/s registros que contienen inmuebles y determina desde qué registro hasta qué otro registro se extiende el conjunto de inmuebles que pertenecen a dicha zona.

Para realizarlo es importante el campo TEMP_DATOS.id, esté ordenado ascendentemente.

Column Name	Data Type	Nullable	Data Default	COLUMN ID	Primary Key
ID	NUMBER(10,0)	No		1	
ID_TO_DATOS_TEMP	NUMBER(15,0)	No		2	
IDZONA	NUMBER(5,0)	Yes		3	
CODI	VARCHAR2(255 Bytes)	Yes		4	
DESDE	NUMBER(15,0)	Yes		5	
HASTA	NUMBER(15,0)	Yes		6	

Fig. 32 : Imagen de la tabla TEMP_INTERVALZONAS.

A continuación una imagen del concepto de área de influencia de una zona sobre datos de la tabla TEMP_DATOS

ID	CODI	ZONA	TIPUS	PLANTA	DORMITORIS	METRES	CAMPO1	CAMPO2	CAMPO3	CAMPO4
1657	VW0000000789241	43-123-03-004-05-XXX	pisos	2	1	75	168000	168000		
1658	VW0000000802741	43-123-03-004-05-XXX	pisos	1	4	113	270000	270000	270000	270000
1659	VW0000000803680	43-123-03-004-05-XXX	pisos	1	4	100	264445	258435	258435	258435
1660	VW0000000823926	43-123-03-004-05-XXX	pisos	4	2	60	138500	138500	138500	138500
1661	# ::: tarragona ::: baix camp ::: área de reus ::: m...									
1662	VC0000000632246	43-XXX-03-003-XX-X...	chal...		4	135	450759	450759	450759	450759
1663	VC0000000906015	43-123-03-004-03-XXX	pisos	1	4	120	280000	280000	280000	
1664	VW0000000452136	43-XXX-03-003-XX-X...	chal...		4	195	462000	462000	462000	462000
1665	VW0000000728252	43-123-03-004-03-XXX	pisos	3	4	124	294500	294500	294500	294500
1667	# ::: tarragona ::: baix camp ::: área de reus ::: c...									
1668	VC0000000453536	43-XXX-03-003-XX-X...	pisos	2	3	65	112000	112000	112000	112000
1669	VC0000000751307	43-123-03-004-01-XXX	pisos	2	3	92	242000	242000	242000	242000
1670	VC0000000751645	43-123-03-004-01-XXX	pisos	3	2	55	132000	132000	132000	132000

Fig. 33 : Imagen del área de influencia de una jerarquía

.

y el correspondiente registro en TEMP_INTERVALZONAS

ID	ID_TO_DATOS_TEMP	IDZONA	CODI	DESDE	HASTA
634	1661	629	# ::: tarragona ::: baix camp ::: área de reus ::: mestral	1662	1665

Fig. 34 : Imagen de la tabla TEMP_INTERVALZONAS correspondiente.

c) Procedimientos de procesado.

En este apartado describiré el procesado que se ha llevado a cabo de los datos en los correspondientes procedimientos almacenados utilizados.

Como se ha comentado anteriormente , el proceso se centraliza en el procedimiento almacenado ProcesaDatosCargados que llama de forma secuencial a los procedimientos:

SetUpDatosyUmbrales,
SetUpTEMPINTERVALOS,
SetUpZonas,
InsertaInmuebles,
CargaFechasOferts y
SetOfertas.

En el anexo a este documento se incluye el código de los mismos.

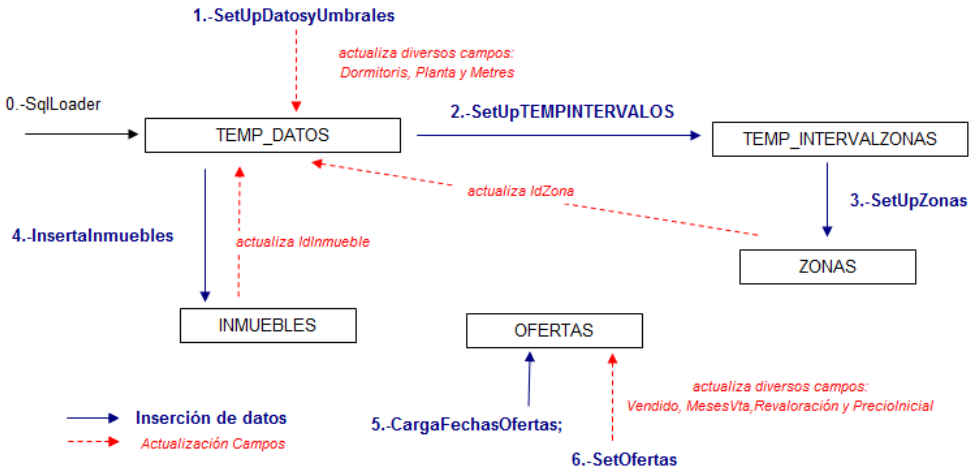

Fig. 35 : Imagen de las partes del procesado de datos

- SetUpDatosyUmbrales

Actualiza aquellos registros que no tienen información coherente o aquellos que presentan valores en sus campos que se consideran fuera de rango. Así por ejemplo:

- para los estudios el campo Dormitoris se pone a null, pues no tiene sentido un estudio con dormitorios.

- para los chalets el campo Planta se pone a null

- el campo metres se pone a null para los siguientes casos:

inmuebles con metres<20
chalets con metres>2500
estudios con metres>100
dúplex con metres >900
áticos y pisos con metres >600

Todos estos casos serán revisables por el cliente quién podrá decidir su modificación ó bien añadir nuevos casos.

- SetUpTEMPINTERVALOS

Utiliza la tabla TEMP_DATOS. Este procedimiento establece el conjunto de registros consecutivos de inmuebles que se encuentran en una

determinada zona. Para ello, localiza el registro de zona correspondiente, que es aquel inmediatamente anterior al primer registro de inmueble (donde el campo codi comienza por el carácter 'V').

Posteriormente se establece el área de influencia de la zona como el mínimo y el máximo identificador de registro de inmueble, antes del primer registro de zona siguiente (aquel registro que comienza por el carácter '#`).

- SetUpZonas

A partir de la zona obtenida en la tabla temp_intervalzonas, divide su contenido en varios niveles considerando a éstos como la parte entre los caracteres ' **::::** '.

El conjunto de niveles establecido será por este orden: provincia, comarca, municipio, distrito y barrio, además de un nivel país.

En los casos en los que el número de niveles exceda o quede por debajo de los establecidos, se procederá a concatenar los niveles de la parte derecha o a rellenar éstos con los valores <Provincia>, <Comarca>, <Municipio>, <Distrito> y <Barrio>.

```
#  ::: ibiza-formentera ::: área de sant antoni
#  ::: barcelona ::: barcelonès ::: barcelona ::: sarrià-sant gervasi ::: vallvidrera-les planes
```

IDZONA	PAIS	PROVINCIA	COMARCA	MUNICIPIO	DISTRITO	BARRIO
214	ESPAÑA	IBIZA-FORMENTERA	ÁREA DE SANT ANTONI	<MUNICIPIO>	<DISTRITO>	<BARRIO>
269	ESPAÑA	BARCELONA	BARCELONÈS	BARCELONA	SARRIÀ-SANT GERVASI	VALLVIDRERA-LES PLANES

Fig. 36 : Imagen de ejemplo de zonas

- InsertaInmuebles

Inserta los inmuebles desde la tabla TEMP_DATOS. Aquellos que ya se encontrasen en dicha tabla (mismo código) tan sólo actualizarán sus campos superficie, planta, dormitorios y tipo.

- CargaFechasOfertas

El proceso de la carga de las ofertas correspondientes a cada mes para cada uno de los inmuebles, se realiza mediante un bucle que recoge el contenido del primer registro (que es parte de la cabecera del fichero .csv ó .txt importado) de los campos Campo1, Campo2, … Campo25 de la tabla TEMP_DATOS.

Los valores de este primer registro para dichos campos que no sean nulos, corresponderán a columnas con precios de la oferta para ese mes para los inmuebles.

De esta manera, en cada vuelta del bucle se inspecciona en primer lugar el contenido de ese primer registro para una columna dada y se extrae el mes y año que serán insertados, si no lo han sido ya, en la tabla TIEMPOS de la que se recuperará el identificador asociado al par mes y año.

A continuación se insertan los registros de los campos idTiempo, idInmueble, idZona y CampoN en la tabla OFERTAS, seleccionado dichos valores de la tabla TEMP_DATOS correspondientes a los inmuebles.

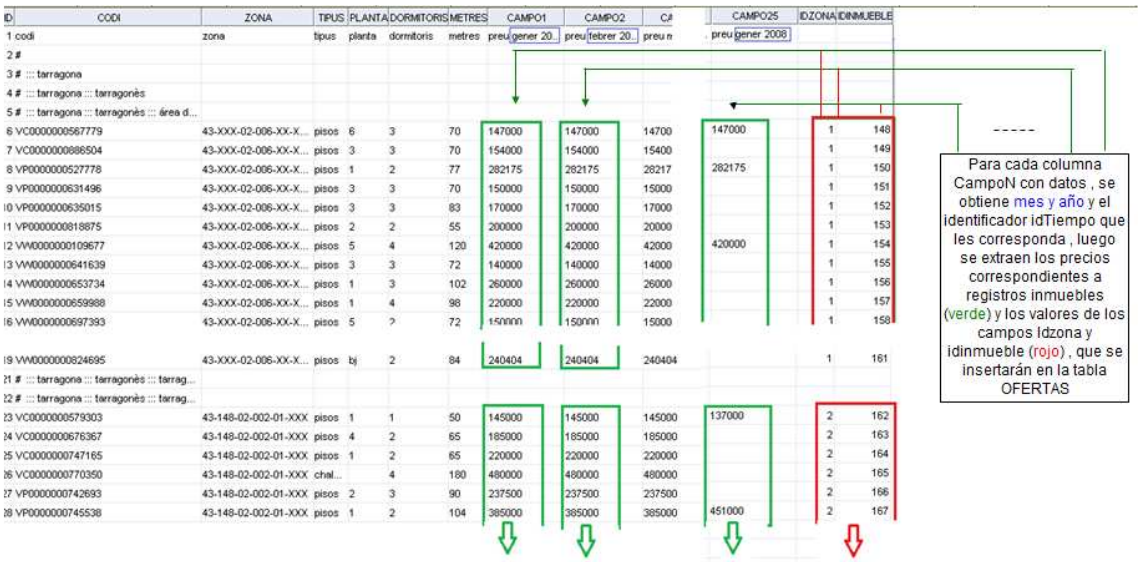

Fig. 37 : Imagen proceso de extracción de ofertas de inmuebles

- SetOfertas

Para cada uno de los inmuebles, se obtienen sus ofertas de forma <u>ordenada en el tiempo</u>, y se calcula para <u>la totalidad</u> de las mismas el valor de los campos PrecioInicial, MesesVenta, Vendido y Revaloración según corresponda.

En la imagen inferior se muestra un ejemplo de las ofertas de un inmueble con identificador 107 y de la totalidad de las ofertas existentes en el almacén de datos. Se puede comprobar que el campo revaloración se actualiza únicamente cuando cambia el precio de un inmueble de un mes a otro, y que los campos vendido, precio inicial y meses venta, lo hacen en el último registro del inmueble

	IDINMUEBLE	IDTIEMPO	IDZONA	PRECIO	PRECIOINICIAL	VENDIDO	MESESVTA	REVALORACIO
151	107	1	6	400000				
152	107	2	6	400000				
153	107	3	6	400000				
154	107	4	6	400000				
155	107	5	6	400000				
156	107	6	6	390000				-10000
157	107	7	6	390000				
158	107	8	6	390000				
159	107	9	6	390000				
160	107	10	6	390000				
161	107	11	6	390000				
162	107	12	6	390000				
163	107	13	6	390000				
164	107	14	6	390000				
165	107	15	6	390000				
166	107	16	6	370000				-20000
167	107	17	6	370000				
168	107	18	6	370000				
169	107	19	6	370000				
170	107	20	6	370000				
171	107	21	6	370000				
172	107	22	6	370000				
173	107	23	6	370000				
174	107	24	6	370000				
175	107	25	6	360000	360000	NO	25	-10000

Fig. 38 : Imagen de ejemplo ofertas para un inmueble

La duración aproximada del proceso total incluyendo la inserción en la tabla TEMP_DATOS y la fase siguiente de preparación e inserción en el almacén de datos es aproximadamente de **15 minutos** sobre la máquina virtual aportada y configurada según el cliente sobre un host a 2,3 Ghz con 6 GB de RAM y S.O. Windonws7.

Anexo C: "Acceso a Oracle Discoverer Administration"

Se ha creado en la base de datos un usuario/esquema eul_usr/1234, con tablespace por defecto USERS y tablespace temporal TEMP, al que se le han otorgado los privilegios RESOURCE y CONNECT, así como privilegios de consulta sobre las tablas del almacén de datos: TIEMPO, ZONAS, INMUEBLES y OFERTAS.

Con dicho usuario se accede a la herramienta y se crean el EUL y un área de negocio nueva AN_Ventas.

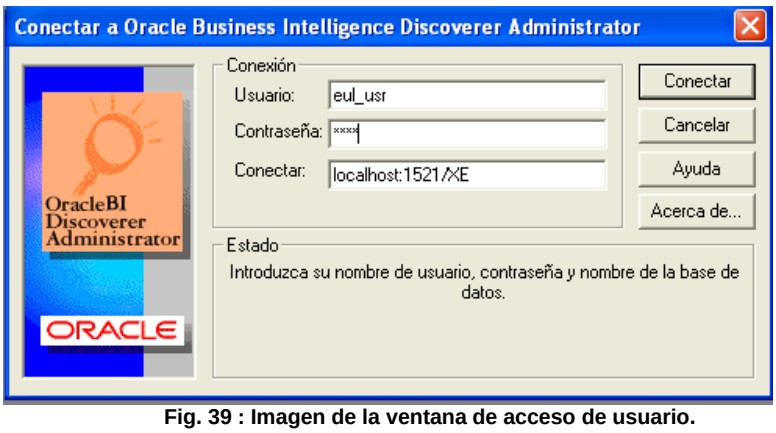

Fig. 39 : Imagen de la ventana de acceso de usuario.

Anexo D: "Acceso a Oracle Discoverer Desktop"

Una vez creado el área de negocio por el usuario eul_usr/1234, a través de la herramienta Oracle Discoverer Administrator, se le ha autorizado al usuario uoc_usr/1234 acceso a dicho área de negocio a través del menú Privilegios (se utilizan las opciones por defecto):

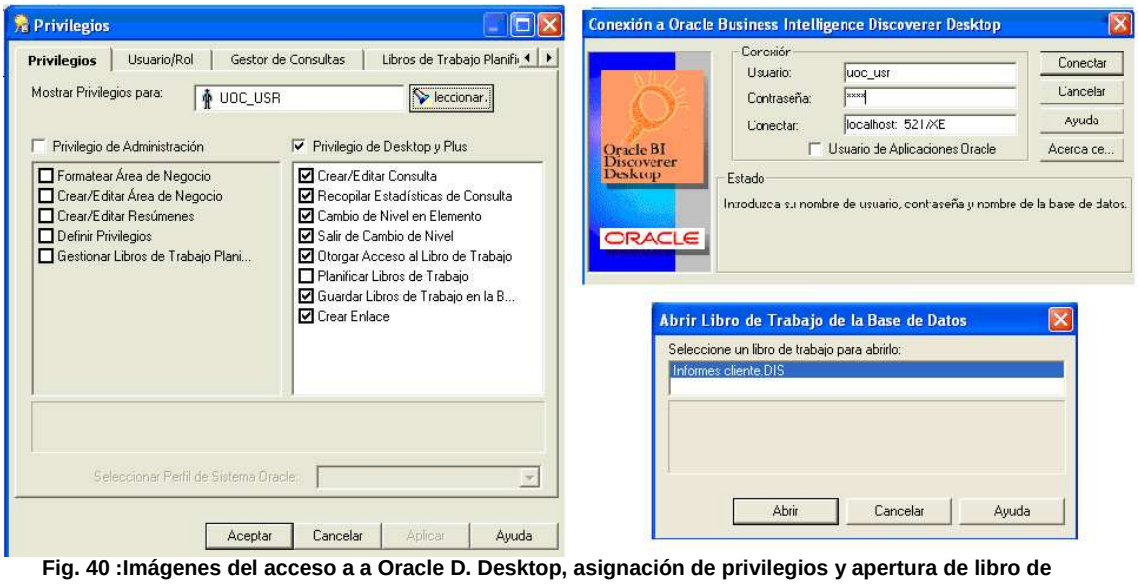

Fig. 40 :Imágenes del acceso a a Oracle D. Desktop, asignación de privilegios y apertura de libro de trabajo.

Con uoc_usr/1234 se podrá acceder al área de negocio a través de la herramienta Oracle Discoverer Desktop para el diseño y creación de informes.
El libro creado se ha guardado en la base de datos con el nombre de Infomes cliente.DIS.

Tabla de usuarios y contraseñas

System	uoc	Base de datos	Administrador base de datos
uoc_usr	1234	Base de datos y AN_Ventas	propietario esquema almacén datos y acceso al AN a través O.D. Desktop
eul_usr	1234	AN_Ventas	propietario AN y acceso a través de O.D. Administrato

Anexo E: "Entorno de desarrollo"

E.1 Hardware

Se dispone de un portátil en donde se llevarán a cabo todas las tareas. Las características de dicha unidad son:

TOSHIBA Satellite L750 2,3 GHz - 4Gb RAM – 320 Gb HD con Windows 7Home Premium

Sobre esta máquina se instala la máquina virtual proporcionada por el cliente.

Eventualmente, y únicamente para labores de documentación, se dispone de los puestos de una biblioteca pública cercana. Las características de dichos puestos son:

PACKARD BELL 733 MHz, 191 Mb RAM - 15 Gb HD con Windows XP

E.2 Software

- Para la redacción de la documentación Office Word, Office Excel y Office PowerPoint 2003.
- Para la elaboración de la planificación : OpenProj.
- De los productos Oracle se utilizarán : Oracle 10g para la base de datos y Oracle Discoverer para la explotación de la información y reportes.
- Además se también se utilizará: SQL Loader para la cargas de datos, y SQL Developer como entorno de programación que nos debe de permitir realizar todas las acciones sobre los datos.
- Camtasia Studio 7