

Memoria del proyecto

Construcción de un almacén de datos
para la empresa “Sostre per a tothom”

Maties Tramullas Mesquida

ETIG

Consultor

José Ángel Martín Carballo

8 de Junio de 2008

Para Luisa, sin tu apoyo y cariño nada de esto hubiera sido posible.

Para Matías y Ariadna, nuestro proyecto de vida...

1 Resumen

El proyecto consiste en el diseño e implementación de un almacén de datos para la empresa inmobiliaria “Sostre per a tothom”. La implementación se realizará sobre el producto Oracle Discoverer.

El resultado principal de proyecto es un almacén de datos junto a una serie de informes sobre dicho almacén. Los informes a implementar son:

1. Número de inmuebles en venta por zona, tipología de la vivienda y características de las mismas.
2. Listado de inmuebles que han bajado de precio. Mostrar la evolución de los precios por zona, tipología de la vivienda y características de las mismas
3. Precios máximos, mínimos i medios de las viviendas en venta por zona, tipología de la vivienda y características de las mismas.
4. Metros cuadrados máximos, mínimos y media por zona, tipología de la vivienda y características de las mismas.
5. Tiempo medio de venta de las viviendas por zona, tipología de la vivienda y características de las mismas.
6. Número de inmuebles vendidos y en venta por zona, tipología de la vivienda y características de las mismas.

Índice

1	Resumen	3
2	Introducción	7
2.1	Justificación y contexto.....	7
2.2	Objetivos.....	7
2.3	Enfoque y método seguido.....	8
2.4	Planificación del proyecto.....	9
2.5	Productos obtenidos	10
2.6	Contenido de los siguientes capítulos	10
3	Análisis	11
3.1	Diseño conceptual	11
3.1.1	Dimensiones	11
3.1.1.1	Zona	11
3.1.1.2	Tiempo.....	12
3.1.1.3	Número de habitaciones.....	12
3.1.1.4	Superficie	12
3.1.1.5	Tipo de vivienda	13
3.1.2	Hechos.....	14
3.1.3	Poner en venta	14
3.1.3.1	En Venta	15
3.1.4	Establecimiento de la granularidad de las dimensiones en las celdas.	16
3.1.5	Definición de las medidas.....	17
3.1.6	Definición de las celdas	17
3.1.6.1	Celdas para el Hecho Poner en Venta	18
3.1.6.2	Celdas para el Hecho Venta	19
3.1.7	Observaciones	20

3.2	Diseño técnico.....	21
3.2.1	Conformación.....	21
3.2.2	Modelo de datos.....	21
3.2.2.1	Diseño físico dimensión Tiempo	21
3.2.2.2	Diseño físico dimensión Superficie.....	22
3.2.2.3	Diseño físico dimensión Número habitaciones	22
3.2.2.4	Diseño físico dimensión Zona	23
3.2.2.5	Diseño físico dimensión Tipo Vivienda.....	23
3.2.2.6	Diseño físico de la tabla del hecho “En Venta”	25
3.2.2.7	Diseño físico de la tabla del hecho “Venta”	26
3.2.2.8	Identificador operacional.....	26
3.2.3	Diagramas de datos	27
3.2.3.1	Diagrama de datos para el hecho Inventa	27
3.2.3.2	Diagrama de datos para el hecho Venta	28
3.2.4	Descripción del proceso de carga.....	29
3.2.4.1	Proceso de integración del archivo de texto.	31
3.2.4.2	Descripción de las extracciones	35
3.2.4.3	Estrategia para determinar la vivienda tipo para España y Andorra.....	41

4 Implementación 42

4.1	Modelo de base de datos.....	42
-----	------------------------------	----

4.2	Proceso de importación.....	44
-----	-----------------------------	----

El punto de entrada se encuentra en el procedimiento LOAD.

	Recibe como único parámetro el nombre de archivo que contiene la información.....	45
--	--	----

4.3	Diseño del área de negocio.....	45
-----	---------------------------------	----

4.3.1	Datos.....	46
-------	------------	----

4.3.1.1	Carpetas básicas	46
---------	------------------------	----

4.3.1.2	Carpeta Ventas	47
---------	----------------------	----

4.3.1.3	Carpeta Información Cartera	48
---------	-----------------------------------	----

4.3.1.4	Carpeta Información Ventas	49
---------	----------------------------------	----

4.3.2	Jerarquías	50
4.3.2.1	Jerarquía Geográfica:	50
4.3.2.2	Jerarquía de Fechas	50
4.3.2.3	Jerarquía Superficies	50
4.4	Informes.....	51
4.4.1	Informe total de viviendas en venta.....	51
4.4.2	Informe de viviendas que han bajado de precio.....	53
4.4.3	Informe de Precios mínimos, medios y máximos	55
4.4.4	Informe de superficies mínimas, medias y máximas de las viviendas en venta	56
4.4.5	Informe de tiempos medios de venta	57
4.4.6	Informe de viviendas en venta y vendidas	59
4.4.7	Top N de precios medios.....	60
5	Conclusiones	63
6	Glosario	64
7	Bibliografía	66
8	Anexos	67
8.1	Tabla de ilustraciones.....	67

2 Introducción

2.1 *Justificación y contexto*

Se trata de implementar un pequeño almacén de datos así como un sencillo conjunto de informes sobre el mismo. Así mismo utilizando dicho almacén, se pide contestar a la pregunta de cual sería el piso típico español y andorrano.

2.2 *Objetivos*

Los objetivos principales a conseguir en este proyecto son:

1. Conocer qué son los almacenes de datos, cuales son sus características, como se diseñan y construyen para dar respuesta al negocio (tomando aquí negocio en un sentido amplio)
2. Aprender PL/SQL de Oracle para la implementación de la carga de datos.
3. Aprender algunas nociones de administración de Oracle.
4. Aprender la construcción de áreas de trabajo y a diseñar informes para el productor Oracle Discoverer.

2.3 Enfoque y método seguido

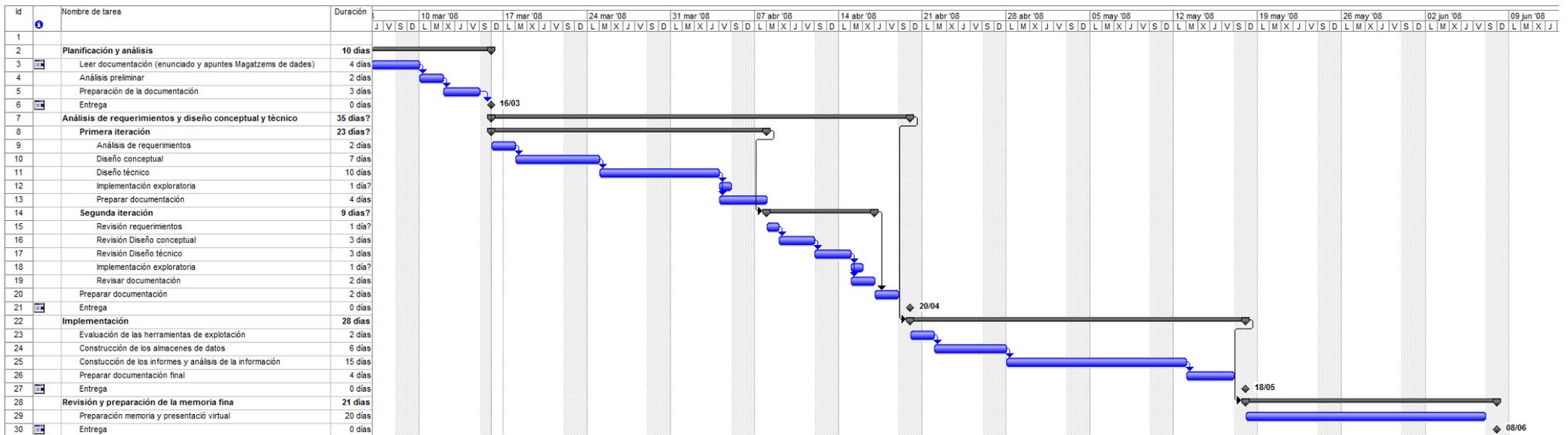
Como se detallará seguidamente, voy a seguir una estrategia basada en dos iteraciones en la fase de análisis y diseño añadiendo dos implementaciones exploratorias. La primera iteración, que como es obvio será la más exhaustiva de las dos preveo que consuma el 65% del tiempo mientras que la iteración siguiente consuma el tiempo restante (35%)

El objetivo para seguir esta estrategia es la de descubrir lo antes posible errores y/o requerimientos parciales para así concentrar el máximo trabajo en esta segunda etapa del TFC. La razón principal es la de adelantar el máximo trabajo evitando que a mitad de la implementación surjan inconsistencias o problemas en el diseño que provoquen, a posteriori, cambios costosos.

La tabla siguiente no pretende detallar todas y cada una de las tareas incluidas en la planificación, simplemente muestra aquellos puntos donde se requiera una explicación.

Id	Tarea	Duración	Fecha Inicio
1	Planificación y análisis inicial	10	6/03/2008
Etapa actual. Básicamente se trata de leer la documentación inicial, enunciado y apuntes auxiliares facilitados, para poder elegir la estrategia a seguir así como para poder elaborar la primera aproximación al análisis de requisitos.			
2	Análisis de requisitos y diseño (funcional y técnico)	35	17/03/2008
Esta es con diferencia la etapa más importante. Como he explicado anteriormente, creo que la mejor estrategia que puedo utilizar se basa en dos iteraciones con sendas fases exploratorias de implementación. Esto ayudará a que la tercera fase minimice los riesgos.			
2.1	Iteración 1	23	17/03/2008
Supone el 65% del tiempo de toda la fase. La implementación exploratoria de esta fase debe contemplar la importación de los datos así como la ejecución de algún informe de los propuestos (probablemente se realizará utilizando Microsoft Access)			
2.2	Iteración 2	9	08/04/2008
No debe suponer más del 35% del tiempo disponible para la realización de esta fase. En el caso que los requisitos y/o análisis no se vean muy afectados, la fase exploratoria de implementación debería encauzarse hacia la creación de las cargas de datos ya sobre Oracle XE.			
3	Implementación	28	20/04/2008
Implementación definitiva de las cargas de datos así como de los filtros de integridad y consistencia de los datos. Se seleccionará la herramienta a utilizar para el reporting y se generaran los informes definidos por el negocio.			
4	Revisión y preparación de la memoria y presentación virtual	21	18/05/2008

2.4 Planificación del proyecto



2.5 Productos obtenidos

El resultado obtenido por el proyecto está compuesto por:

- Esta memoria del proyecto. Contiene la planificación y el análisis lógico y técnico del mismo.
- Una estructura de base de datos para el almacén de datos.
- Unos procesos de carga para los archivos con la información del sistema operacional.
- Finalmente un conjunto de consultas implementadas sobre Oracle Discoverer

2.6 Contenido de los siguientes capítulos

Vamos a resumir brevemente el contenido de los siguientes capítulos:

El capítulo 3 contiene el análisis del proyecto, abarcando ambos niveles: lógico y técnico. En ellos se desgrana tanto la construcción de las estrellas necesarias para la implementación del proyecto como de los procesos para cargar los datos provenientes del sistema operacional del cliente.

El capítulo 4 contiene una descripción de la implementación del proyecto, detallando cada uno de los informes solicitados por el cliente.

3 Análisis

3.1 Diseño conceptual

A continuación diseñaremos las estrellas que van a modelizar la información que deseamos conocer. Para ello y como primer paso vamos a enumerar las dimensiones que vamos a utilizar así como sus atributos. Para estos últimos indicaremos cuales de ellos actúan como descriptores y cuales definen la jerarquía de agregación.

3.1.1 Dimensiones

3.1.1.1 Zona

Esta dimensión nos permite establecer la ubicación geográfica de la vivienda. Establece la siguiente estructura jerárquica:

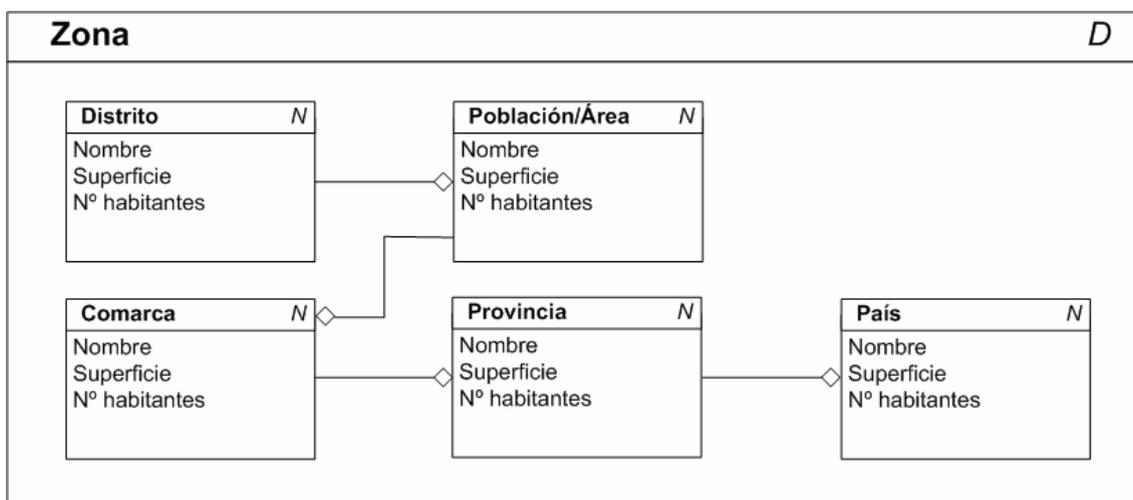


Ilustración 1. Dimensión Zona

Funcionalmente no haremos distinción entre el nivel jerárquico Población y área. En el caso que la ubicación de la vivienda dependa de un área en lugar de una población únicamente afectará en la exactitud de los valores superficie y número de habitantes. En el caso de área estos valores serán estimados

mientras que en el caso de población se basarán en valores geográficos exactos y en un censo oficial.

3.1.1.2 Tiempo

Se trata de una dimensión estándar utilizada prácticamente en cualquier diseño. En nuestro caso restringiremos la jerarquía a los niveles que realmente nos interesen.

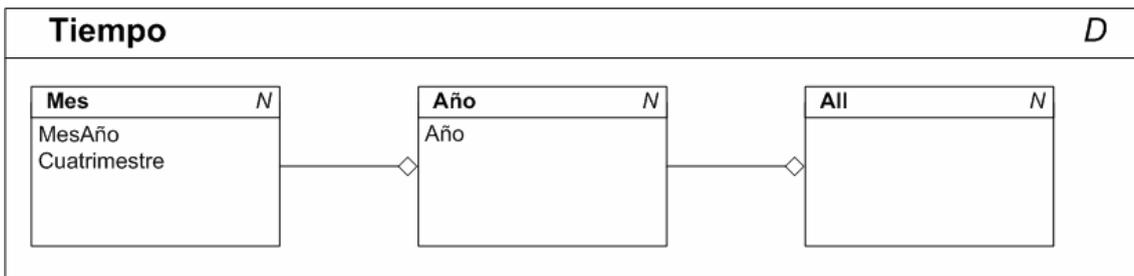


Ilustración 2. Dimensión Tiempo

3.1.1.3 Número de habitaciones

Establece un valor entero positivo mayor que 0. Estableceremos como valor máximo de esta dimensión el valor 10. Al igual que en el resto de dimensiones, a parte de los valores 1..10 dispondremos del valor "All" que representa a todas las anteriores de forma consolidada.

3.1.1.4 Superficie

Para el dominio del valor de superficie estableceremos un recubrimiento de n intervalos de igual tamaño, de tal forma que a la unión de los intervalos contenga el dominio de la dimensión. Además la intersección de cada intervalo con cualquiera de los demás es vacía. A este recubrimiento lo llamaremos **Intervalos**. Finalmente dispondremos de un último nivel "All".

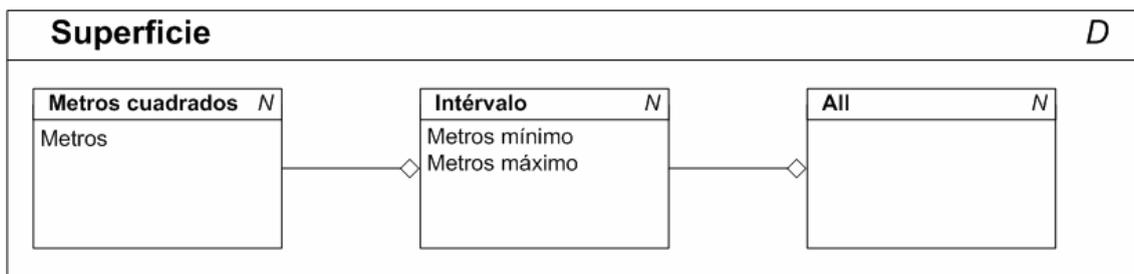


Ilustración 3. Dimensión Superficie

3.1.1.5 Tipo de vivienda

Establece una clasificación de la vivienda según una serie de aspectos morfológicos de la misma. Esta dimensión permite varios diseños jerárquicos. El siguiente es sólo uno de los posibles en función de las características que he considerado más relevantes a la hora de presentar la información en las diversas tamaños de grano.

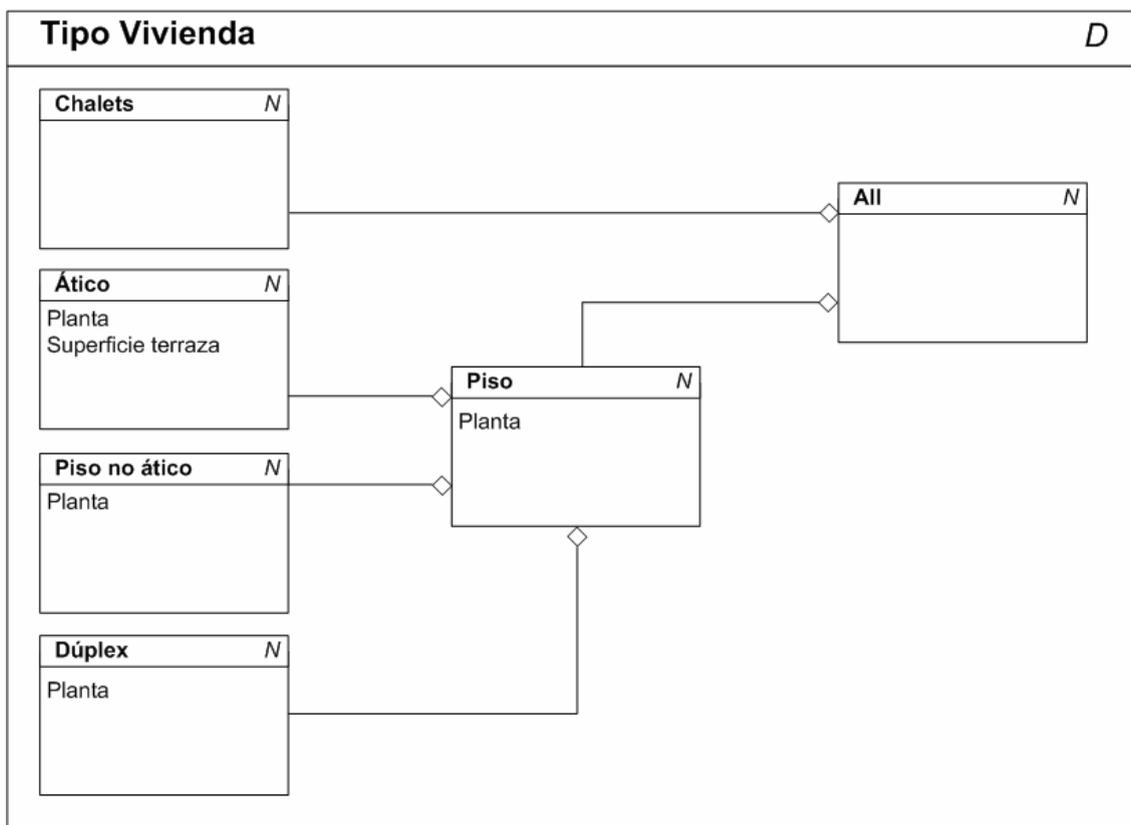


Ilustración 4. Dimensión Tipo Vivienda

3.1.2 Hechos

Cara a determinar los *Hechos*¹ para los cuales diseñaremos las estrellas, inicialmente enumeraremos los procesos de negocio relacionados con ellos.

Analizando la información proporcionada por el cliente, vemos que esta obedece básicamente a dos procesos de negocio:

1. **Poner en venta:** Disponemos de la información referida a las viviendas junto a sus características físicas principales así como de su precio. Hay que tener en cuenta que así como las características físicas son inmutables en el tiempo, el precio puede variar, es decir, puede sufrir incrementos y/o decrementos para así adaptarse a la evolución del mercado y/o incentivar su venta.
2. **Ventas:** Disponemos de la información sobre las viviendas que se han vendido y por que precio. Así mismo disponemos de las características de cada vivienda que ha sido vendida así como la fecha en que esta se ha producido.

3.1.3 Poner en venta

Veamos cual sería su diagrama de estrella.

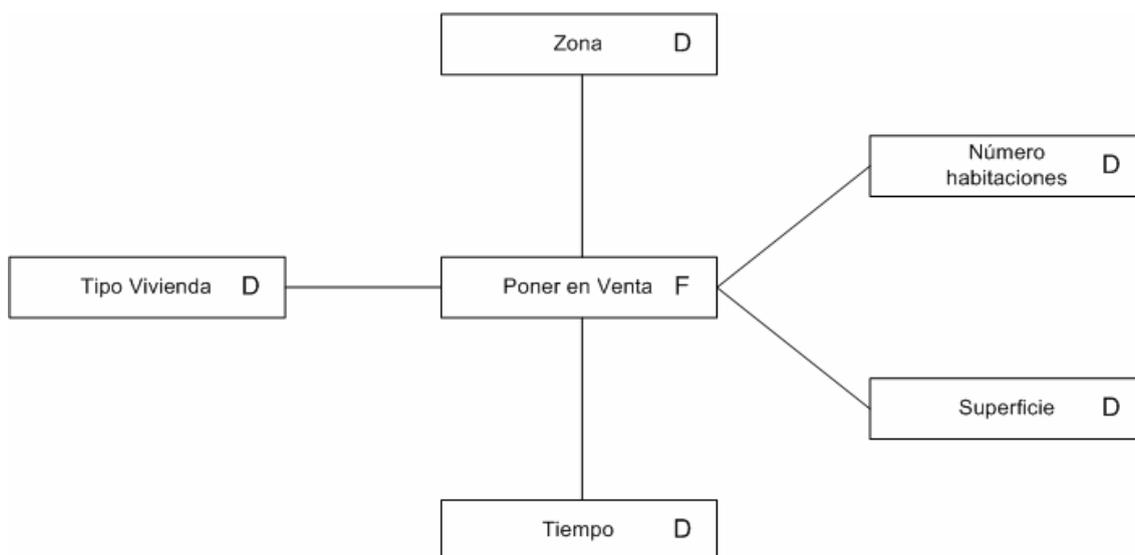


Ilustración 5. Hecho En Venta

¹ Fact en Inglés

3.1.3.1 En Venta

Este hecho se refiere a la acción de vender un inmueble. En los datos proporcionados por el cliente se identifica por que el precio de la vivienda desaparece en el mes siguiente a su venta.

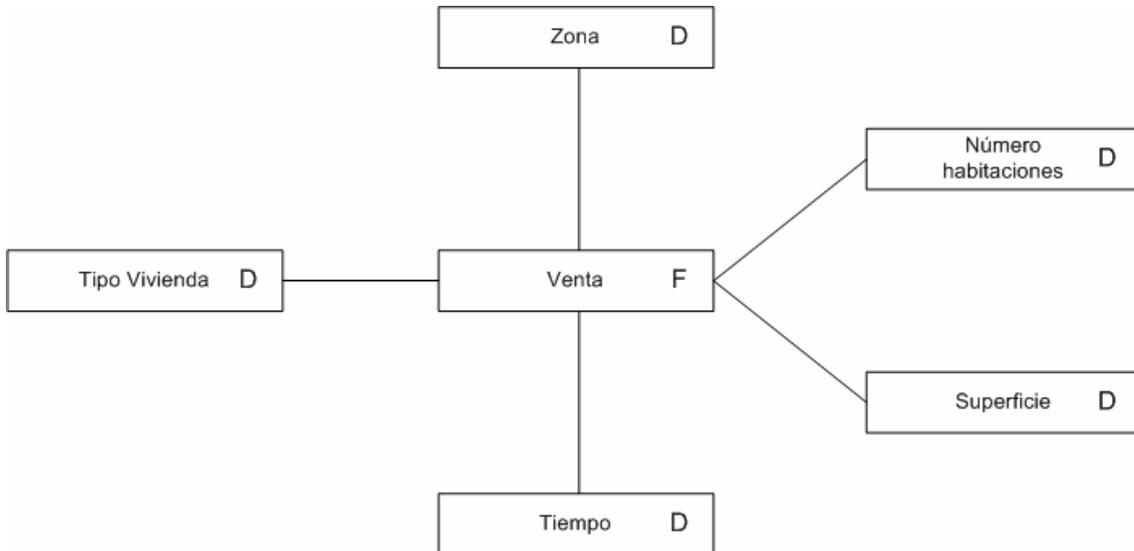


Ilustración 6. Hecho Venta

Como medida principal vamos a tomar el precio de venta definitivo.

3.1.4 Establecimiento de la granularidad de las dimensiones en las celdas.

La norma general indica que debemos situar la granularidad de la información en la unidad atómica, es decir, aquella que ya no podamos descomponer. En nuestro caso, al disponer de una única fuente de datos la granularidad va a quedar de la forma siguiente:

- *Dimensión Tiempo*

La información temporal existente en los datos provenientes del sistema operacional tiene está a nivel de Mes/Año, entonces esta será la granularidad de la dimensión.

- *Dimensión Tipo Vivienda*

El nivel atómico corresponde a los valores: Ático, Pisos no ático, Chalet y Dúplex.

- *Dimensión Zona*

El nivel atómico corresponde a Distrito y es por tanto el nivel que utilizaremos,

- *Dimensión Número de habitaciones*

Utilizaremos el entero positivo 1..10 como valor atómico en esta dimensión.

- *Dimensión Superficie*

Para implementar los informes demandados por el cliente no creo que fuera necesario de tener los metros cuadrados como nivel atómico, pero muy probablemente el cliente, en un futuro, desee complementar los informes con otros que precisen de este nivel, por ejemplo tamaño medio de las viviendas según la zona, número de habitaciones y tipo de vivienda, o precio medio del metro cuadrado por zona, tipo de vivienda, mes, y número de habitaciones...

3.1.5 Definición de las medidas

Ahora es el momento de especificar las medidas disponibles para cada Hecho.

- **Hecho En Venta:**
 - *Precio vivienda:* Es la cantidad monetaria exigida por la venta de la vivienda. Supondremos que todas las cantidades se encuentran expresadas en la misma moneda.
 - *Precio metro cuadrado:* Al conocer tanto el precio como la superficie podemos calcular este valor.
 - *Variación de precio:* Es la cantidad, positiva o negativa según se trate de un incremento o decremento del precio respectivamente.
- **Hecho Venta:**
 - *Precio inicial:* Precio por el que salió a la venta la vivienda.
 - *Precio venta:* Es la cantidad final por la que se vendió la vivienda.
 - *Tiempo en venta:* Valor entero positivo indicando cuantos meses se pasó la vivienda en venta.

3.1.6 Definición de las celdas

Seguidamente se procederá a definir las celdas teniendo en cuenta la granularidad de las mismas.

3.1.6.1 Celdas para el Hecho Poner en Venta

Para este hecho tenemos una única celda con tres medidas definidas en el apartado anterior. Todas las dimensiones participan en su nivel atómico.

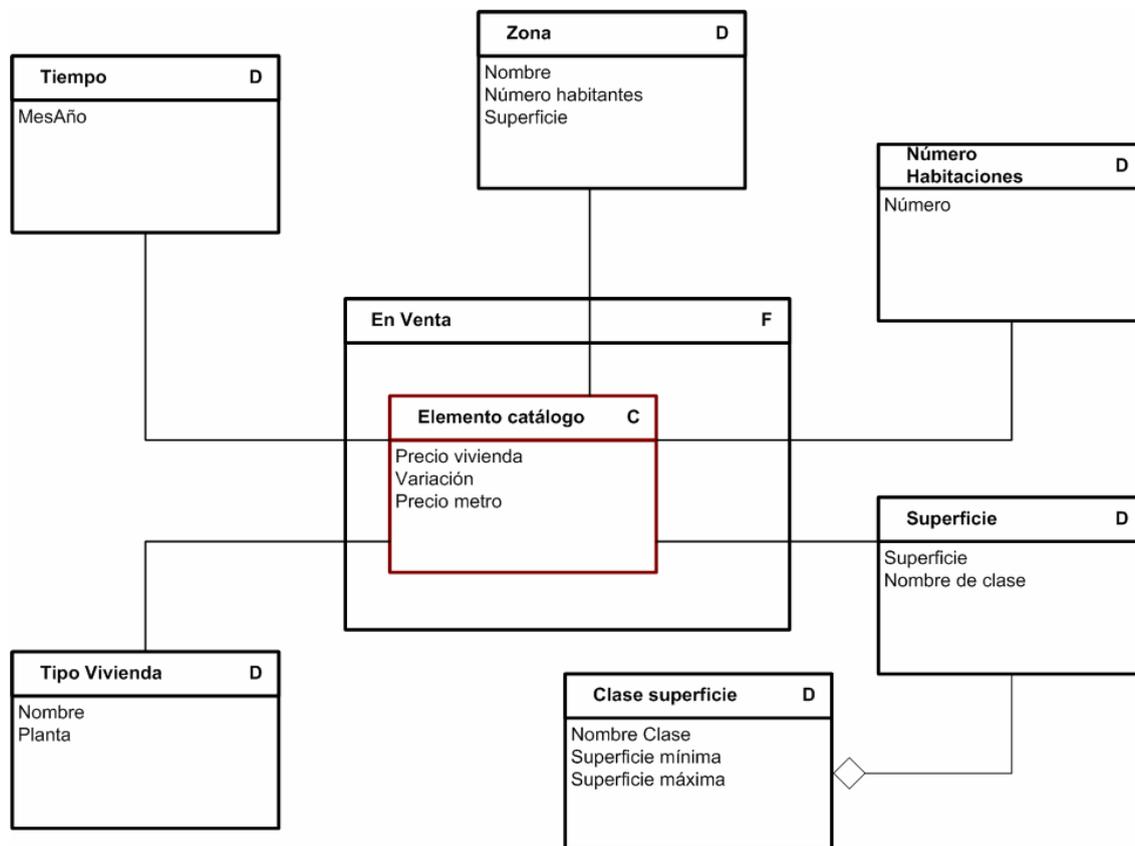


Ilustración 7. Diseño de celda para el Hecho Poner en Venta

Según vemos en el diagrama anterior vemos que el precio de venta es una medida que no es aditiva para ninguna dimensión, lo que quiere decir que en el momento de realizar los informes deberemos realizar cálculos del tipo AVG. Lo mismo sucede con el precio del metro cuadrado. En cambio la medida variación sí es aditiva en la dimensión Tiempo, lo cual nos permite calcular la variación de precio acumulada a medida que pasan los meses.

Para cada entrada en la tabla del hecho se cumple que, manteniendo las otras dimensiones constantes y variando únicamente la temporal, que $precio_vivienda_i = precio_vivienda_{i+1} + variación_{i+1}$.

Aunque el precio metro cuadrado se pueda calcular fácilmente a partir de dos de las medidas presentes a nivel atómico, Precio Vivienda y superficie, proporcionaremos una medida con este valor calculado.

También hemos de tener en cuenta que cuando contabilicemos el número de viviendas en venta, esta medida es semiaditiva ya que no es aditiva en la dimensión temporal.

3.1.6.2 Celdas para el Hecho Venta

De nuevo nos es suficiente con una celda para describir las medidas para este hecho. Todas las dimensiones participan en su nivel atómico.

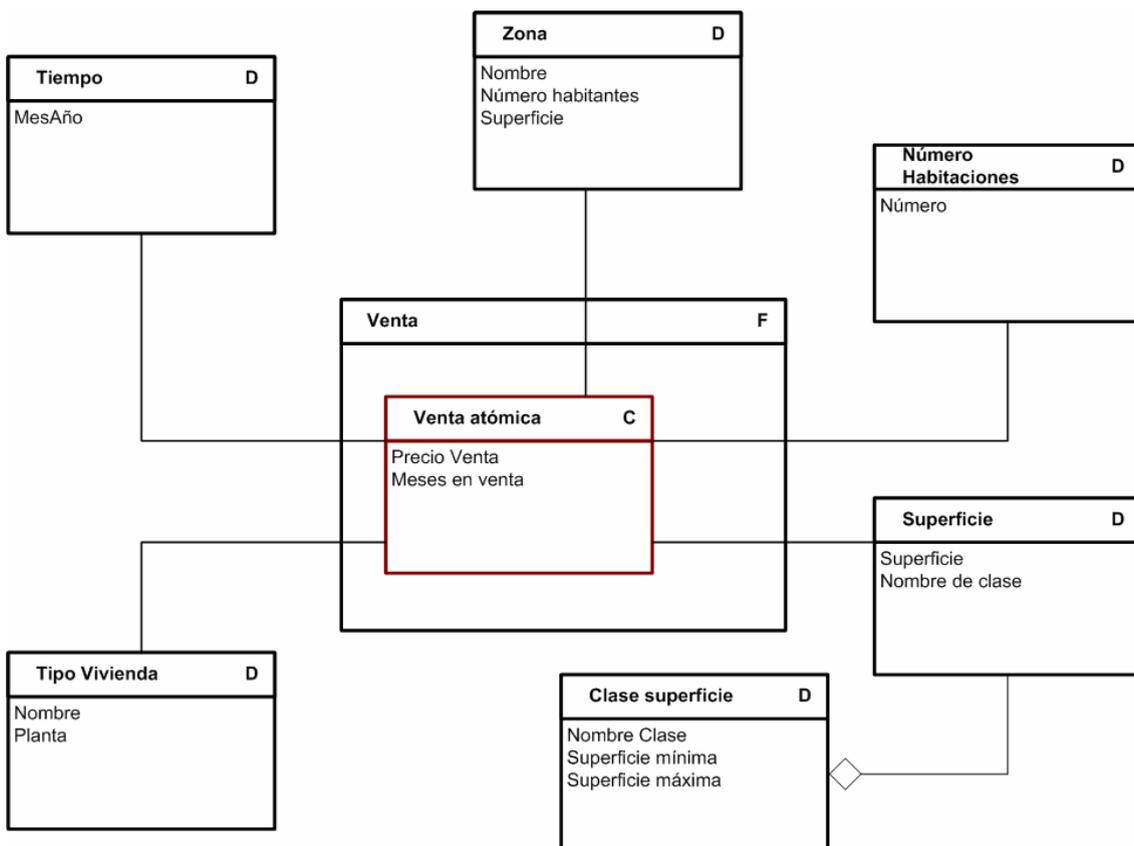


Ilustración 8. Diseño de celda para el Hecho Venta

En este caso la medida Precio de venta es aditiva para todas las dimensiones. La medida meses en venta en cambio es no aditiva. El número de viviendas vendidas es semiaditiva ya que no es aditiva en la dimensión temporal.

3.1.7 Observaciones

Las dos tablas de hechos participan de las mismas dimensiones y a un mismo nivel de granularidad, esto hace posible cruzarlas muy fácilmente lo cual será muy importante para implementar el informe etiquetado con el nº6 y cuyo enunciado es: “Número de inmuebles vendidos y en venta por zona, tipología, y características”

3.2 Diseño técnico

Ha llegado el momento de proponer una implementación para el modelo conceptual resultante de la etapa anterior. Los objetivos a alcanzar en esta etapa son:

- Diseñar la implementación física del almacén de datos, definiendo las tablas, atributos, claves primarias, etc.
- Modelizar los procesos de carga de los datos en el almacén.

3.2.1 Conformación

Los dos hechos a implementar comparten todas las dimensiones al mismo nivel de granularidad, esto permite que las compartan y podamos realizar operaciones “Drill-across” entre ellas (de echo hay un informe que compara el número de viviendas en venta y vendidas que no podría realizarse sin que se diera esta conformidad entre las dimensiones de las dos estrellas)

3.2.2 Modelo de datos

Vamos a iniciar esta etapa estudiando cómo vamos a implementar cada dimensión.

3.2.2.1 Diseño físico dimensión Tiempo

Crearemos una tabla con la estructura siguiente:

Tiempo	
PK	<u>TiempoPk</u>
	MesAño Cuatrimestre

Ilustración 9. Tabla dimensión Tiempo

Proponemos el uso de una clave alternativa para el tiempo para independizar la estructura de cualquier cambio en el formato en la misma.

3.2.2.2 Diseño físico dimensión Superficie

La tabla para la dimensión superficie tendrá la estructura siguiente:

Superficie	
PK	<u>SuperficiePk</u>
	Metros Intervalo

Ilustración 10. Tabla dimensión Superficie

Donde Metros es un valor entero positivo indicando la superficie en metros cuadrados, mientras que intervalo es una cadena de caracteres indicando el intervalo al que pertenece de los posibles:

[0,40), [40,80), [80, 120), [120, 160), [160, 200), [200, 240), [240, 280),
[280, 320), [320, 360], [360, 400), [400, ∞)

3.2.2.3 Diseño físico dimensión Número habitaciones

La tabla de dimensión Número de habitaciones contiene como único atributo un valor entero positivo entre 1..50

Habitaciones	
PK	<u>NumeroHabitaciones</u>

Ilustración 11. Tabla dimensión Habitaciones

3.2.2.4 Diseño físico dimensión Zona

Esta dimensión es la que contiene más atributos de tipo jerárquico. Esta es su estructura

Zona	
PK	<u>ZonaPk</u>
	Distrito Población Comarca Provincia País Superficie Habitantes

Ilustración 12. Tabla dimensión Zona

Los atributos Distrito, Población, Comarca, Provincia y País son de tipo cadena de caracteres. Superficie y habitantes son valores enteros. Estos dos últimos valores no son proporcionados por el sistema operacional sino que deben ser calculados desde información estadística proporcionada por algún organismo oficial.

3.2.2.5 Diseño físico dimensión Tipo Vivienda

Esta dimensión ofrece posibilidades para utilizar un patrón Especialización para su diseño físico. Vamos a estudiar esta posibilidad

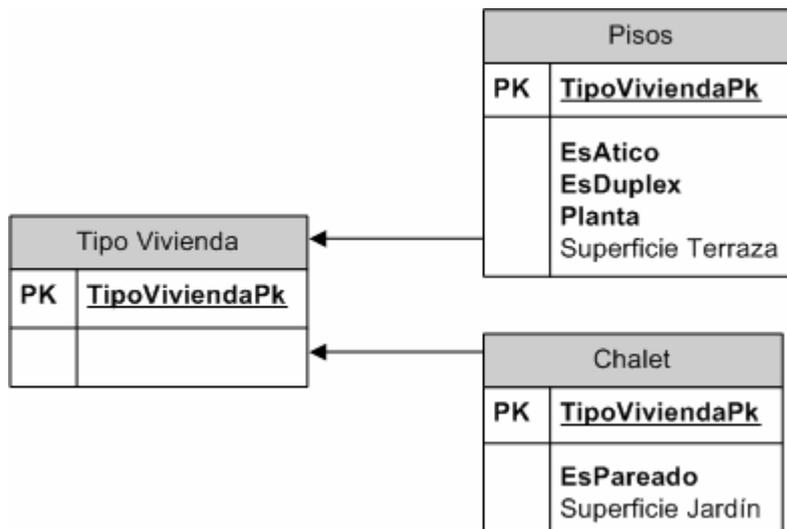


Ilustración 13. Tabla dimensión Tipo vivienda con subclasificación

En nuestro caso la superclase no tiene atributos propios, así que debemos decidir entre implementar las subclases de la dimensión o bien integrarlas en una única tabla. En nuestro caso, “Sostre per a tothom” no provee de los datos de superficie de jardín o terraza, por lo que la subclasificación únicamente contiene campos lógicos indicando el tipo de vivienda, y únicamente el tipo chalet no puede indicar el número de planta. Creo que la pérdida de espacio fruto de este último atributo no es suficientemente importante para complicar el esquema en más de una tabla. Los campos lógicos los vamos a substituir por un atributo enumerado que tome uno de los valores siguientes: “Ático”, “Dúplex”, “Pisos” y “Chalet”. Para este último tipo el valor de planta será establecido a 0 por defecto.

Una vez establecida la estrategia para esta dimensión, veamos como queda la implementación de la tabla

Tipo Vivienda	
PK	<u>TipoViviendaPk</u>
	Tipo Planta

Ilustración 14. Tabla dimensión Tipo Vivienda definitiva

En caso que “Sostre per a tothom” ampliara los atributos propios de cada tipo, la estrategia debería ser reevaluada.

3.2.2.6 Diseño físico de la tabla del hecho “En Venta”

La tabla del hecho va a contener como parte de su clave primaria una clave compuesta por las claves primarias de las dimensiones que participan de su estrella. Además dispondrá de los atributos siguientes: **precio venta** de tipo entero positivo, **variación** de tipo entero positivo o negativo el cual indica la cantidad en que se ha incrementado o reducido respectivamente el precio de venta desde el mes anterior, y finalmente **precio metro**, de tipo numérico que indica el precio por metro cuadrado de la vivienda.

EnVenta	
PK	<u>ZonaPk</u>
PK	<u>TipoViviendaPk</u>
PK	<u>TiempoPk</u>
PK	<u>SuperficiePk</u>
PK	<u>NumeroHabitaciones</u>
	PrecioVenta
	Variacion
	PrecioMetro

Ilustración 15. Tabla hecho En Venta

Restricciones básicas

- El precio siempre es mayor o igual a 0.
- Variación puede ser positivo o negativo. En caso que sea positivo no estableceremos inicialmente ningún límite a ese valor. En caso que sea negativo, restringiremos su valor absoluto a que sea menor o igual a PrecioVenta.
- PrecioMetro será un entero positivo cuyo valor debe ser siempre menor o igual a precioVenta.

3.2.2.7 Diseño físico de la tabla del hecho “Venta”

De nuevo la tabla Venta contendrá como identificador la clave compuesta por las claves primarias de las dimensiones que participan de su estrella, que son exactamente las mismas que en el hecho anterior. Como atributos exclusivos contendrá los siguientes: **Precio** de tipo entero positivo que indica el precio por el que se vendió la vivienda y **TiempoEnVenta**, de tipo entero corto positivo indicando el número de meses que la vivienda estuvo en venta antes de que esta se produjera.

Venta	
PK	<u>ZonaPk</u>
PK	<u>TipoViviendaPk</u>
PK	<u>TiempoPk</u>
PK	<u>SuperficiePk</u>
PK	<u>NumeroHabitaciones</u>
	Precio
	TiempoEnVenta

Ilustración 16. Tabla hecho Venta

Restricciones básicas

- Precio es siempre mayor o igual a 0.
- Tiempo en venta es siempre mayor o igual a 0.

3.2.2.8 Identificador operacional

Junto con los datos que describen las viviendas en venta, disponemos de un valor que identifica la vivienda de forma unívoca en el sistema operacional. Este valor será incorporado como un elemento más de la tabla del hecho, obviamente, sin formar parte de la clave primaria. La consideraremos una dimensión degenerada

3.2.3 Diagramas de datos

3.2.3.1 Diagrama de datos para el hecho Inventa

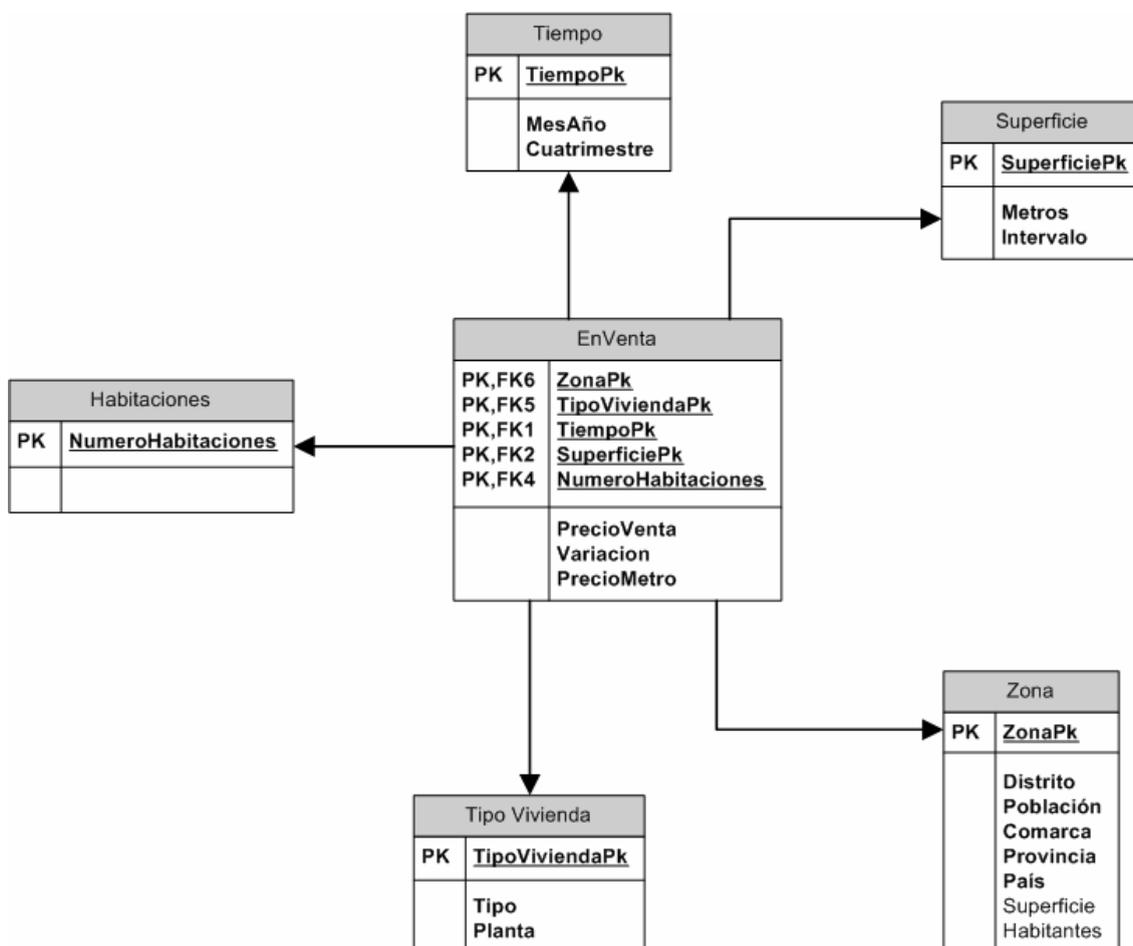


Ilustración 17. Diagrama de datos para el hecho Inventa

3.2.3.2 Diagrama de datos para el hecho Venta

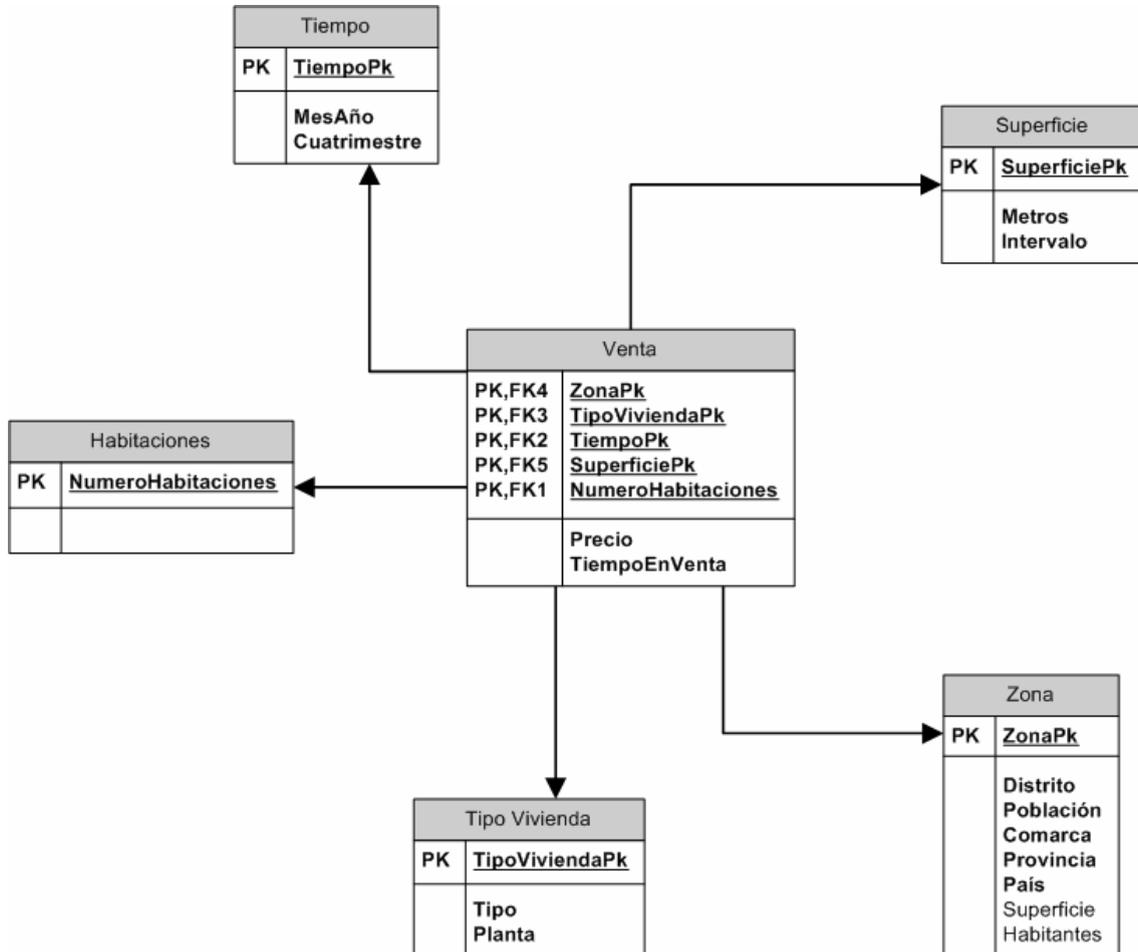


Ilustración 18. Diagrama de datos para el hecho Venta

3.2.4 Descripción del proceso de carga

Como se ha descrito anteriormente, el cliente “Sostre per a tothom” dispone de un archivo de texto, fruto de la unión de la información existente en varias hojas de cálculo, como fuente de datos para la creación del almacén de datos. El procedimiento utilizado para la consolidación de los datos de las diversas hojas de cálculo es manual y por lo tanto inherente a la presencia de errores.

Vamos a subdividir la carga en tres etapas siguiendo el diagrama siguiente:

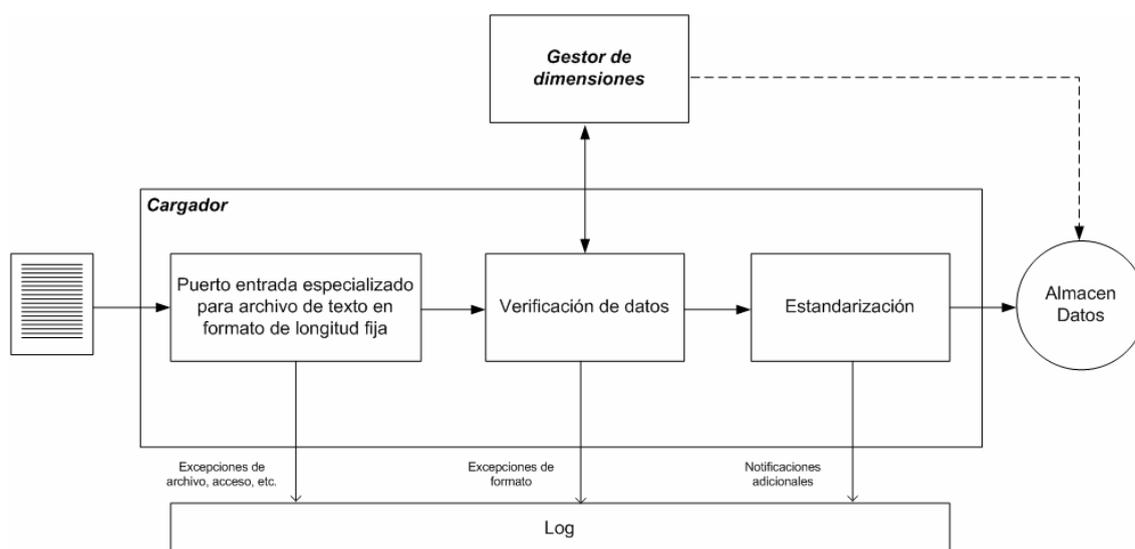


Ilustración 19. Proceso del cargador de datos

El cargador debe trabajar mano con mano con el gestor de dimensiones. Este gestor será el encargado de crear las dimensiones básicas, por ejemplo la dimensión de tiempo, número de habitaciones, superficie, ... Hay pero otras dimensiones que son más dinámicas. Este es el caso de la dimensión Zona. Aunque la mejor estrategia es siempre proveer de todos elementos que conforman la dimensión, en este caso ofreceremos la posibilidad de ir ampliándola a medida que sea necesario (se pueden ir incorporando nuevas zonas) debido a que la fuente de datos para construir la dimensión es exactamente la misma que para rellenar las tablas de hechos.

Aprovecharemos la fase de verificación de datos para, en caso necesario, crear la nueva entrada en la tabla de dimensión, en cuyo caso también escribiremos una anotación en el archivo de log advirtiendo de esta operación.

En lo concerniente al proceso cargador, vamos a distinguir tres partes completamente diferenciadas:

- Por una parte nos encontramos con la tarea responsable de leer el archivo de texto teniendo en cuenta la estructura procedente del sistema operacional, como resultado obtendremos la carga de una serie de tablas que conformaran nuestro sistema ETL². Nuestro objetivo fundamental es obtener el mínimo acoplamiento entre todas las partes, así pues si la estructura del archivo de texto varía (o incluso la forma en que se nos proporciona la información) sólo afectará a este proceso.
- La siguiente tarea será la de realizar una serie de verificaciones sobre los datos (campos requeridos, coherencia de datos) y tal como hemos indicado anteriormente también es parte de la responsabilidad de esta tarea la de verificar las referencias (y en caso que estas no existan actuar en consecuencia, como por ejemplo creándolas delegando en el *gestor de dimensiones*).
- Para finalizar, se procederá a realizar tareas de estandarización de la información. Esta tarea es opcional, depende únicamente de los requerimientos que impongamos sobre los datos. Entre las responsabilidades de este componente se encuentra el cálculo de campos derivados de otros (por ejemplo precio por metro cuadrado a partir de la superficie y del precio de la vivienda, pero también la de que los nombres de zona siempre se guarden en mayúsculas, etc.

² Extraer, Transformar y Cargar (Load en inglés). Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio (Wikipedia)

3.2.4.1 Proceso de integración del archivo de texto.

A continuación presentamos los procedimientos abstractos para integrar el archivo de texto que sirve de fuente de datos.

Proceso CargarArchivo(cadena nombreArchivo)

Cadena datos = ""

Inicio

Si (nombreArchivo <> "")

Prueba

datos = LeerArchivo(nombreArhivo);

Si (datos <> "")

ProcesaDatos(datos);

Sino

EscribeLog("No había información en el archivo.");

FinSi

Excepción

EscribeLog("Hubo un error tratando de leer los datos. Compruebe
Que tiene permiso para realizar la operación.")

FinPrueba

Sino

EscribeLog("Nombre de archivo incorrecto o inexistente.");

FinSi

FinProceso

Proceso ProcesaDatos(cadena datos)

Cadena datosEnProceso = datos

Cadena[] campos

Cadena zonaActual

Cadena datosViviendas

Inicio

campos = ProcesarCabecera(datosEnProceso)

Si (ValidarCampos(campos))

FinalProceso = EvaluarFinalProceso(datosEnProceso)

Mientras (No FinalProceso)

zonaActual = ExtraerZona(datosEnProceso)

datosViviendas = ExtraerViviendas(datosEnProceso)

Si (viviendas <> "")

ProcesarViviendas(datosViviendas, zonaActual, campos)

FinSi

FinMientras

Sino

EscribeLog("Hay una discrepancia entre los campos esperados y los
recuperados")

FinSi

FinProceso

Análisis

Proceso **ProcesarCabecera(cadena datos) : Cadena[]**

```
Bool FinalizarProceso = falso
Cadena campo = ""
Cadena[] resultado
```

Inicio

```
Mientras (No FinalizarProceso)
    FinalizarProceso = EvaluarFinalizarProceso(datos)
    campo = ExtraerCampoCabecera(datos)
    Añadir(campo, resultado)
FinMientras
ProcesarCabecera = resultado
```

FinProceso

Proceso **ExtraerCampoCabecera(cadena datos) : Cadena**

```
Bool FinalizarProceso = falso
Cadena delimitador = " "
Cadena campo = ""
```

Inicio

```
Mientras (No FinalizarProceso)
    FinalizarProceso = EvaluarFinalizarProceso(datos)
    Si (datos[0] <> delimitador)
        Si (datos[0] == "\"")
            delimitador = "\""
        FinSi
        campo = campo + datos[0]
    FinSi
    datos = Subcadena(datos, 1)
FinMientras
ExtraerCampoCabecera = campo
```

FinProceso

El proceso ValidarCampos simplemente debe comparar los campos recibidos con los que espera recibir.

Proceso ExtraerZona(cadena datos) : Cadena

```
Bool FinalizarProceso = falso
Cadena localizador = "#"
Cadena separador = "::::"
Entero contador = 0
Cadena zona = ""
```

Inicio

Mientras (No FinalizarProceso)

```
FinalizarProceso = EvaluarFinalizarProceso(datos)
// Buscamos la marca de principio de definición de zona
Si (datos[0] <> localizador)
datos = Subcadena(datos, 1)
FinSi
Mientras (contador < 3)
  Si (Subcadena(datos, 3) == separador)
    Contador = contador + 1
    datos = Subcadena(datos, 3)
    zona = zona + separador
  Sino
    zona = zona + datos[0]
    datos = Subcadena(datos,1)
  FinSi
```

FinMientras

// Recuperar última parte

```
Mientras (datos[0] <> " " Y datos[0] <> localizador)
  zona = zona + datos[0]
  datos = Subcadena[datos, 1]
```

FinMientras

FinMientras

ExtraerZona = zona

FinProceso

Proceso ExtraerViviendas(cadena datos) : Cadena

```
Bool FinalizarProceso = falso
Cadena localizador = "#"
Cadena datosViviendas = ""
```

Inicio

Mientras (No FinalizarProceso)

```
FinalizarProceso = EvaluarFinalizarProceso(datos)
Si (cadena[0] <> localizador Y NO EOF(cadena))
  datosViviendas = datosViviendas + datos[0]
  datos = SubCadena(datos,1)
```

FinSi

FinMientras

ExtraerViviendas = datosViviendas

FinProceso

Análisis

Proceso ProcesarViviendas(Cadena datos, Cadena zona, Cadena[] campos) : Cadena[]

```
    Bool FinalizarProceso = falso
    Cadena separador = " " // espacio en blanco
    Cadena noHayDato = "\t" // tabulador
    Cadena camposVivenda[]
    Cadena campo = ""
```

Inicio

```
    Para (entero i=0, i<Len(campos), i++)
        // inicializamos el campo
        campo = ""
        Mientras (No FinalizarProceso)
            FinalizarProceso = EvaluarFinalizarProceso(datos)
            Si (datos[0] <> separador Y datos[0] <> noHayDato)
                campo = campo + datos[0]
            FinSi
            datos = SubCadena(datos, 1)
        FinMientras
        Añadir(campo, camposVivenda)
    FinPara
    // Guardar la información procesada
    SalvarInformacion(zona, campos, camposVivenda);
```

FinProceso

3.2.4.2 Descripción de las extracciones

Finalmente vamos a describir a alto nivel y utilizando sintaxis SQL'99 como se recuperaría la información necesaria para la implementación de los informes solicitados.

1. Número de inmuebles en venta por zona, tipología de la vivienda y características de las mismas.

```
SELECT D1.Distrito,
       D2.Tipo,
       D3.MesAño,
       D4.Metros,
       D5.NumeroHabitaciones,
       COUNT(*) NumeroViviendas
FROM   EnVenta H,
       Zona D1,
       TipoVivienda D2,
       Tiempo D3,
       Superficie D4,
       Habitaciones D5
WHERE  H.ZonaPk = D1.ZonaPk
      AND H.TipoViviendaPk = D2.TipoViviendaPk
      AND H.TiempoPk = D3.TiempoPk
      AND H.SuperficiePk = D4.SuperficiePk
      AND H.Habitaciones = D5.Habitaciones
GROUP BY CUBE( D1.Distrito,
              D2.Tipo,
              D4.Metros,
              D5.NumeroHabitaciones )
          D3.MesAño,
ORDER BY D1.Distrito,
        D2.Tipo,
        D3.MesAño,
        D4.Metros,
        D5.NumeroHabitaciones
```

En el cubo no tenemos en cuenta la dimensión tiempo debido que en este contexto es la única dimensión no aditiva.

2. Listado de inmuebles que han bajado de precio. Mostrar la evolución de los precios por zona, tipología de la vivienda y características de las mismas

```
SELECT D1.Distrito,
       D2.Tipo,
       D3.MesAño,
       D4.Metros,
       D5.NumeroHabitaciones,
       AVG(H.PRECIO) PRECIO_MEDIO_FINAL,
       AVG(H.PRECIO_METRO) PRECIO_MEDIO_METRO_FINAL,
FROM   EnVenta H,
       Zona D1,
       TipoVivienda D2,
       Tiempo D3,
       Superficie D4,
       Habitaciones D5
WHERE  H.ZonaPk = D1.ZonaPk
      AND H.Variacion < 0
      AND H.TipoViviendaPk = D2.TipoViviendaPk
      AND H.TiempoPk = D3.TiempoPk
      AND H.SuperficiePk = D4.SuperficiePk
      AND H.Habitaciones = D5.Habitaciones
GROUP BY CUBE( D1.Distrito,
              D2.Tipo,
              D4.Metros,
              D5.NumeroHabitaciones )
          D3.MesAño
ORDER BY  D1.Distrito,
          D2.Tipo,
          D3.MesAño,
          D4.Metros,
          D5.NumeroHabitaciones
```

En este informe, el dominio lo forman todas aquellas viviendas que han bajado de precio. Aunque en este caso podríamos incluir la dimensión Tiempo dentro de la función CUBE, al ser requerimiento ver la evolución, debemos considerarla siempre.

3. Precios máximos, mínimos i medios de las viviendas en venta por zona, tipología de la vivienda y características de las mismas.

```
SELECT D1.Distrito,
        D2.Tipo,
        D3.MesAño,
        D4.Metros,
        D5.NumeroHabitaciones,
        MIN(H.PRECIOVENTA) MIN_PRECIO_VENTA,
        AVG(H.PRECIOVENTA) AVG_PRECIO_VENTA,
        MAX(H.PRECIOVENTA) MAX_PRECIO_VENTA,
FROM EnVenta H,
        Zona D1,
        TipoVivienda D2,
        Tiempo D3,
        Superficie D4,
        Habitaciones D5
WHERE H.ZonaPk = D1.ZonaPk
AND H.TipoViviendaPk = D2.TipoViviendaPk
AND H.TiempoPk = D3.TiempoPk
AND H.SuperficiePk = D4.SuperficiePk
AND H.Habitaciones = D5.Habitaciones
GROUP BY    CUBE( D1.Distrito,
                    D2.Tipo,
                    D4.Metros,
                    D5.NumeroHabitaciones )
                    D3.MesAño,
ORDER BY    D1.Distrito,
                D2.Tipo,
                D3.MesAño,
                D4.Metros,
                D5.NumeroHabitaciones
```

De nuevo la dimensión Tiempo queda fuera de la función CUBE debido a que no es aditiva y proporcionaría información falsa.

4. Metros cuadrados máximos, mínimos y media por zona, tipología de la vivienda y características de las mismas.

```
SELECT D1.Distrito,
        D2.Tipo,
        D3.MesAño,
        D4.Metros,
        D5.NumeroHabitaciones,
        MIN(D4.METROS) MIN_SUPERFICIE,
        AVG(D4.METROS) AVG_SUPERFICIE,
        MAX(D4.METROS) MAX_SUPERFICIE,
FROM EnVenta H,
        Zona D1,
        TipoVivienda D2,
        Tiempo D3,
        Superficie D4,
        Habitaciones D5
WHERE H.ZonaPk = D1.ZonaPk
AND H.TipoViviendaPk = D2.TipoViviendaPk
AND H.TiempoPk = D3.TiempoPk
AND H.SuperficiePk = D4.SuperficiePk
AND H.Habitaciones = D5.Habitaciones
GROUP BY CUBE( D1.Distrito,
                D2.Tipo,
                D4.Metros,
                D5.NumeroHabitaciones )
                D3.MesAño,
ORDER BY D1.Distrito,
            D2.Tipo,
            D3.MesAño,
            D4.Metros,
            D5.NumeroHabitaciones
```

En este caso el valor de superficie se encuentra dentro de la tabla de dimensión Superficie.

5. Tiempo medio de venta de las viviendas por zona, tipología de la vivienda y características de las mismas.

```
SELECT D1.Distrito,
        D2.Tipo,
        D3.MesAño,
        D4.Metros,
        D5.NumeroHabitaciones,
        AVG(H.TIEMPOENVENTA) AVG_TIEMPO_VENTA,
FROM Venta H,
        Zona D1,
        TipoVivienda D2,
        Tiempo D3,
        Superficie D4,
        Habitaciones D5
WHERE H.ZonaPk = D1.ZonaPk
        AND H.TipoViviendaPk = D2.TipoViviendaPk
        AND H.TiempoPk = D3.TiempoPk
        AND H.SuperficiePk = D4.SuperficiePk
        AND H.Habitaciones = D5.Habitaciones
GROUP BY    CUBE( D1.Distrito,
                    D2.Tipo,
                    D3.MesAño,
                    D4.Metros,
                    D5.NumeroHabitaciones )

ORDER BY    D1.Distrito,
                D2.Tipo,
                D3.MesAño,
                D4.Metros,
                D5.NumeroHabitaciones
```

Debemos tener en cuenta que el dominio de esta consulta la forman las viviendas que se han vendido, es posible que si realizamos esta consulta sobre las viviendas en venta encontremos un valor diferente (sobre el tiempo medio que una vivienda se encuentra en venta)

6. Número de inmuebles vendidos y en venta por zona, tipología de la vivienda y características de las mismas.

```
SELECT D1.Distrito,
       D2.Tipo,
       D3.MesAño,
       D4.Metros,
       D5.NumeroHabitaciones,
       COUNT(H1.*) VIVIENDAS_EN_VENTA,
       COUNT(H2.*) VIVIENDAS_VENDIDAS
FROM   EnVenta H1,
       Venta H2,
       Zona D1,
       TipoVivienda D2,
       Tiempo D3,
       Superficie D4,
       Habitaciones D5
WHERE  H1.ZonaPk = D1.ZonaPk
      AND H1.TipoViviendaPk = D2.TipoViviendaPk
      AND H1.TiempoPk = D3.TiempoPk
      AND H1.SuperficiePk = D4.SuperficiePk
      AND H1.Habitaciones = D5.Habitaciones
      AND H2.ZonaPk = D1.ZonaPk
      AND H2.TipoViviendaPk = D2.TipoViviendaPk
      AND H2.TiempoPk = D3.TiempoPk
      AND H2.SuperficiePk = D4.SuperficiePk
      AND H2.Habitaciones = D5.Habitaciones
GROUP BY      CUBE( D1.Distrito,
                   D2.Tipo,
                   D3.MesAño,
                   D4.Metros,
                   D5.NumeroHabitaciones )

ORDER BY      D1.Distrito,
              D2.Tipo,
              D3.MesAño,
              D4.Metros,
              D5.NumeroHabitaciones
```

Las consultas anteriormente descritas no contemplan todos los casos, únicamente los de mayor granularidad. En cualquier caso la adaptación de las consultas a otros niveles jerárquicos dentro de las dimensiones es trivial.

3.2.4.3 Estrategia para determinar la vivienda tipo para España y Andorra.

La pregunta que se nos formula aquí es bastante diferente a las planteadas hasta el momento ya que no puede ser contestada directamente operando sobre los datos.

Lo que vamos a hacer es intentar determinar dimensión a dimensión³ aquellos valores que consolidados nos darían como resultado la vivienda más común en cada uno de esos países.

- Para la dimensión Número de habitaciones realizaremos la media aritmética sobre el número de habitaciones para cada zona (en su nivel jerárquico País) independientemente de las demás características.
- Para la dimensión Superficie actuaremos de la misma forma.
- Finalmente, para la dimensión Tipo de vivienda contaremos cuantas hay de cada tipo y ofreceremos la que obtenga mayor representación como la que consideramos que identifica al país.

³ Sólo aquellas que tengan sentido, por ejemplo Superficie, Número de habitaciones y Tipo Vivienda.

4 Implementación

Iniciamos pues la construcción del almacén de datos así como de los informes solicitados por los clientes. Empezaremos con el diseño del almacén de datos.

4.1 Modelo de base de datos

Dispondremos de cuatro tablas: Viviendas, Zonas, Superficies y Precios. La tabla de superficies se genera una única vez, mientras que las tres restantes dependen del proceso de importación del archivo de texto proporcionado por los clientes.

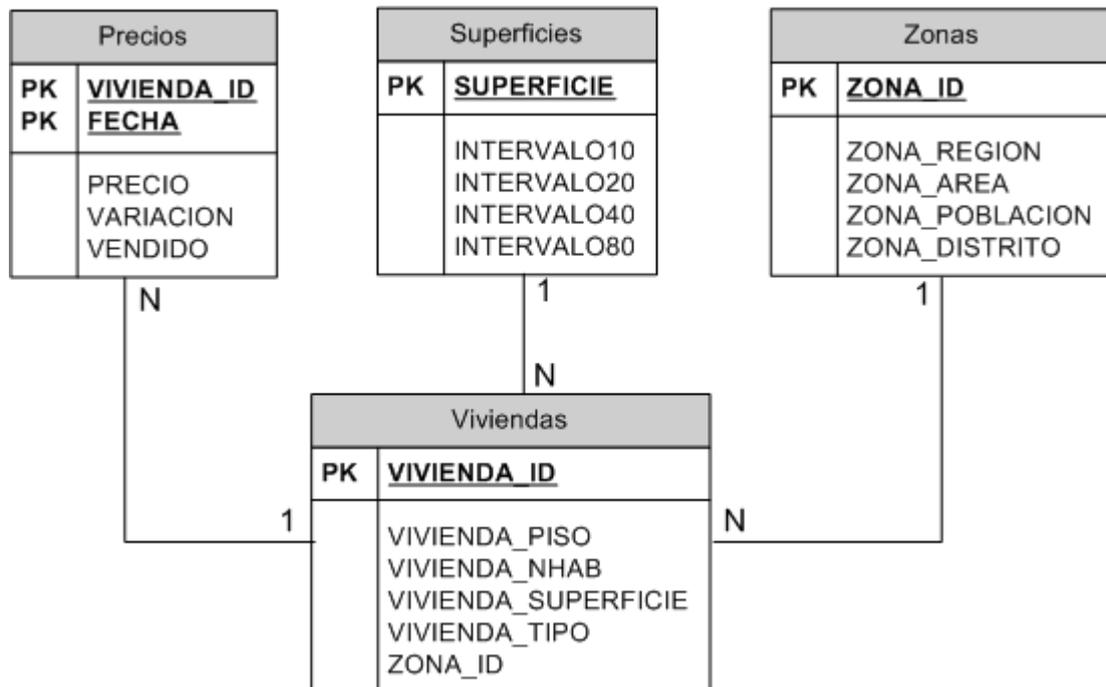


Ilustración 20. Modelo de datos

- **Tabla Viviendas**

Contiene las características generales de la vivienda: planta, número de habitaciones, superficie y tipología.

- **Tabla Precios**

Contiene el precio de la vivienda en cada mes. Así mismo se va calculando la variación de precio respecto al mes anterior. El campo Vendido nos permite identificar el mes y el precio por el que se vendió la vivienda.

- **Tabla Superficies**

Contiene los valores individuales de superficie, y para cada uno de ellos calcula en que intervalo de ancho 10, 20 , 40 y 80 m² se encuentra contenido. Así por ejemplo al valor atómico 147 m² le corresponden los intervalos: [140, 150) para un ancho 10, [140, 160) para un ancho 20, [120, 160) para un ancho 40 y [80, 160) para un ancho 80.

- **Tabla Zonas**

Contiene los valores correspondientes al nombre de la región, área, población y distrito en el que se encuentra ubicada una vivienda.

4.2 Proceso de importación

Hemos creado una Package de Oracle conteniendo el proceso de importación de los archivos proporcionados por los clientes para montar el almacén de datos.

```
CREATE OR REPLACE PACKAGE PROYECTE AS
/* TAREA Introducir declaraciones de paquetes (tipos excepciones, métodos, etc) aquí */
TYPE TABLA_FECHAS_PRECIOS IS TABLE OF DATE INDEX BY BINARY_INTEGER;
TYPE PRECIO_FECHA IS RECORD
(
    FECHA DATE,
    IMPORTE NUMBER(8,0),
    VARIACION NUMBER(8,0),
    VENDIDO SMALLINT
);
TYPE TABLA_PRECIOS IS TABLE OF PRECIO_FECHA INDEX BY BINARY_INTEGER;
TYPE ZONA IS RECORD
(
    REGION VARCHAR(100),
    AREA VARCHAR(100),
    POBLACION VARCHAR(100),
    DISTRITO VARCHAR(100)
);
TYPE VIVIENDA IS RECORD
(
    IDENT VARCHAR2(15),
    ZONA VARCHAR2(50),
    TIPO VARCHAR2(15),
    PLANTA VARCHAR(10),
    DORMITORIOS INTEGER,
    METROS INTEGER,
    PRECIO TABLA_PRECIOS
);

log_handle UTL_FILE.FILE_TYPE;
fechas_precios TABLA_FECHAS_PRECIOS;

PROCEDURE LOAD(file_name IN VARCHAR2);
FUNCTION PROCESA_CABECERA(dades IN VARCHAR2) RETURN TABLA_FECHAS_PRECIOS;
FUNCTION PROCESA_ZONA ( area_data IN VARCHAR2) RETURN ZONA;
FUNCTION PROCESA_VIVIENDA(datos_vivienda IN VARCHAR2) RETURN VIVIENDA;
FUNCTION COMPLEX_TRIM(cadena IN VARCHAR2) RETURN VARCHAR2;
FUNCTION ALMACENAR_ZONA(zona ZONA) RETURN NUMBER;
PROCEDURE ALMACENAR_VIVIENDA(vivienda VIVIENDA, zona_id NUMBER);

END PROYECTE;
```

Ilustración 21. Cabecera del Package de Importación

Implementación

El punto de entrada se encuentra en el procedimiento LOAD. Recibe como único parámetro el nombre de archivo que contiene la información.

El archivo debe encontrarse en un directorio para el cual se haya aplicado la sentencia siguiente:

```
-- Habilitar acceso al directorio desde el cual se procederá a
importar los datos...
CREATE OR REPLACE DIRECTORY DIR_TEMP AS 'F:\FILES';
-- Necesitamos permisos de escritura debido a que generamos en ese
mismo directorio
-- Un log de importación.
GRANT READ, WRITE ON DIRECTORY DIR_TEMP TO <<owner_datawarehouse>>;
```

4.3 Diseño del área de negocio

Hemos creado un área de negocio denominada “Sostre per a tothom”. Dentro de ella hemos creado los objetos siguientes:

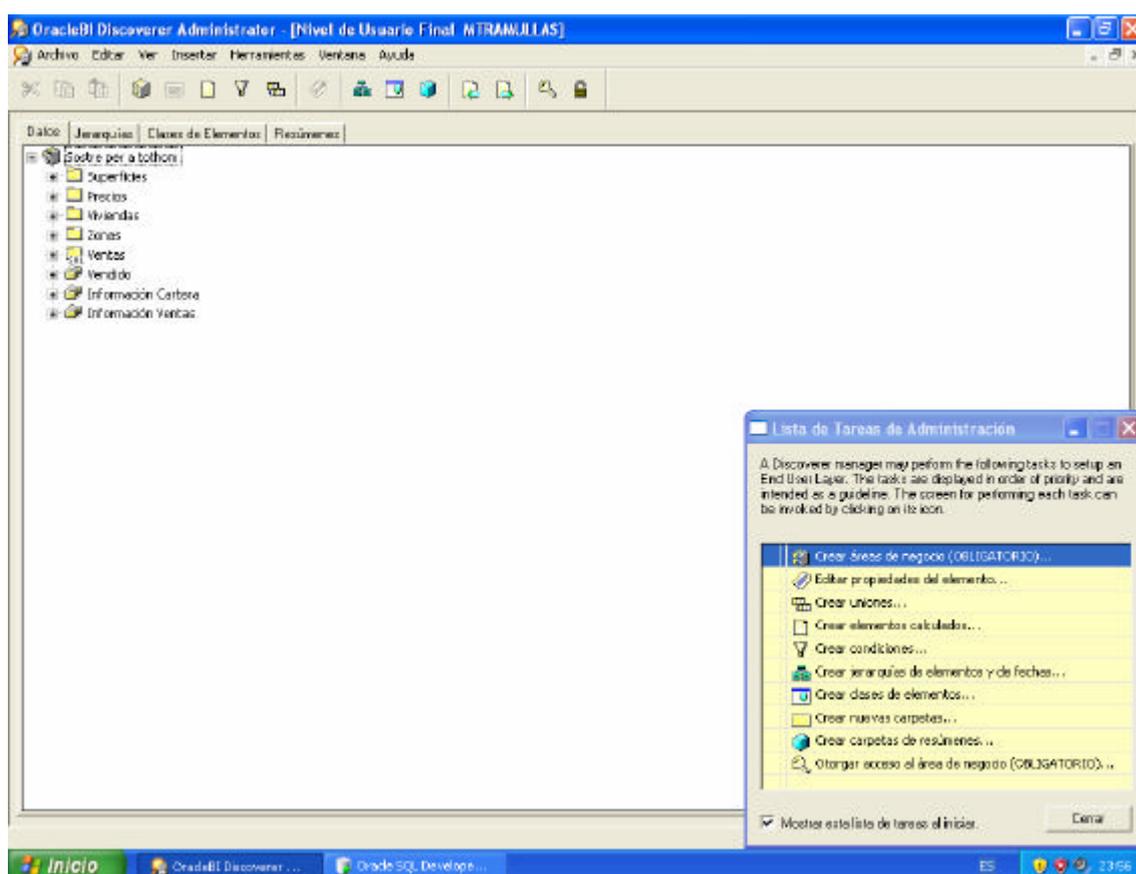


Ilustración 22. Diseño del área de negocio

4.3.1 Datos

4.3.1.1 Carpetas básicas

Hemos importado las 4 tablas que conforman nuestro almacén de datos. A esas tablas hemos añadido nuevos elementos como por ejemplo Fecha Month y Fecha Year.

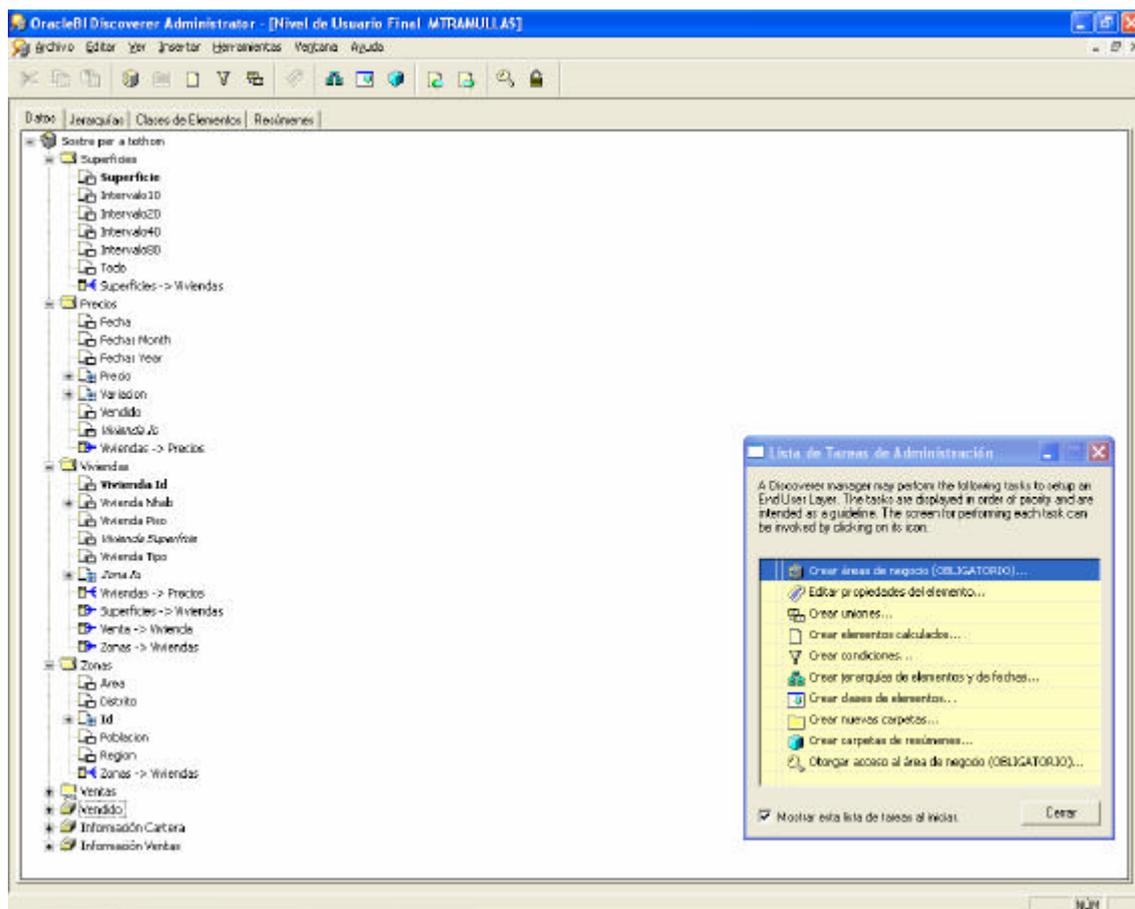


Ilustración 23. Carpetas básicas

4.3.1.2 Carpeta Ventas

Hemos creado una carpeta adicional llamada ventas que nos permite conocer ciertos valores sobre las viviendas vendidas. He optado por una carpeta personalizada basada en la siguiente consulta

```
SELECT V.VIVIENDA_ID, COUNT(P.FECHA) MESES, V.FECHA FECHA_VENTA, V.PRECIO PRECIO_VENTA
FROM PRECIOS P, (SELECT VIVIENDA_ID, FECHA, PRECIO FROM PRECIOS WHERE VENDIDO=1) V
WHERE P.VIVIENDA_ID = V.VIVIENDA_ID
GROUP BY V.VIVIENDA_ID, V.FECHA, V.PRECIO
```

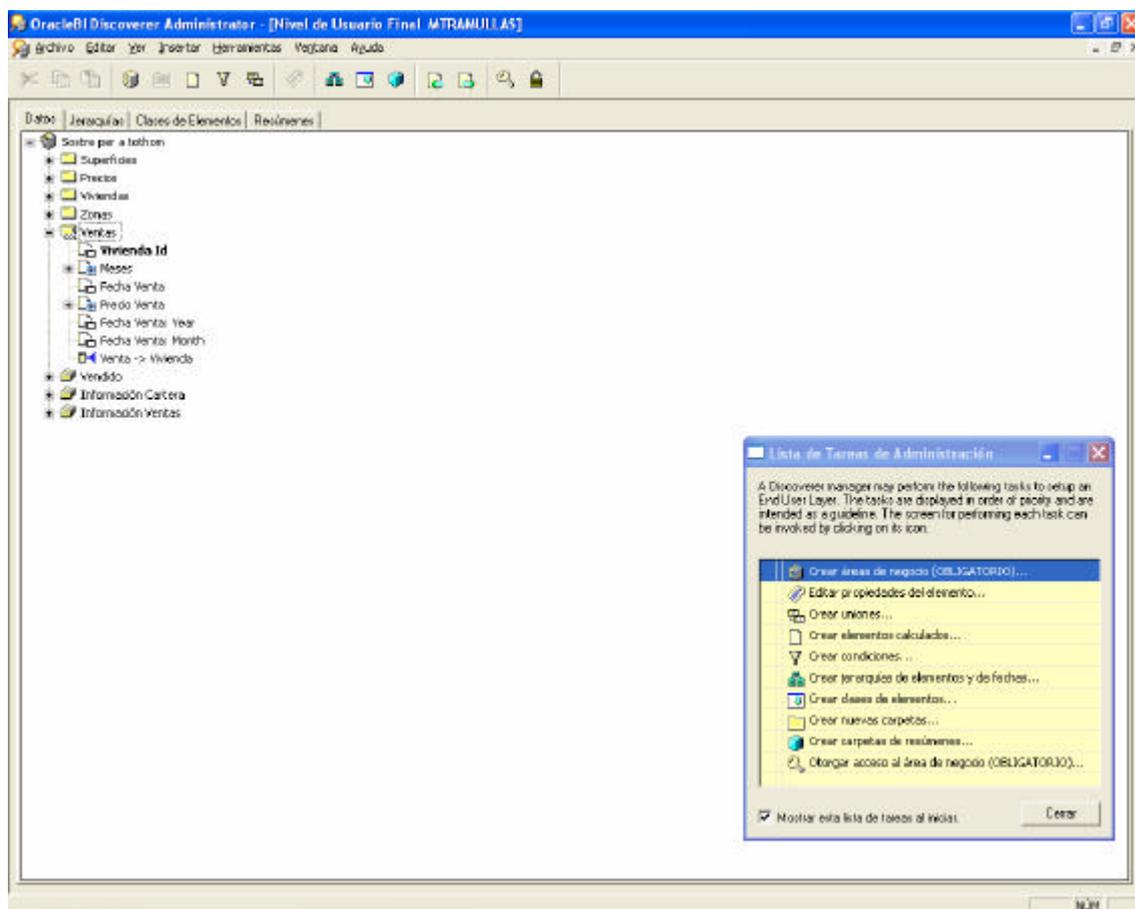


Ilustración 24. Carpeta Ventas

4.3.1.3 Carpeta Información Cartera

Esta carpeta contiene toda la información necesaria para las consultas que se refieren a las viviendas disponibles para la venta

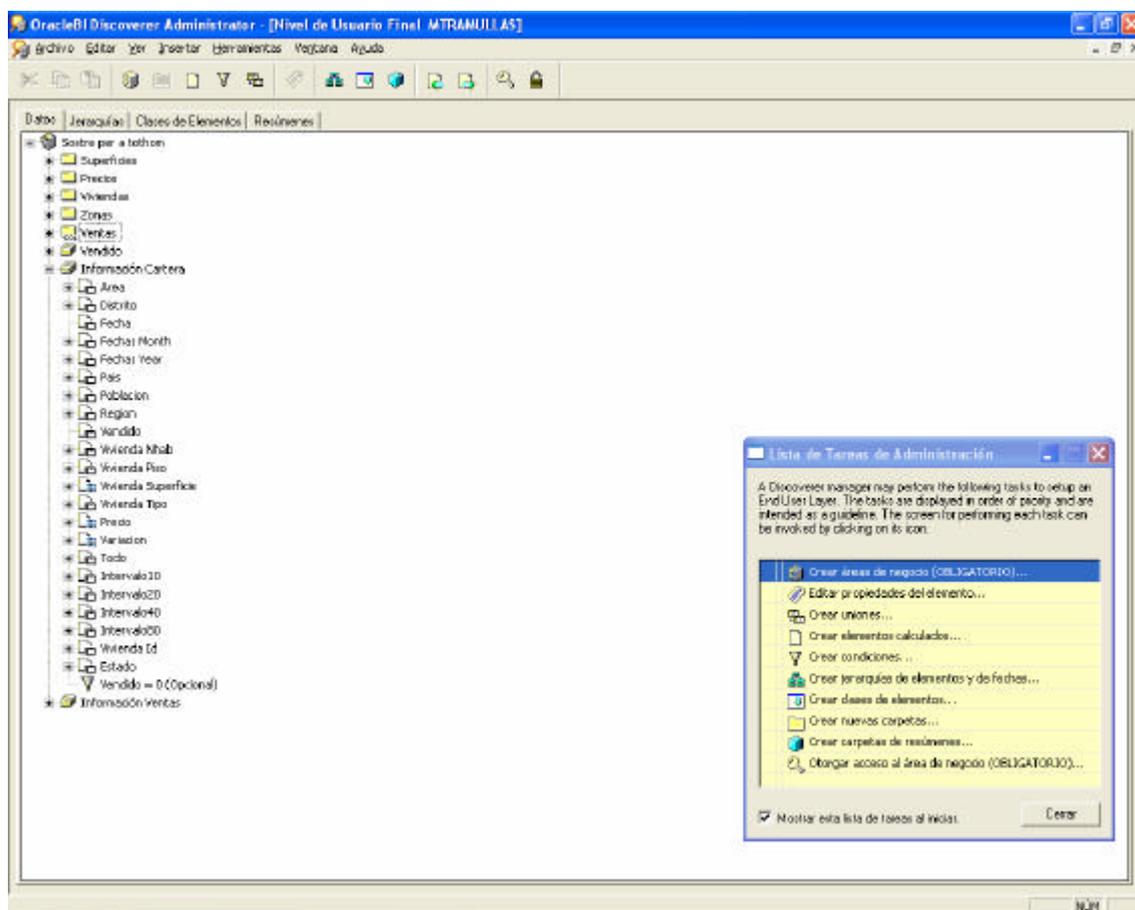


Ilustración 25. Carpeta Información Cartera

Implementación

Tanto esta carpeta como la siguiente (Información Ventas) disponen de un elemento calculado adicional llamado País que nos permite establecer un elemento jerárquico superior para las zonas.

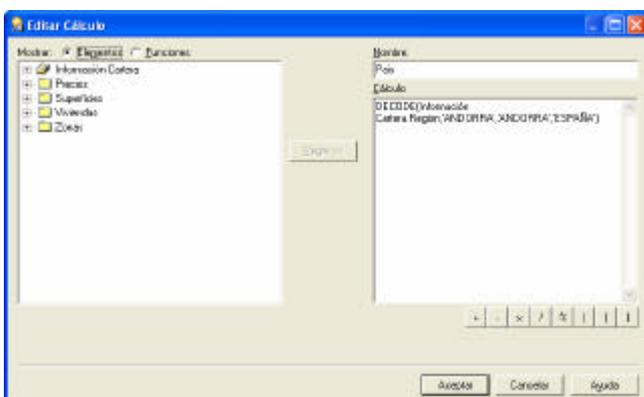


Ilustración 26. Definición del campo calculado País

4.3.1.4 Carpeta Información Ventas

Esta carpeta contiene todos los elementos que pueden ser utilizados por los informes sobre las ventas de viviendas.

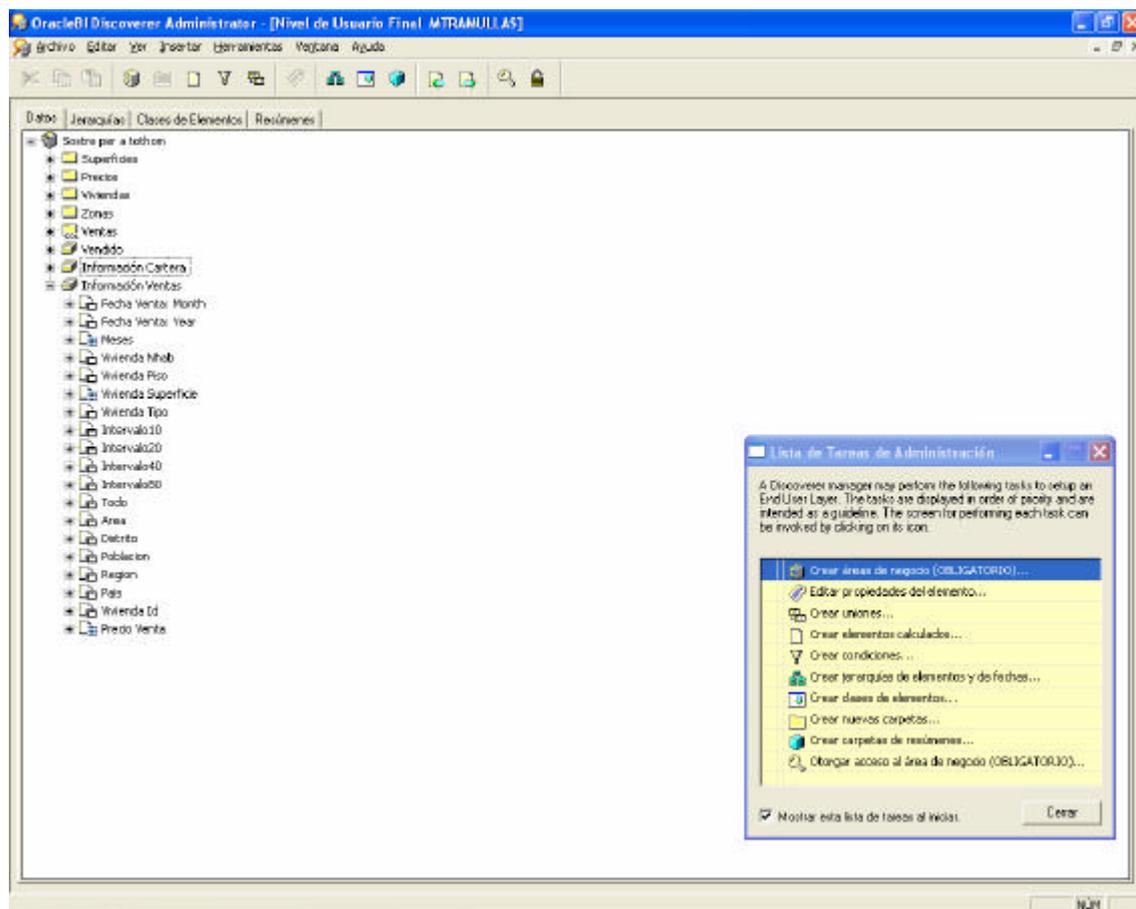


Ilustración 27. Carpeta Información Ventas

4.3.2 Jerarquías

Hemos definido las jerarquías siguientes tanto para la carpeta de Información Cartera como para la de información Ventas

4.3.2.1 Jerarquía Geográfica:

Esta formada por los elementos siguientes (de mayor a menor) País, Región, Área, Población y Distrito.

4.3.2.2 Jerarquía de Fechas

Está formada por los elementos siguientes (de mayor a menor) Fecha Year, Fecha Month

4.3.2.3 Jerarquía Superficies

Está formada por los elementos siguientes (de mayor a menor) Todo, Intervalo80, Intervalo40, Intervalo20, Intervalo10, Superficie.

4.4 Informes

Seguidamente pasamos a detallar los informes que han sido implementados a petición de los clientes.

4.4.1 Informe total de viviendas en venta.

Para esta consulta nos hemos basado en la carpeta “Información Cartera” configurada de la forma siguiente:

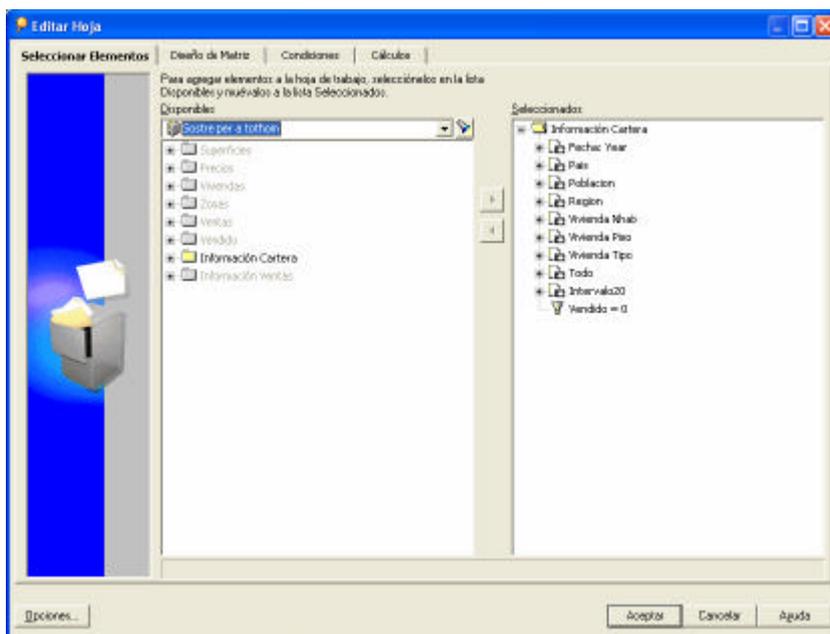


Ilustración 28. Selección de elementos para la consulta Total Viviendas en venta

Implementación

Para el diseño de la consulta hemos establecido el modelo siguiente:

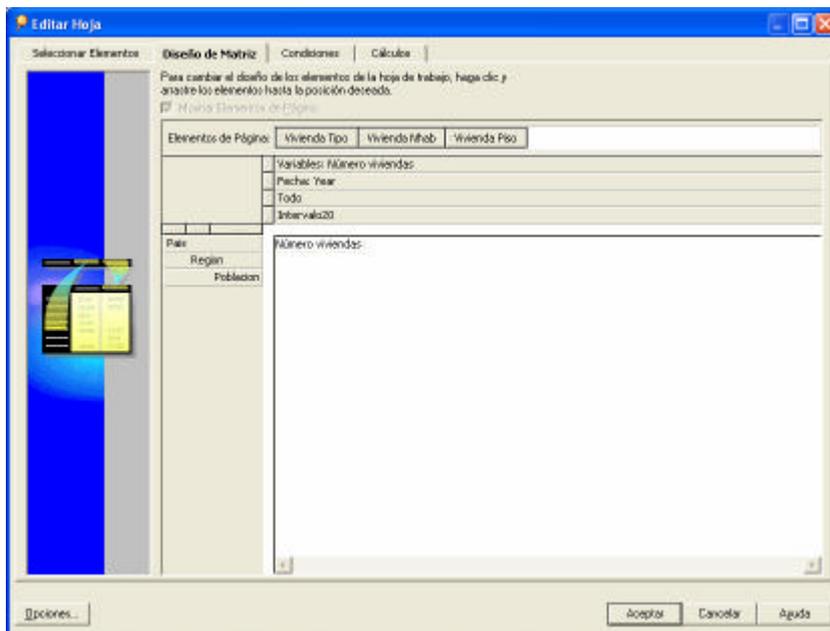


Ilustración 29. Diseño Matriz consulta Total Viviendas en Venta.

Así mismo hemos establecido como condición que el campo Vendido sea 0 y hemos realizado el cálculo llamado Número de viviendas.

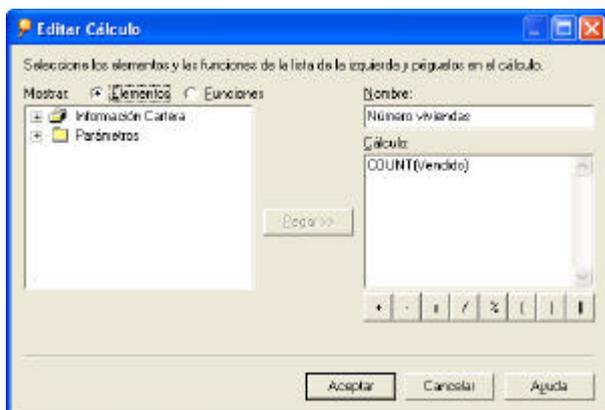


Ilustración 30. Campo calculado Total Ventas

Implementación

Este sería el formato de informe construido.

The screenshot shows a report window titled "Viviendas en Venta". The report is a pivot table with the following structure:

- Column Headers:** Año (2000), Todas Superficies, Intervalo0, Intervalo1, Intervalo2, Intervalo3, Intervalo4, Intervalo5, Intervalo6, Intervalo7, Intervalo8, Intervalo9, Intervalo10, Intervalo11, Intervalo12.
- Row Headers:** País (Andalucía, Castilla - Alicante), Región, Población.
- Table Content:** A grid of data points representing the number of houses sold in each region, categorized by price intervals.

Ilustración 31. Informe Total Viviendas en Venta

4.4.2 Informe de viviendas que han bajado de precio.

Para este informe hemos partido de nuevo de la carpeta Información Cartera. Esta es la configuración del mismo

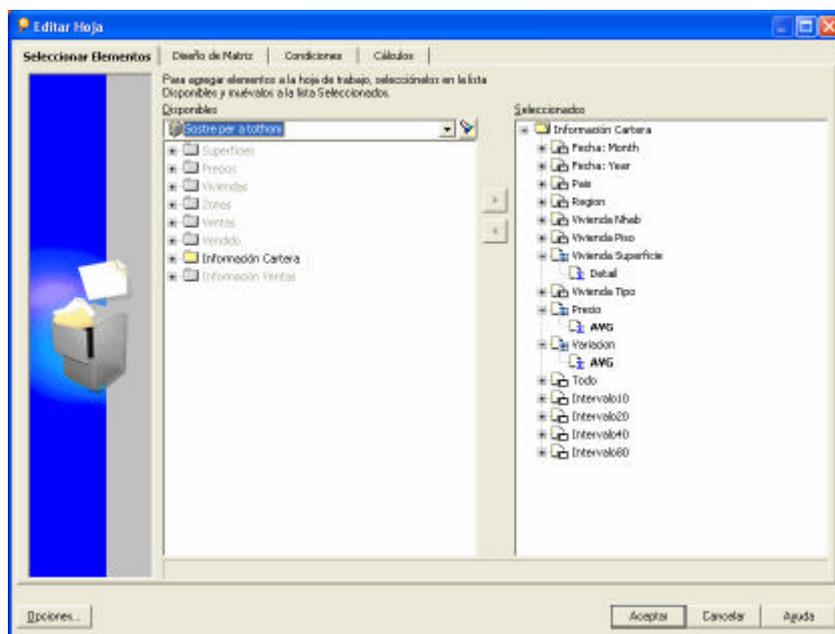


Ilustración 32. Selección de los elementos Viviendas que han bajado de precio

Implementación

La configuración de la matriz es muy parecida a la de la consulta anterior. Cabe destacar que la información se encuentra filtrada por la condición Variación < 0.

Este es un ejemplo de este informe:

The screenshot shows a report titled "Viviendas que han bajado de precio" (Houses that have decreased in price). The report is displayed in a BI tool interface. The data is presented in a matrix format with the following structure:

- Columns:** Months (Mes) from May (May) to December (Dic).
- Rows:**
 - Year (Año) 2010
 - Price Range (Rango de clasificación de precio) with categories:
 - Indice Superior 40
 - Superficie Base 10
 - Superficie Base 40
 - Superficie Base 20
 - Superficie Base 10
 - Province (Provincia) with categories:
 - ALICANTE
 - BARCELONA
 - CANILLÓN
 - GERONA
 - ISLA-IBIZAS
 - LEIVA
 - MALLORCA
 - MEMORIA
 - TARAGONA
 - VALENCIA

The data cells contain numerical values representing the change in price for each category and month. For example, in the 'ALICANTE' row, the values for months May through December are: -3475, -2474, -2909, -3085, -3551, -2589, -3470, -3758, -2581.

Ilustración 33. Informe Viviendas que han bajado de precio

4.4.3 Informe de Precios mínimos, medios y máximos

Continuamos utilizando la carpeta “Información Cartera” para implementar este informe. Mantiene una estructura muy pareja al informe anterior. Veamos los elementos que hemos seleccionado.

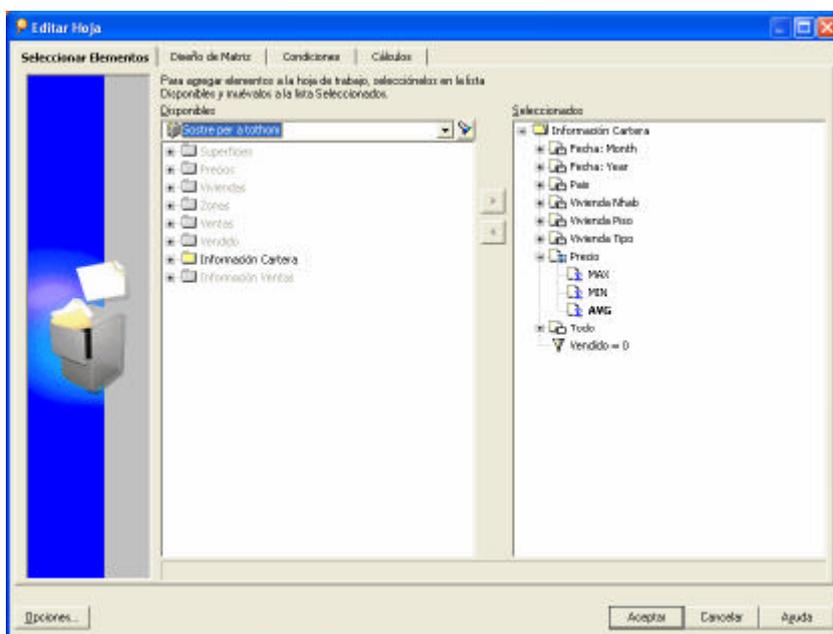


Ilustración 34. Selección de elementos para precios mínimos, medios y máximos

Mantendremos la condición de Vendido = 0.

	Precios Máximos, Medios y Mínimos															
	2000		Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	2007		
País	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Superficies	45.29.000	62.79.000	62.99.000	61.09.000	62.29.000	62.29.000	62.29.000	62.29.000	62.29.000	62.29.000	62.29.000	62.29.000	62.29.000	62.29.000	62.29.000	62.29.000
Reserva	64	45.800	45.800	45.000	45.000	45.199	45.199	45.199	45.199	45.199	45.199	45.199	45.199	45.199	45.199	45.199
Español																

Ilustración 35. Informe Precios mínimos, medios y máximos

4.4.4 Informe de superficies mínimas, medias y máximas de las viviendas en venta

Este informe simplifica ligeramente el diseño de la matriz ya que desaparecen tanto la superficie como todos los intervalos de superficie de la cabecera de la misma. Veamos pues los elementos seleccionados para la elaboración de este informe.

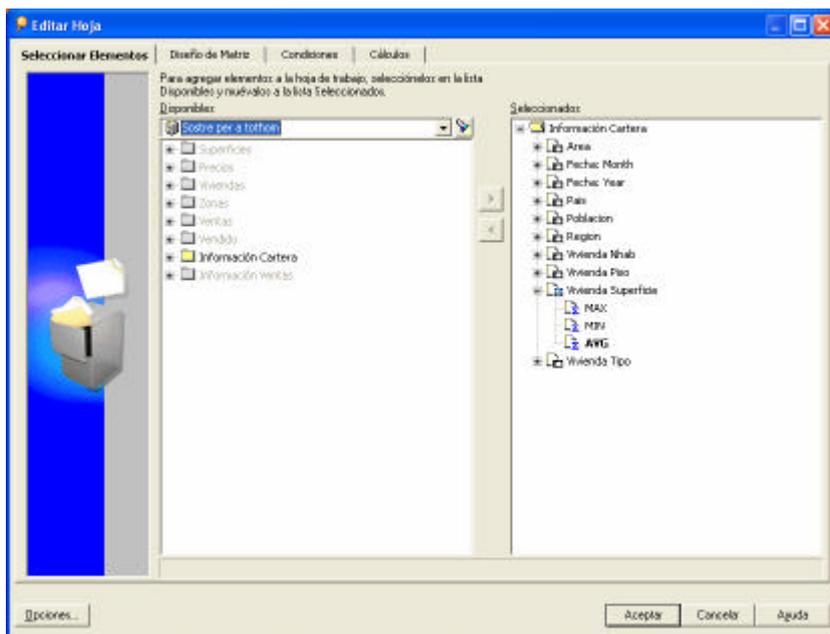


Ilustración 36. Selección de elementos para el informe de superficies

La estructura de la matriz será la siguiente:

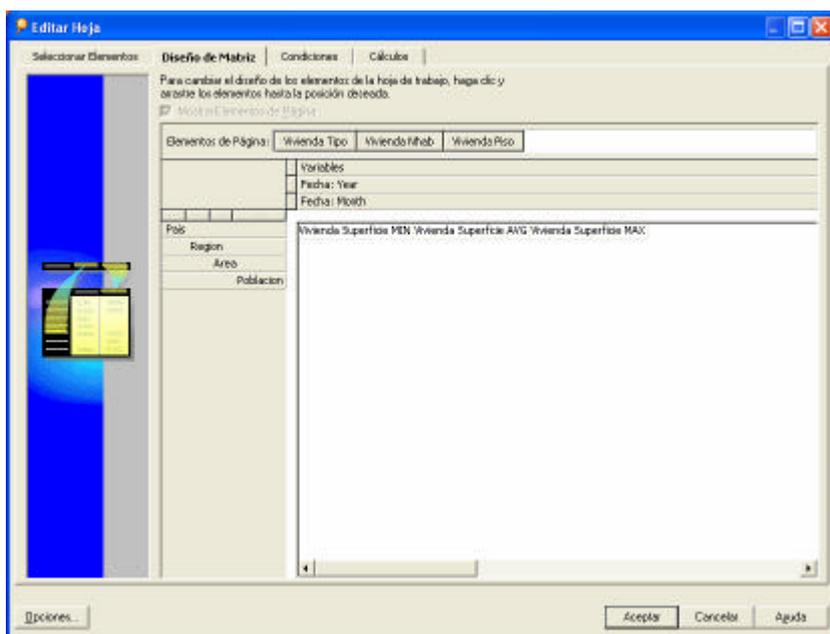


Ilustración 37. Diseño de la matriz del informe de superficies

País	Región	Área	Población	Meses											
				Año	Jan	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov
Andorra	Andorra	Encamp - Canillo - Ordino									107	107	54	54	54
		Mascard - Andorra - Escaldes													
España	Alicante	Área De Incaubate Del Segura		55	55	55	55	55	55	55	55	55	55	55	55
		Área De La Mancha		47	47	47	47	47	47	47	47	47	47	47	47
		Área De Orihuela		65	65	65	65	65	65	65	65	65	65	65	65
		Área De Orihuela Costa		32	32	32	32	32	32	32	32	32	32	32	32
		Área De San Miguel De Salinas		44	44	44	44	44	44	44	44	44	44	44	44
		Santoveja		43	43	43	43	43	43	43	43	43	43	43	43
	Balears	Devolente		88	88	88	88	88	88	88	88	88	88	88	88
		Elías Eix		88	88	188	188	188	188	188	188	188	188	188	188
		París De Elche Norte		88	88	88	88	88	88	88	88	88	88	88	88
	Cataluña	Comtal		44	44	44	44	44	44	44	44	44	44	44	44
		Llabardit		44	44	44	44	44	44	44	44	44	44	44	44
	Islas Baleares	Alcanar / Alcanar		44	44	44	44	44	44	44	44	44	44	44	44
Escorç			44	44	44	44	44	44	44	44	44	44	44	44	
El Campello			47	47	47	47	47	47	47	47	47	47	47	47	
Alanya / Alanya			242	242	242	242	242	242	242	242	242	242	242	242	
Mutxamel			78	78	78	78	78	78	78	78	78	78	78	78	
San Juan De Alicante			44	44	44	44	44	44	44	44	44	44	44	44	
San Vicente Del Raspeig			44	44	44	44	44	44	44	44	44	44	44	44	
Área De Alcoy / Alcoy			78	78	78	78	78	78	78	78	78	78	78	78	
Madrugada	Área De Benissa		78	78	78	380	380	380	380	380	380	380	380	380	
	Área De Calpe		88	88	88	88	88	88	88	88	88	88	88	88	
	Área De Orihuela		104	104	104	104	104	104	104	104	104	104	104	104	
	Área De Turis		44	44	44	44	44	44	44	44	44	44	44	44	
Madrugada	Dama		47	47	47	47	47	47	47	47	47	47	47	47	
	Alanya / Alanya		88	88	88	88	88	88	88	88	88	88	88	88	
	Área De Albuñol		47	47	47	47	47	47	47	47	47	47	47	47	
	Área De Callosa D'En Sarrià		44	44	44	44	44	44	44	44	44	44	44	44	
Madrugada	Área De Incaubate		44	44	44	44	44	44	44	44	44	44	44	44	
	Área De La Vila Nueva / Villavieja		72	72	72	72	72	72	72	72	72	72	72	72	
	Dandiana		44	44	44	44	44	44	44	44	44	44	44	44	
	Área De Albuñol		47	47	47	47	47	47	47	47	47	47	47	47	

Ilustración 38. Informe de superficies mínimas, medias y máximas

4.4.5 Informe de tiempos medios de venta

Para implementar este informe he utilizado elementos de la carpeta de "información Ventas". Veamos los elementos seleccionados:

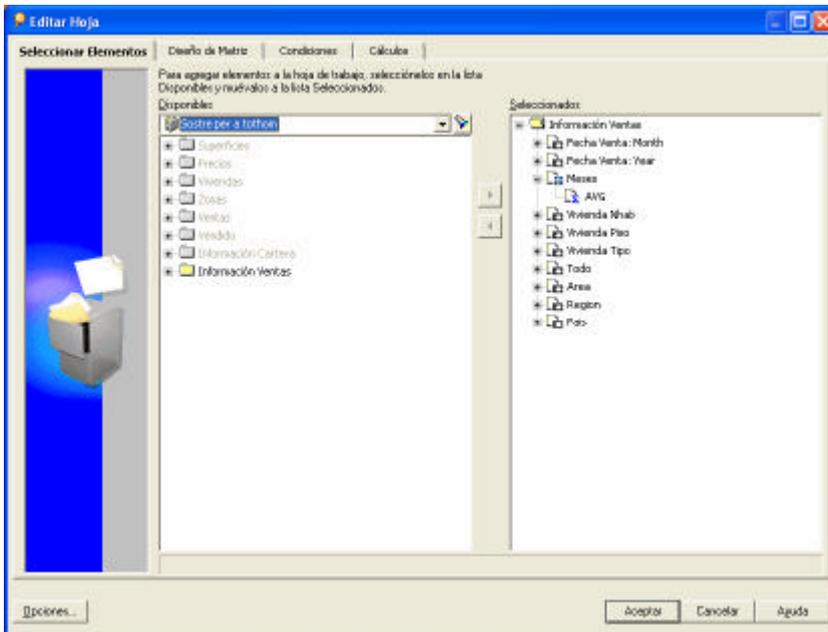


Ilustración 39. Elementos seleccionados informe de tiempos medios de venta

Implementación

El diseño de la matriz queda como sigue:

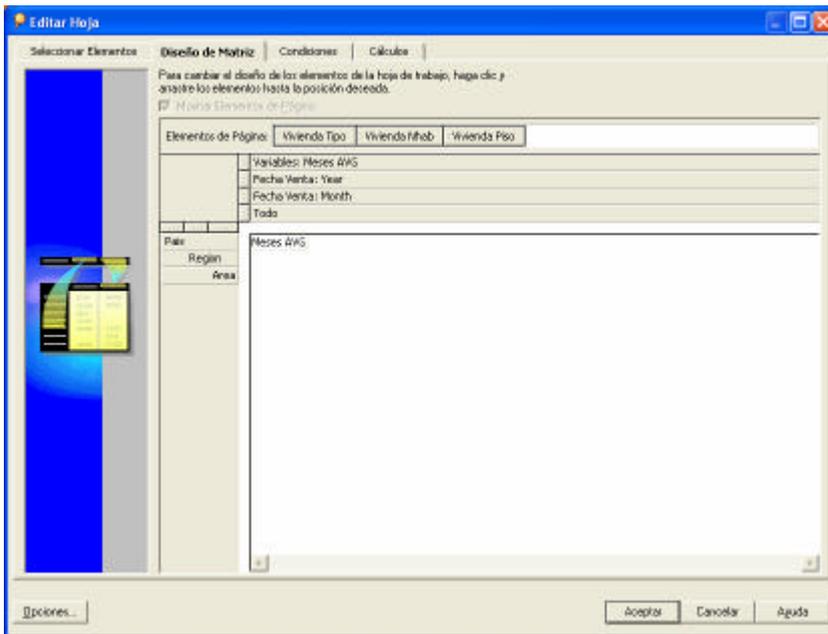


Ilustración 40. Diseño matriz del informe de tiempos medios de venta

Veamos finalmente un ejemplo del resultado del informe.

			2007											
			Jan	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Pais	Region	Area	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB
Andalucía	Andalucía		32	31	32	31	32	31	32	31	32	31	32	31
Castilla	Alicante		30	30	31	32	31	32	31	32	31	32	31	32
	Castellón		7	32	32	8	4	15	0	30	14	8	13	32
	Gerona		33	33	33	33	33	33	34	34	34	33	37	36
	Ibiza-Formentera						11							18
Murcia	Centro			8		8	12	14	11	1		8	13	33
	Este				4	17				1				33
	Este-Murcia		8		2	15				4		0		38
	Norte		33							11			8	24
	Palma de Mallorca		7	31	32	32	32	14	7	0	0	20	33	34
Murcia	Este													33
	Este-Murcia		33											33
Murcia	Este		33	30	30	31	8	8	30	8	18	18	17	34
	Este		38	31	32	33	32	33	33	33	33	33	33	33

Ilustración 41. Informe de tiempos medios de venta

4.4.6 Informe de viviendas en venta y vendidas

Informe basado en los elementos de la carpeta “Informe Cartera”. Aprovechamos el valor del elemento *Estado* para distinguir las viviendas vendidas de las que se encuentran en venta. La selección de elementos es la que se muestra.

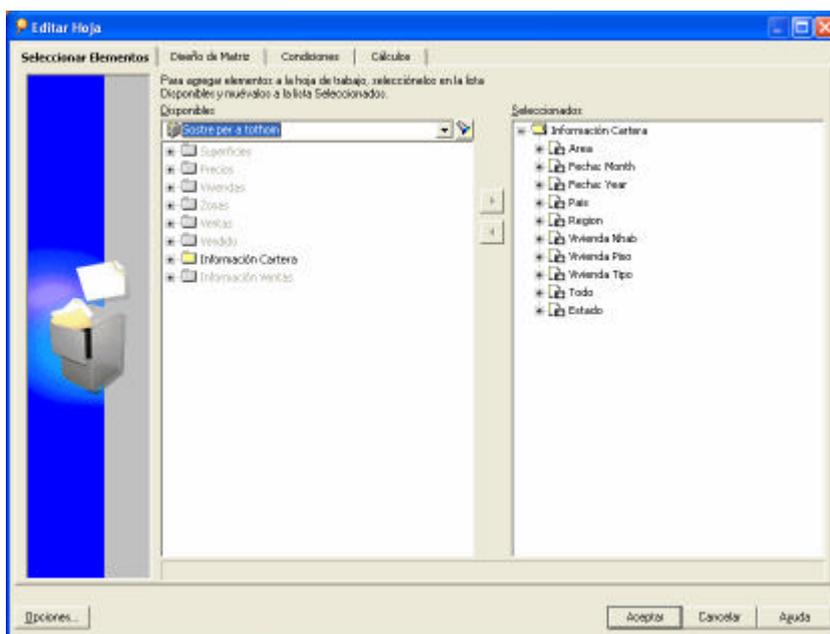


Ilustración 42. Elementos seleccionados informe de viviendas en venta y vendidas

El diseño de la matriz es el siguiente:

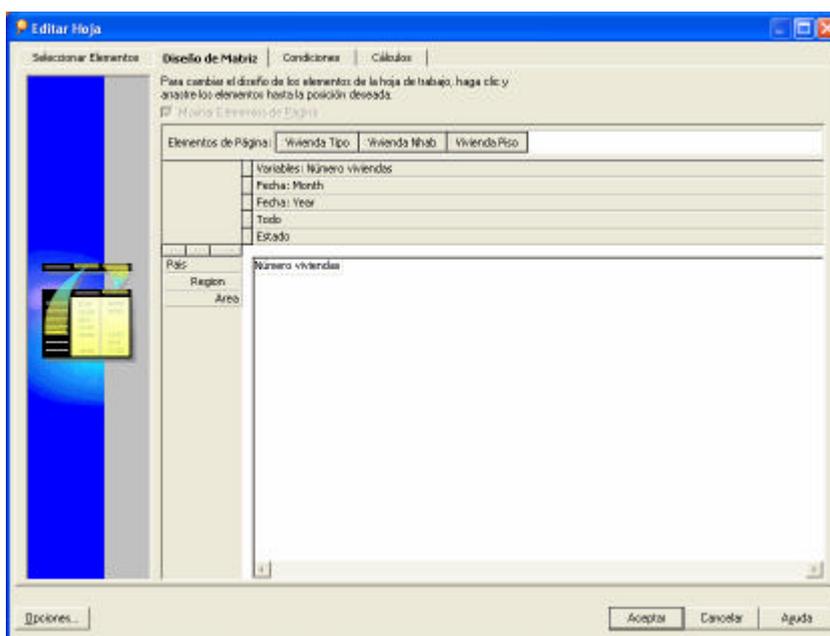


Ilustración 43. Diseño de la matriz del informe de viviendas vendidas y en venta

Implementación

Este es un ejemplo del resultado de dicho informe

Viviendas en Venta y Vendidas

Tipología: Vivienda DUPLEX | Situaciones: Estado | Planes: <Todos>

		Número viviendas en venta y vendidas											
		Ene 2006		Feb 2006		Mar 2006		Abr 2006		May 2006		Jun 2006	
		En Venta	Vendidas	En Venta	Vendidas	En Venta	Vendidas	En Venta	Vendidas	En Venta	Vendidas	En Venta	Vendidas
País	Región	Área											
	Andorra												
	España												
		Almería	281	7	288	7	292	8	295	8	298	23	298
		Baleares	38	1	18		12	15	1	12	21	19	21
		BarCELONA	28	1	23	3	18	21	2	28	28	2	28
		BIZKAIA	11		36	2	9	54	2	02	36	25	1
		BIZKAIA (CORONA) NORO	9		8		8	18		10	12	8	18
		BIZKAIA (CORONA) SUD	142	1	159	15	135	18	155	8	160	5	198
		BIZKAIA (CP) S	156	2	212	2	188	4	214	8	188	8	188
		BIZKAIA (CP) N	7		12		7	18	2	11		3	18
		CASTELLÓN	21		18	3	20	2	28	2	24	1	24
		CANARIAS	26		125	3	16	130	4	17	5	100	1
		CEUTA	2		18		5	18	1	5	12	3	12
		CANARIAS (CORONA) SUD	220	8	368	11	225	18	368	8	379	4	225
		CANARIAS (CORONA) N	51		88	2	78	1	82	2	88	4	88
		CANARIAS (CORONA) E	58	1	81	8	52	68	8	52	1	82	2
		Castellón	58		138	2	78	2	148	4	143	5	81
		CANARIAS (CORONA) N	33		18	3	12	18		11	1	30	11
		CANARIAS (CORONA) S	28	1	48	2	38	1	44	2	38	1	48
		CANARIAS (CORONA) E	35	1	38	2	38	1	38	1	40	2	40
		CANARIAS (CORONA) W	31		13		11	12	8	2	14	8	14
		CANARIAS (CORONA) S	46	1	71	2	47	2	74	1	60	1	74
		CANARIAS (CORONA) N	278	8	397	9	279	9	290	18	304	23	290

Ilustración 44. Informe de viviendas en venta y vendidas

4.4.7 Top N de precios medios

Es un informe parametrizado que requiere que introduzcamos el número de mayores precios medios que deseamos visualizar.

Asistente de Parámetros

Introduzca los valores de los siguientes parámetros:

* indica un campo necesario

N *

Descripción:

Terminar Cancelar Ayuda

Ilustración 45. Introducción parámetro

Implementación

Para ello en el diseño de la consulta he especificado la condición siguiente:

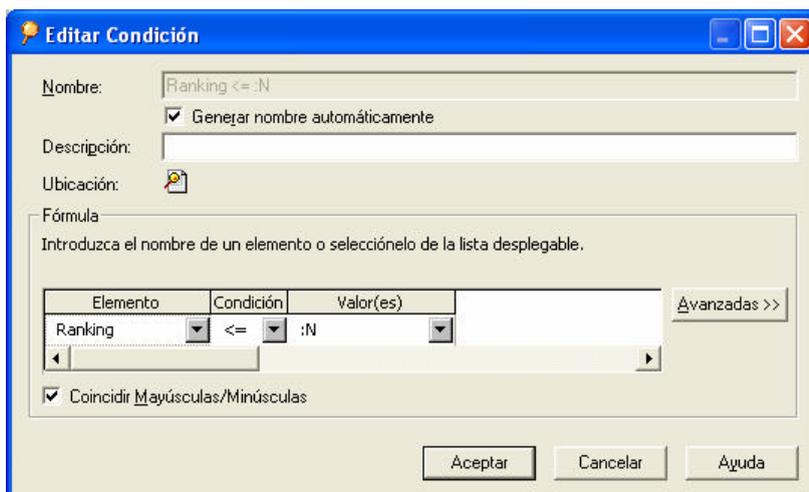


Ilustración 46. Condición parametrizada

Además hemos añadido el siguiente elemento calculado para poder obtener el ranking deseado:

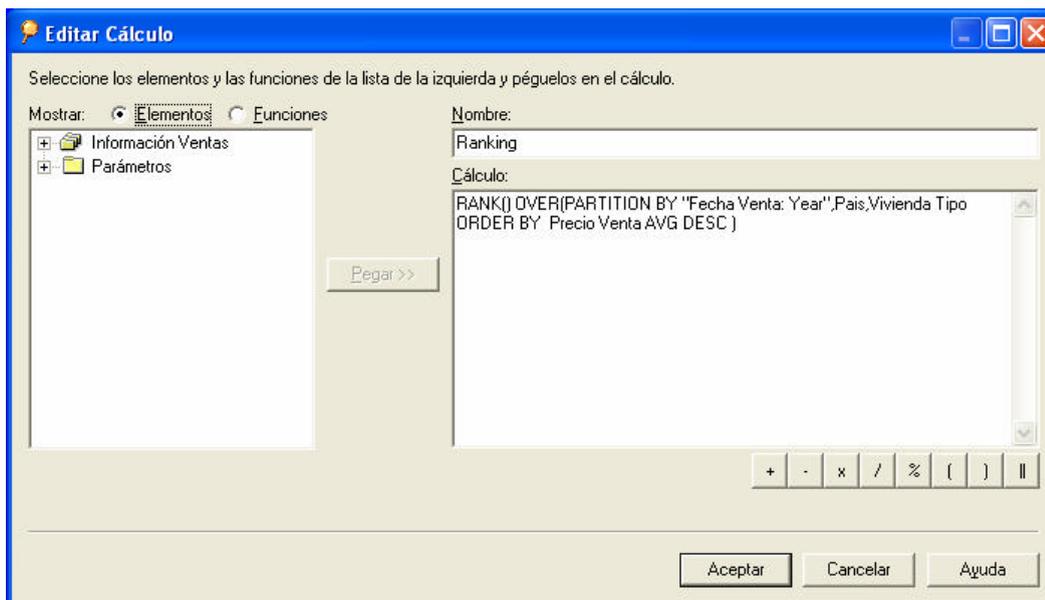


Ilustración 47. Campo calculado en función de un parámetro

Implementación

El resultado de esta consulta se muestra en el ejemplo siguiente:

País	Región	Medio de Precio Venta				
Argentina		420.403	41.884.412	498.238	41.781.881	430.482
		420.403	41.884.412	498.238	41.781.881	430.482
España		420.403	41.884.412	498.238	41.781.881	430.482
		420.403	41.884.412	498.238	41.781.881	430.482
Italia		420.403	41.884.412	498.238	41.781.881	430.482
		420.403	41.884.412	498.238	41.781.881	430.482
México		420.403	41.884.412	498.238	41.781.881	430.482
		420.403	41.884.412	498.238	41.781.881	430.482
Venezuela		420.403	41.884.412	498.238	41.781.881	430.482
		420.403	41.884.412	498.238	41.781.881	430.482

Ilustración 48. Informe Top N de precios medios

5 Conclusiones

- He utilizado una estrategia basada en dos iteraciones, la primera de las cuales sólo incluía una fase de implementación exploratoria de los procesos de carga.
- He aplicado los conocimientos adquiridos durante mis estudios para llevar a cabo el proyecto.
- He desarrollado sobre un entorno completamente desconocido y sobre una materia que no tenía conocimientos previos.
- Se han alcanzado los objetivos marcados al principio del proyecto

6 Glosario

Base de datos operacional

Base de datos destinada a gestionar el día a día de una organización, almacena la información de a la operativa diaria.

Almacén de datos

Base de datos orientadas a áreas de interés de la empresa que integran datos de diferentes fuentes con información histórica y no volátil que tienen por objetivo principal dar soporte a la toma de decisiones.

Transacción

Conjunto de operaciones de lectura y/o actualización de la base de datos que acaba confirmando o cancelando los cambios que se han llevado a cabo.

Dimensión

Punto de vista utilizado en el análisis de un hecho.

Drill-across

Operación multidimensional que, dado un espacio, cambia los datos que muestra. Cambia de un tema de análisis a otro.

Drill-down

Operación multidimensional que muestra los datos más detallados.

Hecho

Objeto del análisis.

Granularidad

Tamaño de un objeto respecto a un altro.

Jerarquía de agregación

Conjunto de relaciones entre las instancias de una dimensión que indica como agrupamos unas para obtener otras.

Medida

Dato numérico asociado a un hecho que queremos analizar.

Nivel de agregación

Conjunto de instancias de la misma granularidad dentro de una dimensión.

7 Bibliografia

- *Magatzems de Dades i models multidimensionals. Àngels Rius Gavídia i Montse Serra Vizern (UOC)*
- *The Data Warehouse Toolkit. (2nd Edition). Ralph Kimball and Margy Ross.*
- *Building the Data Warehouse (3rd Edition). W. H. Inmon.*
- *The Data Warehouse ETL Toolkit. Joe Caserta and Ralph Kimball.*
- *Documentación Oracle XE*

8 Anexos

8.1 *Tabla de ilustraciones*

Ilustración 1. Dimensión Zona	11
Ilustración 2. Dimensión Tiempo	12
Ilustración 3. Dimensión Superficie.....	12
Ilustración 4. Dimensión Tipo Vivienda	13
Ilustración 5. Hecho En Venta	14
Ilustración 6. Hecho Venta	15
Ilustración 7. Diseño de celda para el Hecho Poner en Venta	18
Ilustración 8. Diseño de celda para el Hecho Venta	19
Ilustración 9. Tabla dimensión Tiempo.....	21
Ilustración 10. Tabla dimensión Superficie.....	22
Ilustración 11. Tabla dimensión Habitaciones	22
Ilustración 12. Tabla dimensión Zona.....	23
Ilustración 13. Tabla dimensión Tipo vivienda con subclasificación	24
Ilustración 14. Tabla dimensión Tipo Vivienda definitiva	24
Ilustración 15. Tabla hecho En Venta	25
Ilustración 16. Tabla hecho Venta.....	26
Ilustración 17. Diagrama de datos para el hecho Inventa	27
Ilustración 18. Diagrama de datos para el hecho Venta	28
Ilustración 19. Proceso del cargador de datos.....	29
Ilustración 20. Modelo de datos.....	42
Ilustración 21. Cabecera del Package de Importación	44
Ilustración 22. Diseño del área de negocio	45
Ilustración 23. Carpetas básicas	46
Ilustración 24. Carpeta Ventas	47
Ilustración 25. Carpeta Información Cartera	48
Ilustración 26. Definición del campo calculado País	49
Ilustración 27. Carpeta Información Ventas	49

Ilustración 28. Selección de elementos para la consulta Total Viviendas en venta	51
Ilustración 29. Diseño Matriz consulta Total Viviendas en Venta	52
Ilustración 30. Campo calculado Total Ventas.....	52
Ilustración 31. Informe Total Viviendas en Venta	53
Ilustración 32. Selección de los elementos Viviendas que han bajado de precio	53
Ilustración 33. Informe Viviendas que han bajado de precio.....	54
Ilustración 34. Selección de elementos para precios mínimos, medios y máximos	55