

TREBALL FI DE CARRERA



Construcció i explotació d'un magatzem de dades per a l'anàlisi d'informació immobiliària

MEMÒRIA DEL PROJECTE

Alumne: Jesús Maria Vicente Pérez

Consultor: Pere Juanola Juanola

Enginyeria Tècnica en Informàtica de Gestió
GENER 2012



RESUMEN

L'empresa immobiliària *Sostre per Tothom* ens ha fet l'encàrrec de construir un magatzem de dades a partir d'un fitxer de dades en format text amb la informació del seu mercat immobiliari durant un cert període de temps, amb la finalitat de, mitjançant consultes i eines de *Business Intelligence*, explotar aquesta informació.

Les passes seguides per tal d'assolir aquest objectiu han estat:

- Pel disseny conceptual de la base de dades s'han triat el *fet* i les *dimensions*, la granularitat amb la que es volen fer les anàlisis i s'han determinat els atributs, descriptors, jerarquies i mesures.
- Pel disseny lògic s'ha triat un model multidimensional en forma d'estrella amb els preus dels immobles com a taula de *fet* i el temps, la zona i les característiques dels immobles com a taules de dimensions.
- Pel disseny físic s'ha establert l'estructura de les taules que conformen el magatzem de dades i els índex per tal d'optimitzar els temps de resposta de les consultes.
- Per al sistema de gestió de la base de dades s'ha utilitzat *Oracle 10g Express Edition* i s'han realitzat els processos de creació de taules, extracció, transformació i càrrega de dades utilitzant els llenguatges *SQL i PL/SQL*.
- Per a l'obtenció de la informació s'ha fet servir l'eina d'explotació de dades *Oracle Discoverer* amb la que s'han construït l'àrea de negoci i el llibre de treball amb la finalitat de generar les consultes necessàries per cobrir les necessitats descrites per *Sostre per Tothom*

A partir de les consultes generades s'han explotat les dades en forma de diferents informes amb la informació sol·licitada:

- Immobles que han baixat de preu, els que han evolucionat els preus de venda i d'oferta al llarg del temps, els preus i les superfícies màxims, mínims i mitjans, els temps mitjans de venda, els immobles disponibles i venuts classificats per zona, tipologia i característiques...
- I també s'ha pogut determinar quines són les característiques tipus d'un immoble model tant espanyol com andorrà

PARAULES CLAU

Magatzem de dades, TFC, OLAP, Anàlisi multidimensional, ETL, Gestió immobiliària, Oracle, Consultes, Discoverer, SQL, PL/SQL



ÍNDEX

RESUMEN	2
1. INTRODUCCIÓ	5
1.1 El projecte	5
1.2 Justificació	5
1.3 Objectius	6
1.3.1. Objectius concrets del projecte	6
1.4 Enfocament i metodologia	7
1.5 Planificació temporal	8
1.5.1 Descripció de tasques	8
1.5.2 Calendari previst	9
1.5.3 Diagrama de Gantt	11
1.6 Productes obtinguts	12
1.7 Descripció dels altres apartats de la memòria	13
2. ANÀLISIS i DISSENY	14
2.1 Descripció	14
2.2 Informes a realitzar	14
2.3 Dades d'origen	15
2.4 Especificacions tècniques	18
2.5 Elements per a l'anàlisi multidimensional	19
2.6 Disseny conceptual	20
2.6.1 Triar el fet	20
2.6.2 Trobar el grànul escaient	20
2.6.3 Escollir les dimensions que s'utilitzaran en l'anàlisi	21
2.6.4 Trobar els atributs de cada dimensió	22
2.6.5 Distingir entre descriptors i jerarquies d'agregació	22
2.6.6 Decidir quines són les mesures que interessin	22
2.6.7 Definir cel·les	23
2.6.8 Explicitar les restriccions d'integritat	23
2.6.9 Estudiar la viabilitat	23
2.6.10 Disseny lògic	25
2.6.11 Disseny físic	26
2.6.12 Procediments ETL	27
3. IMPLEMENTACIÓ	28
3.1 Accés a la Base de Dades	28
3.2 Accés a l'eina Business Intelligence	29
3.2.1 Accés a Oracle Discoverer Administrator	29
3.2.2 Accés a Oracle Discoverer Desktop	31
3.2.3 Usuaris i paraules de pas	32
3.3 Procés de càrrega	32
3.4. Informes	41
3.4.1 Nombre d'immobles per zona, tipologia i característiques per mes i any	41
3.4.2 Llistat d'immobles que han baixat de preu per mes i any	42
3.4.3 Preus màxims, mínims i mitjans per zona, tipologia i característiques per mes i any	43
3.4.4 Metres quadrats màxims, mínims i mitjans per zona, tipologia i característiques per mes i any	45
3.4.5 Temps mitjans de venda per zona, tipologia i característiques per mes i any	46
3.4.6 Nombre d'immobles venuts i existents per zona, tipologia i característiques per mes i any	47
3.4.7 Llistat d'immobles que han baixat de preu versus el mes anterior i que s'han venut per mes i any	49
3.4.8 Pis-tipus espanyol i andorrà	50
4 CONCLUSIONS	52
5 LÍNIES DE FUTUR	53
6 GLOSSARI	54
7 BIBLIOGRAFIA	56



Índex de figures

<i>Figura 1. Diagrama de Gantt</i>	11
<i>Figura 2. fitxer que recull el catàleg d'immobles</i>	15
<i>Figura 3. Possibles valors del camp Tipus</i>	16
<i>Figura 4. Possibles valors del camp Planta</i>	16
<i>Figura 5. Possibles valors del camps preus per diferents mesos</i>	17
<i>Figura 6. Exemple agrupador territorial per 4 nivells</i>	17
<i>Figura 7. Arquitectura del programari</i>	18
<i>Figura 8. Arquitectura del maquinari</i>	19
<i>Figura 9. Passos a seguir per modelitzar el disseny</i>	20
<i>Figura 10. Nivell jeràrquic de les dimensions</i>	21
<i>Figura 11. Diagrama ER</i>	24
<i>Figura 12. Configuració usuari STP</i>	28
<i>Figura 13. Entorn de treball d'Oracle</i>	29
<i>Figura 14. Connexió amb l'usuari EUL</i>	29
<i>Figura 15. Selecció de l'entorn de treball de l'usuari EUL</i>	30
<i>Figura 16. Entorn de treball amb l'usuari EUL</i>	30
<i>Figura 17. Privilegis atorgats a l'usuari SPT</i>	31
<i>Figura 18. Connexió per l'usuari SPT</i>	31
<i>Figura 19. Assistent per l'entorn de treball amb l'usuari SPT</i>	32
<i>Figura 20. Fitxer de dades dels immobles catalog.txt</i>	32
<i>Figura 21. Disseny de l'informe 5.1</i>	41
<i>Figura 22. Exemple de resultat de l'informe 5.1</i>	42
<i>Figura 23. Disseny de l'informe 5.2</i>	42
<i>Figura 24. Exemple de resultat de l'informe 5.2</i>	43
<i>Figura 25. Disseny de l'informe 5.3</i>	44
<i>Figura 26. Exemple de resultat de l'informe 5.3</i>	44
<i>Figura 27. Disseny de l'informe 5.4</i>	45
<i>Figura 28. Exemple de resultat de l'informe 5.4</i>	45
<i>Figura 29. Disseny de l'informe 5.5</i>	46
<i>Figura 30. Exemple de resultat de l'informe 5.5</i>	47
<i>Figura 31. Disseny de l'informe 5.6</i>	48
<i>Figura 32. Exemple de resultat de l'informe 5.6</i>	48
<i>Figura 33. Disseny de l'informe 5.7</i>	49
<i>Figura 34. Exemple de resultat de l'informe 5.7</i>	49
<i>Figura 35. Resultat consulta de tipus d'inmoble model espanyol</i>	50



1. INTRODUCCIÓ

1.1 El projecte

El projecte que constitueix la base de aquest *Treball Fi de Carrera (TFC)*, consisteix en la construcció d'un magatzem de dades, com a eina de treball, per als tècnics de l'empresa immobiliària *Sostre per a tothom* que compta amb un extens catàleg d'immobles en venda a tot Espanya i Andorra.

Degut a la crisi del sector, *Sostre per a tothom* considera que li seria molt útil disposar d'un sistema d'informació que li permetés comparar diferents zones, tipologies de pisos i preus per intentar ajustar l'oferta i la demanda, obtenint així un avantatge competitiu respecte altres empreses del sector.

El problema amb el que es troba l'empresa és que, malgrat el gran volum d'informació del que disposa, no té un sistema informàtic prou òptim i el catàleg es troba emmagatzemat en un conjunt de fulls Excel.

Aquest magatzem ha de permetre automatitzar la recollida de la informació provinent d'una fulla de text i explotar aquesta informació amb eines de *Business Intelligence* adients.

1.2 Justificació

La justificació d'aquest projecte es basa en la petició realitzada per l'empresa immobiliària *Sostre per a tothom* que ha encarregat el disseny i la realització d'un magatzem de dades, que li serveixi d'ajuda en la presa de decisions que li permetin ajustar l'oferta i la demanda tenint com a base l'anàlisi de les zones geogràfiques, les tipologies dels pisos oferts, els preus de venda,...

Cal tenir en compte però, la importància que té la qualitat de la informació en els processos de presa de decisions i que les empreses que desitgen ser competitives han d'orientar les seves decisions en funció d'informació concreta i fiable.

És habitual que les empreses comptin amb diferents bases de dades organitzades per departaments o àrees de negoci, però segons augmenta el volum d'aquestes dades es fa cada vegada més difícil extreure informació global i resumida.

Es fa per tant necessari unificar les diferents fonts d'informació en un únic repositori, al qual s'incorpori només informació rellevant, ben organitzada, estructurada, integrada, i de fàcil explotació. Aquesta unificació s'aconsegueix amb la implantació de magatzems de dades.

El *Data Warehouse* o magatzem de dades sol·licitat i que gira entorn d'aquest projecte, pot definir-se com una col·lecció de dades orientat a temes, integrat, no volàtil, de temps variant, que s'utilitza fonamentalment per donar suport al procés de presa de decisions .

Segons la definició de *Bill Inmon*, un *Data Warehouse* ha de tenir les següents característiques:

INTEGRAT

Les dades que resideixen en el *Data Warehouse* han d'estar integrades en una estructura



consistent, el que implica que les possibles inconsistències existents entre els diversos sistemes operacionals han de ser eliminades.

TEMÀTIC

Les dades s'han d'organitzar per temes per facilitar el seu accés i permetre el seu enteniment per part dels usuaris finals, i només les dades necessàries per al procés de generació del coneixement del negoci en qüestió s'integren des de l'entorn operacional.

HISTÒRIC

El magatzem de dades es carrega amb els diferents valors que pren una variable en el temps, de manera que es generen històrics que són com fotografies dels diferents moments pels que ha s'ha passat. Això és fonamental per poder realitzar anàlisi sobre les tendències.

NO VOLÀTIL

El magatzem de dades pot ser llegit però no modificat. És a dir, la incorporació dels últims valors que poden prendre les diferents variables contingudes en ell, no ha d'influir sobre la informació que ja existia prèviament. En els següents apartats d'aquest document es descriuen les fases i els processos seguits en l'elaboració del magatzem de dades sol·licitat.

1.3 Objectius

L'objectiu principal del projecte és adquirir experiència en el disseny, construcció i explotació d'un magatzem de dades a partir de la informació disponible en una base de dades transaccional, així com la utilització d'eines que facilitin la generació d'informes.

En concret, es pretén aprofundir en els coneixements d'un sistema de gestió de bases de dades *Oracle 10g*, emprant el llenguatge *PL/SQL* mitjançant l'eina del *SQLDeveloper*, així com aprendre el tractament i control d'errors produïts en els procediments realitzats.

D'altra banda, es tindran en compte altres objectius necessaris en l'elaboració d'un TFC mostrant l'assoliment d'aprenentatge que s'ha dut a terme al llarg dels estudis de l'Enginyeria Tècnica en Informàtica de Gestió.

- Analitzar un problema complex de tipus pràctic transformant-lo en un projecte informàtic.
- Planificar i estructurar el desenvolupament del projecte mitjançant l'elaboració d'un pla de treball aplicant una metodologia adient.
- Treballar a fons els aspectes formals del desenvolupament de projectes.
- Sintetitzar una solució viable i realista al problema proposat, basada en l'anàlisi de requeriments.
- Elaborar una memòria del projecte segons una estructura prefixada.
- Elaborar una presentació del desenvolupament i resultats finals del projecte.

1.3.1. Objectius concrets del projecte

L'objectiu principal d'aquest projecte es pot descompondre en objectius més senzills, pràctics i fàcilment abordables:

PRIMER OBJECTIU

Crear un procés automatitzat d'importació de les dades facilitades per Sostre per a tothom que es



trobessin en format text, al sistema de gestió de bases de dades elegit *Oracle 10g*. Aquest procés d'importació ha de seleccionar només la informació d'interès per a les possibles consultes per tal de no sobrecarregar el sistema i ha d'evitar les dades que siguin clarament inconsistents o erronis. Per aconseguir aquest objectiu serà necessari una anàlisi prèvia de les dades que ajudarà a determinar la seva disposició i els possibles errors recurrents que es puguin trobar.

SEGON OBJECTIU

Dissenyar i posar en marxa un magatzem de dades que aprofiti la informació de partida facilitada per *Sostre per a tothom* i faciliti la generació de determinats informes que constitueixin una bona eina de presa de decisions.

TERCER OBJECTIU

L'aplicació desenvolupada ha de ser capaç de generar en un temps acceptable una sèrie d'informes automatitzats, que mostren determinats aspectes referents a les propietats en venda i a les ja venudes tal com es detalla més endavant en un altre apartat d'aquest mateix document. A més d'aquests informes preestablerts l'aplicació ha d'estar construïda de tal manera que faciliti la creació de consultes *ad-hoc*.

1.4 Enfocament i metodologia

La implantació d'un magatzem de dades en qualsevol àmbit en què sigui necessària la presa de decisions basades en informes, es pot descompondre una sèrie de passos o fases clarament diferenciades, algunes de les quals poden variar segons l'autor que es consulti.

En el desenvolupament d'aquest projecte es segueix la metodologia descrita en www.1keydata.com adaptant-la als requeriments i la temporalitat imposats per la signatura:

ANÀLISIS DE REQUERIMENTS

Aquesta primera fase del projecte consisteix a recollir els requeriments dels usuaris, que generalment no tenen coneixements tècnics i amb ells elaborar el document de requisits. D'aquesta primera fase ha de sortir també una planificació temporal i una aproximació al cub *OLAP* que serveix de base als informes.

PREPACIÓ SUPORT FÍSIC

En aquesta fase es configura i es deixa en funcionament tot el suport físic del magatzem de dades. Això inclou el maquinari necessari i el sistema de gestió de bases de dades a utilitzar.

GENERACIÓ DEL MODEL DE DADES

Dels requeriments dels usuaris s'obté un model lògic que és traslladat a un model físic de dades. Una part important d'aquest pas és la correcta identificació de les fonts de dades.

ETL

La fase d'*ETL* (Extraction, Transformation, Loading) és una de les fases més llargues i comprèn l'extracció de les dades, la seva transformació i depuració i la càrrega dels mateixos en la base de dades.

DISSENY DEL CUB OLAP

En aquesta fase es defineix ja de forma més precisa el cub *OLAP* que es deriva dels requeriments dels usuaris i dels informes que s'han de generar.

DESENVOLUPAMENT DEL FRON-END

Durant aquesta fase es defineix el *Front End* a emprar. Generalment se sol optar per una eina



comercial, encara que també pot ser dissenyat a mida. En aquest projecte s'ha optat per emprar *Oracle Discoverer*.

AJUST DEL RENDIMENT

En aquesta fase es verifiquen i ajusten els rendiments dels processos que més temps i potència de càlcul poden consumir. Generalment l'anàlisi se centra en els processos *ETL*, les consultes i els informes.

ASSEGUAMENT DE LA QUALITAT

En aquesta fase l'equip de desenvolupament, juntament amb els representants dels futurs usuaris verifica que el programari desenvolupat compleix les funcions desitjades.

POSTA EN MARXA

En aquesta fase es posa en marxa el magatzem de dades en condicions de producció.

MANTENIMENT

Inclou les tasques de manteniment posteriors a la posada en producció.

MILLORES DE FUTUR

Contempla els possibles afegits no recollits en el projecte inicial i que caldria tenir en compte curt o llarg termini.

1.5 Planificació temporal

Per poder realitzar adequadament la planificació del projecte, cal conèixer en primer lloc les tasques a realitzar, distribuir correctament i adaptar-les a les dates de desenvolupament proposades.

En aquest apartat prenent com a base les fases típiques de la implantació d'un magatzem de dades comentades en el punt anterior, es comparen i combinen aquestes amb les fases proposades en el programa de l'assignatura i finalment, s'adapten i integren les fases resultants a les dates claus previstes al del pla docent. Finalment es realitza una distribució de la càrrega de treball dins de cada fase per aconseguir una planificació equilibrada.

Disposar d'una planificació temporal és vital a l'hora de l'elaboració de qualsevol projecte. Dóna una visió global de l'abast temporal del mateix. En aquest punt es fa un estudi de cadascuna d'aquestes tasques, el temps disponible i, en funció d'aquestes dades i de les que presenta el treball (dates de lliurament) es marquen les fites que s'hauran d'assolir.

1.5.1 Descripció de tasques

En resum, les tasques que s'ha dut a terme el desenvolupament d'aquest projecte, dividit en Proves d'Avaluació Continuada, *PAC*'s. han estat:

PAC1

- Descripció del projecte: es realitza una descripció del projecte a partir de l'enunciat donat.
- Descripció de les tasques a realitzar: es descriuen les tasques a realitzar.
- Planificació temporal: es planifiquen les tasques a realitzar així com la seva durada.
- Anàlisi preliminar de requeriments: es prepara un document d'anàlisi preliminar (no detallat) amb l'enumeració i breu descripció dels elements d'anàlisi identificats.
- Revisió del document: es revisen els documents a entregar.



PAC2

- Revisió requeriments: es revisen els requeriments per a la creació del disseny conceptual.
- Disseny conceptual: es realitza un disseny conceptual de la base de dades.
- Disseny lògic: es fa una transformació del disseny conceptual a un disseny lògic.
- Preparació programari (*Oracle 10g*): es prepara el programari *Oracle 10g*, per a la seva utilització.

PAC3

- Construcció del magatzem de dades: base de dades, càrregues, etc.
- Instal·lació de l'eina d'exploració de dades.
- Construcció dels informes i anàlisi de la informació.

ENTREGA FINAL

- Redacció de la memòria: es redacta la memòria final del projecte.
- Creació de la presentació virtual (*Power Point*): es crea la presentació virtual del projecte.
- Revisió i modificació escaient: es revisa tot el projecte i es modifiquen aquells punts que en l'espai temporal siguin factibles de ser modificats i millorats.

1.5.2 Calendari previst

A la següent taula es presenten les tasques anteriorment descrites, amb el detall de la seva durada, data d'inici i data fi. De la mateixa manera queden reflectides el resum dels dies dels principals esdeveniments i s'afegeix quina precedència necessita cadascuna de les fites.

Cal destacar que aquesta taula presenta una aproximació bastant detallada de l'abast del projecte però cal tenir en compte que en determinats moments calgui reajustar-la.

ID	Nom de la tasca	Durada	Inici	Fi	Precedents
1	1. Inici del projecte	3 dies	26/09/11	28/09/11	
2	1.1 Descàrrega i estudi del material de l'assignatura	1 dies	26/09/11	26/09/11	
3	1.2 Cerca bibliogràfica	2 dies	27/09/11	28/09/11	
4	2. Elaboració del' esborrany de la PAC1	6 dies	28/09/11	03/10/11	
5	2.1. Descàrrega i lectura de l'enunciat del projecte	1 dia	28/09/11	28/09/11	
6	2.2. Recerca d'informació	1 dia	28/09/11	28/09/11	5
7	2.3. Consulta de dubtes sobre l'enunciat	1 dia	28/09/11	28/09/11	5
8	2.4. Instal·lació del programari necessari	1 dia	29/09/11	29/09/11	2
9	2.5. Elaboració de l'esborrany del Pla de treball	2 dies	30/09/11	01/10/11	
10	2.6. Trobada presencial	1 dia	01/10/11	01/10/11	
11	2.7. Elaboració de l'esborrany de l'anàlisi preliminar de requeriments	2 dies	01/10/11	02/10/11	



12	2.8.Lliurament de l'esborrany de la PAC1	1 dia	03/10/11	03/09/11	9
13	3.Elaboració de la PAC1	2 dies	04/10/11	05/10/11	
14	3.1.Revisió de les correccions indicades a l'esborrany de la PAC1	1 dia	04/10/11	04/09/11	
15	3.3.Elaboració del Pla de treball definitiu	1 dia	04/10/11	04/10/11	
16	3.4.Elaboració de l'anàlisi preliminar de requeriments	1 dia	04/10/11	04/10/11	
17	3.5.Lliurament de la PAC1	1 dia	05/10/11	05/10/11	12,15,16
18	4.Elaboració de l'esborrany de la PAC2	28 dies	06/10/11	01/11/11	
19	4.1.Lectura detallada dels enunciats	1 dia	06/10/11	06/10/11	
20	4.2.Redacció de la introducció del projecte	2 dies	07/10/11	08/10/11	
21	4.3.Anàlisi i estudi de les dades d'origen	4 dies	09/10/11	12/10/11	
22	4.4.Cerca d'informació sobre el disseny tècnic	3 dies	12/10/11	14/10/11	
23	4.5.Elaboració de l'esborrany del document d'anàlisi de requeriments, basat en l'anàlisi preliminar de la PAC1	18 dies	15/10/11	01/11/11	19
24	4.6.Elaboració de l'esborrany de disseny tècnic (conceptual i físic)	18 dies	15/10/11	01/11/11	19
25	4.7.Redacció dels objectius del projecte	6 dies	27/10/11	01/11/11	
26	4.8.Lliurament del esborrany de la PAC2	1 dia	01/11/11	01/11/11	23,24
27	5.Elaboració de la PAC2	8 dies	02/11/11	09/11/11	
28	5.1.Revisió de les correccions indicades a l'esborrany de la PAC2	1 dia	02/11/11	02/11/11	
29	5.2.Elaboració del document definitiu d'anàlisi de requeriments i de disseny tècnic	6 dies	03/11/11	08/11/11	28
30	5.3.Lliurament de la PAC2	1 dia	09/11/11	09/11/11	29
31	6.Elaboració de l'esborrany de la PAC3	34 dies	10/11/11	13/12/11	
32	6.1.Lectura detallada dels enunciats	1 dia	10/11/11	10/11/11	
33	6.2.Construcció del magatzem de dades: bases de dades, càrregues, etc.	34 dies	10/11/11	13/12/11	
34	6.3.Instal·lació de l'eina d'explotació de dades	2 dies	11/11/11	12/11/11	
35	6.4.Construcció dels informes i anàlisi de la informació	34 dies	10/11/11	13/12/11	
36	6.5.Elaboració de l'esborrany de la PAC3	7 dies	07/12/11	13/12/11	32,35
37	6.6.Redacció del Glossari i bibliografia del projecte	7 dies	07/12/11	13/12/11	
38	6.7.Lliurament de l'esborrany de la PAC3	1 dia	13/12/11	13/12/11	36
39	7.Elaboració de la PAC3	7 dies	14/12/11	21/12/11	
40	7.1.Revisió de les correccions indicades a l'esborrany de la PAC3	1 dia	14/12/11	14/12/11	
41	7.2.Depuració dels informes	6 dies	15/12/11	20/12/11	40



42	7.3.Elaboració del document definitiu de la PAC3	6 dies	15/12/11	20/12/11	
43	7.4.Elaboració de la presentació	6 dies	15/12/11	20/12/11	
44	7.5.Lliurament de la PAC3	1 dia	21/12/11	21/12/11	41,42
45	8. Elaboració de la Memòria Final i Presentació	20 dies	22/12/11	10/01/12	
46	8.1.Elaboració de la memòria final i presentació	18 dies	22/12/11	08/12/11	44
47	9.2.Revisió i format final	1 dia	09/01/12	09//01/12	
48	9.3 Entrega final (memòria i presentació)	1 dia	10/01/12	10/01/12	46,47
49	9. Debat	4 dies	23/01/12	26/01/12	
50	9.1.Intervenció en el debat	4 dies	23/01/12	26/01/12	

Taula que emmagatzema els tipus d'immobles objecte de l'estudi.

1.5.3 Diagrama de Gantt

A continuació es mostra un diagrama de *Gantt*, amb totes les tasques reflectides, i el temps estimat per desenvolupar cadascuna de les tasques, així com el temps total del projecte.

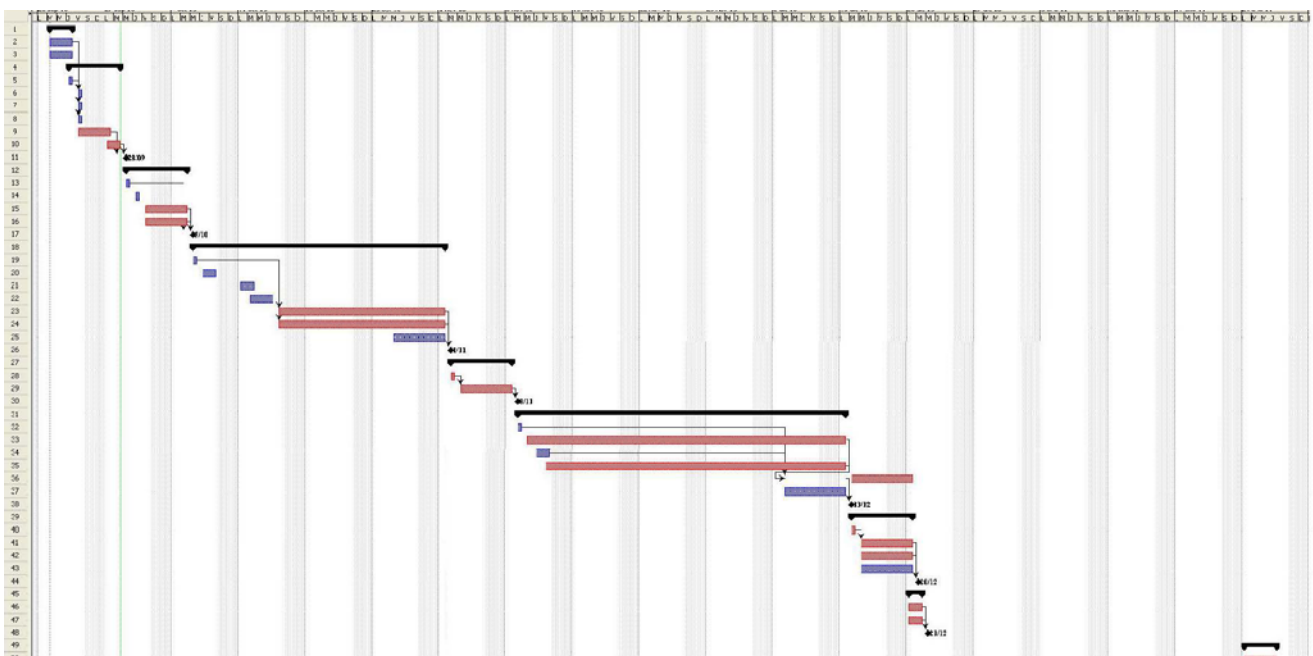


Figura 1. Diagrama de Gantt



1.6 Productes obtinguts

Al llarg del projecte s'han elaborat els següents documents i productes:

Pla de treball: Descripció general del projecte i els seus objectius, metodologia a seguir per desenvolupar el projecte i la seva planificació temporal, recursos de maquinari, programari bibliogràfics disponibles, possibles incidències que es puguin produir al llarg del projecte i les solucions i precaucions preses.

Anàlisi preliminar de requeriments: Descripció i anàlisi prèvia de les dades proporcionades, primera identificació dels elements d'anàlisi i descripció preliminar dels informes que es lliuraran.

Anàlisi de requeriments: Anàlisi de les dades, identificació dels elements d'anàlisi definitius i descripció dels informes que es lliuraran al client.

Disseny: Disseny conceptual, lògic i físic del model multidimensional emprat per desenvolupar el magatzem de dades, descripció del procés d'extracció, transformació i càrrega de dades (ETL) i primera aproximació de la metodologia de construcció de consultes.

Implementació: Construcció de la base de dades i elaboració dels informes.

Memòria final: El document actual. Descripció detallada del projecte, els seus objectius, la seva planificació, la metodologia de desenvolupament emprada, els resultats obtinguts i les conclusions.

Presentació: Presentació gràfica del projecte que dona una visió global i resumida de la feina desenvolupada.

Producte: Carpeta que conté tots els elements necessaris per a l'execució del projecte.

- *crear_tables.sql*: script SQL per a la creació de les taules del magatzem de dades.
- *crear_taula_temp.sql*: script SQL per a la creació de la taula temporal, *CATALEG_TEMP* prèvia a la càrrega final.
- *catalog.ctl*: fitxer amb les instruccions per a la càrrega de la taula *CATALEG_TEMP*. S'utilitza juntament amb el fitxer de text proporcionat per la immobiliària que ha de portar el nom *catalog.txt*
- *catalog.txt*: catàleg d'immobles proporcionat per l'empresa. S'utilitza, juntament amb el fitxer *catalog.ctl*, per carregar la taula *CATALEG_TEMP*.
- *correccio_errors.sql*: script SQL de correcció d'errors detectats a les dades del fitxer *catalog.dat*. S'executa sobre les dades carregades a la taula *CATALEG_TEMP*
- *carrega_dades.pls*: script PL/SQL de càrrega de dades al magatzem de dades
- *jesvic_mv.vdi*: disc dur virtual amb la base de dades i l'entorn de treball instal·lat.
- *TFC-Sostre per a tothom.eex*: espai de negoci.
- *Explotacio_STP.DIS*: llibre de treball.



1.7 Descripció dels altres apartats de la memòria

ANÀLISIS I DISSENY

En aquest apartat es fa una primera descripció del fitxer de dades proporcionat per l'empresa, *catleg.txt*. A partir de la informació d'aquest fitxer es realitza el disseny conceptual del model multidimensional on es tria el *fet* que es vol analitzar, les *dimensions* o punts de vista des dels que s'analitzen el *fet*, la granularitat i es defineixen els atributs, descriptors, jerarquies i mesures.

A partir del model conceptual es fa un disseny lògic on els fets i les dimensions s'estructuren en taules i es defineixen les seves claus primàries i foranes, que relacionen unes taules amb altres. També es fa el disseny físic, on es decideix com estructurar la informació en les taules per tal d'optimitzar el temps de resposta a les consultes.

IMPLEMENTACIÓ

A part del disseny de la base de dades, en aquest apartat es descriu la seva implementació, que consta de la creació de taules i l'extracció, transformació i càrrega de dades.

Finalment es fa una descripció del disseny de les consultes i informes que donen resposta als requeriments plantejats per l'empresa *Sostre per Tothom* i es presenten els resultats obtinguts.



2. ANÀLISIS I DISSENY

2.1 Descripció

Com ja s'ha comentat *Sostre per a tothom* voldria disposar d'un sistema d'informació que li permetés comparar diferents zones, tipologies de pisos i preus per intentar ajustar l'oferta i la demanda, obtenint així un avantatge competitiu respecte altres empreses del sector.

Per aquest fet, se'ns ha encarregat la construcció d'un magatzem de dades, com a eina de treball, per als tècnics de l'empresa immobiliària que compta amb un extens catàleg d'immobles en venda a tot Espanya i Andorra.

Aquest magatzem ha de permetre automatitzar la recollida de la informació provinent d'una fulla de text i explotar aquesta informació amb eines de *Business Intelligence* adients.

2.2 Informes a realitzar

Els informes que ha demanat *Sostre per tothom* que caldria implementar serien:

- Dintre d'una temporalitat a nivell de mes i any.
 - Nombre d'immobles per zona, tipologia i característiques
 - Llistat d'immobles que han baixat de preu. Mostrar l'evolució de preus per zona, tipologia i característiques
 - Preus màxims, mínims i mitjans per zona, tipologia i característiques
 - Metres quadrats màxims, mínims i mitjans per zona, tipologia i característiques
 - Temps mitjans de venda per zona, tipologia i característiques
 - Nombre d'immobles venuts i existents per zona, tipologia i característiques
- Un conjunt predefinit d'informes on es mostri la informació demanada i qualsevol altra que creiem que pot resultar útil per a l'empresa.
- Quin és el pis-tipus espanyol i andorrà, és a dir, quina tipologia, característiques i preu representa el pis estàndard d'aquests.



2.3 Dades d'origen

Sostre per a tothom proporciona el seu catàleg d'immobles en un únic fitxer de text resultat de la concatenació dels diversos fulls *Excels* exportat a un fulla de text, amb el que es treballarà a partir d'ara. Cal advertir que tant el manteniment dels fulls com la concatenació dels mateixos s'ha realitzat mitjançant un procediment manual i podrien haver-hi errors. També s'indica que cada línia del fitxer de text conté les següents columnes:

- Codi i Zona: codificació de la territorialitat
- Tipus: tipologia de l'immoble
- Planta: planta de l'immoble (si es tracta d'un pis)
- Dormitoris: nombre de dormitoris
- Metres: metres quadrats de l'immoble

I una llista de camps del tipus preu mes any, que conté el preu de l'immoble per al mes i anys detallats. Si el camp està en blanc, significa que l'immoble no està a la venda en aquell període; això vol dir, que un immoble que tenia preu en un mes-any anterior al que es troba en blanc, ha estat venut en aquell mes i any.

Finalment, també expliquen que cada conjunt de línies d'immobles d'una zona es troben agrupats amb unes capçaleres que descriuen textualment la zona. Aquesta zona es compon d'una jerarquia de 4 grups territorials que s'agrupen d'esquerre cap a la dreta. En el magatzem de dades voldran poder agregar i desglossar la informació segons aquesta jerarquia territorial de 4 nivells.

Si importem el fitxer a una fulla d'*Excel* es veu la distribució de les dades (figura 2)

	A	B	C	D	E	F	G	H	I	J	K	L
1	codi	zona	tipus	planta	dormitoris	metres	preu gener 2	preu febrer 2	preu març 2	preu abril 2	preu maig 2	preu juny 2
2	#											
3	#	::: tarragona										
4	#	::: tarragona ::: tarragonès										
5	#	::: tarragona ::: tarragonès ::: àrea de sant salvador										
6	VC0000000567779	43-XXX-02-006-XX-XX	pisos	6	3	70	147000	147000	147000	147000	147000	147000
7	VC0000000886504	43-XXX-02-006-XX-XX	pisos	3	3	70						
8	VP0000000527778	43-XXX-02-006-XX-XX	pisos	1	2	77	282175	282175	282175	282175	282175	282175
9	VP0000000631496	43-XXX-02-006-XX-XX	pisos	3	3	70	150000	150000	150000	150000	150000	150000
10	VP0000000635015	43-XXX-02-006-XX-XX	pisos	3	3	83	170000	170000	170000	170000	170000	170000
11	VP0000000818875	43-XXX-02-006-XX-XX	pisos	2	2	55	200000	200000	200000	200000	200000	200000
12	VW0000000109677	43-XXX-02-006-XX-XX	pisos	5	4	120	420000	420000	420000	420000	420000	420000
13	VW0000000641639	43-XXX-02-006-XX-XX	pisos	3	3	72			140000	140000	140000	140000
14	VW0000000653734	43-XXX-02-006-XX-XX	pisos	1	3	102				260000	260000	260000

Figura 2. fitxer que recull el catàleg d'immobles.

Cal destacar però, que el manteniment dels fulls com la concatenació dels mateixos s'ha realitzat mitjançant un procediment manual i podrien haver-hi errors.

Cada línia del fitxer de text conté les següents columnes:

- Codi : codificació territorial.
 - Seria un valor alfanumèric: ex: VC0000000567779
- Zona: codificació de la territorialitat
 - Seria un valor alfanumèric, ex: 43-XXX-02-006-XX-XXX
- Tipus: tipologia de l'immoble
 - Els valors poden ser: *àtics*, *dúplexs*, *chalets*, *estudios*, *pisos* o sense tipologia

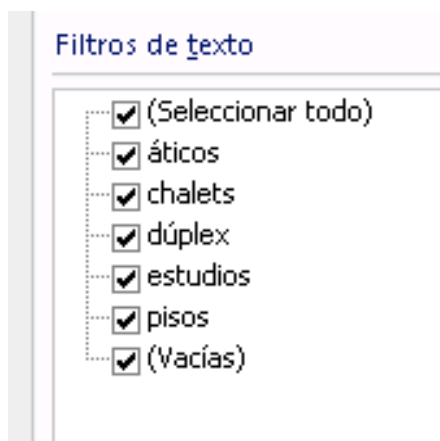


Figura 3. Possibles valors del camp Tipus

- Planta: planta de l'immoble (si es tracta d'un pis)
 - Els valors serien alfanumèrics, anirien des de la planta 1 fins la 99 i resta de nivells, *baixos*, *entresol*, ... i fins tot sense valor.

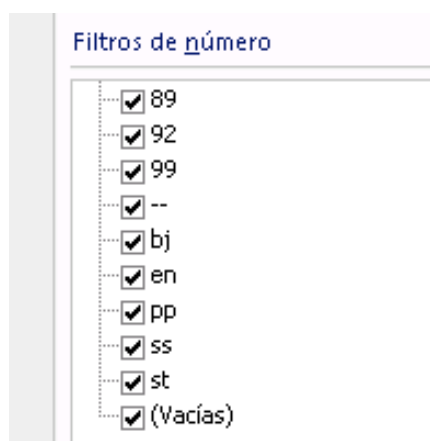


Figura 4. Possibles valors del camp Planta

A tenir en compte:

- ➔ *El valor 99 es pot considerar com error*
- ➔ *S'adopta posar el valor "-- " i '99' als valors buits*

- Dormitoris: nombre de dormitoris
 - Seria un valor numèric enter, comprès entre 0 i 99 que pot estar també sense valor.

A tenir en compte:

- ➔ *El valor 99 es pot considerar com error*



- Metres: metres quadrats de l'immoble
 - Seria un valor numèric enter, comprès entre 0 i 99999 i que també pot estar buit.

A tenir en compte:

➔ El valor més alt es de trobat es 77885 m2

- Una llista de camps del tipus preu mes any, que conté el preu de l'immoble per al mes i anys detallats. Si el camp està en blanc, significa que l'immoble no està a la venda en aquell període; això vol dir, que un immoble que tenia preu en un mes-any anterior al que es troba en blanc, ha estat venut en aquell mes i any.
 - Seria valors enters i que anirien cronològicament des de el "preu gener 2006" fins el "preu gener 2008". Per tant hauria 25 possibles valors de preus.

	G	H	I	J	K	L	M	N	O	P
1	preu gener 2006	preu febrer 2006	preu març 2006	preu abril 2006	preu maig 2006	preu juny 2006	preu juliol 2006	preu agost 2006	preu setembre 2006	preu octubre 2006
2										
3										
4										
5										
6	147000	147000	147000	147000	147000	147000	147000	147000	147000	147000
7										
8	282175	282175	282175	282175	282175	282175	282175	282175	282175	282175
9	150000	150000	150000	150000	150000	150000	150000	150000	150000	150000
10	170000	170000	170000	170000	170000	170000	170000	170000	170000	170000
11	200000	200000	200000	200000	200000	200000	200000	200000	200000	200000
12	420000	420000	420000	420000	420000	420000	420000	420000	420000	420000
13			140000	140000	140000	140000	140000	140000	140000	140000
14				260000	260000	260000	260000	260000	260000	260000
15					220000	220000	220000	220000	220000	220000
16								150000	150000	150000
17										

Figura 5. Possibles valors del camps preus per diferents mesos

- Cada conjunt de línies d'immobles d'una zona es troben agrupats amb unes capçaleres que descriuen textualment la zona. Aquesta zona es compon d'una jerarquia de 4 grups territorials que s'agrupen d'esquerre cap a dreta. En el magatzem de dades voldran poder agrupar i desglossar la informació segons aquesta jerarquia territorial de 4 nivells.
 - Seria un camp de text que caldrà tractar per extreure aquests 4 nivells

46	VW0000000835608	43-148-02-002-01->XXX	pisos	1	1
47	VW0000000859202	43-148-02-002-01->XXX	pisos	3	3
48	# ::: tarragona ::: tarragonès ::: tarragona ::: eixample				
49	VC0000000350131	43-148-02-002-02->XXX	pisos	1	1
50	VC0000000556342	43-148-02-002-02->XXX	pisos	7	1
51	VC0000000705230	43-148-02-002-02->XXX	pisos	1	4

Figura 6. Exemple agrupador territorial per 4 nivells

A tenir en compte:

- ➔ Els primer agrupador territorial admet 9 valors. Representa les províncies de Catalunya (*Barcelona, Tarragona, Lleida i Girona*), les províncies del País Valencià (*Castellón, Valencia i Alicante*), Andorra y Mallorca
- ➔ Hi ha agrupacions fins a 6 nivells, s'adopta només emmagatzemar 4 nivells, com es veurà a l'apartat 3, Implementació.

2.4 Especificacions tècniques

Per dur a terme el projecte i garantir el seu funcionament l'eina desenvolupada així com el sistema de gestió de bases de dades ha de disposar, al menys, sobre *Windows XP Professional* i *Windows 2003 Server*.

ARQUITECTURA DEL PROGRAMARI

Els programaris utilitzats han estat:

- Base de dades *Oracle 10g Express Edition*: Per a la gestió de la base de dades.
- *SQL*Loader*: Per a la càrrega de dades a la taula CATALEG a partir del fitxer *catalog.dat* proporcionat per l'empresa.
- *SQL Developer*: Per a la construcció de taules i l'execució del procés ETL sobre la base de dades.
- *Oracle BI Discoverer Administrator*: Per a la construcció de l'àrea a de negoci.
- *Oracle BI Discoverer Desktop*: Per a la construcció del llibre de treball.

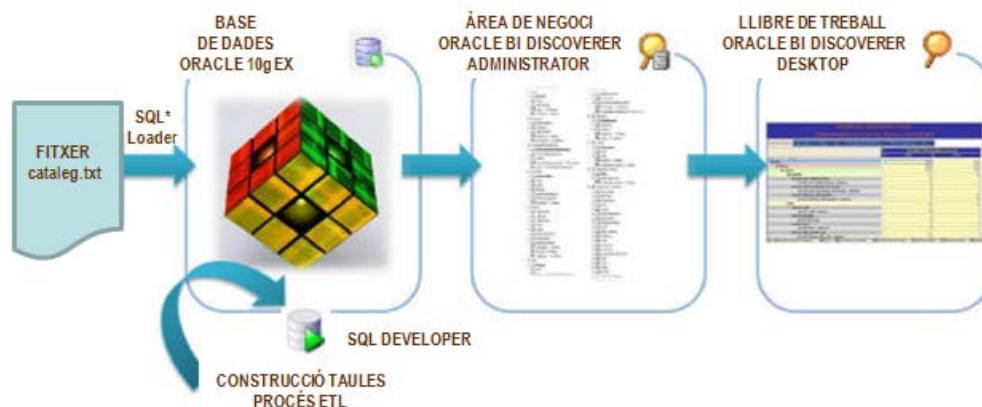


Figura 7. Arquitectura del programari

ARQUITECTURA DEL MAQUINARI

Per la implantació del magatzem de dades a l'empresa *Sostre per a tothom*, la solució més adequada seria centralitzar la base de dades a un servidor de manera que els diferents treballadors de l'empresa s'hi poguessin connectar des dels seus punts de treball per realitzar les consultes pertinents. En el següent diagrama (*Figura 8*) es mostra l'arquitectura del programari necessari per portar a terme aquesta implantació.



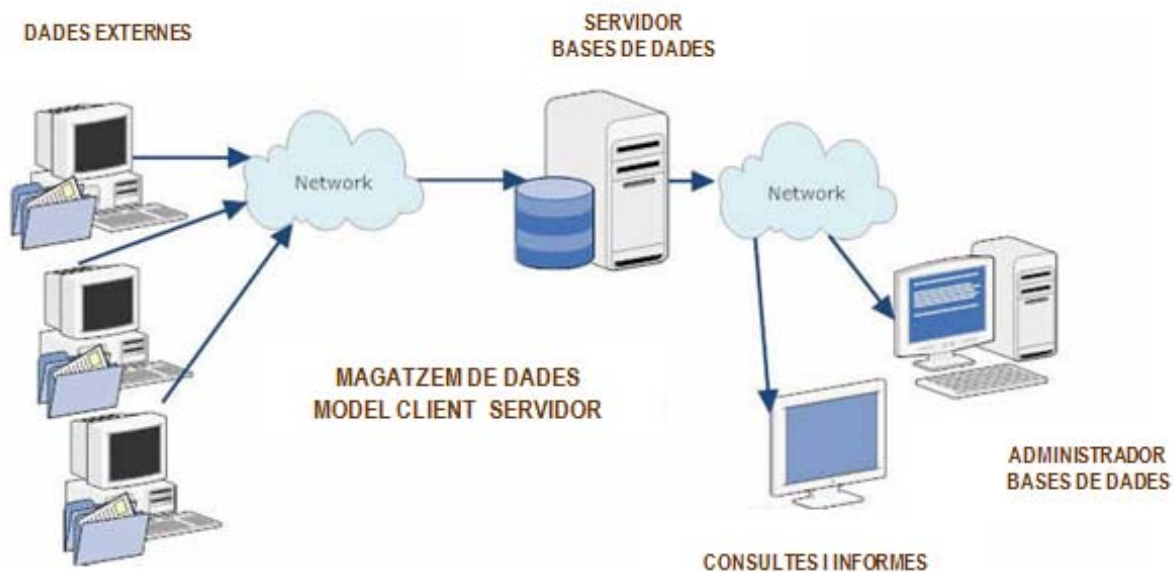


Figura 8. Arquitectura del maquinari

2.5 Elements per a l'anàlisi multidimensional

Dels requeriments de *Sostre per tothom*, més el conjunt d'informes sol·licitats, així com de l'anàlisi de les dades subministrades, podem extreure les següents aproximacions:

Fets: podem considerar com a fet les dades en una data determinada d'un immoble

Mesures: la mesura del fet serà el preu, que estarà en els camps:

- Preu
- Mes
- Any

Dimensions: segons la nostra taula de fets, podem considerar 3 dimensions

- Data: període de temps de la consulta
- Grups territorials: capçaleres de la codificació territorial
- Immobles: tipus i característiques de l'immoble

Descriptors: per a cada dimensió tenim els següents atributs o descriptors

- Data: mes i any
- Grups territorials: 4 capçaleres que descriuen la zona
- Tipus: no té atribut específic



2.6 Disseny conceptual

En aquest apartat es descriu la metodologia seguida que aborda el disseny conceptual i es detallen els passos seguits per a obtenir l'estrella que modelitzarà el disseny, en el qual s'estructuraran les dades que es recullen dels fulls d'*Exce/s*.

Es tracta d'un procés iteratiu de consta de nou passos.



Figura 9. Passos a seguir per modelitzar el disseny

2.6.1 Triar el fet

Com a primer pas cal identificar el fet que constitueix el nucli del model que es definirà. En aquest projecte es tracta de la creació d'un magatzem de dades que permeti automatitzar la recollida de la informació provinent de fulls d'*Exce/s* i explotar aquesta informació. Per tant, s'haurà de saber quines dades volem mostrar als informes. Un primer anàlisi permet indicar que el nostre fet es pot definir com preus d'immobles en un període determinat, que pertanyen a una zona determinada.

L'anàlisi de funcionalitats s'obté, en un principi, de la informació que ens demana *Sostre per tothom*, i que fonamentalment es:

- Llistat d'immobles que han baixat de preu. Mostrar l'evolució de preus per zona, tipologia i característiques
- Preus màxims, mínims i mitjans per zona, tipologia i característiques
- Nombre d'immobles venuts i existents per zona, tipologia i característiques
- Quin és el pis-tipus espanyol i andorrà, és a dir, quina tipologia, característiques i preu representa el pis estàndard d'aquests.

Així, podem considerar com a fet les dades a una data determinada d'un immoble i des de aquí desgranar tota la informació que ens sigui sol·licitada a cada informe.

2.6.2 Trobar el grànul escaient

El grànul és l'individu últim que volem analitzar, la cel·la més petita que volem tenir disponible. És molt important definir bé aquest nivell de detall, ja que triar un grànul massa gran representa perdre informació. Però triar-lo massa petit pot representar malbaratar espai o arribar a fer inviable el projecte per excés de dades.

Les dades se'ns presenten com un conjunt de fulls d'*Exce/s* amb els preus dels immobles durant un període comprès des de gener del 2006 fins gener de 2008.

Aquest és el nivell de granularitat idoni per la nostra taula de fets tenint en compte que l'hem de separar per períodes de temps.



2.6.3 Escollir les dimensions que s'utilitzaran en l'anàlisi

Les dimensions en un magatzem de dades es defineixen com la informació general complementària a cada un dels registres de la taula de fets.

D'acord amb l'anàlisi preliminar de requeriments, lliurat a la PAC1, podem identificar tres dimensions:

- **Data:** període de temps objecte de la consulta que es vol realitzar que conté les dades emmagatzemades a la nostra taula de fets. Estarà formada per anys i mesos, compresos entre gener del 2006 i gener del 2008.
- **Grup territorial:** conjunt d'agrupacions que formen el total de dades a tractar. Hi ha fins a 4 nivells.
- **Immables:** conjunt de característiques de cada immoble.

El nivell jeràrquic de les dades definit per les dimensions descrites seria:

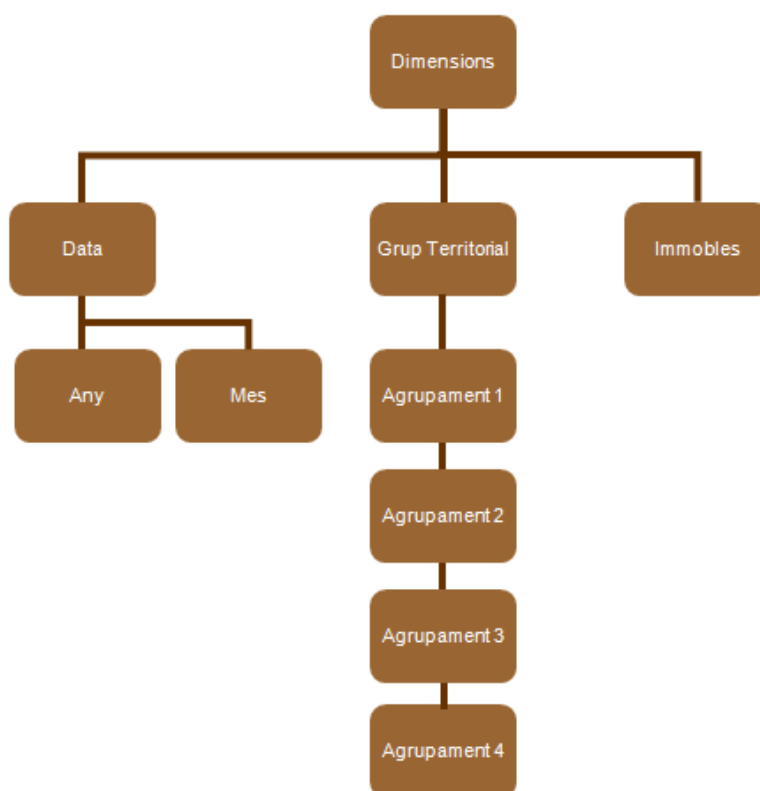


Figura 10. Nivell jeràrquic de les dimensions



2.6.4 Trobar els atributs de cada dimensió

Un cop identificades les dimensions caldrà estudiar els atributs d'aquestes dimensions:

- **Data:** es disposa de dades del període gener 2006 fins a gener del 2008, agrupades per mesos, per tant l'atribut característic serà la data, arribant a nivell de mes, encara que a mesura que anem avançant en el desenvolupament podem considerar altres períodes de temps necessaris per l'extracció dels informes, per exemple trimestres o semestres.
- **Grup territorial:** contindrà els atributs de 4 nivells d'agrupament, donat que es té la informació estructurada 'un a molts' i això ens proporcionarà la funcionalitat d'extreure els informes agrupats pels 4 nivells d'agrupament.
- **Tipus:** no tindrà cap atribut específic, a banda d'un codi identificador.

2.6.5 Distingir entre descriptors i jerarquies d'agregació

Entre els atributs que hi ha en una dimensió, es distingeixen de dos tipus: els que s'utilitzen per a agrupar i els que serviran simplement per a seleccionar. En aquesta fase del projecte s'estableixen quins atributs actuaran com a descriptors, permetent fer seleccions basades en els seus valors i quins permetran fer agrupacions i conformar jerarquies.

Així, en la dimensió **Data** podem establir una jerarquia d'agregació en que la base siguin els mesos i aquests s'agrupin en anys.

Igualment, **Grup territorial** presenta una jerarquia clarament marcada, en 4 nivells d'agrupació.

Agrupament 1: Província
Agrupament 2: Comarca
Agrupament 3: Cuitat
Agrupament 4: Districte

→ Pot haver-hi casos que coincideixen els agrupaments, la província, la comarca i la ciutat.

2.6.6 Decidir quines són les mesures que interessin

Podem considerar com a mesures del nostre projecte les següents:

- Preu
 - Mes
 - Venda
- Metres quadrats
- Número de dormitoris

Aquesta informació es presentarà en la taula de fets ja que és una mesura que serà variable, doncs vindrà definida, sobretot, per la data objecte de l'estudi.



2.6.7 Definir cel·les

En aquest pas es defineixen quines cel·les resultants de l'anàlisi poden ser interessants per al nostre model, i, per tant, han de ser emmagatzemades.

En aquest cas només es contempla una cel·la que reflecteix el fet analitzat. S'utilitza una cel·la que representarà les dades d'un immoble en una data determinada.

2.6.8 Explicar les restriccions d'integritat

En aquest pas es defineixen les Bases i estableixen les restriccions d'integritat. El conjunt inicial de Dimensions (*Data*, *Grup territorial* i *Inmobles*), que es troben després de definir el gràdul ha de donar lloc a una Base. Substituint Dimensions en aquesta segons les dependències funcionals que hi ha amb la resta de Dimensions, s'obtenen les altres Bases de l'espai.

2.6.9 Estudiar la viabilitat

Un cop un finalitzat el disseny conceptual, cal veure si l'*Estrella* és pot implementar o no, i per això s'estima l'espai que ocuparan totes les dades una vegada emmagatzemades en el sistema.

Inmobles: es parteix d'unes dades inicials que contenen 9 províncies amb dades referides a comarques, ciutats i districtes que ens donen un total de 92.555 registres. Si pensem amb una longitud de registre de 256 bytes, aproximadament, estem parlant d'uns 0,80 MB, amb un increment anual (si ens continuen passant un fitxer mensual) d'un 10%. Es a dir, una quantitat que no ens tindria que preocupar en absolut.

Rendiment: és un volum de dades molt petit amb el que el rendiment no té que ser un problema per l'explotació de les sol·licituds que podrem fer a l'*Estrella*.

En conclusió, l'*estrella* a nivell de volum de dades i rendiment és totalment viable.



D'acord amb els passos anteriors, el model conceptual resultant seria el següent:

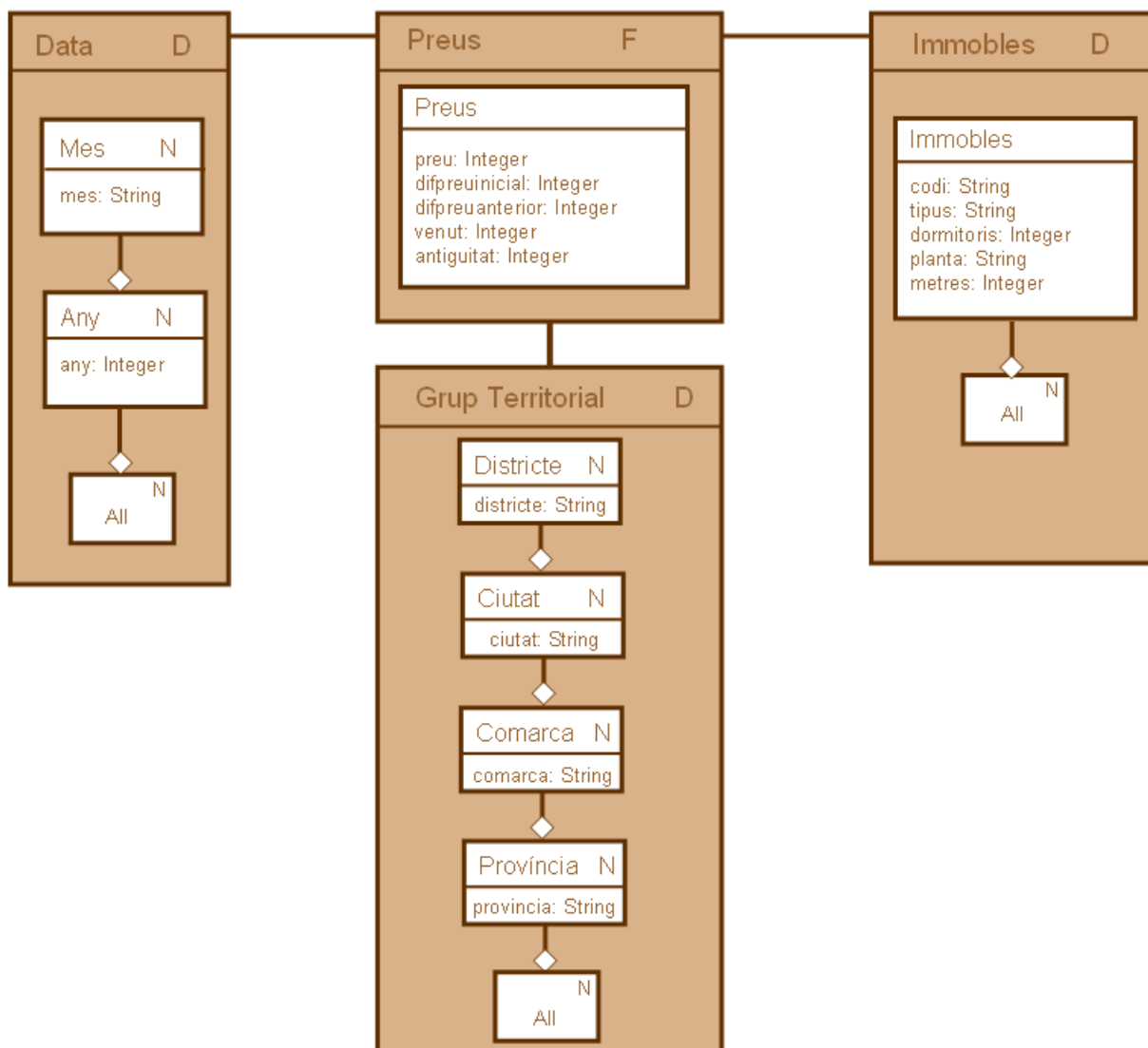


Figura 11. Diagrama ER



2.6.10 Disseny lògic

Un cop un plantejat el disseny conceptual, es presenta ara el disseny lògic. Partint de l'anàlisi de l'estrella definida en els punts anteriors es necessita una taula per al fet (en què cada fila representa una cel·la de l'espai multidimensional) i una taula més per a cadascuna de les dimensions que formen el model.

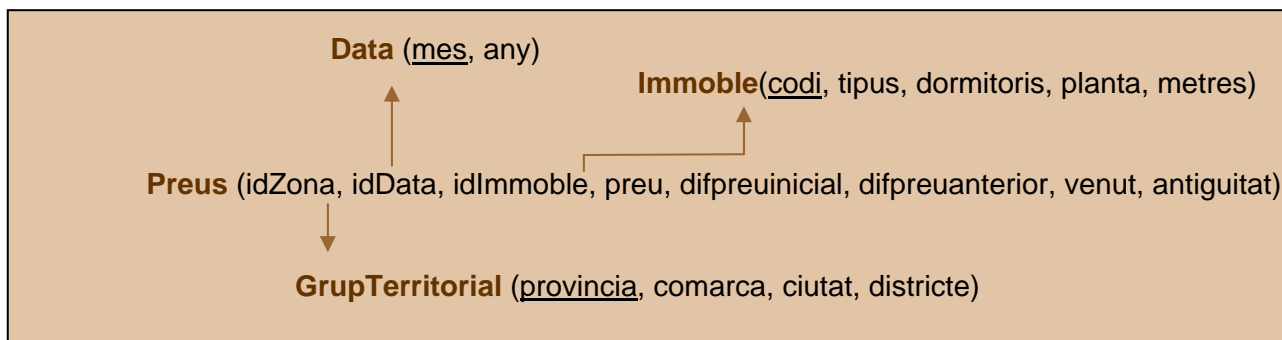
La taula de fet estarà lligada per claus foranes amb les taules de dimensió. Cada clau forana apunta de la taula de fet cap a una de les taules de les dimensions. Aplicant aquesta metodologia descrita s'obté el model lògic següent:

Preus (idZona, idData, idImmoble, preu, difpreuinicial, difpreuanterior, venut, antiguitat)

Data (mes, any)

GrupTerritorial (província, comarca, ciutat, districte)

Immoble(codi, tipus, dormitoris, planta, metres)



2.6.11 Disseny físic

Un cop ja tenim les relacions que componen la nostra base de dades i també sabem quin tipus de consultes volem executar, farem el disseny físic tenint al cap que cal obtenir un bon temps de resposta a les consultes.

Les taules obtingudes a partir del disseny lògic i que poden ser gestionades per un motor de bases de dades, són les següents:

Preus

Nom Columna	Tipus de dades	Nul	Valor per defecte	Clau primària
idPreus	NUMBER(10,0)	No	-	1
idZona	NUMBER(10,0)	No	-	-
idData	NUMBER(10,0)	No	-	-
idImmoble	NUMBER(10,0)	No	-	-
preu	NUMBER(10,0)	No	-	-
venut	NUMBER(1,0)	Yes	-	-
antiguitat	NUMBER(5,0)	No	-	-
dpreuanterior	NUMBER(10,0)	Yes	-	-
difpreuinitial	NUMBER(10,0)	Yes	-	-

Taula que emmagatzema els preus per dates objecte de l'estudi.

Data

Nom Columna	Tipus de dades	Nul	Valor per defecte	Clau primària
idData	NUMBER(4,0)	No	-	-
mes_data	VARCHAR(20)	No	-	-
any_data	NUMBER(4,0)	No	-	-

Taula que emmagatzema les dates objecte de l'estudi.

Immables

Nom Columna	Tipus de dades	Nul	Valor per defecte	Clau primària
idImmoble	NUMBER(10,0)	No	-	1
codi	VARCHAR(150)	No	-	-
tipus	VARCHAR(20)	No	-	-
dormitoris	NUMBER(5,0)	No	-	-
planta	VARCHAR(10)	No	-	-
metres	NUMBER(5,0)	No	-	-

Taula que emmagatzema els immables objecte de l'estudi

Grup Territorial

Nom Columna	Tipus de dades	Nul	Valor per defecte	Clau primària
idGt	NUMBER(4,0)	No	-	1
provincia	VARCHAR(150)	No	-	-
comarca	VARCHAR(150)	No	-	-
ciutat	VARCHAR(150)	No	-	-
districte	VARCHAR(150)	No	-	-

Taula que emmagatzema els grups territorials objecte de l'estudi.



2.6.12 Procediments ETL

ETL (Extracció, Transformació i Càrrega) és el procediment de manipulació de dades consistent en l'extracció dels mateixos a partir de diverses fonts, la seva transformació per tal d'adaptar-los al model i finalment la seva càrrega en les taules que conformen el magatzem de dades. A continuació es comenten, a grans trets, aquests tres apartats.

Extracció

Aquesta part, com s'explica a l'anàlisi de requeriments, anirà a càrrec de Sostre per tothom, encarregada de facilitar els llistats mensuals de les dades dels immobles que conformen Sostre per tothom

Aquests fitxers, subministrats en format de fulles d'Excels, han estat analitzats a l'anàlisi de requeriments, arribant a l'acord de depurar els errors i fer arribar, en les properes entregues, les dades en el format adequat.

Transformació

El principal error que es detecta als fitxers *Excel*, prèviament analitzat a l'anàlisi de requeriments, és la diversitat d'immobles en la que manquen dades, en la primera càrrega es depuraran els errors i s'intentarà consensuar que per càrregues futures es faciliti un fitxer amb els immobles correctament escrits i amb el seu grup territorial ben especificat.

En quant a la resta d'errors, de moment, per aquesta primera càrrega, es farà una transformació automatitzada mitjançant *PL/SQL*. Alguns dels principals errors detectats i que es presenten de forma ampliada a l'apartat 3.*Implmentació*, que es tractaran són:

- Es detecten més de 4 grups territorials, però només es carreguen 4 tal i com indica els requeriments del projecte
- Exemple: # :: alicante :: baix segura :: torrevieja :: centro ciudad :: puerto deportivo
- No es té en compte el cinquè nivell, "puerto deportivo"
- Transformar els valors nuls o '99' de 'planta' en valor '--'
- Transformar els valors nuls o '99' de 'dormitoris' en '0'
- Transformar els valors nuls de 'metres' en '0'

Càrrega

Un cop transformades les dades i realitzades les correccions, es procedeix a la càrrega de les mateixes. En el cas d'un magatzem de dades, aquesta càrrega esdevindrà tenint en compte que les dades no s'han de sobre escriure sinó acumular-se, ja que ens interessa en tot moment mantenir un històric.

El procés a seguir seria el següent:

- Llegir les dades de l'arxiu de text amb *PL/SQL*
- Inserir les dades en una taula temporal *Oracle*
- Transformar les dades segons les indicacions de l'apartat anterior
- Fer la inserció en les taules *Oracles* definitives



3. IMPLEMENTACIÓ

Una cop finalitzada la fase de disseny, a continuació es presenta la fase d' implementació, on es crearà la base de dades del magatzem de dades (taules, índexs, restriccions, seqüències, disparadors, etc.), el model amb l'eina de *Business Intelligence Oracle Discoverer*, la construcció dels processos de càrrega (ETL) i la càrrega de dades al magatzem de dades, així com la creació d'informes segons els indicadors sol·licitats per *Sostre per tothom*

3.1 Accés a la Base de Dades

Per crear el magatzem de dades s'utilitza l'usuari *SYSTEM* i la paraula de pas proporcionada per *Sostre per tothom*. Es crea l'usuari/esquema de la base de dades *Sostre per tothom*, (SPT) amb paraula de pas *uoc* i *tablespace* per defecte *USERS* i *tablespace* temporal *TEMP*. A aquest usuari se li atorguen els rols *CONNECT* i *RESOURCE*. Per accedir a la base de dades del magatzem, la cadena de connexió és *spt@//uoc:1521/XE*.

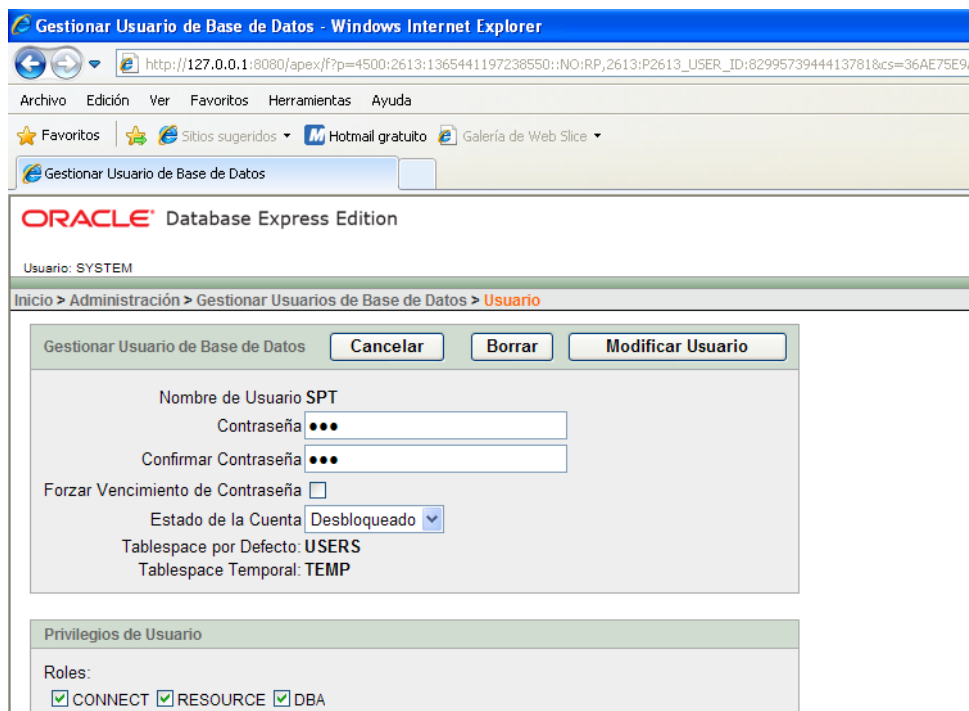


Figura 12. Configuració usuari STP

Un cop connectat, ja es pot accedir a tots els objectes que componen el model físic del magatzem de dades (taules, índex, seqüències, disparadors, paquets, procediments emmagatzemats, etc.).



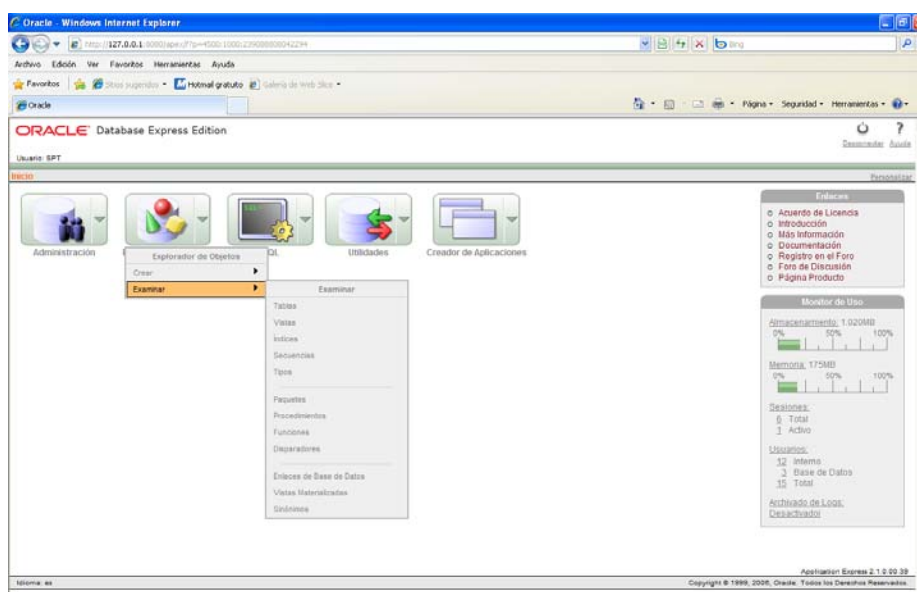


Figura 13. Entorn de treball d'Oracle

3.2 Accés a l'eina Bussines Intelligence

3.2.1 Accés a Oracle Discoverer Administrator

Per crear la capa final d'usuari (EUL), s'ha creat un usuari/esquema de bases dades *EUL* amb la paraula de pas *uoc* amb *tablespace* per defecte *USERS* i *tablespace* temporal *TEMP*. A aquest usuari se li han atorgat els rols de *CONNECT* i *RESOURCE*.

Aquest usuari s'utilitza única i exclusivament per la connexió a *Oracle Discoverer Administrator* i gestionar les diferents àrees de negoci o crear-ne de noves.

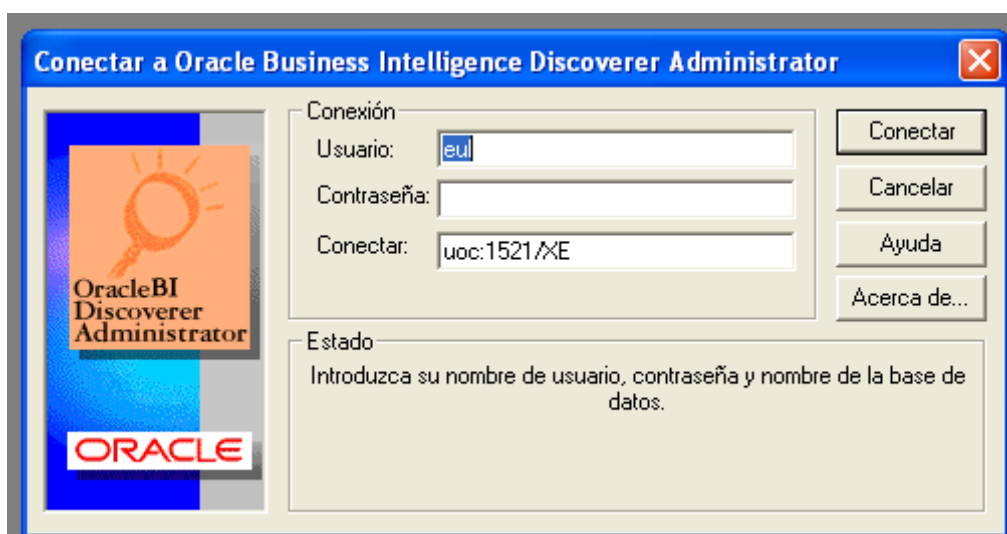


Figura 14. Connexió amb l'usuari EUL

En el cas d'aquest projecte s'ha creat l'àrea de negoci *Sostre per Tothom* on s'ha definit tot el model d'exploració del magatzem de dades (dades i jerarquies).

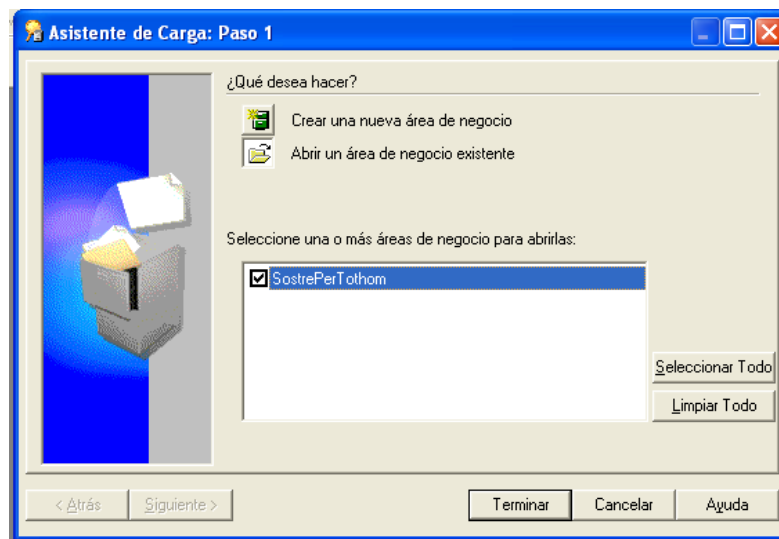


Figura 15. Selecció de l'entorn de treball de l'usuari EUL

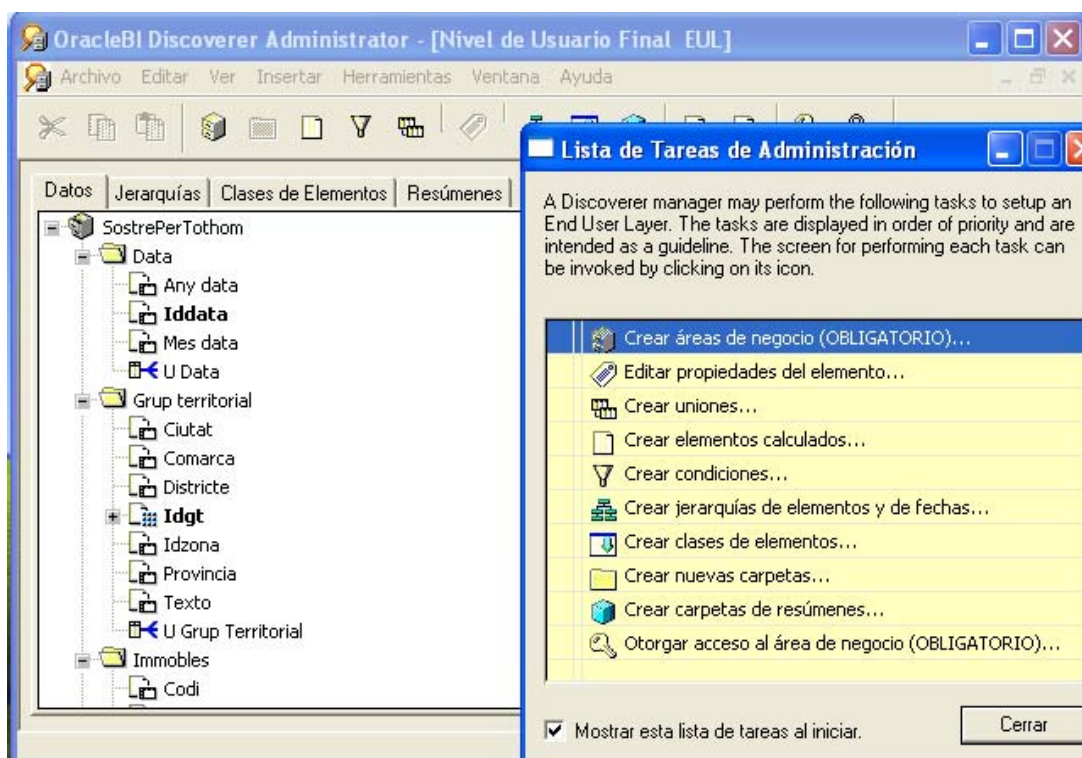


Figura 16. Entorn de treball amb l'usuari EUL

Per a aquest fet s'ha atorgat a l'usuari *EUL* el privilegi de *SELECT* sobre les taules *stp.grup_territorial*, *stp.inmoble*, *stp.data* i *stp.preus* ja que l'àrea de negoci utilitza aquest informació tal i com es mostrarà a continuació.



Per últim, s'han atorgat permís d'accés a l'àrea de negoci *Sostre per Tothom* a l'usuari *SPT* amb els privilegis que es mostren a continuació.

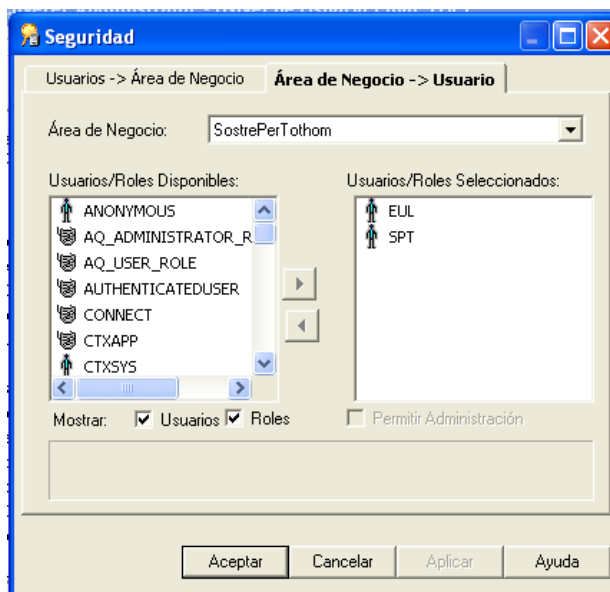


Figura 17. Privilegis atorgats a l'usuari SPT

3.2.2 Accés a Oracle Discoverer Desktop

Per accedir a Oracle Discoverer Desktop, s'utilitza l'usuari *SPT*.

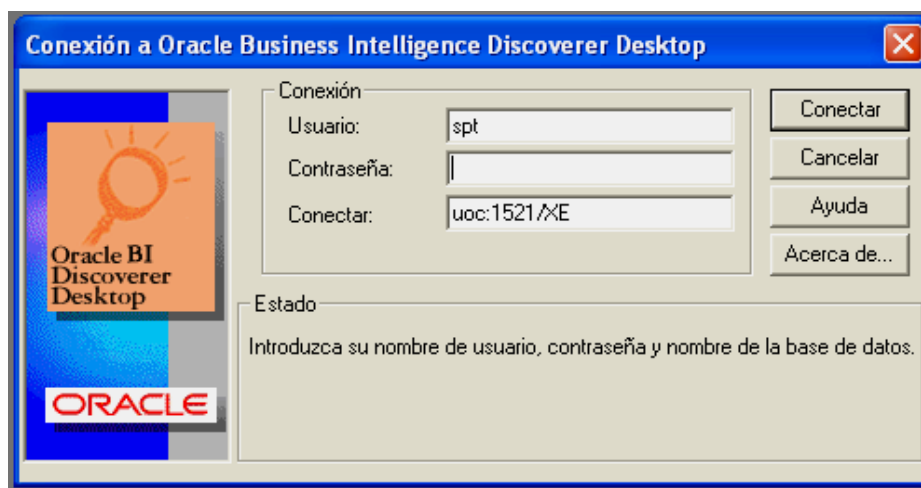


Figura 18. Connexió per l'usuari SPT

Un cop connectat, es pot crear el llibre de treball o seleccionar un d'existent.



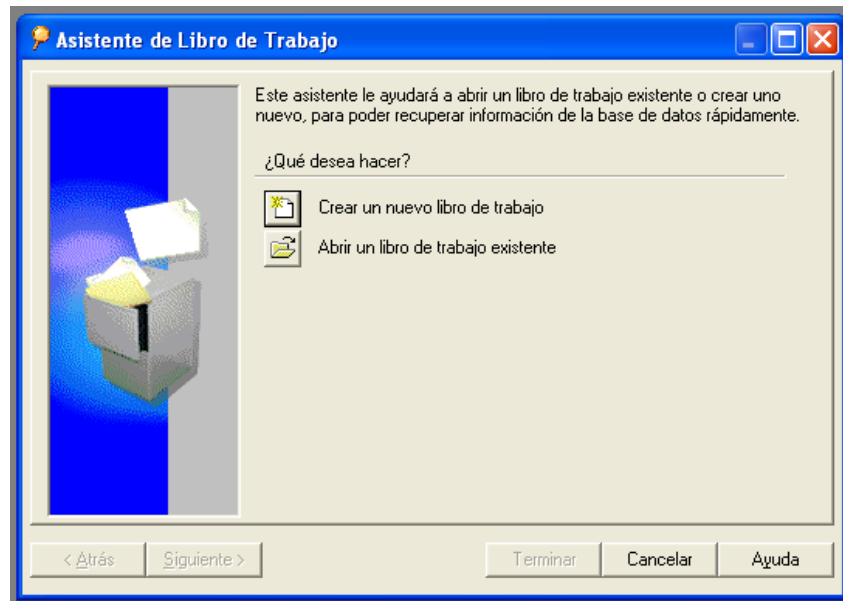


Figura 19. Assistent per l'entorn de treball amb l'usuari SPT

3.2.3 Usuaris i paraules de pas

Usuario	Paraula de pas	Observacions
system	uoc	Usuari administrador de la BBDD
eul	uoc	Usuari Oracle Discoverer Administrator
spt	uoc	Usuari BBDD i Oracle Discoverer Desktop

3.3 Procés de càrrega

El procés d'extracció, transformació i càrrega es divideix en els següents passos:

PAS 1

El primer pas es carregar les dades que proporciona *Sostre per Tothom* mitjançant el fitxer *cataleg.txt* en una taula temporal anomenada *CATALEG_TEMP*.

codi	zona	tipus planta	dormitoris	metres	"preu gener 2006"	"preu febrer 2006"	"preu març 2006"	"preu abril 2006"	"preu maig 2006"	"preu juny 2006"
#	:::	tarragona								
#	:::	tarragona	:::	tarragonès						
#	:::	tarragona	:::	tarragonès	:::	área de sant salvador				
VC0000000567779	43-XXX-02-006-XX-XXX	pisos 6	3	70	147000	147000	147000	147000	147000	147000
VC0000000886504	43-XXX-02-006-XX-XXX	pisos 3	3	70						
VP0000000527778	43-XXX-02-006-XX-XXX	pisos 1	2	77	282175	282175	282175	282175	282175	282175
VP0000000631496	43-XXX-02-006-XX-XXX	pisos 3	3	70	150000	150000	150000	150000	150000	150000
VP0000000635015	43-XXX-02-006-XX-XXX	pisos 3	3	83	170000	170000	170000	170000	170000	170000
VP0000000818875	43-XXX-02-006-XX-XXX	pisos 2	2	55	200000	200000	200000	200000	200000	200000

Figura 20. Fitxer de dades dels immobles *cataleg.txt*

- Es crea la taula i un índex **IDCODI** que enumera els registres de la taula de forma autonumèrica (i es crea una seqüència i un disparador per dur a terme aquesta acció)


```
(CODI CHAR "ltrim(rtrim(:CODI))",  
ZONA CHAR "ltrim(rtrim(:ZONA))",  
TIPUS CHAR "ltrim(rtrim(:TIPUS))",  
PLANTA CHAR "ltrim(rtrim(:PLANTA))",  
DORMITORIS,  
METRES,  
PREU_GENER_2006,  
PREU_FEBRER_2006,  
PREU_MARÇ_2006,  
PREU_ABRIL_2006,  
PREU_MAIG_2006,  
PREU_JUNY_2006,  
PREU_JULIOL_2006,  
PREU_AGOST_2006,  
PREU_SETEMBRE_2006,  
PREU_OCTUBRE_2006,  
PREU_NOVEMBRE_2006,  
PREU_DESEMBRE_2006,  
PREU_GENER_2007,  
PREU_FEBRER_2007,  
PREU_MARÇ_2007,  
PREU_ABRIL_2007,  
PREU_MAIG_2007,  
PREU_JUNY_2007,  
PREU_JULIOL_2007,  
PREU_AGOST_2007,  
PREU_SETEMBRE_2007,  
PREU_OCTUBRE_2007,  
PREU_NOVEMBRE_2007,  
PREU_DESEMBRE_2007,  
PREU_GENER_2008)
```

PAS 2

El següent pas és transformar la a taula els valors estranys, o incoherents.

- Àtics que es troben en una planta baixa (bj): el valor de la planta s'ha substituït pel valor indefinit (--)

```
UPDATE CATALEG_TEMP  
SET PLANTA = '--'  
WHERE TIPUS = 'àtics'  
AND PLANTA = 'bj';
```

- Xalets que no tenen la planta a 'st': El camp planta indica la planta en la que es troba l'immoble. Els xalets no es troben a cap planta per tant aquesta informació no és aplicable als xalets i per aquesta raó s'ha posat el valor de la planta a '- -' en els xalets.

```
UPDATE CATALEG_TEMP  
SET PLANTA = '- -'  
WHERE TIPUS = 'chalets'  
AND PLANTA = 'st';
```



- Immobles amb més de 20 dormitoris: S'han trobat alguns casos amb valors exageradament alts al camp dormitori. S'ha considerat 20 com el valor màxim de dormitoris i per tant tots el valor l'excedien s'han posat a NULL.

```
UPDATE CATALEG_TEMP  
SET DORMITORIS = NULL  
WHERE DORMITORIS >20;
```

- Immobles amb superfícies de menys de 10 m2 o més de 3000 m2: Els valors de superfície exageradament baixos o exageradament alts s'han considerat erronis i per tant s'han posat a NULL.

```
UPDATE CATALEG_TEMP  
SET METRES = NULL  
WHERE METRES < 10;  
  
UPDATE CATALEG_TEMP  
SET METRES = NULL  
WHERE METRES >1000  
AND ((TIPUS ='pisos')  
OR (TIPUS ='àtics')  
OR (TIPUS ='estudios'));
```

Executant el fitxer adjunt *transformacio_tables.sql* des d'*Oracle SQL Developer* aquest errors es poden corregir de manera automàtica.

Abans d'executar cada script de correcció d'errors es fa un recompte dels registres erronis. Aquest mateix recompte executat després de la correcció de l'error ha de retornar el valor 0 si la correcció s'ha realitzat amb èxit.

PAS 3

A continuació és creen les taules GRUP TERRITORIAL, IMMOBLE, PREU i DATA amb el fitxer *taules.sql*

```
CREATE TABLE "GRUP_TERRITORIAL"  
( "IDGT" NUMBER NOT NULL ENABLE,  
  "PROVINCIA" VARCHAR2(150),  
  "COMARCA" VARCHAR2(150),  
  "CIUTAT" VARCHAR2(150),  
  "DISTRICTE" VARCHAR2(150),  
  "IDZONA" VARCHAR2(150),  
  "TEXT0" VARCHAR2(150),  
  CONSTRAINT "PK_GRUP_TERRITORIAL" PRIMARY KEY ("IDGT")  
ENABLE  
)  
  
CREATE TABLE "IMMOBLES"  
( "IDIMMOBLE" NUMBER(10,0) NOT NULL ENABLE,  
  "CODI" VARCHAR2(150),  
  "TIPUS" VARCHAR2(20),  
  "PLANTA" VARCHAR2(10),  
  "DORMITORIS" NUMBER(5,0),  
  "METRES" NUMBER(5,0),
```



```
        CONSTRAINT "PK_IMMOBLES" PRIMARY KEY ("IDIMMOBLE") ENABLE
    )

CREATE TABLE "PREUS"
( "IDPREUS" NUMBER(10,0),
  "IDZONA" NUMBER(10,0),
  "IDDATA" NUMBER(10,0),
  "IDIMMOBLE" NUMBER(10,0),
  "PREU" NUMBER(10,0),
  "VENUT" NUMBER(1,0),
  "ANTIGUITAT" NUMBER(5,0),
  "DPREUANTERIOR" NUMBER(10,0),
  "DIFPREUINICIAL" NUMBER(10,0)
)

CREATE TABLE "DATA"
( "IDDATA" NUMBER(10,0) NOT NULL ENABLE,
  "MES_DATA" VARCHAR2(20),
  "ANY_DATA" NUMBER(4,0),
  CONSTRAINT "PK_DATA" PRIMARY KEY ("IDDATA") ENABLE
)
```

I a la taula **DATA** es carreguen els valors dels 25 mesos que es disposa de dades mitjançant el fitxer *carregar_data.sql*

```
insert into DATA values (1, 'GENER', 2006);
insert into DATA values (2, 'FEBRER', 2006);
insert into DATA values (3, 'MARÇ', 2006);
insert into DATA values (4, 'ABRIL', 2006);
insert into DATA values (5, 'MAIG', 2006);
insert into DATA values (6, 'JUNY', 2006);
insert into DATA values (7, 'JULIOL', 2006);
insert into DATA values (8, 'AGOST', 2006);
insert into DATA values (9, 'SETEMBRE', 2006);
insert into DATA values (10, 'OCTUBRE', 2006);
insert into DATA values (11, 'NOVEMBRE', 2006);
insert into DATA values (12, 'DESEMBRE', 2006);
insert into DATA values (13, 'GENER', 2007);
insert into DATA values (14, 'FEBRER', 2007);
insert into DATA values (15, 'MARÇ', 2007);
insert into DATA values (16, 'ABRIL', 2007);
insert into DATA values (17, 'MAIG', 2007);
insert into DATA values (18, 'JUNY', 2007);
insert into DATA values (19, 'JULIOL', 2007);
insert into DATA values (20, 'AGOST', 2007);
insert into DATA values (21, 'SETEMBRE', 2007);
insert into DATA values (22, 'OCTUBRE', 2007);
insert into DATA values (23, 'NOVEMBRE', 2007);
insert into DATA values (24, 'DESEMBRE', 2007);
insert into DATA values (25, 'GENER', 2008);
```

PAS 4

Una vegada transformades les dades contingudes a la taula **CATALEG_TEMP** s'omplen les taules **GRUP TERRITORIAL**, **IMMOBLE** i **PREU** executant l'script *carrega_temp.pls* des d'Oracle SQL Developer.



Amb aquest script es fa un recorregut pels 93.360 registres de la taula *CATALEG_TEMP*. Segons si el registre indica una zona o un immoble es carregen les dades en unes taules o altres.

Si el registre indica una zona (indicador #), el primer pas és dividir la cadena en els diferents subnivells, prenent la seqüència ':::' com a indicador des de on s'ha de fer la divisió. Si la cadena té menys de 4 subnivells, aquests s'omplen amb les paraules província, comarca i ciutat respectivament, sempre que el següent registre no sigui ampliat de l'anterior. Si en canvi la cadena té més de 4 subnivells, els últims es fusionen en un de sol per denominar el subgrup territorial.

CAS 1: 2 grups territorials seguits

```
# ::: tarragona ::: tarragonès ::: tarragona  
# ::: tarragona ::: tarragonès ::: tarragona ::: part alta
```

→ Es descarta la primera perquè la que ve a continuació amplia l'anterior.

CAS 2: Només 3 grups territorials

```
# ::: tarragona ::: tarragonès ::: àrea de torreforta
```

→ En aquest cas es carrega les dades de la província, comarca i ciutat i al districte es repeteix el valor de la ciutat.

CAS 3: Només 2 grups territorials

```
# ::: tarragona ::: conca de barberà
```

→ En aquest es carrega la província i la comarca, i es repeteix la ciutat i el districte com la comarca.

CAS 4: Només 1 grup territorial

```
# ::: andorra
```

→ En aquest cas totes es guarden els quatre grups territorials amb el mateix nom.

CAS 5: 5 grups territorials

```
# ::: alicante ::: alacantí ::: alicante / alacant ::: distrito 3 ::: san gabriel / la florida
```

→ En aquest cas es s'uneix el 5è grup territorial amb el 4r separat per un guionet '-'

Cal també destacar que en el moment d'inserir les dades a la taula *GRUP_TERRITORIAL* es substitueix el caràcter Ž per un espai ja que s'intueix que era l'apòstrof i per no tenir problemes amb les cometes senzilles amb l'SQL.

```
ciutat:= REPLACE(SUBSTR (cadena, 1, INSTR (cadena, ':::',2, 1)-2),'Ž',' ');
```

Nota: Aquesta transformació es podria realitzar en la taula de *CATALEG_TEMP* però fent-ho en aquest pas s'estalvia recórrer la taula sencera un altre cop.



Si el registre indica un immoble, primer s'omple la taula **IMMOBLE** amb les dades de codi, tipus, planta, nombre de dormitoris i metres quadrats. Després es fa un recorregut pels camps que indiquen el preu de l'immoble per a cada mes i s'introdueix un registre d'oferta corresponent a aquest immoble en la taula **PREU**. Hi haurà tants registres d'oferta per un mateix immoble com mesos estigui en oferta aquest immoble. Quan es llegeix un primer preu, s'emmagatzema aquest valor com a preu inicial per poder fer els càlculs de diferència de preu respecte al preu inicial i respecte el mes anterior. Quan es llegeix l'últim preu s'indica que l'immoble s'ha venut.

Una vegada finalitzat el procés de càrrega, tenim 25 registres a la taula **DATA**, 618 a la taula **GRUP TERRITORIAL**, 92.555 registres a la taula **IMMOBLE** i 1.091.388 registres a la taula **PREUS**.

Si es realitza una comprovació de elements no repetits a les taules es comprova que a la taula **GRUP TERRITORIAL** hi ha 5 zones que estan repetides (per tant la taula es quedarà amb 613 registres) Gràcies a aquesta comprovació, s'eliminen els duplicats i es tornen a calcular els índexs mitjançant el fitxer

Exemple:

```
select count(*), provincia, comarca,ciutat,districte from
grup_territorial group by provincia,comarca,ciutat,districte
having count(*) >1
```

```
2 barcelona baix llobregat sud àrea de sant vicenç dels horts
àrea de sant vicenç dels horts
2 barcelona vallès occidental àrea de st.llorenç savall -
castellar vallès àrea de st.llorenç savall - castellar vallès
2 barcelona baix llobregat sud àrea de santa coloma de
cervelló àrea de santa coloma de cervelló
2 barcelona vallès occidental àrea de ripollet - montcada i
reixac àrea de ripollet - montcada i reixac
2 barcelona vallès occidental àrea de barberà del vallès -
badia àrea de barberà del vallès - badia
```

```
delete from grup_territorial where idgt=330
delete from grup_territorial where idgt=432
delete from grup_territorial where idgt=318
delete from grup_territorial where idgt=412
delete from grup_territorial where idgt=425
```

```
update preus set idzona=329 where idzona=330
update preus set idzona=417 where idzona=425
update preus set idzona=426 where idzona=432
update preus set idzona=411 where idzona=412
update preus set idzona=301 where idzona=318
```

Per últim el codi del procediment de càrrega seria:

```
create or replace procedure "CARREGA_TEMP"
is
cursor c_f is
select idcodi, codi from cataleg_temp order by idcodi;
idcodiT cataleg_temp.idcodi%type;
codiT cataleg_temp.codi%type;
cadena cataleg_temp.codi%type;
```



```

codiseguent cataleg_temp.codi%type;
zonaT cataleg_temp.zona%type;
provinciaT grup_territorial.provincia%type;
comarcaT grup_territorial.comarca%type;
ciutatT grup_territorial.ciutat%type;
districteT grup_territorial.districte%type;
codiaux cataleg_temp.idcodi%type;
zonaaux cataleg_temp.zona%type;
cont number;
cont_im cataleg_temp.idcodi%type;
continua_cadena boolean;
idcodiseguent cataleg_temp.idcodi%type;
preu_gener_2006T cataleg_temp.preu_gener_2006%type;
dormitorisT cataleg_temp.dormitoris%type;
metresT cataleg_temp.metres%type;
plantaT cataleg_temp.planta%type;
tipusT cataleg_temp.tipus%type;
type arr_preus is table of number(10,0) index by binary_integer;
preusv arr_preus;
i cataleg_temp.idcodi%type;
venutT preus.venut%type;
difpreuinitialT preus.difpreuinitial%type;
antiguitatT preus.antiguitat%type;
dpreuanteriorT preus.dpreuanterior%type;
preuinitial number;
continuar boolean;

begin
cont:=0;
cont_im:=0;
open c_f;
loop
fetch c_f into idcodiT, codiT;
exit when c_f%notfound;
continua_cadena:=true;
if SUBSTR(codiT,1,1) = '#' then
idcodiseguent:=idcodiT+1;
select codi into codiseguent from cataleg_temp where idcodi=idcodiseguent;
if SUBSTR(codiseguent,1,1) <> '#' then
cont:=cont+1;
cadena:=codiT;
if (INSTR (cadena, ':::', 7, 1)-7)> 0 then
provinciaT:=REPLACE(SUBSTR (cadena, 7, INSTR (cadena, ':::', 7, 1)-
7),'Ž', ' ');
else
provinciaT:=REPLACE(SUBSTR (cadena, 7, LENGTH(cadena)),'Ž', ' ');
insert into grup_territorial values (cont, provinciaT, provinciaT,
provinciaT, provinciaT, idcodiT, codiT);*/
continua_cadena:=false;
end if;
if continua_cadena then
cadena := SUBSTR (cadena, INSTR (cadena, ':::', 7, 1) + 4);
if (INSTR (cadena, ':::', 7, 1))> 0 then
comarcaT:=REPLACE(SUBSTR (cadena, 1, INSTR (cadena, ':::',2, 1)-
2),'Ž', ' ');
else
comarcaT:=REPLACE(SUBSTR (cadena, 1, LENGTH(cadena)),'Ž', ' ');
insert into grup_territorial values (cont, provinciaT, comarcaT,
comarcaT, comarcaT, idcodiT, codiT);
continua_cadena:=false;
end if;
end if;
if continua_cadena then
cadena := SUBSTR (cadena, INSTR (cadena, ':::', 7, 1) + 4);
if (INSTR (cadena, ':::', 7, 1))> 0 then
ciutatT:=REPLACE(SUBSTR (cadena, 1, INSTR (cadena, ':::',2, 1)-
2),'Ž', ' ');
else
ciutatT:=REPLACE(SUBSTR (cadena, 1, LENGTH(cadena)),'Ž', ' ');
insert into grup_territorial values (cont, provinciaT, comarcaT,
ciutatT, ciutatT, idcodiT, codiT);
continua_cadena:=false;
end if;
end if;
end if;
end if;

```

```
        if continua_cadena then
            cadena := SUBSTR (cadena, INSTR (cadena, '::::', 7, 1) + 4);
            districteT:=REPLACE(REPLACE(SUBSTR (cadena, 1,
            LENGTH(cadena)), '::::', '-'), 'Ž', ' ');
            insert into grup_territorial values (cont, provinciaT, comarcaT,
            ciutatT, districteT, idcodiT, codiT);
        end if;
    end if;
commit;

else
cont_im:=cont_im+1;
select tipus, planta, dormitoris, metres,
preu_gener_2006, preu_febrer_2006, preu_març_2006, preu_abril_2006, preu_maig_2006,
preu_juny_2006, preu_juliol_2006, preu_agost_2006, preu_setembre_2006, preu_octubre_2006,
preu_novembre_2006, preu_desembre_2006, preu_gener_2007, preu_febrer_2007, preu_març_2007,
preu_abril_2007, preu_maig_2007, preu_juny_2007, preu_juliol_2007, preu_agost_2007,
preu_setembre_2007,preu_octubre_2007,preu_novembre_2007,preu_desembre_2007,preu_gener_2008
into tipusT, plantaT, dormitorisT, metresT, preusv(1), preusv(2), preusv(3), preusv(4),
preusv(5), preusv(6), preusv(7), preusv(8), preusv(9), preusv(10), preusv(11), preusv(12),
preusv(13), preusv(14), preusv(15), preusv(16), preusv(17), preusv(18), preusv(19),
preusv(20), preusv(21), preusv(22), preusv(23), preusv(24), preusv(25)
from cataleg_temp
where idcodi=idcodiT;
insert into immobles values (cont_im, codiT, tipusT, plantaT, dormitorisT, metresT);
continuar:=true;
preuinicial:= null;
antiguitatT:= 0;
venutT:= 0;
for i in 1..25 loop
    if preusv(i) is not null then
        if preuinicial is null then
            preuinicial:= preusv(i);
            dpreuanteriorT:=null;
            difpreuinicialT:=null;
        else
            dpreuanteriorT:= preusv(i) - preusv(i-1);
            difpreuinicialT:= preusv(i) - preuinicial;
        end if;
        antiguitatT:=antiguitatT+1;
        insert into preus values(i+25*(cont_im-1), cont, i, cont_im, preusv(i), venutT,
        antiguitatT, dpreuanteriorT,difpreuinicialT);
    else
        if preuinicial is not null then
            update preus set venut = 1 where idpreus= i + 25*(cont_im-1)-1;
            continuar:=false;
        end if;
    end if;
end loop;

end if;
end loop;
close c_f;

end;
```

PAS 5

El darrer pas, seria, si escau esborrar la taula CATALEG_TEMP, procés *esborrar_temp.sql*

```
DROP table CATALEG_TEMP cascade constraints;
```

I també s'haurien d'esborrar les columnes *idzona* i *texto* de la taula GRUP_TERRITORIAL ja que s'han utilitzat com ajut de comprovació en el moment d'inserir i tallar la cadena de caràcters.

```
ALTER TABLE GRUP_TERRITORIAL
DROP idzona, texto;
```



3.4. Informes

A continuació es presenten els informes demanats per *Sostre per Tothom*.

3.4.1 Nombre d'immobles per zona, tipologia i característiques per mes i any

En aquest informe es demana, que es dissenyi un informe amb el número d'immobles per cada grup territorial (que com té quatre subnivells seran els que es mostraran), per tipologia i característiques. A més cal que estigui per mesos i per anys. El disseny seria una matriu en la qual en les files es tindria els 4 grups territorials en forma de jerarquia, per elements de pàgina de filtre i a mode de selecció la resta de variables. Per cada cel·la s'utilitza un comptador que recull el nombre d'immobles.

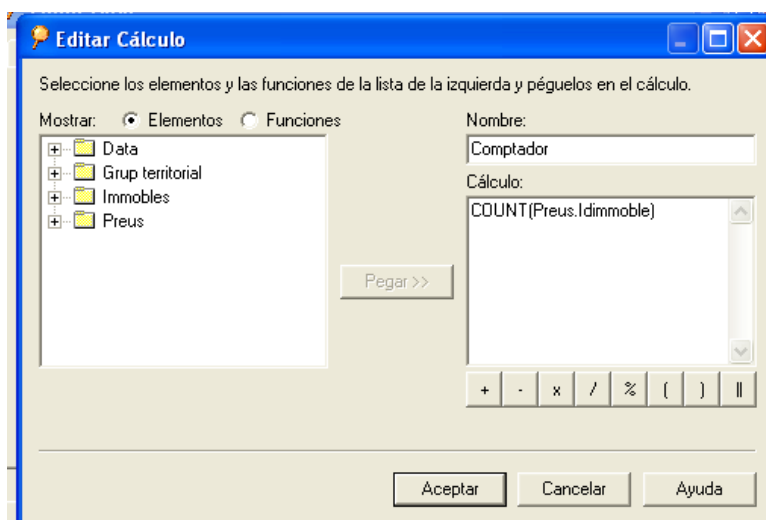
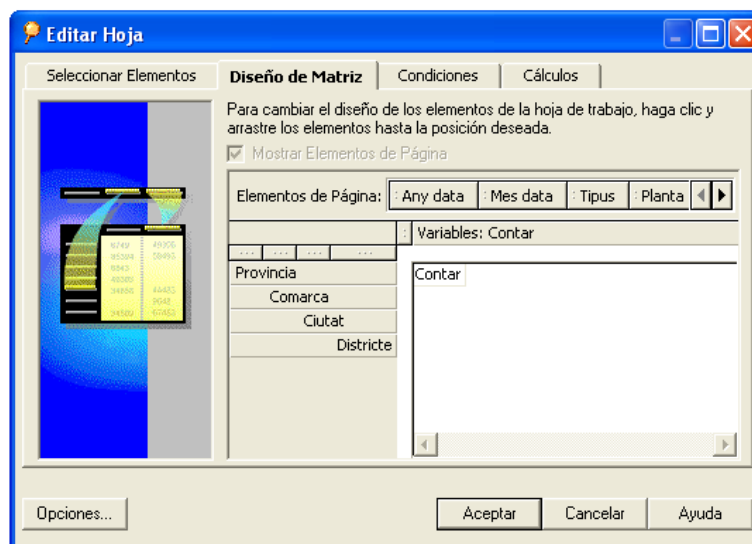


Figura 21. Disseny de l'informe 5.1



I un exemple del resultat amb els paràmetres: any 2007, al mes d'Abril i tipus igual a pisos seria:

Zona	Tipus	Quantitat
Comptador		
alicante		55
baix segura		13
àrea de guardar del segura		2
àrea de guardar del segura		2
àrea de orihuela costa		1
àrea de orihuela costa		1
torrevieja		10
centro ciudad		2
centro ciudad - puerto deportivo		1
la mata		2
los balcones		1
playas de los naufragos		2
playas de levante - playa de los locos		1
playas de levante - playa del cura		1
baix vinalopó		5
elche/elx		1
centro		1
partidas de elche-norte		3
partidas de elche-norte		3
santa pòla		1
calas de santiago bernabeu		1
l'alacantí		18
alicante / alacant		11
distrito 2 - altozano / carolinas / pla del bon repós		2
distrito 2 - san agustín / los ángeles / tómbola		1
distrito 3 - benalúa / san blas		4
distrito 4		1
distrito 6 - cabo de las huertas		1
distrito 6 - la albufereta		2
el campello		3
el campello		3
mutxamel		2
mutxamel		2
san juan de alicante		1
san juan de alicante		1
san vicente del raspeig		1
san vicente del raspeig		1
marina alta		12
marina alta		1

Figura 22. Exemple de resultat de l'informe 5.1

3.4.2 Llistat d'immobles que han baixat de preu per mes i any

En aquest informe es demana el llistat d'immobles que han baixat el preu per mes i per any. En aquest cas s'utilitzarà una taula, en la que es tindrà tant registres com immobles es disposa i es col·loquen els elements de pàgina de filtre tant el mes i l'any com a l'informe anterior. Caldrà afegir, però, la condició que diferència del preu inicial i la de cada mes sigui negativa.

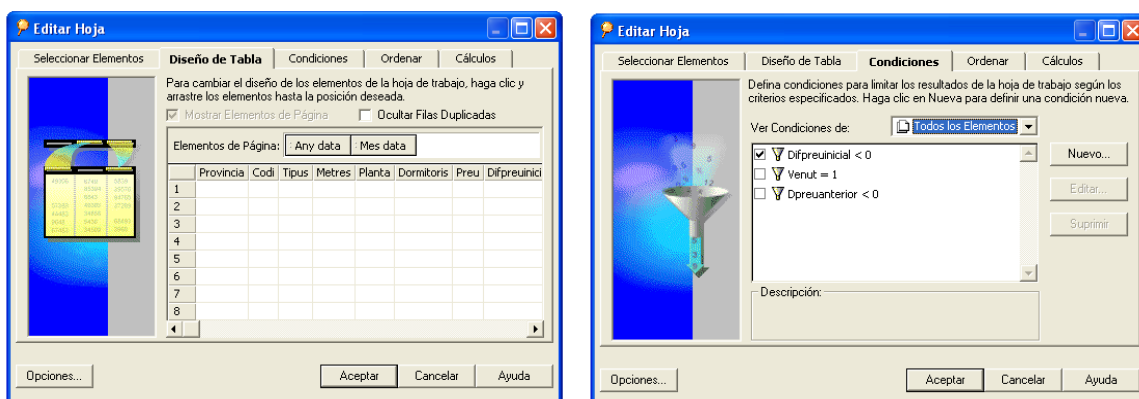


Figura 23. Disseny de l'informe 5.2

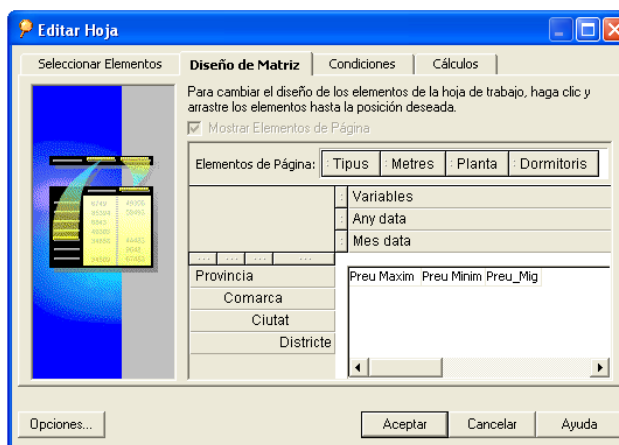
El resultat de l'informe per l'any 2007 al mes d'Agost seria:

	Provincia	Codí	Tipus	Metres	Planta	Dormitoris	Preu	Dif.preuinal
1	alicante	VW0000000613618	pisos	60	2	2	117500	-7000
2	alicante	VP00000000716189	pisos	90	6	3	148000	-17000
3	alicante	VP00000000886761	pisos	75	6	2	149000	-21000
4	alicante	VW00000000547861	chalets	55	--	2	120000	-13000
5	alicante	VW00000000401303	chalets	210	--	3	240000	-6000
6	alicante	VW00000000615020	chalets	50	--	2	95000	5000
7	alicante	VW00000000539254	pisos	60	2	2	116000	-4000
8	alicante	VW00000000673211	chalets	78	--	2	96000	-10000
9	alicante	VP00000000628504	pisos	104	6	3	283000	-1250
10	alicante	VW00000000519547	dúplex	90	bj	2	190000	-10000
11	alicante	VW00000000637457	pisos	90	bj	2	145000	-5000
12	alicante	VW00000000632031	pisos	150	4	3	408688	-11312
13	alicante	VP00000000914029	pisos	130	1	3	270000	-5000
14	alicante	VW00000000763294	pisos	131	3	3	312000	-6000
15	alicante	VW00000000610752	chalets	150	--	4	295000	-35000
16	alicante	VP00000000620113	òtics	160	0	4	351500	-50
17	alicante	VW00000000309125	chalets	310	--	1	155000	15000
18	alicante	VP00000000764838	pisos	80	4	3	126000	-15238
19	alicante	VW00000000395075	chalets	200	--	4	240000	-30000
20	alicante	VW00000000570023	pisos	67	2	3	107000	-11300
21	alicante	VW00000000734673	òtics	115	3	3	423000	57810
22	alicante	VW00000000715815	pisos	103	2	2	200000	-11000
23	alicante	VP00000000832408	pisos	90	4	2	252000	-10500
24	alicante	VP00000000770256	pisos	69	--	2	104460	-26000
25	alicante	VW00000000666009	pisos	95	1	2	105000	-5000
26	alicante	VW00000000447953	pisos	101	4	2	243000	6000
27	alicante	VW00000000556100	òtics	127	6	2	234000	-6000
28	alicante	VW00000000631843	pisos	100	3	3	193000	-2000
29	alicante	VW00000000501700	chalets	147	--	3	240000	-17000
30	alicante	VW00000000594909	òtics	110	2	2	348000	-13000
31	alicante	VW00000000569489	chalets	150	--	4	240000	5000
32	alicante	VW00000000455311	pisos	56	2	1	99899	-20101
33	alicante	VW00000000596559	pisos	65	4	2	130000	-13000
34	alicante	VP00000000699652	pisos	103	--	3	161210	-9290
35	alicante	VP00000000503367	pisos	110	3	3	340000	-20000
36	alicante	VP00000000503378	pisos	110	4	4	222000	-20000
37	alicante	VP00000000734887	pisos	95	bj	3	385000	-35000
38	alicante	VW00000000730999	pisos	100	3	2	162000	-6000
39	alicante	VW00000000303763	òtics	115	5	2	144000	-12500
40	alicante	VP0000000043899	pisos	70	1	2	100000	10000
41	alicante	VP0000000049477	òtics	65	6	2	139000	-4600

Figura 24. Exemple de resultat de l'informe 5.2

3.4.3 Preus màxims, mínims i mitjans per zona, tipologia i característiques per mes i any

En aquest informe es demanen els preus màxims i mínims i mitjans per zona, tipologia i característiques dels immobles, i per mes i any. Per modelitzar-lo s'utilitza de nou una matriu en la que a l'igual que a l'informe 5.1 a les files es venen donats el grups territorials. En aquest cas els elements de pàgina de filtres són la tipologia i les característiques. A les files es col·loquen l'any i el mes i per cada cel·la es calcula el preu màxim amb la funció MAX, el preu mínim amb la funció MIN, i el preu mig amb la funció AVG, definits a l'apartat de *Condicions*



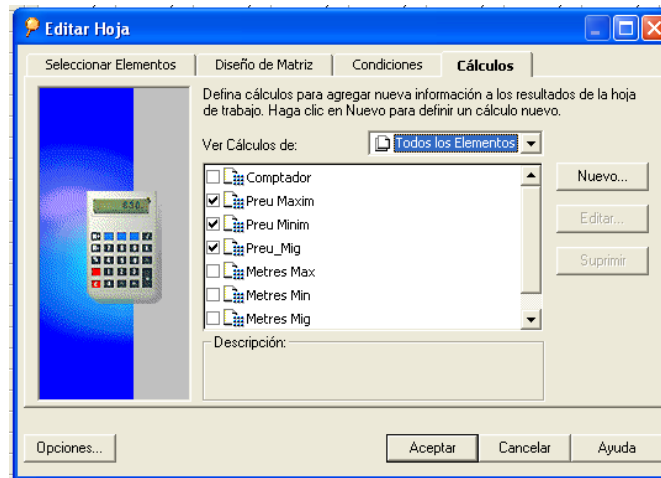


Figura 25. Disseny de l'informe 5.3

I un exemple del resultat de l'informe a Lleida i Mallorca, als estudis presentarien el següent preu mig al 2007 i 2008.

Oracle Business Intelligence Discoverer Desktop - [Explotacio_STP.DIS]

Preus màxims, mínims i mitjans per zona, tipologia característiques per mes i any

Elementos de Página: Tipus: estudios | Metres: <Todo> | Planta: <Todo> | Dormitoris: <Todo>

	2007												Preu_Mig
	ABRIL	AGOST	DESEMBRE	FEBRER	GENER	JULIOL	JUNY	MAIG	MARÇ	NOVEMBRE	OCTUBRE	SETEMBRE	2008
lleida													
alt urgell - cerdanya				318536,00					120000,00	234768,00			
alt urgell - cerdanya													
alt urgell - cerdanya													
alta ribagorça				318536,00					120000,00	318536,00			
alta ribagorça				318536,00					120000,00	318536,00			
alta ribagorça				318536,00					120000,00	318536,00			
pallars sobirà										151000,00			
pallars sobirà										151000,00			
pallars sobirà										151000,00			
mallorca	114961,75	97250,00	106668,86	100000,00	105500,00	98375,00	178000,00	102000,00	101800,00	9091,00	114500,00	117250,00	106000,00
centro													
centro													
centro													
levante	73000,00								55000,00	76500,00	72000,00		72000,00
levante	73000,00								55000,00	76500,00	72000,00		72000,00
levante	73000,00								55000,00	76500,00	72000,00		72000,00
llucmajor	153247,00												
llucmajor	153247,00												
llucmajor	153247,00												
nord		90000,00	100000,00			100000,00							
nord		90000,00	100000,00			100000,00							
nord		90000,00	100000,00			100000,00							
palma de mallorca	100000,00	92000,00	107727,33	100000,00	118000,00	77500,00	178000,00	126500,00	128000,00	108182,00		94500,00	
aragó - manacor - pol·levant - pere garau						77500,00							
aragó - manacor - pol·levant - pere garau						77500,00							
ciutat antiga													
ciutat antiga													
molinar - platja de palma - coll d'en rebassa													
molinar - platja de palma - coll d'en rebassa													
paseo marítimo-ponent	100000,00	92000,00	107727,33	100000,00	118000,00		178000,00	126500,00	128000,00	108182,00		94500,00	
paseo marítimo-ponent	100000,00	92000,00	107727,33	100000,00	118000,00		178000,00	126500,00	128000,00	108182,00		94500,00	
sta.catalina-son espanyolet-son dameto-son cotoner													
sta.catalina-son espanyolet-son dameto-son cotoner													
poniente	133600,00	115000,00	107833,33		93000,00	108000,00		100000,00	100000,00		114500,00	140000,00	140000,00
poniente	133600,00	115000,00	107833,33		93000,00	108000,00		100000,00	100000,00		114500,00	140000,00	140000,00
poniente	133600,00	115000,00	107833,33		93000,00	108000,00		100000,00	100000,00		114500,00	140000,00	140000,00

Figura 26. Exemple de resultat de l'informe 5.3

3.4.4 Metres quadrats màxims, mínims i mitjans per zona, tipologia i característiques per mes i any

Aquest informe és molt semblant a l'anterior l'únic que varia que en lloc dels preus, per cada cel·la es mostra els metres quadrats màxims, mínims i mitjans.

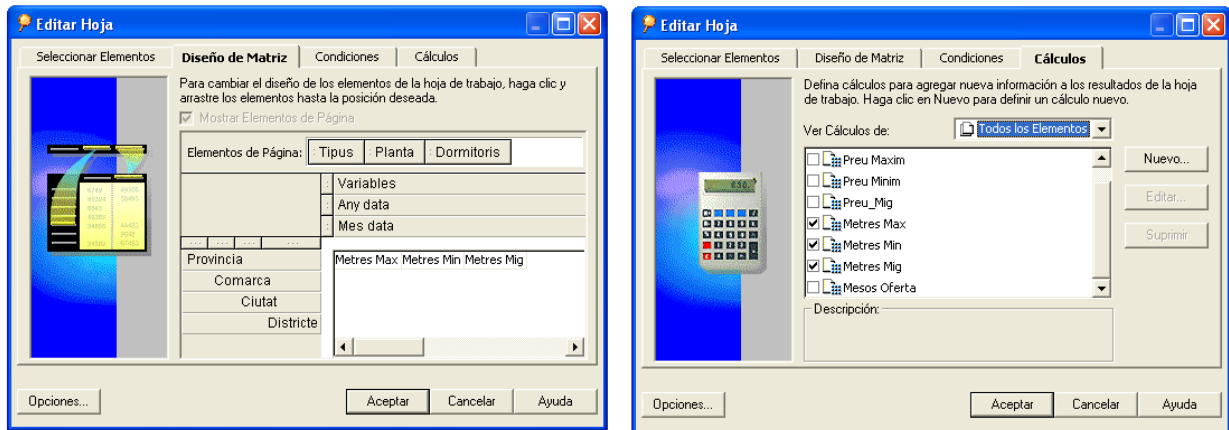


Figura 27. Disseny de l'informe 5.4

I el següent exemple mostraria els metres quadrats màxims, mínims i mig de l'any 2006, per a la tipologia pisos.

	2006												2007		
	ABRIL	AGOST	DESEMBRE	FEBRER	GENER	JULIOL	JUNY	MAIG	MARÇ	NOVEMBRE	OCTUBRE	SEPTEMBRE	ABRIL	AGOST	
menorca			50,00			75,00									106,00
àrea de ciutadella		71,00	50,00			75,00									106,00
àrea de ciutadella		71,00	50,00			75,00									106,00
àrea de esmercadal		65,00				75,00									
àrea de esmercadal		65,00													
àrea de esmercadal		65,00													
àrea de mahón															
àrea de mahón															
àrea de mahón															
larragona															
all camp															
àrea de valls								105,00							51,00
àrea de valls								105,00							51,00
baix camp									74,00			125,00			
àrea de cambriès				55,00											127,00
àrea de cambriès				55,00											127,00
àrea de mont-roig del camp		55,00						81,00		70,00					68,00
àrea de mont-roig del camp		55,00						81,00		70,00					68,00
àrea de reus			80,00			65,00					47,00	125,00			70,00
castre						60,00					125,00				70,00
levant															
ponent			80,00								47,00				
àrea de riudoms						96,00									50,00
àrea de riudoms						96,00									50,00
baix ebre		140,00		106,00	97,00										
àrea de l'ampolla					97,00							97,00			75,00
àrea de l'ampolla					97,00							97,00			75,00
àrea de tortosa		140,00		106,00											117,00
àrea de tortosa		140,00		106,00											117,00
baix penedès								70,00		69,00	60,00				
àrea de calafell			112,00					70,00		52,00	60,00				74,00
àrea de calafell			112,00					70,00		52,00	60,00				74,00
àrea de el vendrell		80,00								120,00	69,00				86,00
àrea de el vendrell		80,00								120,00	69,00				86,00
àrea de la bisbal del penedès															

Figura 28. Exemple de resultat de l'informe 5.4

3.4.5 Temps mitjans de venda per zona, tipologia i característiques per mes i any

El següent informe demana els temps mitjans de venda de cada immoble també per zona, tipologia característiques, mes i any. Per dissenyar-lo utilitzem una altra cop una matriu on els nivells d'agrupament serien els grups territorials, els elements de filtre de pàgina serien la tipologia i les característiques, a les columnes mostrarien les variables de temp i par cada cel·la es mostraria la variable *antiguitat* que recull el temps mitjà en mesos que ha trigat en vendre l'immoble.

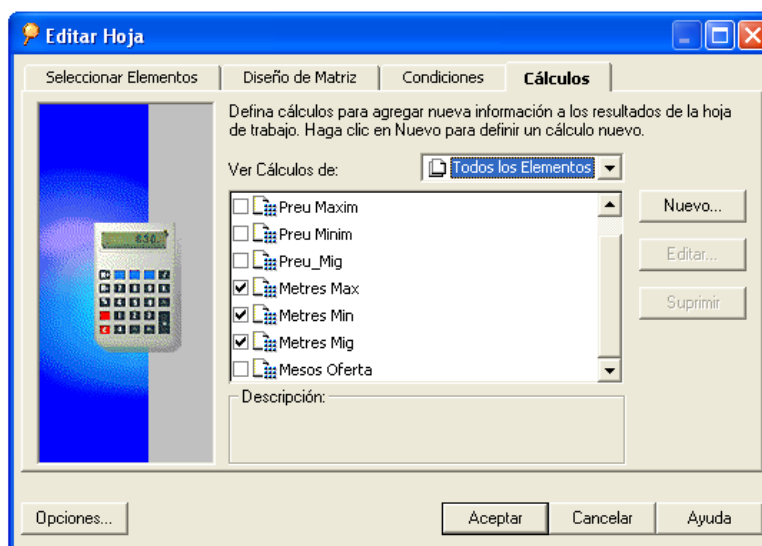
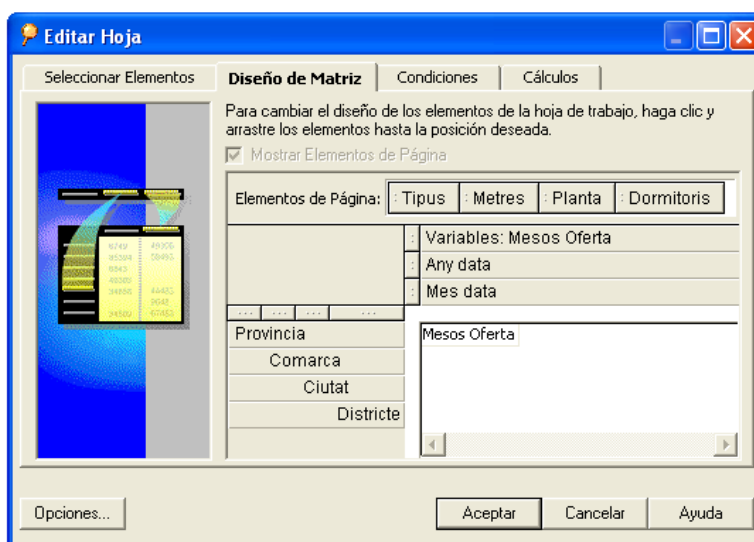


Figura 29. Disseny de l'informe 5.5



I un exemple del resultat seria, per als pisos al 2007, amb 2 dormitoris i a *Ibiza-Formentera* i Lleida tal i com es distribueix en la següent figura, on per cada cel·la indicaria els mesos en promig que ha trigat un pis en vendre's. Per exemple a l'abril en *Ibiza-Formetera*, el temps promig per vendre un pis de 2 habitacions era de 10,38 mesos

The screenshot shows a pivot table titled "Temps mitjans de venda per zona, tipologia i característiques per mes i any". The table is filtered for "Tipus: pisos", "Metres: <Todo>", "Planta: <Todo>", and "Dormitoris: 2". The columns represent months from April 2007 to January 2008. The rows list various zones and typologies, such as "ibiza-formentera", "lleida", and "segarra".

	2007												Mesos Oferta Mig 2008	
	ABRIL	AGOST	DESEMBRE	FEBRER	GENER	JULIOL	JUNY	MAIG	MARÇ	NOVEMBRE	OCTUBRE	SETEMBRE	2008 GENER	
ibiza-formentera	10,38	11,29	15,29	7,00	7,00	10,25	14,00	17,00	15,00	14,20	6,33	19,40	19,00	
àrea de eivissa	9,00	11,00	24,00	2,00		8,50	6,00		15,00	15,00	7,50	16,00		
àrea de eivissa	9,00	11,00	24,00	2,00		8,50	6,00		15,00	15,00	7,50	16,00		
àrea de eivissa	9,00	11,00	24,00	2,00		8,50	6,00		15,00	15,00	7,50	16,00		
àrea de sant antoni	16,00		21,00								4,00	18,00		
àrea de sant antoni	16,00		21,00								4,00	18,00		
àrea de sant antoni	16,00		21,00								4,00	18,00		
àrea de sant josep		9,00	6,33	14,00	2,00	14,67	18,00		15,00	13,00		21,00		
àrea de sant josep		9,00	6,33	14,00	2,00	14,67	18,00		15,00	13,00		21,00		
àrea de sant josep		9,00	6,33	14,00	2,00	14,67	18,00		15,00	13,00		21,00		
àrea de santa eulària	9,80	15,00	21,50	6,00	8,67	9,50	18,00	17,00	15,00	14,20		21,00	19,00	
àrea de santa eulària	9,80	15,00	21,50	6,00	8,67	9,50	18,00	17,00	15,00	14,20		21,00	19,00	
àrea de santa eulària	9,80	15,00	21,50	6,00	8,67	9,50	18,00	17,00	15,00	14,20		21,00	19,00	
formentera													2,00	
formentera													2,00	
formentera													2,00	
lleida	11,00	8,14	8,50	11,20	4,14	12,00	8,43	10,17	8,67	8,45	10,70	9,86	11,67	
alt urgell - cerdanya		8,00		6,00	1,00						10,00		2,00	
alt urgell - cerdanya		8,00		6,00	1,00						10,00		2,00	
alt urgell - cerdanya		8,00		6,00	1,00						10,00		2,00	
alta ribagorça														
alta ribagorça									2,00		6,00		6,00	
alta ribagorça									2,00		6,00		6,00	
alta ribagorça									2,00		6,00		6,00	
garrigues			10,00								6,00		7,00	
garrigues			10,00								6,00		7,00	
garrigues			10,00								6,00		7,00	
noguera	16,00			14,00		6,00		4,00	2,00					
noguera	16,00			14,00		6,00		4,00	2,00					
noguera	16,00			14,00		6,00		4,00	2,00					
pallars jussà	16,00	16,00				17,67			15,00	4,00				
pallars jussà	16,00	16,00				17,67			15,00	4,00				
pallars jussà	16,00	16,00				17,67			15,00	4,00				
pallars sobirà	6,00	7,00	5,00		6,00		10,50			12,33	7,00			
pallars sobirà	6,00	7,00	5,00		6,00		10,50			12,33	7,00			
pallars sobirà	6,00	7,00	5,00		6,00		10,50			12,33	7,00			
segarra	12,00	6,00							1,00	5,00				

Figura 30. Exemple de resultat de l'informe 5.5

3.4.6 Nombre d'immobles venuts i existents per zona, tipologia i característiques per mes i any

En aquest darrer informe demanen el nombre d'immobles que s'han venut i els que havia en aquell moment per grup territorial, tipologia i característiques, any i mes. Per elaborar-ho es fa servir una matriu, que com en casos anteriors, per files es presentaran els grups territorials, per elements de pàgina de filtre la tipologia i les característiques, per columnes les variables temporals i la variable **venut** que indica amb un 1 si l'immoble s'ha venut i amb un 0 en el cas contrari. Per cada cel·la es mostra el número d'immobles.

El disseny seria:

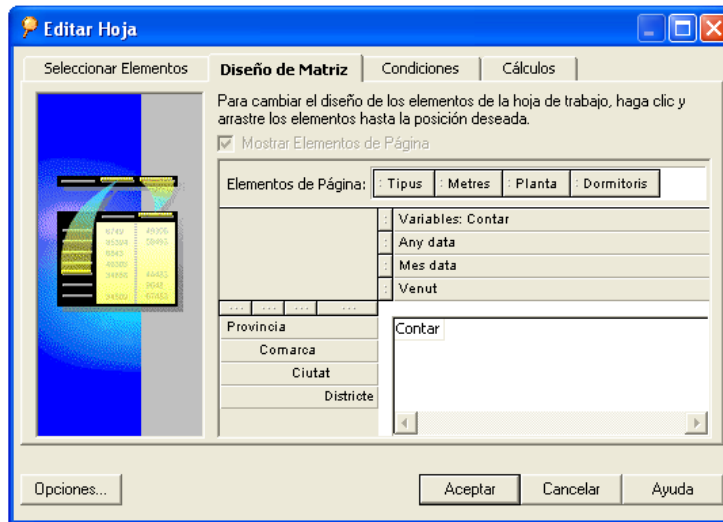


Figura 31. Disseny de l'informe 5.6

Aquest exemple mostraria els àtics que s'han venut i els que hi havia a Barcelona al 2006

	2006																	
	ABRIL		AGOST		DESEMBRE		FEBRER		GENER		JULIOL		JUNY		MAIG		MARÇ	
	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
barcelona	11	102	2	100	2	123	4	81	3	97	6	75	2	108	4	107	8	86
alt penedès		2				1		1										
àrea de sant sadurní d'anoia		1				1		1										
àrea de sant sadurní d'anoia		1				1		1										
àrea de santa margarida i els monjos																		
àrea de santa margarida i els monjos																		
àrea de vilafrañca del penedès		1								1								1
àrea de vilafrañca del penedès		1								1								1
anoia			1								1							1
àrea de igualada											1							1
centre											1							1
llevant											1							1
àrea de piera			1															
àrea de piera			1															
bages				1		1		2				1		1				
àrea de manresa				1		1		2				1		1				
àrea de manresa															1			
centre									1									
nord									1									
ponent				1					1									
àrea de sallent							1											
àrea de sallent							1											
àrea de sant joan de vilatorrada												1						
àrea de sant joan de vilatorrada												1						
baix llobregat nord				1														1
àrea de abrera																		
àrea de abrera																		
àrea de esparraguera																		
àrea de esparraguera																		
àrea de martorell					1													1
àrea de martorell					1													1
àrea de palma de cervelló-el papiol																		
àrea de palma de cervelló-el papiol																		
àrea de san andreu de la barca																		
àrea de san andreu de la barca																		

Figura 32. Exemple de resultat de l'informe 5.6

3.4.7 Llistat d'immobles que han baixat de preu versus el mes anterior i que s'han venut per mes i any

Un informe que no es demanava però que pot ser també interessant seria veure quins pisos s'han venut al baixar el preu. Durant el procediment de càrrega de les dades es va enregistrar una variable que guardava la diferència de preus entre mesos. Per elaborar aquest informe s'utilitza aquesta variable en el cas que tingui un valor negatiu juntament amb la condició que l'immoble s'hagi venut. En aquest cas s'utilitzaria una taula, on apareixerien tants registres com immobles compleixin les condicions anteriors, com a elements de filtre de pàgines s'utilitzen les variables temporals i per cada cel·la es presenta la informació demanada.

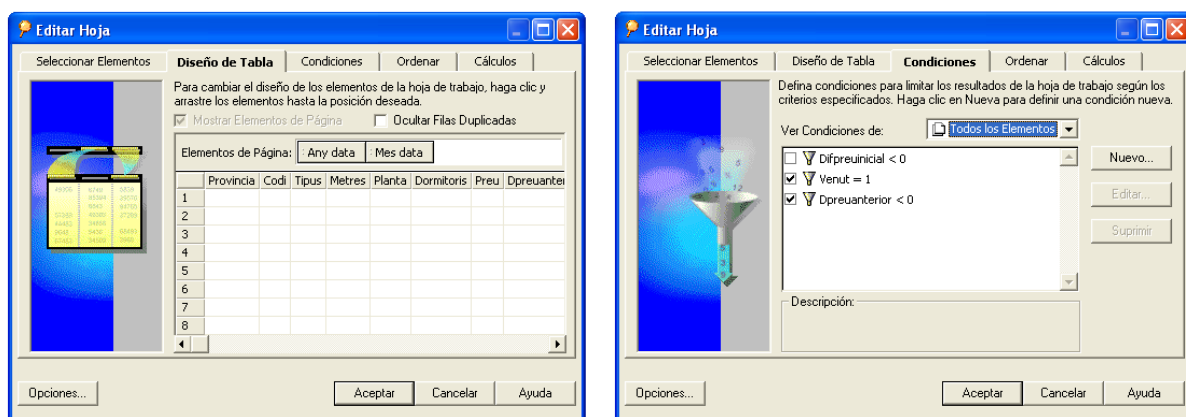


Figura 33. Disseny de l'informe 5.7

Per al mes d'agost del 2007 els immobles que s'haurien venut perquè hagin baixat el preu serien:

Oracle Business Intelligence Discoverer Desktop - [Explotació_STP.DIS]

Archivo Editor Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma

Llistat d'immobles que han baixat de preu respecte el mes anterior i que s'han venut per mes i any

Elementos de Página: Any data: 2007 Mes data: AGOST

	Provincia	Codi	Tipus	Metres	Planta	Dormitoris	Preu	Dpreu anterior
1	alicante	VP000000995311	pisos	92	4	3	117000	-1000
2	alicante	VP000000989694	pisos	70	4	2	110000	-2000
3	alicante	VP000000996117	pisos	90	4	3	126000	-1300
4	alicante	VW000000556202	pisos	85	3	3	279000	-3500
5	alicante	VP000001031460	pisos	95	1	3	221500	-1000
6	alicante	VP000000946099	pisos	150	1	3	407000	-10700
7	barcelona	VP000000840670	pisos	67	em	4	288500	-9000
8	barcelona	VP000001030512	pisos	90	3	3	192324	-54091
9	barcelona	VP000000957411	pisos	68	1	3	299905	-601
10	barcelona	VP000000912889	pisos	144	bj	3	574000	-80000
11	barcelona	VP000000966500	pisos	95	2	3	369000	-10000
12	barcelona	VP000000867733	pisos	50	3	3	228400	-9000
13	barcelona	VP000001021797	pisos	55	2	2	237400	-6010
14	barcelona	VP000000889933	pisos	90	1	3	500000	-10000
15	barcelona	VP000000851493	pisos	70	em	3	323946	-15436
16	barcelona	VP000000921046	pisos	80	5	2	1190000	-70000
17	barcelona	VP000000810303	pisos	65	1	3	336600	-6000
18	barcelona	VW000000754259	pisos	75	8	3	339600	-18000
19	barcelona	VP000000912501	chalets	301	bj	3	1200000	-150000
20	barcelona	VP000000901045	pisos	82	8	2	1360000	-78000
21	barcelona	VP000000936796	chalets	200	---	4	1362000	-102350
22	barcelona	VW0000005416523	chalets	180	---	4	450000	-3000
23	barcelona	VW000000525985	pisos	105	4	4	312000	-6000
24	barcelona	VP000000949963	pisos	73	1	2	256435	-9015
25	barcelona	VP000000970387	pisos	44	1	3	186314	-16030
26	barcelona	VP000000613823	estudios	25	bj	0	108000	-4000
27	barcelona	VP000000901636	pisos	78	8	1	1040000	-65000
28	tarragona	VW000000544105	pisos	77	2	3	126213	-21005
29	tarragona	VP000000836095	pisos	51	4	2	203028	-6000
30	valencia	VP000000846828	ibicos	115	5	2	244884	-7510
31	valencia	VP000000824299	pisos	84	8	2	176000	-15000
32	valencia	VP000000912316	ibicos	88	5	3	114000	-6000
33	valencia	VP000000916638	pisos	90	2	2	103600	-6400
34	valencia	VP000000975877	pisos	65	4	3	99999	-1
35	valencia	VP000000972721	pisos	111	2	3	294500	-54100

Figura 34. Exemple de resultat de l'informe 5.7

3.4.8 Pis-tipus espanyol i andorrà

Per trobar el pis-tipus espanyol i andorrà es pot fer de diferents maneres. La primera més visual seria fer els gràfics de cada variable descriptiva de l'immoble i aquell valor més comú o que es repeteixi més seria el que interessaria. En estadística aquest valor se'l coneix com la moda. Una altra forma de trobar-ho seria fent promig de totes aquestes variables, però a l'hora d'interpretar pot ser una mica confós dir que per exemple el pis tipus a Barcelona té 3,4 habitacions. Altres opcions també passarien per triar altres estadístics com per exemple, la mediana, la mitja ponderada, la desviació estàndard que també ens ajustarien altres pisos models. També es podia fer algun tipus d'anàlisi estadístic multivariant (anàlisi de correspondències múltiples, o clúster...)

No obstant la forma més directa seria intentar calcular la moda directament a les variables que ens interessa. Com *Oracle* incorpora la funció `STATS_MODE` es pot aplicar directament sobre les taules i tenir els resultats desitjats.

Per exemple, si volem calcular la moda del tipus d'immoble espanyol la consulta seria:

```
select stats_mode(immobles.tipus) as ModaTipus
from immobles
where immobles.idimmoble in
  ( select distinct (preus.idimmoble)
    from preus, grup_territorial
    where
      grup_territorial.provincia<>'andorra' and
      preus.idzona=grup_territorial.idgt
  );
```

El resultat seria `Modatipus = pisos`

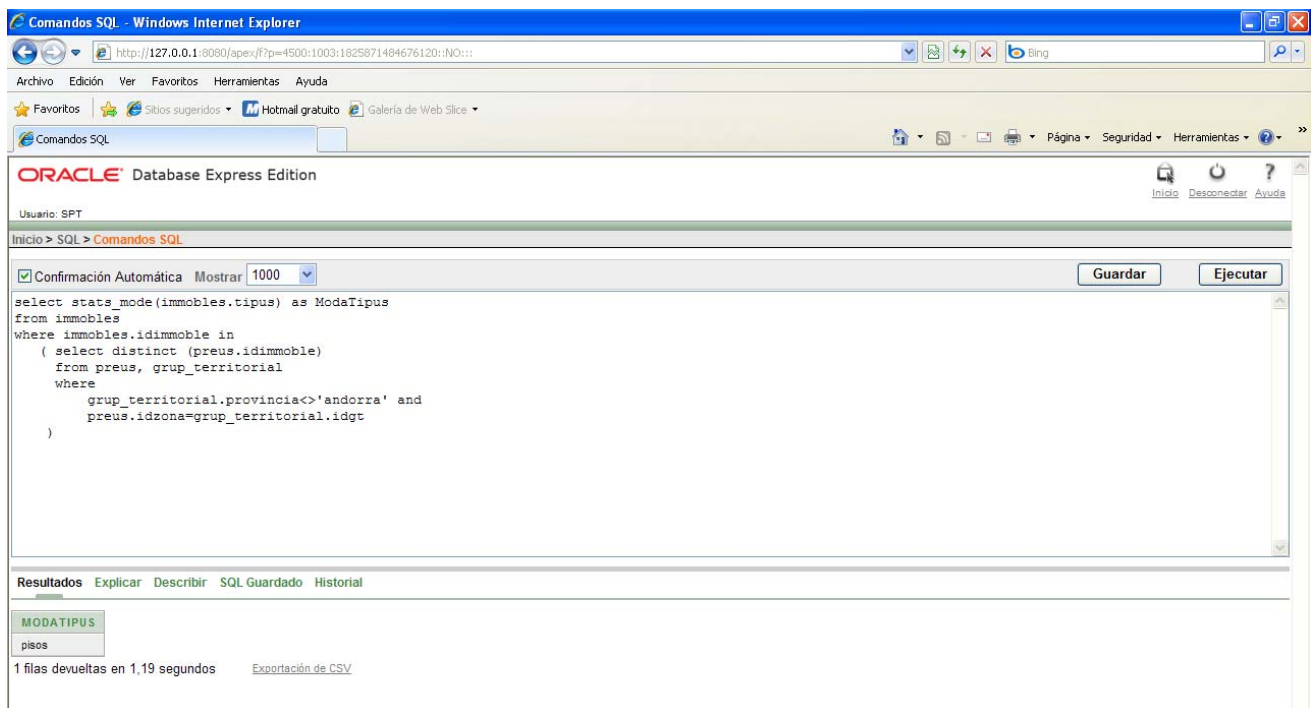


Figura 35. Resultat consulta de tipus d'immoble model espanyol

Si apliquem la mateixa consulta, a les variables metres, planta, dormitoris tant a l'Estat espanyol com a Andorra s'obté:

	Tipus	Metres	Dormitoris	Planta	Preu inicial
Espanya	pisó	90 m	3	1 *	240000
Andorra	pisó	100 m	3	1	210000

- On el valor que pren la Moda seria el '--' però es determina posar el segon valor més repetit per donar perfil al model
- I la consulta per la moda del preu inicial seria:

```
select stats_mode(preus.preu)
  from preus, grup_territorial
  where
    grup_territorial.provincia<>'andorra' and
    preus.idzona=grup_territorial.idgt;
```

Es pot concloure per tant, que l'immoble model a l'Estat Espanyol és un pis de 3 dormitoris de 90 metres quadrats, situat a la primera planta i amb un preu inicial de 240.000 euros. En canvi a Andorra l'immoble model seria una mica més gran de 100 metres quadrats, de 3 dormitoris, situat també a la primera planta però una mica més econòmic, d'uns 210.000 euros.



4 CONCLUSIONS

- S'ha realitzat el procés complet de creació d'un magatzem de dades.
- S'ha fet l'anàlisi de requeriments, el disseny del model multidimensional i finalment la implementació del magatzem pròpiament dit.
- S'han pogut realitzar totes les consultes necessàries per generar els informes que han donat resposta als requeriments de *Sostre per a tothom*
- S'ha pogut extreure la informació necessària per representar les característiques que defineixen l'immoble estàndard espanyol i andorrà.

Pel cas d'Espanya seria, els pisos de 90 m2 amb 3 dormitoris, que es troben a la primera planta i amb un preu de 240.000 euros, i pel cas d'Andorra, els pisos de 100 m2 amb 3 dormitoris, que es troben a la primera planta i amb un preu de 210.000 euros

- S'han assolit per tant, els objectius proposats.



5 LÍNIES DE FUTUR

Cal tenir en compte que, per tal que l'explotació del magatzem de dades proporcioni realment informació útil i aprofitable de cara a la presa de decisions, és imprescindible actualitzar aquestes dades i proporcionar al client la formació necessària perquè pugui realitzar aquesta actualització de forma periòdica.

Com a línies de futur es poden destacar:

ACTUALITZACIONS

Per a l'actualització i inserció de dades noves només caldrà carregar a la taula DATA els mesos que es vagin incorporant.

ADAPTACIÓ DELS INFORMES

Seria adient que *Sostre per a tothom* en previsió de noves necessitats pogués crear les seves pròpies consultes.

ENTRADA DE DADES

Tot i que no s'han detectar masses incoherències en el procés d'entrada de dades, caldrà tenir-ho en compte en les possibles recollides d'informació futures.

TEMPS D'EXECUCIÓ

Seria recomanable millorar el temps d'execució de les consultes mitjançant l'ús de resums una vegada que la base de dades estigui definitivament instal·lada al servidor del client.

WEB

Es podria plantejar l'opció de l'ús de *Oracle Discoverer Plus* ja que les consultes es podrien realitzar des d'un navegador instal·lat en qualsevol equip.



6 GLOSSARI

Ad-hoc: Consulta que no ha estat dissenyada en el moment de determinar les consultes. Es construeix de forma dinàmica i amb eines dissenyades a tal efecte.

Agregació: L'agregació de dades és el procés mitjançant el qual es re defineixen les dades mitjançant suma o integració basat en algunes regles o criteris.

Atribut: Qualsevol detall que serveix per qualificar, identificar, classificar, quantificar o expressar l'estat d'una entitat.

BI (Business Intelligence): Intel·ligència Empresarial: Conjunt d'estratègies i eines orientades a l'administració i creació de coneixement mitjançant l'anàlisi de dades existents a una organització o empresa.

cel·la: Mesurament que fa referència al mateix esdeveniment dins una mateixa estructura

Cel·la: Conjunt de cel·les del mateix fet que estan associades a instàncies del mateix nivell per a cadascuna de les dimensions.

Clau forana: Una o més columnes de la taula de una base de dades relacional que implementa una relació *varis a un* que aquesta taula té amb altres taules o amb si mateixa.

Clau primària: Una o més columnes de la taula de una base de dades els valors de les quals, combinats entre si, han de ser únics dins la taula.

Descriptor: Atribut que es troba en un nivell. Conté informació no jeràrquica i està definit sobre un domini discret.

Dimensió: Punt de vista des del que s'analitza un determinat fet.

Discoverer Administrator: Producte que serveix per a crear, mantenir i administrar dades en la capa d'usuari final (EUL) i per a definir com els usuaris interactuen amb les dades.

Discoverer Desktop: L'usuari final, utilitza aquets component per a executar consultes *ad-hoc* i generar informes.

Estrella: Organització de dades en una base de dades relacional. Al mig es troba la taula del fet amb les columnes que són les mesures multidimensionals. Les branques de l'estrella són les dimensions.

ETL (Extract, Transform and Load): Procés d'extracció, transformació i càrrega dades que permet obtenir dades de múltiples fonts, formatjar-les i netejar-les i carregar-les en una base o magatzem de dades per a la seva posterior anàlisi.

Fet: Activitat de l'empresa objecte d'anàlisi.

Granularitat: Grandària d'un objecte respecte a un altre.

Índex de combinació (join index): Índex definit sobre una clau forana, de manera que té els valors d'una taula i apunta a una altra.



Jerarquia d'agregació: Conjunt de relacions entre les instàncies d'una dimensió que indica com s'agrupen unes per formar d'altres.

Magatzem de dades (*Data warehouse*): Sistema d'informació que reuneix la informació històrica generada pels diferents departaments d'una organització, unificada i depurada d'errors que permet la consulta d'informació de forma àgil i flexible.

Mesura: Atribut numèric d'una Cel·la, normalment additiu.

Metadada: Es solen definir com "dades de dades", ja que es tracta de dades que descriuen quina és la estructura de dades que s'emmagatzemarà i com es relacionen entre elles.

Model dimensional: Tipus de model de dades de varies dimensions que s'utilitza en al magatzem de dades. En un model dimensional existeixen dos tipus de taules: de dimensions i de fets.

Nivell: Conjunt d'instàncies d'una dimensió que tenen la mateixa granularitat.

Oracle: és un sistema de gestió de bases de dades relacional (o *RDBMS* per l'acrònim en anglès de *Relational Data Base Management System*), fabricat per *Oracle Corporation*.

OLAP: és l'acrònim en anglès de processament analític en línia. És una solució utilitzada en el camp de la intel·ligència de negocis, que consisteix en consultes a estructures multidimensionals (o *Cubs OLAP*) que contenen dades resumits de grans bases de dades o *Sistemes Transaccionals (OLTP)*. S'utilitza en informes de negocis de vendes, màrqueting, informes de direcció, mineria de dades i àrees similars.

PL/SQL: Llenguatge de programació que suporta totes les consultes i manipulació de dades que es fan servir en *SQL*, però inclou característiques com el maneig de variables, les estructures modulars, les estructures de control de flux i presa de decisions i el control d'excepcions.

Sistema de gestió de bases de dades: programari que gestiona i controla bases de dades. Les seves funcions principals són les de facilitar l'ús simultani a molts usuaris de tipus diferents, independitzar l'usuari del món físic i mantenir la integritat de les dades.

SQL (*Structured Query Language*): Llenguatge de consulta estructurat utilitzat per crear, modificar i recuperar dades d'un sistema de base de dades relacional.

Taula de dimensions: Taula que emmagatzema els registres relatius a una dimensió concreta. No s'emmagatzemen fets en les taules de dimensions.

Taula de fets: Una de les taules que constitueix el model dimensional. Una taula de fets típica conté dos tipus de columnes, els fets en sí i les claus relatives a les taules de dimensions.



7 BIBLIOGRAFIA

Llibres de text

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.*

Building a Data Warehouse Using Oracle OLAP Tools. (1997) *Oracle Technical Report Satish Mahajan*

Data Warehouse Database Design (2001) *M. Lea Shaw.*

Getting Started with Data Warehouse and Business Intelligence (1999) *Maria Sueli Almeida, Missao Ishikawa, Joerg Reinschmidt, Torsten Roeber*

Oracle Data Warehouse Management. Secrets of Oracle Data Warehousing (2003) *Mike Ault*

Publicacions

Abelló Gamazo, A. (2003). Disseny multidimensional. A: Magatzems de dades i models multidimensionals (Rius Gavídia, A. i Serra Vizern, M. (coord.)). **Fundació per a la Universitat Oberta de Catalunya, Barcelona.** ISBN: 84-8429-437-4.

Enllaços d'interès a Internet

Data Warehousing Processes. DataData Warehousing Processes. [en línia].
<http://www.1keydata.com/datawarehousing/processes.html>. [data de consulta: 12/12/2011]

Data Warehousing Review - Data, Data Everywhere!! [en línia].
<http://www.dwreview.com> [data de consulta: 03/11/2011]

Kimball Group: Data Warehouse Training, Consulting, and Kimball University. [en línia].
<http://www.rkimball.com> . [data de consulta: 02/11/2011]

ORACLE (1996-2005). Oracle® Business Intelligence Discoverer Desktop User's Guide. [en línia].
http://download.oracle.com/docs/html/B13917_03/1intro.htm#sthref17
[data de consulta: 08/12/2011]

ORACLE FAQ'S (2010). SQL*Loader FAQ. [en línia].
http://www.orafaq.com/wiki/SQL*Loader_FAQ [data de consulta: 29/10/2011].

Welcome to the Oracle Database 10g Express Edition Tutorial!. [en línia].
<http://st-curriculum.oracle.com/tutorial/DBXETutorial/index.htm> [data de consulta: 12/12/2011]

WIKIPEDIA (2010). PL/SQL. [en línia].
<http://es.wikipedia.org/wiki/PL/SQL> [data de consulta: 08/12/2011].

