

Construcció i explotació d'un magatzem de dades per a l'anàlisi de vendes d'una cadena de supermercats

Memòria

Salvador Moragues Martin
Enginyeria Tècnica d'Informàtica de Sistemes
Consultor: Carles Llorach Rius
Treball Final de Carrera
Curs 2012-2013
07/01/2013

RESUM

Aquest document il·lustra el procés dut a terme alhora de realitzar el treball final de carrera en els estudis d'Enginyeria Tècnica en Informàtica de Sistemes l'àrea de Magatzem de Dades.

Inclou les diferents fases del projecte, des de una introducció inicial on s'explica que és el *Data Warehouse*, la justificació i objectius del projecte, els requeriments del client, fins el seu anàlisi, disseny i implementació.

Aquesta última part, ja és molt més tècnica, i depèn de la tecnologia emprada, en el nostre cas, la major part del programari utilitzat és de la suite Pentaho.

Ressaltaria dos conceptes, que en part, són els que causen que el *Data Warehouse* des de fa anys sigui tan vital en la major part de les empreses.

Per un cantó tenim la integració de dades. En qualsevol empresa existeix una gran dispersió de dades, producte dels diferents sistemes utilitzats arreu. Per una altra banda, la extensió geogràfica de les empreses, que fa que estiguin ubicades a punts molt llunyans entre si.

Per tant, per tenir una visió general de l'empresa, és imprescindible aquesta integració de dades que ens brinden els magatzems de dades.

Aquesta integració a més fa possible el segon concepte, l'anàlisi d'aquestes dades per tal d'extreure coneixement útil per la posterior presa de decisions empresarials.

Índex

1	Introducció	6
1.1	Justificació del Projecte	6
1.1.1	Magatzems de Dades.....	6
1.2	Objectius	7
1.3	Requeriments.....	8
1.4	Funcionalitats	8
2	Resultats esperats.....	9
2.1	Planificació inicial vs. planificació final	9
2.1.1	Tasques	9
2.1.2	Diagrama de Gantt	11
2.1.3	Planificació Final.....	11
2.2	Anàlisi de riscos	12
2.2.1	Equips Informàtics	12
2.2.2	Absències	12
2.2.3	Altres assignatures	12
2.2.4	Imprevistos laborals.....	13
2.2.5	Greus endarreriments.....	13
2.2.6	Aprenentatge Eines	13
3	Anàlisi i Dissenys	14
3.1	Requeriments Funcionals	14
3.1.1	Casos d'ús	14
3.2	Requeriments No Funcionals	16
3.3	Model Conceptual	16
3.3.1	Mesures	16
3.3.2	Dimensions	17
3.3.3	Taula de Fets.....	17
3.3.4	Origen de les dades	18
3.3.5	Base de dades Temporal (<i>Staging Area</i>).....	19
3.4	Disseny de la BD – Diagrama E-R	20
3.4.1	Diagrama E-R.....	21
3.5	Procés ETL	21
3.6	Errors de Càrrega	23
3.6.1	Productes.....	23
3.6.2	Establiments	23
3.6.3	Vendes.....	23
3.6.4	Generals	24
4	Desenvolupament	25
4.1	Programari Utilitzat	25
4.2	Procés ETL	25
4.2.1	Preparació Fitxers	26
4.2.2	Carrega Temporal	26
4.2.3	Càrrega Final.....	36
4.2.4	Treballs	44
4.2.5	Observacions.....	44
4.3	Model Multidimensional	45
4.4	Consultes	48
4.4.1	Informes i Anàlisi	49
5	Treball Futur	59
6	Conclusions.....	60
7	Llista de referències	60
8	Annexos.....	61
8.1	Annex A1 - Bibliografia	61
8.1.1	Enllaços d'interès a Internet.....	61

Taula d'il·lustracions

Il·lustració 1 Taula de Tasques	10
Il·lustració 2 Planificació Inicial	11
Il·lustració 3 Diagrama de Gantt	11
Il·lustració 4 Cas d'ús Usuari.....	15
Il·lustració 5 Cas d'ús Administrador.....	15
Il·lustració 6 Nombre de registres	18
Il·lustració 7 Taula de Fets	18
Il·lustració 8 Origen de les dades.....	19
Il·lustració 9 <i>Staging Area</i>	20
Il·lustració 10 Disseny Final	21
Il·lustració 11 Diagrama E-R	21
Il·lustració 12 Procés ETL	22
Il·lustració 13 Programari Utilitzat	25
Il·lustració 14 Data - Format de camps.....	27
Il·lustració 15 Data - Sortida a Taula.....	27
Il·lustració 16 Hores - Transformació	28
Il·lustració 17 Establiments - Transformació.....	28
Il·lustració 18 Productes - Transformació	29
Il·lustració 19 Productes - Càrrega fitxers.....	29
Il·lustració 20 Productes - Format Camps	30
Il·lustració 21 Productes - Obtenir any.....	30
Il·lustració 22 Productes - Filtrar registres	31
Il·lustració 23 Productes - Mapa Venda_impuls	31
Il·lustració 24 Productes - Homogeneïtzació Proveïdors	32
Il·lustració 25 Productes - Preu Cost Nul.....	32
Il·lustració 26 Productes - Afegir clau	33
Il·lustració 27 Productes - Sortida Taula.....	33
Il·lustració 28 Vendes - Transformació	34
Il·lustració 29 Vendes - Nom del fitxer	34
Il·lustració 30 Vendes - Nom del establiment	35
Il·lustració 31 Vendes - Ordenar registres	35
Il·lustració 32 Vendes - Unir Capçalera i detall.....	35
Il·lustració 33 Vendes - Sortida Taula	36
Il·lustració 34 Depuració registres.....	37
Il·lustració 35 Taula de Fets - Transformació	37
Il·lustració 36 Taula de Fets – Càlculs	38
Il·lustració 37 Taula de Fets - Recerca producte.....	38
Il·lustració 38 Taula de Fets - Calcular Mesures	39
Il·lustració 39 Taula de Fets - Netejar Camps	39
Il·lustració 40 Taula de Fets - Sortida Taula	40
Il·lustració 41 Refinament.....	40
Il·lustració 42 Afegir Num.Setmana	41
Il·lustració 43 BD Taula de Fets.....	41
Il·lustració 44 BD Hores	42
Il·lustració 45 BD Data.....	42
Il·lustració 46 BD Productes.....	42
Il·lustració 47 BD Establiments	43
Il·lustració 48 Model	43
Il·lustració 49 Treball	44
Il·lustració 50 Cub - Dimensions	46
Il·lustració 51 Cub - Mesures	47
Il·lustració 52 PAD - Agregacions	48

Il·lustració 53 Informe - Rànquing d'establiments I	50
Il·lustració 54 Informe - Rànquing d'establiments II	50
Il·lustració 55 Informe - Top Ten de productes.....	51
Il·lustració 56 Informe - Top Ten de productes II.....	51
Il·lustració 57 Informe - Top Ten de productes III.....	52
Il·lustració 58 Informe - Import mitjà de compra per soci i establiment.....	52
Il·lustració 59 Informe - Total vendes i marge net del grup	53
Il·lustració 60 Informe - Marques Blanques	53
Il·lustració 61 Informe - Categorització Clients ABC I	54
Il·lustració 62 Informe - Categorització Clients ABC II	55
Il·lustració 63 Informe - Categorització Clients ABC III	55
Il·lustració 64 Anàlisi - Preus Màxims i Mínims	56
Il·lustració 65 Anàlisi - Percentatge Vendes	57
Il·lustració 66 Anàlisi - Distribució Vendes.....	57
Il·lustració 67 Anàlisi - Compra per impuls I	58
Il·lustració 68 Anàlisi - Compra per impuls II	58

1 Introducció

Aquest document és la memòria pel Treball Final de Carrera (TFC) del semestre 2012-2013, en l'àrea de Magatzem de Dades.

El document inclou una breu descripció del projecte, els requeriments demanats pel client, així com la seva planificació en tasques, un anàlisi de riscos amb el seu pla de contingència i els dissenys inicials.

Després veurem com s'ha realitzat el desenvolupament i si s'ha ajustat o no a les planificacions inicials.

Per últim conclourem amb les millores possibles, les línies que es poden desenvolupar, on hi ha més marge de millora.

Començarem indicant el perquè es creu interessant la realització d'aquest projecte, emmarcant-lo en la situació actual i els objectius que es persegueixen durant el seu desenvolupament.

1.1 Justificació del Projecte

El projecte es basa en crear un Magatzem de Dades, seguint els requeriments i necessitats facilitades pel client.

Així el primer que cal veure, és un esbós de en què consisteix un Magatzem de Dades (*Data Warehouse*).

1.1.1 Magatzems de Dades

És una eina de suport a la presa de decisions corporativa i a la gestió de dades. Així mateix aquesta presa de decisions pot afectar la planificació estratègica de la companyia, d'aquí rau la seva gran importància en l'actualitat.

Algunes de les seves principals característiques són:

- Orientat als usuaris
- Integrat
- Variable en el temps
- No volàtil
- Conté dades resumides i detallades

Orientat als usuaris

La missió del DW és presentar la informació de la manera més efectiva per la correcta presa de decisions. Per tant, la seva primera característica és que ha de ser útil als usuaris, permetent-los presentar la informació de forma entenedora.

Integrat

Habitualment, en qualsevol empresa de certa mida, existeix certa dispersió de les dades. Poden existir diferents aplicacions i bases de dades. Un sistema DW normalitza

i homogeneïtza aquest ventall de dades i les emmagatzema en un lloc comú. Veurem que per això fa servir les eines ETL (Extracció, Transformació i Càrrega).

Variable en el temps

Es registren els canvis produïts amb el temps, perquè els informes estiguin actualitzats.

No volàtil

La informació continguda en el sistema no pateix variacions, és tan sols de lectura. Ni es modifica, ni s'elimina des del sistema DW.

Conté dades resumides i detallades

La informació emmagatzemada té diferents nivells de agregació de manera que permet navegar pels informes per obtenir més detall d'un punt en concret.

Així el DW consisteix en el procés de recollir les dades de un entorn heterogeni, i no normalitzat de dades, per normalitzar-lo i extreure informació útil i fiable perquè l'usuari pugui prendre una decisió correcta.

Degut al gran volum de dades existent en l'actualitat, els sistemes transaccional ja no són útils per aquesta tasca. Estem en l'era de l'anomenada "Big Data".

Per tal d'oferir al client tota aquesta informació es fa servir un estructura multidimensional de dades. De forma que les consultes d'extracció d'informació siguin òptimes.

Hem intentat reforçar l'idea del magatzem de dades des d'un punt de vista de la funcionalitat que ofereix al client i no des de la vessant tecnològica, que malgrat ser força interessant, és tan sols un mitjà, no el fi en si mateix.

Amb tot aquestes explicacions queda clara la idoneïtat del projecte, i la seva importància en el món empresarial actual, i en general en qualsevol pressa de decisions basada en quantitats ingents de dades.

Per això parlem també de la intel·ligència de negoci, com la capacitat d'extreure coneixement de les dades, on els Magatzem de Dades són unes de les seves eines principals.

1.2 Objectius

Tal com ens indica el pla docent, 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.

A més aprendre o perfeccionar la gestió i seguiment de projectes, des de les fases inicials, menys la pressa de requeriments, que en aquest cas ja ens ve donada.

Tot aquest aprenentatge serà mesurable, ja que el coneixement adquirit es veurà plasmat en la realització d'un producte final, a més d'aquesta memòria i la presentació que l'acompanya.

1.3 Requeriments

En l'enunciat facilitat pel client es veuen els requeriments que ha de complir el producte final.

A part el propi treball final de carrera (TFC) té uns requeriments addicionals com són la confecció d'aquesta memòria on s'explica el treball realitzat, així com una presentació que resumirà els punts de major interès i transcendència.

L'accés al programari serà a través del portal, on hi haurà penjat els informes requerits pel client. Així doncs, l'interfície serà mitjançant una web, cosa que permetrà estalviar-nos desenvolupar una nova interfície d'usuari, i tenir que formar els usuaris en el seu ús.

Cal definir unes proves d'acceptació per tal de validar el sistema, així com els informes, que són la representació final, dels requeriments per part del client. Aquestes validacions per una banda es realitzaran acotant els informes, per tal de poder validar el seus resultats mitjançant el sistema transaccional clàssic.

Finalment, cal preparar el sistema per tal de futures ampliacions, en forma de nous establiments o una actualització de la càrrega de dades.

També tenir present les possibles correccions de les dades en les fases inicials.

1.4 Funcionalitats

Amb el sistema ja en funcionament hem de ser capaços de brindar a l'usuari tota aquesta informació que es podria dividir en tres grans blocs.

Vendes

- Total vendes i marge net del grup
- Distribució setmanal i estacionalitat de vendes
- % de vendes respecte el total del grup (per establiment i per demarcació)
- Rànquing d'establiments per nombre de vendes i volum total

Productes

- Preus màxims i mínims per tipologia d'establiment i característiques del producte.
- "Top ten" de productes
- % de vendes de "marques blanques" per habitants

Compres i Clients

- Import mitjà de compra per soci i establiment
- Categorització de Clients A/B/C
- Anàlisi de compra per impuls

2 Resultats esperats

S'espera obtenir un producte que compleixi tots els requeriments demanats pel client, amb les seves funcionalitats, complint els temps pactats amb el client. Per tal de controlar el desenvolupament del projecte i que no hi hagi demores, s'estableix una planificació, per tal de poder aplicar els plans de contingència en cas de necessitat.

Per fer més controlable aquesta planificació es determinen unes fites intermèdies.

2.1 Planificació inicial vs. planificació final

Veiem la planificació inicial realitzada.

2.1.1 Tasques

S'adjunta una taula amb les tasques i fites principals del projecte, així com la seva planificació. Tan les tasques com les fites s'han ajustat al calendari pactat amb el client.

És a dir, cada fita es correspon amb una entrega al client, menys en el cas de la primera, on es tracta tan sols de recollir la informació necessària pel posterior desenvolupament del projecte.

La segona fita i primera entrega al client inclou aquesta planificació, per pactar amb ell el calendari del projecte i els recursos que s'hi assignaran. Així mateix s'elabora un pla de contingència per cadascú dels riscos detectats.

Després es realitza un anàlisi i disseny del producte que s'elaborarà.

Un cop finalitzat, i entregat el disseny, el desenvolupem. Des de la base de dades, els processos de càrrega, fins finalitzar amb els informes que consultarà el client.

Per últim es procedeix a entregar la documentació i presentació al client, així com esperar totes les preguntes que pugui suscitar.

Nº	Fita	Data inici	Data final	Tasca	Precedències
1		19/09/2012	23/09/2012	Recollir Informació Projecte	
2		19/09/2012	27/09/2012	Obtenir bibliografia	
3		19/09/2012	27/09/2012	Descarregar i obtenir les eines	
	A	19/09/2012	27/09/2012	Preparació inicial del projecte	
4		25/09/2012	02/10/2012	Recollir informació	
5		27/09/2012	30/09/2012	Elaborar llista de tasques	
6		01/10/2012	01/10/2012	Ordenar i planificar	5
7		26/09/2012	28/09/2012	Zones vermelles, riscos	
	B	25/09/2012	02/10/2012	Elaboració del Pla de Treball	
8		03/10/2012	06/11/2012	Estudi DW - Lectura i consulta Bibliografia	
9		14/10/2012	21/10/2012	Elaboració Memòria capítols introductoris	
10		14/10/2012	16/10/2012	Estudi Màquina Virtual i SW	

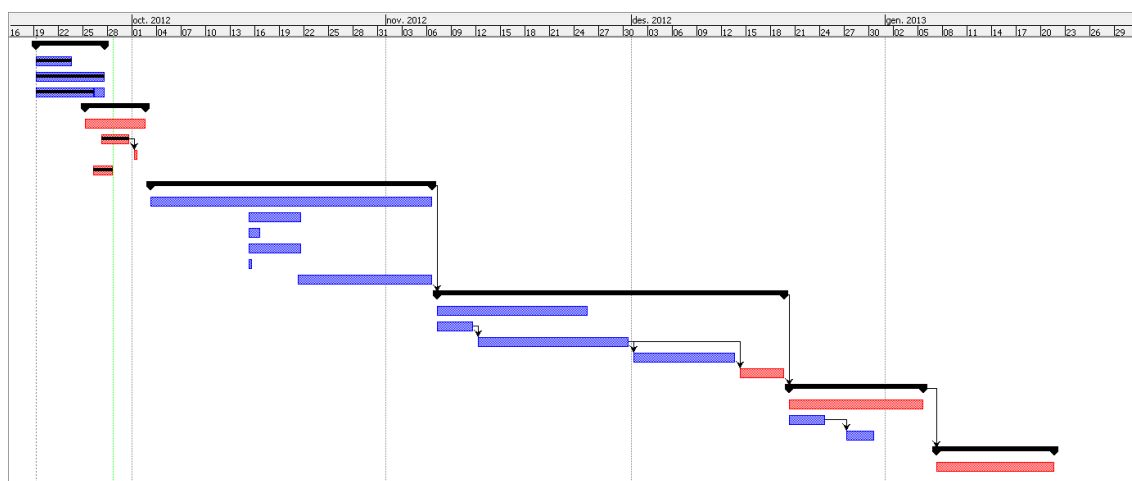
11		14/10/2012	21/10/2012	Estudi Dades facilitades i disseny BBDD	
12		14/10/2012	15/10/2012	Obtenció Dades IDESCAT	
13		21/10/2012	04/11/2012	Disseny Magatzem de dades	
	C	03/10/2012	06/11/2012	PAC2 Anàlisi i disseny	
14		07/11/2012	25/11/2012	Elaboració Memòria Capítol Planificació i Dissenys Inicials	
15		07/11/2012	11/11/2012	Creació BBDD	
16		12/11/2012	30/11/2012	Procés ETL	
17		01/12/2012	13/12/2012	Elaboració Consultes DW	
18		14/12/2012	19/12/2012	Elaboració Informes Addicionals	
	D	07/11/2012	19/12/2012	PAC3	
19		20/12/2012	05/01/2013	Elaboració Memòria Resta Capítols	
20		20/12/2012	24/12/2012	Disseny Presentació	
21		27/12/2012	30/12/2012	Implementar presentació	
	E	20/12/2012	07/01/2013	Lliurament final Memòria , Presentació	
22		07/01/2013	21/01/2013	Contestar Tribunal	
	F	07/01/2013	21/01/2013	Debat	

II-lustració 1 Taula de Tasques

2.1.2 Diagrama de Gantt

	🕒	Name	Duration	Start	Finish	Predecessors
1		📁 Preparació inicial del projecte	9 days?	19/09/12 08:00	27/09/12 17:00	
2	✅	Recollir Informació Projecte	5 days	19/09/12 08:00	23/09/12 17:00	
3	✅	Obtenir bibliografia	9 days?	19/09/12 08:00	27/09/12 17:00	
4		Descarregar i obtenir les eines	9 days?	19/09/12 08:00	27/09/12 17:00	
5		📁 Elaboracio del Pla de Treball	8 days?	25/09/12 08:00	02/10/12 17:00	
6	📅	Recollir informació	8 days?	25/09/12 08:00	02/10/12 17:00	
7	📅✅	Elaborar llista de tasques	4 days?	27/09/12 08:00	30/09/12 17:00	
8	📅	Ordenar i planificar	1 day?	01/10/12 08:00	01/10/12 17:00	7
9	📅✅	Zones vermelles, riscos	3 days	26/09/12 08:00	28/09/12 17:00	
10	📅	📁 PAC2 Anàlisi i disseny	20 days?	03/10/12 08:00	06/11/12 17:00	
11		Estudi DW - Lectura i consulta Bibli	20 days?	03/10/12 08:00	06/11/12 17:00	
12	📅	Elaboració Memoria capítols introd	7 days?	14/10/12 08:00	21/10/12 17:00	
13	📅	Estudi Maquina Virtual i SW	2 days?	14/10/12 08:00	16/10/12 17:00	
14	📅	Estudi Dades facilitades i disseny f	7 days?	14/10/12 08:00	21/10/12 17:00	
15	📅	Obtenció Dades IDESCAT	1 day?	14/10/12 08:00	15/10/12 17:00	
16	📅	Disseny Magatzem de dades	13 days	21/10/12 08:00	06/11/12 17:00	
17		📁 PAC3	39 days?	07/11/12 08:00	19/12/12 17:00	10
18		Elaboració Memoria Capítol Planific	19 days?	07/11/12 08:00	25/11/12 17:00	
19		Creació BBDD	5 days?	07/11/12 08:00	11/11/12 17:00	
20	📅	Proces ETL	19 days?	12/11/12 08:00	30/11/12 17:00	19
21	📅	Elaboració Consultes DW	9 days?	01/12/12 08:00	13/12/12 17:00	20
22	📅	Elaboració Informes Addicionals	6 days?	14/12/12 08:00	19/12/12 17:00	20
23		📁 Lliurament final Memoria , Pre	14 days?	20/12/12 08:00	05/01/13 17:00	17
24		Elaboració Memoria Resta Capítol	14 days?	20/12/12 08:00	05/01/13 17:00	
25		Disseny Presentació	5 days?	20/12/12 08:00	24/12/12 17:00	
26	📅	Implementar presentació	4 days?	27/12/12 08:00	30/12/12 17:00	25
27		📁 Debat	15 days?	07/01/13 08:00	21/01/13 17:00	23
28		Contestar Tribunal	15 days?	07/01/13 08:00	21/01/13 17:00	

II-lustració 2 Planificació Inicial



II-lustració 3 Diagrama de Gantt

2.1.3 Planificació Final

La planificació s'ha anat complint fins la fita d'entrega de la PAC2. En la realització de la PAC3 ens hem trobat amb més problemes per complir els temps que ens havíem marcat.

En l'apartat corresponent al procés ETL ha hagut un petit retràs en els terminis previstos, doncs s'ha complert un dels riscos inclosos en el pla de contingència, l'aprenentatge d'una nova eina. Inicialment l'eina utilitzada no era del tot intuïtiva que hom desitjaria.

De totes formes mitjançant l'ús del llibre de María Carina Roldán (2010), s'ha solucionat aquest risc, sobretot un cop s'ha entès quina era la filosofia del programa.

Això deixa palès que caldria haver-ho previst en la planificació inicial amb més temps per familiaritzar-se en aquestes eines, i haver creat una tasca pel aprenentatge i parametrització de totes elles.

Finalment, degut en part als problemes de rendiment, la major part de retràs acumulat s'ha degut a l'elaboració d'informes i l'aprenentatge del llenguatge MDX.

2.2 Anàlisi de riscos

En aquest apartat veurem els possibles riscos detectats i el seu pla de contingència.

2.2.1 Equips Informàtics

El principal problema i risc al que ens enfrontem durant el desenvolupament de qualsevol tasca informàtica és que un equip o perifèric com ara el disc dur que fem servir es faci malbé.

Les causes per les que això pot passar són molt diverses, una pujada de tensió, apagar l'ordinador sense fer servir els passos correctes, un virus, etcètera...

Per tant, haurem de tenir una còpia de seguretat el màxim d'actualitzada per tal de poder estar més tranquils. Aquesta còpia es faria en un dispositiu extern tal com un disc dur extern o una memòria USB.

A més guardar còpies dels documents en la xarxa, com ara en el correu electrònic o un servei com *dropbox*, d'emmagatzematge a Internet.

2.2.2 Absències

Alhora de planificar cal tenir en compte les diferents absències ja conegudes, i fer la planificació d'acord amb elles.

Actualment hi ha varis períodes on ja sabem que no estarem.

4 al 14 Octubre

1 al 4 de Novembre

6 al 9 de Desembre

A més cal estar en guàrdia per absències provocades per altres motius com una malaltia.

2.2.3 Altres assignatures

El fet d'estar matriculat de varies assignatures pot fer que un endarreriment en una assignatura, pugui afectar la planificació del nostre projecte.

És acceptable un petit endarreriment en la planificació, però en casos més greus hem de donar prioritat al projecte.

2.2.4 Imprevistos laborals

Un altre punt a tenir en compte, és que un projecte laboral provoqui que tinguin que abocar-hi més esforços, o un viatge imprevist.

Caldrà llavors realitzar una nova planificació del projecte, i advertir al consultor.

2.2.5 Greus endarreriments

Es pot donar el cas que durant el desenvolupament del treball final de carrera es detecti un greu endarreriment, que pot ser provocat tan per una mala planificació inicial, com per un problema sorgit durant el treball amb el que no es comptava.

En aquest cas, caldrà advertir al consultor del problema, i planificar-ho de nou el més aviat possible perquè reflecteixi la situació actual. També pot ocasionar dedicar més recursos dels inicialment previstos al projecte.

2.2.6 Aprenentatge Eines

Qualsevol eina nova comporta al principi una corba d'aprenentatge més plana, amb pocs progressos pel temps invertit.

En aquest projecte existeixen moltes eines, la major part, amb les que es treballa per primer cop, per lo que aquest risc ha de ser tingut en compte.

Per combatre'l tan sols hi ha dos maneres, dedicar-hi més temps del previst, o fer ús de la bibliografia o dels articles a Internet, per tal d'accelerar l'aprenentatge.

3 Anàlisi i Dissenys

3.1 Requeriments Funcionals

L'empresa *Grup Líder en Distribució de Proximitat (GLDP)* vol crear un magatzem de dades sobre les seves vendes, per tal de poder centralitzar i analitzar aquesta informació i millorar-ne els seus resultats.

Explícitament ens demana una sèrie d'informes:

- **Total vendes i marge net del grup:** Un agrupació de les vendes per tots els establiments. El marge net és el total de vendes per establiment, menys els seus costos fixos.
- **Import mitjà de compra per soci i establiment.** Els socis són aquells clients habituals que disposen de targeta del establiment. El número de soci és per establiment.
- **% de vendes respecte el total del grup** (per establiment i per demarcació). El client ens indica que el nivell de demarcació desitjat és la Província.
- **Preus màxims i mínims** per tipologia d'establiment i característiques del producte.
- **Rànquing d'establiments** per nombre de vendes i volum total.
- **"Top ten" de productes** : És a dir els productes més venuts.
- **% de vendes** de "marques blanques" per habitants. S'entén per marques blanques, tots els productes que el seu proveïdor sigui el propi client, GLPD.
- Categorització de **Clients A/B/C**. La categorització pot ser dinàmica per tal que l'usuari triï, quins percentatges indiquen un client A, B ó C.
- Anàlisi **de compra per impuls**. Són aquells productes, que s'ofereixen el dia del soci, al costat de les caixes, per provocar la compra impulsiva del producte.
- Estudi de **temporalitat de vendes**, poder saber quan és ven més i quin tipus de producte.

Així mateix ja ens indica amb quin agrupament i en quines dimensions voldrà aquesta informació, a saber, temporalitat mensual, per demarcació territorial, per tipus d'establiment i per família de productes.

De totes formes, en els propis informes ja ens indica les dimensions per les que està interessat veure la informació.

Podem veure per exemple, termes com socis, habitants, clients...

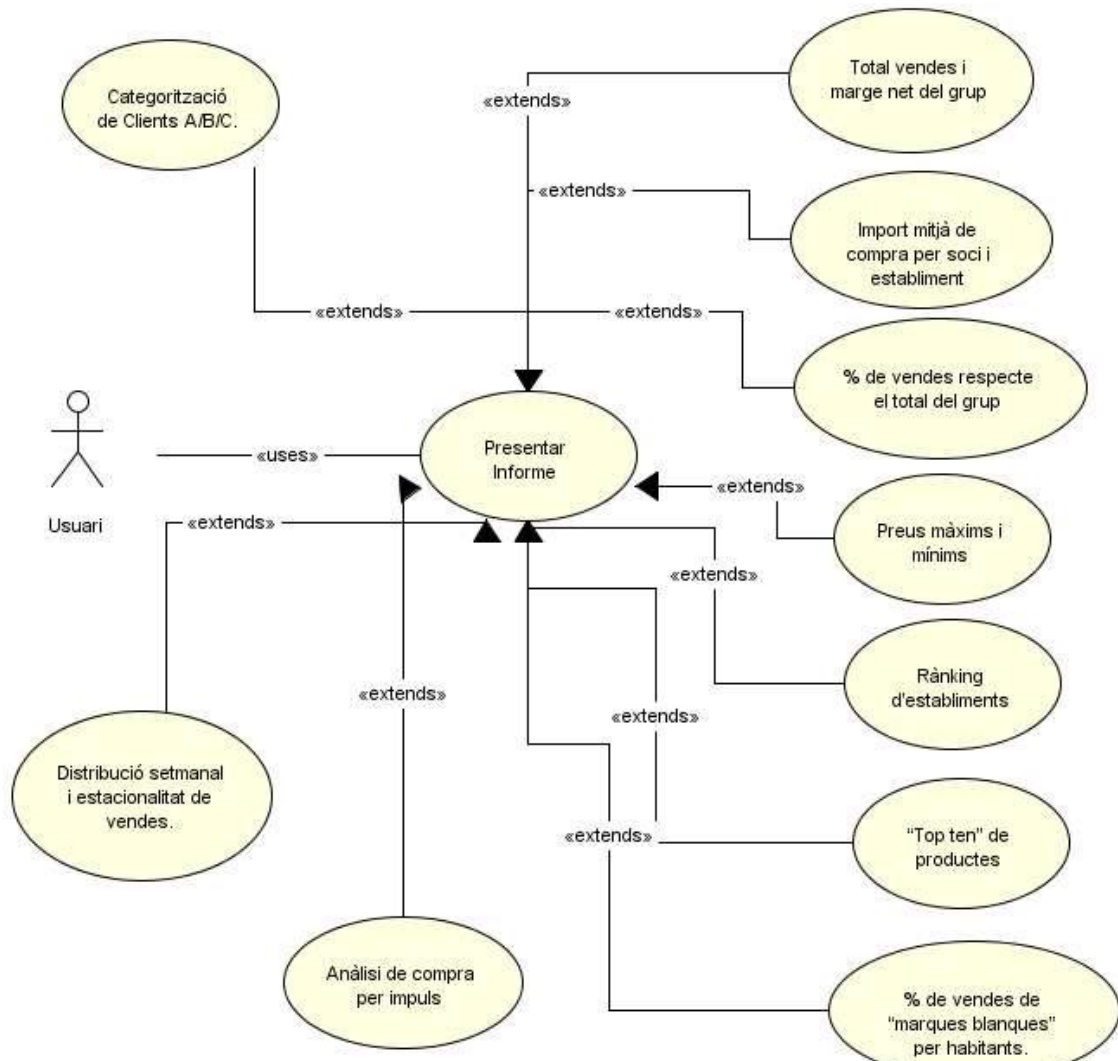
Veurem més endavant, que canviarem la temporalitat, per tal que si en el futur volen aprofundir més en la informació puguin fer-ho sense necessitat de modificar substancialment el model. De fet l'últim informe que ens demana l'usuari ja demana poder veure un informe per setmana o estació, és a dir el marc temporal no seria suficient tenir-lo a nivell mensual.

D'igual manera durant l'anàlisi de la informació facilitada pel client, és possible que incorporem alguna dimensió més a les dades, per tal de millorar-ne l'ús.

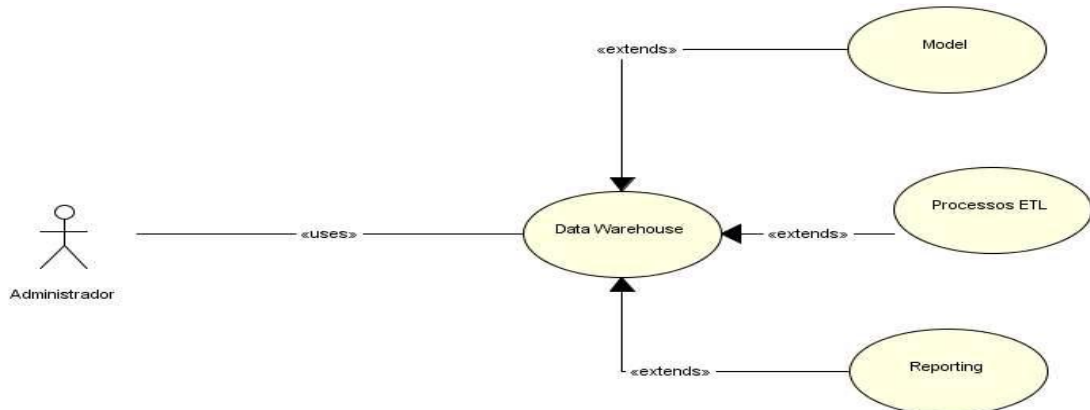
3.1.1 Casos d'ús

En aquest projecte tan sols hi ha dos actors, l'usuari que fa servir els nostres informes, definits en el punt anterior i un administrador encarregat de gestionar l'aplicació i generar nous informes segons les necessitats dels primers, noves jerarquies, dimensions, etc...

Es podria crear un nou tipus d'usuari, més avançat, per tal que sigui qui manipuli els informes dinàmics (ROLAP).



II-lustració 4 Cas d'ús Usuari



II-lustració 5 Cas d'ús Administrador

3.2 Requeriments No Funcionals

En aquest cas es vol una aplicació àgil, on l'usuari pugui consultar els informes indicats, amb unes prestacions de rendiment correctes, per lo qual cal fer un estudi del maquinari adient, segons les necessitats que es determinin.

Aquesta eina ha de permetre l'accés concurrent per diferents usuaris. Tenim dos tipus d'usuari. Un amb perfil administrador al que se li atorgaran permisos per tal de generar nous informes, modificar les dimensions actuals, i generar noves jerarquies. I un segon qui tan sols podrà fer servir els informes presents al sistema.

Actualment no està previst restringir informes segons l'usuari, però a petició del client es podria estudiar. Es tractaria que un client X, tan sols tingués accés als informes del seu establiment per exemple.

O definir diferents tipus d'usuari, un usuari base, i un usuari analista que treballaria amb els informes OLAP, dinàmics, mentre que els usuaris base tan sols farien ús dels reports estàtics.

Al tenir tota la informació en una base de dades estàndard com MySQL, no serà costós efectuar una portabilitat, de fet el propi Kettle de Pentaho, que es fa servir pel procés ETL, es pot utilitzar per realitzar una migració.

A més durant l'elaboració del projecte, es generarà tot de documentació, en aquest cas, seran les diferents entregues i la memòria final.

3.3 Model Conceptual

Segons els requeriments esmentats podem obtenir les mesures i dimensions. Després veurem les taules de fets, així com el seu nivell de detall, la seva granulació.

3.3.1 Mesures

- **Preu Venda** : Preu unitari de venda del producte.
- **Preu Cost**: Preu unitari de cost del producte.
- Volum Vendes (número de vendes), **Unitats**: Unitats venudes en un tiquet de venda.
- **Marge**: Diferència entre preu de venda i preu de cost.
- **Preu Cost total**: Preu de cost per totes les unitats de la línia.
- **Preu Venda total**: Preu de venda per totes les unitats de la línia.
- **Base Imposable Unitària**: Preu unitari sense IVA.
- **Base Imposable Total**: Preu sense IVA per totes les unitats de la línia.
- **IVA Unitari**: IVA per unitat de producte.
- **IVA Total**: IVA per total d'unitats de producte.

- **Descompte Total:** Import Descomptat total.
- **Descompte Unitat:** Import Descomptat Unitat.

El resta de mesures ràtios %, categorització de clients, seran calculades a posteriori, en temps d'execució, o bé en forma de taules agrupades, en cas que es vegi que el rendiment no fos correcte.

Totes aquestes altres mesures, no es poden incorporar directament, doncs són calculades, i no agregables.

3.3.2 Dimensions

- **Temporal:** La data amb la que es realitza la compra. Ho tindrem a nivell diari, doncs així podem establir qualsevol tipus de temporalitat desitjada. Estació, Dia de la setmana, Setmana del any, etc...
- **Horària:** Ens pot servir per tenir relacions de torns, matí o tarda. Es posa per futures ampliacions, doncs actualment no caldria, o inclús podria tractar-se d'un camp calculat.
- **Establiment:** A més de la informació de l'establiment, com la seva tipologia o els seus costos, tindrem les dades de la província on s'ubica.
- **Producte:** Totes les dades vinculades al producte venut, el seu proveïdor, el seu tipus, etc...

A més hi haurà varies dimensions a la taula de fets, que no tindran correspondència amb cap taula de dimensions, són les que es coneixen com dimensions degenerades. Un exemple seria la forma de pagament.

Un dubte, que hem tingut és si el client és una dimensió degenerada o té la seva pròpia taula de dimensions. Com realment tan sols tenim el seu identificador i el establiment al que pertany, i a més la seva categoria hem dit que serà calculada de forma dinàmica, hem decidit tenir-ho com dimensió degenerada, encara que es codificarà de nou perquè faci esment a l'establiment, ja que un mateix identificador pot fer referència a clients diferents.

3.3.3 Taula de Fets





La nostre taula de fets, a nivell de granularitat, es correspon amb la línia del detall de venda, de forma que no fem cap mena d'agrupació prèvia. Per saber el volum de dades al que fem referència mirem la quantitat d'informació actual.

Establiment	Nombre de Registres
Tarragona	231.848
Girona	151.920
Figueres	154.441
Poble Sec	112.931
Terrassa	169.031
Sants	63.932

Olot	63.932
Vielha	64.069
Lleida	74.345
Tortosa	69.407
Reus	83.808
Total	1.239.664

II-lustració 6 Nombre de registres

Un nombre força acceptable. Així un disseny de la nostre taula de fets podria ser la següent:

 Id_Venda	LONG INTEGER	NOT NULL
 Id_Linia	LONG INTEGER	NOT NULL
 Id_Establiment	INTEGER	NOT NULL
Id_Producte	LONG INTEGER	NOT NULL
Data	DATE	NULL
Hora	TIME	NULL
Forma_Pagament	TEXT	NULL
Id_client	LONG INTEGER	NULL
IVA	INTEGER	NULL
Descompte	INTEGER	NULL
Unitats	INTEGER	NOT NULL
Preu_Venda	SMALL FLOAT	NULL
Preu_Cost	SMALL FLOAT	NULL
Marge	SMALL FLOAT	NULL
Preu Cost total	FLOAT	NULL
Preu Venda total	FLOAT	NULL
Base Imposable Unitària	SMALL FLOAT	NULL
Base Imposable Total	FLOAT	NULL
IVA Unitari	SMALL FLOAT	NULL
IVA Total	FLOAT	NULL
Descompte Unitat	SMALL FLOAT	NULL
Descompte Total	FLOAT	NULL
DiaClient	TEXT	NULL
 Primary Key (Id_Establiment, Id_Linia, Id_Venda)		

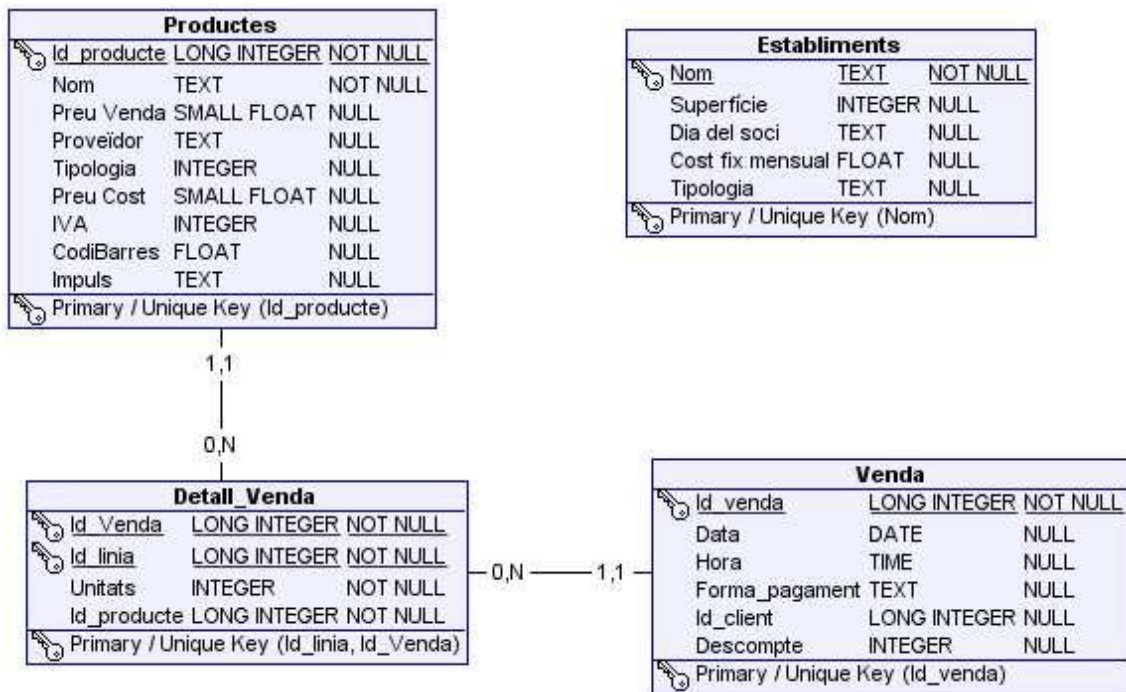
II-lustració 7 Taula de Fets

3.3.4 Origen de les dades

Les dades ens venen en tres formats diferents.

- Microsoft Excels: Informació sobre els Establiments.
- Arxius csv: Informació dels productes.
- Microsoft Access: Informació sobre les vendes de cada establiment.

Els diferents arxius contenen la següent informació:



II-lustració 8 Origen de les dades

Hi ha un fitxer de productes per cada any. El Client ens ha facilitat els últims tres anys. Existeix també una base de dades per cada establiment, indicant les vendes d'aquest establiment.

A més cal recuperar del IDESCAT el nombre d'habitants de les diferents demarcacions. L'idea és guardar la mitjana per a cada establiment. La demarcació es vol a nivell de província. Com es pot veure en la figura anterior, aquesta dada no apareix en el fitxer origen facilitat pel client. La seva incorporació donat el baix número d'establiments està previst fer-ho de forma manual.

3.3.5 Base de dades Temporal (*Staging Area*)

Aquestes dades d'origens diferents es carregaran en un base de dades del nostre servidor MySQL amb caràcter temporal, per poder manipular-ho més fàcilment amb posterioritat. A aquesta base de dades temporal se'l anomena *Staging Area*, que és on s'emmagatzemen temporalment les dades, per tal de fer les transformacions necessàries abans de fer la càrrega cap al magatzem de dades.

L'estructura que tindrem és molt semblant a la vista en el punt anterior, amb algunes petites modificacions. S'incorporarà l'any a la taula de productes, i l'identificador de l'establiment a les de vendes.



A més, aprofitant que les dades dels establiments ens venen en un fitxer Excel consultarem IDESCAT i guardarem dos camps més al Excel amb la informació referent a la demarcació i al número de habitants, calculat amb el promig desitjat.



Les vendes quedaran inserides en un sola taula.





D'igual manera a la taula d'establiments se li afegirà un camp clau numèric.

Analitzant les dades dels productes, veiem que la clau que identifica al producte és el codi de barres, i que d'un any a un altre no canvien els preus, tan sols el tipus de IVA. Però hem de tenir en compte que un any pot haver un canvi de tarifes. De totes maneres, en la carrega temporal no farem aquestes operacions. Però si afegirem un camp clau per la taula de productes. En el següent punt s'explica amb més detall el perquè s'han afegit aquestes claus.

Així finalment tindríem tres taules:

Productes		
 <u>ID</u>	INTEGER	NOT NULL
Any	SMALL INTEGER	NOT NULL
Id_producte	LONG INTEGER	NOT NULL
Nom	TEXT	NOT NULL
Preu Venda	SMALL FLOAT	NULL
Proveïdor	TEXT	NULL
Tipologia	INTEGER	NULL
Preu Cost	SMALL FLOAT	NULL
IVA	INTEGER	NULL
CodiBarres	FLOAT	NULL
Impuls	TEXT	NULL
 Primary / Unique Key (ID)		

Establiments		
 <u>ID Establiment</u>	INTEGER	NOT NULL
Nom	TEXT	NOT NULL
Superfície	INTEGER	NULL
Dia del soci	TEXT	NULL
Cost fix mensual	FLOAT	NULL
Tipologia	TEXT	NULL
Demarcació	TEXT	NULL
Habitants	LONG INTEGER	NULL
 Primary / Unique Key (ID_Establiment)		

Venda		
 <u>Id Venda</u>	LONG INTEGER	NOT NULL
 <u>Id Linia</u>	LONG INTEGER	NOT NULL
 <u>Id Establiment</u>	INTEGER	NOT NULL
Id_Producte	LONG INTEGER	NOT NULL
Data	DATE	NULL
Hora	TIME	NULL
Forma_Pagament	TEXT	NULL
Id_client	LONG INTEGER	NULL
Descompte	INTEGER	NULL
Unitats	INTEGER	NOT NULL
 Primary Key (Id_Establiment, Id_Linia, Id_Venda)		

II-lustració 9 Staging Area

A més d'aquestes tres taules que ens venen directament de la informació facilitada pel client, ens crearem uns fitxers amb la informació de les futures dimensions de data i temps.

Per tal d'aprofitar les diferents formules oferides per Microsoft Excel, crearem aquests dos fitxers com Excel, amb una sèrie de formules per obtenir els camps desitjats, com per exemple si un dia és cap de setmana o no, quin dia de la setmana era, etc... Ho exportarem a un fitxer de tipus csv i ho importarem en dues taules temporals més.

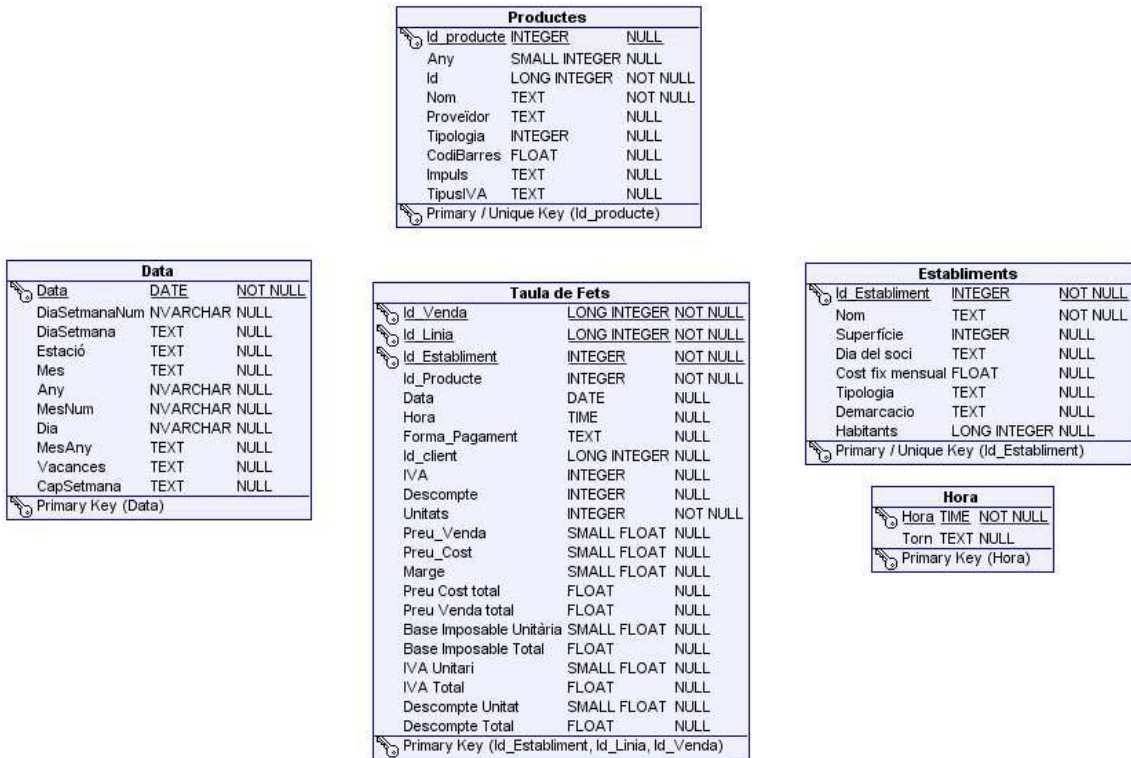
3.4 Disseny de la BD – Diagrama E-R

Respecte el disseny de la BD Temporal hi haurà forces modificacions. Es reorganitza tota la informació per tal que vagi a la taula de fets, o la taula de dimensions que pertoca.

En aquest pas, com els dels camps que actualment tenim en algunes de les taules temporals, passaran a la taula de fets, mentre altres aniran a les de dimensions.

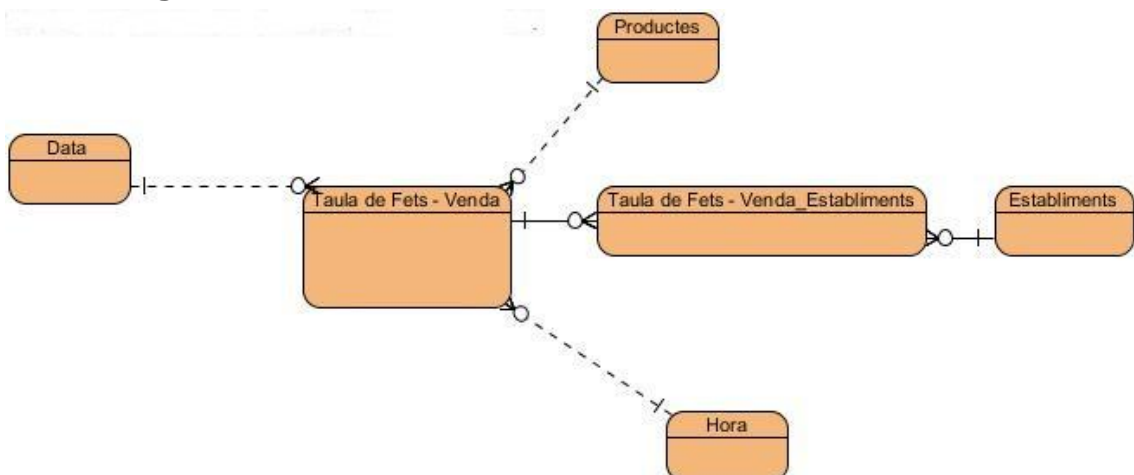
A més es calcularan tots els camps que requereixen algun tipus d'operació.

Seguint els consells de Ralph Kimball i Margy Ross (2002) la majoria de relacions entre la taula de fets i les taules de dimensions, es fan amb claus substituïdes, l'anomenant en anglès *Surrogate key*. Això es fa per motius d'eficiència a nivell d'espai, i per independència de les dades facilitades pels clients.



II-lustració 10 Disseny Final

3.4.1 Diagrama E-R



II-lustració 11 Diagrama E-R

3.5 Procés ETL

Les sigles corresponen a *Extraction, Transformation and Loading*. És a dir, extracció, transformació i càrrega.

Es tracta del procés on partint de les fonts de dades origen, arribem a carregar la nostra base de dades destí, fent totes les transformacions que ens interessin.

En el nostre cas, com a programari utilitzarem el que ens facilita l'empresa Pentaho, PDI o *Pentaho Data Integration*, amb el seu motor *Kettle* i el seu entorn gràfic *Spoon*.

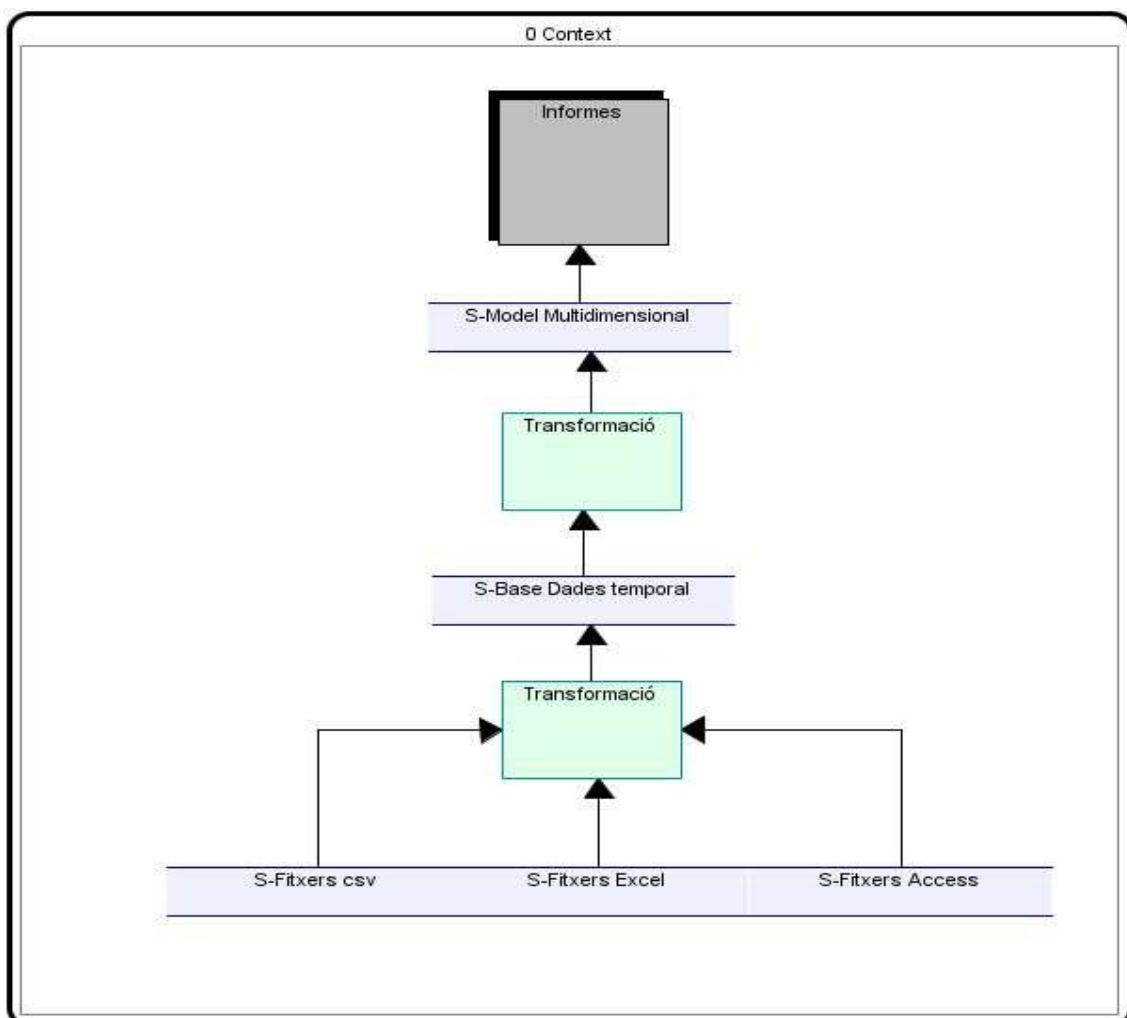
En el nostre cas, tenim uns fitxers origen, que voldrem carregar en unes taules temporals a una base de dades MySQL, ja havent realitzat algunes transformacions.

I al final carregar les taules del nostre model multidimensional.

A banda d'aquest procés, cal veure dos temes de molta importància, l'automatització de tot el procés, de forma que es pugui fer sense la nostre intervenció.

I com cal tractar les actualitzacions, quan es generi per exemple un nou any, o les vendes. La discussió d'aquestes actualitzacions ha de incloure, la seva periodicitat.

Això cal parlar-ho amb el client, per si està en disposició de facilitar la informació.



II-Il·lustració 12 Procés ETL

3.6 Errors de Càrrega

En qualsevol procés de carrega un dels punts importants és el tractament dels errors, inevitables quan tractem amb grans volums de dades.

Al detectar qualsevol error, més enllà de la solució que s'empri per la seva resolució, o minimització, caldrà avisar al client, per tal que gradualment es vagin netejant les dades. O en alguns casos, ens confirmi o no, si las suposicions inicials són correctes.

3.6.1 Productes

Durant el anàlisi previ dels fitxers de productes, s'han detectat algunes anomalies, a saber:

- Registres sense proveïdor. (Any 2010 identificador 651)

En el cas que no tinguem un proveïdor informat s'informarà per defecte amb el text "Sense Proveïdor".

- Registres sense preu cost.

Si ens trobem un producte sense un tipus de preu, preu de cost o preu de venda, l'informarem amb el preu informat. És a dir en el cas que no tinguem preu cost, informarem aquest camp amb el preu de venda. Si cap dels dos tenen preu s'informarà amb preu zero.

- Registres amb preu cost zero.

Aquells registres amb preu cost zero, com desconeixem si podrien ser obsequis o mostres per part del proveïdor, es mantindran amb aquest preu.

- Registres amb nom proveïdor incorrecte.

Finalment s'intentarà homogeneïtzar els noms dels proveïdors, de forma que per culpa d'una errada ortogràfica no tinguem dos proveïdors diferents, quan es tracta d'un sol.

- Preus de cost per sobre els preus de Venda

Es permet vendre productes per sota el seu cost, per liquidar estoc per exemple? Si el client no ens diu lo contrari, suposarem que si.

3.6.2 Establiments

Durant el anàlisi previ no s'ha detectat cap anomalia.

3.6.3 Vendes

- Identificador de línia zero

No es realitza cap acció, es carrega igualment el registre. Tan sols es fa servir com part de la clau i per identificar el registre origen.

- Identificadors de línies sense continuïtat dins un mateix tiquet de venda.

No es realitza cap acció, es carrega igualment el registre. Tan sols es fa servir com part de la clau i per identificar el registre origen.

- 0 Unitats

No té cap sentit, s'elimina el registre.

- Unitats negatives

El sistema permet devolucions? Si el client no ens diu lo contrari suposarem que sí.

- Vendes d'anys anteriors al horitzó fixat (2009 per exemple)

No es carregaran aquests registres.

- Identificador client en blanc

S'informarà el camp amb un zero.

- Detalls venda sense correspondència amb cap capçalera.

No es carregaran aquests registres.

- Capçalera sense cap línia.

No es carregaran aquests registres.

3.6.4 Generals

- Venda fa referència a producte inexistent o Producte en blanc.

Es crearà un producte genèric anomenat "Producte Inexistent".

En general a la taula de fets no es deixarà cap dimensió sense informar, ni amb nuls ni amb blancs, si és necessari es crearà un registre a la corresponent taula de dimensions amb un literal com Sense Dimensió, per exemple el cas del Producte inexistent. O bé si es tracta d'una dimensió degenerada sobre la pròpia taula de fets com en el cas dels proveïdors. "Sense Proveïdor".

4 Desenvolupament

En aquest apartat s'explicarà el treball realitzat per tal d'assolir les fites proposades.

Explica els passos realitzats pel desenvolupament del projecte. Un cop finalitzat l'anàlisi i el disseny, s'ha d'implementar realment.

El procés no és ben bé lineal, doncs a mesura que treballem, veiem que ens fa falta refinar o modificar un pas anterior. De forma, que el model original es va modificant per tal de millorar-lo.

4.1 Programari Utilitzat

Aquesta implementació es realitza sobre una màquina virtual amb Windows XP, servidor de base de dades MySQL i la suite de Pentaho preinstal·lada.

Primer es realitza el procés ETL per incorporar les diferents fonts de dades al nostre sistema. Aquesta tasca es realitza principalment amb l'eina Pentaho Data Integration (PDI o Kettle).

Després s'implementa el Cub amb el que treballarem, amb les seves dimensions i mesures necessàries. Així com altres parametrizacions al sistema. Fem servir entre altres el Schema Workbench també de Pentaho.

Per últim es fa front a la creació dels informes usant el Pentaho Report Designer (PRD) i les eines d'anàlisi facilitades pel portal de Pentaho, JPivot.

Programa	Propòsit
Oracle VM Virtualbox	Executar la màquina virtual.
MV Windows XP	Màquina virtual amb sistema Operatiu Windows XP.
MySQL	Base de dades
Microsoft Excel	Per manipular els fitxers enviats pel client. I Generar els nostres.
Microsoft Word	Elaboració de la documentació.
OpenProj	Planificació del projecte, Diagrama de Gantt
Open Model Sphere	Elaboració diagrames, casos d'ús, E-R, UML.
Pentaho Data Integration	Procés ETL
Schema Workbench	Creació del Cub
Pentaho Aggregation Designer	Per crear les taules agregades
Pentaho Report Designer	Elaboració informes estàtics
JPivot	Elaboració informes OLAP
Mozilla Firefox	Per accedir al portal, i a la consola d'administració.
Tomcat	Servidor Web.
Camtasia	Per realitzar la presentació.

Il·lustració 13 Programari Utilitzat

4.2 Procés ETL

Acrònim de *Extract, Transform, Load*. Extracció, Transformació i càrrega.

La tasca consisteix en integrar la informació distribuïda, en diferents formats al nostre *data warehouse* fent les transformacions necessàries.

4.2.1 Preparació Fitxers

Comencem donant una ullada a les diferents dades que el client ens ha facilitat. Tenim bases de dades Access amb les vendes de cada establiment, un fitxer Excel amb les dades d'aquests, i uns fitxers de tipus csv (valors separats per comes) amb els productes de cada any.

Segons el nostre anàlisi previ necessitarem també carregar un calendari i un fitxer de torns, per lo que preparem dos fitxers Excel amb les dades que estímem necessàries. (dia, mes, vacances, cap de setmana, torn, etc..)

A més el programa PDI amb el que realitzem els processos ETL, no admet el fitxer Excel de nova generació, amb lo que és grava amb una versió anterior que sigui compatible.

Així mateix aprofitant que hem de modificar el fitxer, incorporem les dades de població de IDESCAT, així com la demarcació de cada establiment. també ho posem de forma més còmode pel seu posterior tractament, una fila per cada establiment, amb les dades com columnes.

4.2.2 Carrega Temporal

En aquest punt volem passar tota la informació heterogènia que tenim a la nostre base de dades MySQL. Per això utilitzarem el programa *Pentaho Data Integration*. El programa permet recuperar la informació tan de fitxers Excel, Access o csv. Anomena transformacions a cada procés que generem per realitzar la homogeneïtzació de les dades, poden incloure tan l'extracció, com la transformació i/o la càrrega final. Aquestes transformacions són guardades com fitxers XML. Poden ser guardades tan en disc, com en un catàleg en Base de dades. Hem triat la primera opció per tal de poder realitzar totes les còpies de seguretat parcials (veure pla de contingència) amb més senzillesa.

Tal com hem explicat al pensar el disseny fem ús de les claus substituïdes. En cada càrrega generem un ID únic per emmagatzemar en les taules que seran les que pertanyen a cada dimensió, cada taula amb el seu propi ID, i també la referència a la taula de fets.

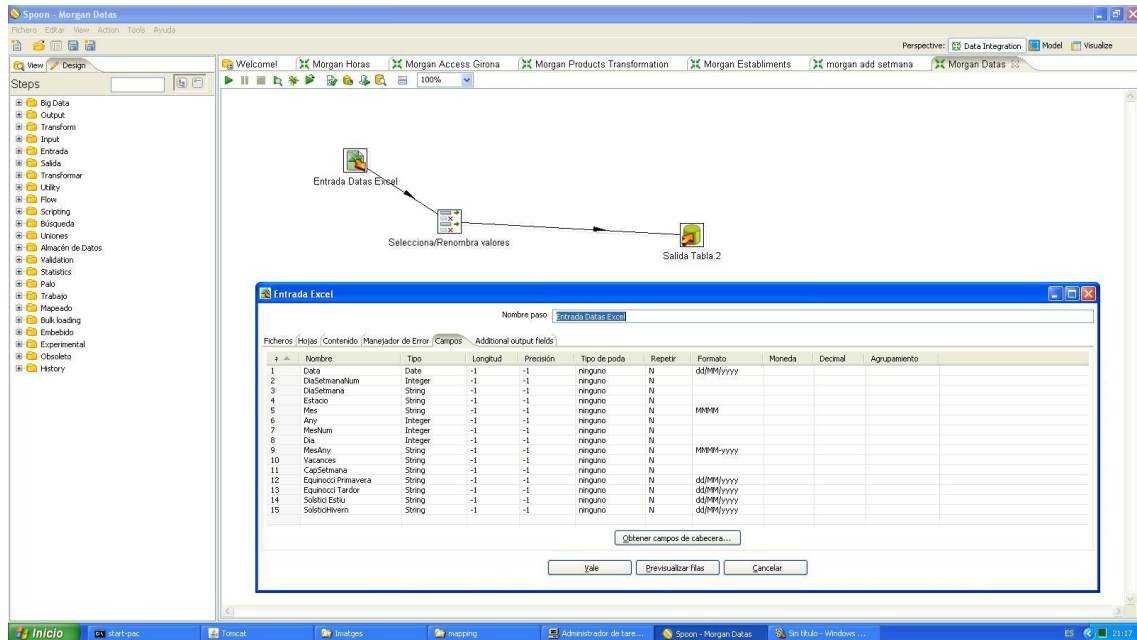
En el fitxer de vendes recuperem el nom del establiment automàticament del nom del fitxer, de manera anàloga en el de productes, recuperem l'any de vigència. En el de vendes, a més, ho gravem ja unificant la informació de capçalera amb les diferents línies. La resta de carregues són més senzilles i directes.

Examinem-ho amb més detall començant amb aquestes últimes, més senzilles.

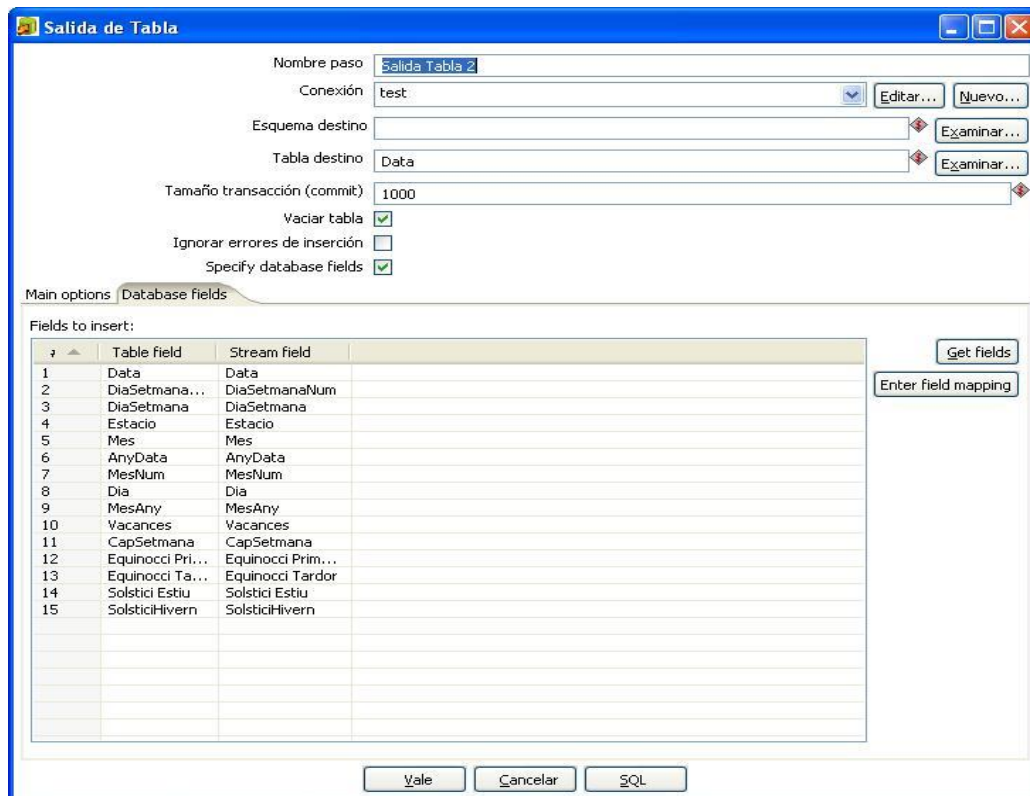
Veiem en la pantalla adjunta, el programa *Spoon*, al que ens hem referit també com *Pentaho Data Integration* o PDI. A l'esquerra es veuen els diferents tipus de funcions que ens ofereix per realitzar les transformacions, i a la dreta, la transformació que ens carregarà a la base de dades la taula **DATA**.

Com veiem, fa servir un connector per obrir el fitxer Excel, que prèviament hem generat amb tota la informació que creiem necessitar. En el pantalla es veu també, tots els camps, així com el format que s'aplica en algun d'ells.

Hi ha un pas per donar-li un altre nom a un camp, el camp Any, per tal que sigui AnyVenda. Això és fa degut a que el nom "Any" és també una paraula clau en alguns sistemes. Per evitar confusions.

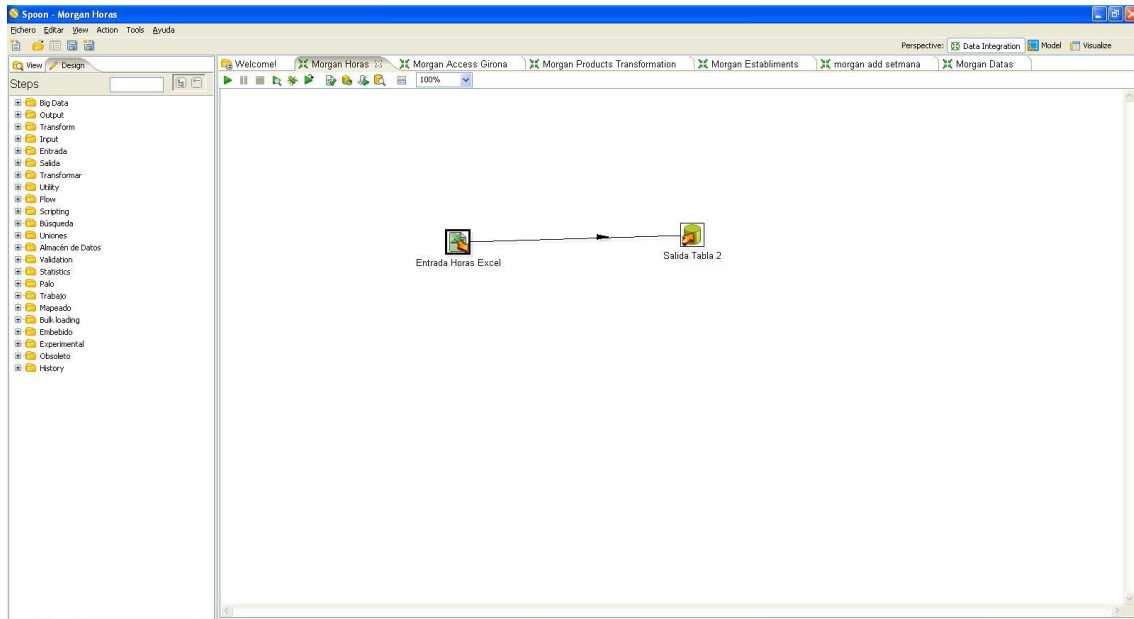


II-Il·lustració 14 Data - Format de camps



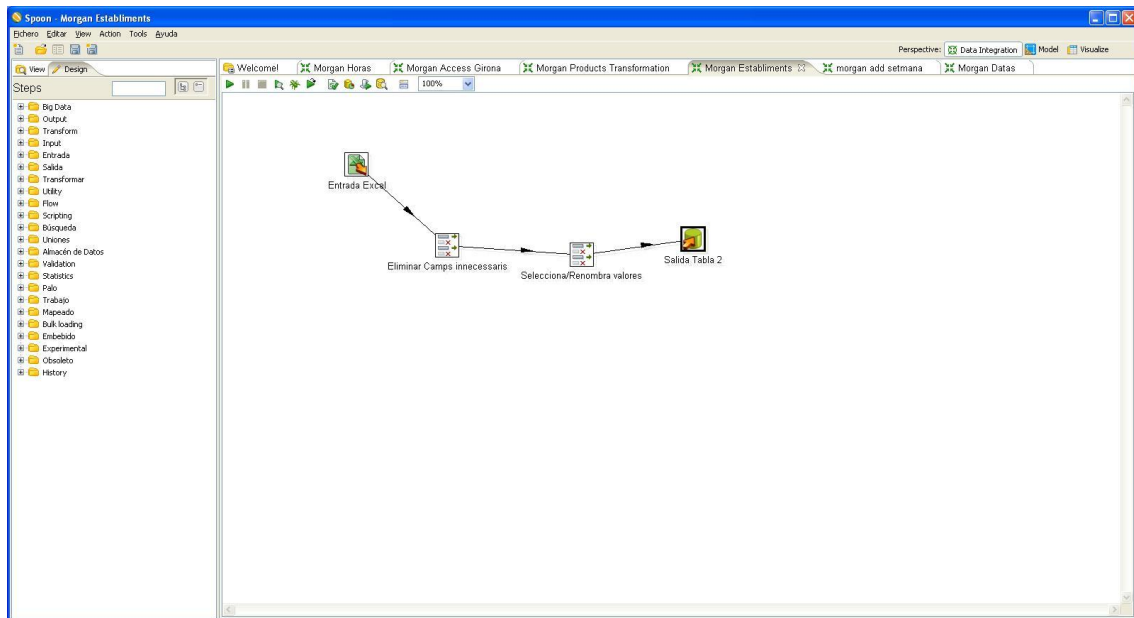
II-Il·lustració 15 Data - Sortida a Taula

A continuació tenim la transformació que ens carregarà les hores, que com es veu encara és més senzilla, sense cap manipulació, obrim el fitxer Excel i és volca la informació a la base de dades en la taula **HORAS**.



Il·lustració 16 Hores - Transformació

Veurem ara la pantalla que correspon a la taula **ESTABLIMENTS**. L'única diferència amb les pantalles anteriors, és el pas Eliminar Camps innecessaris, i el seu nom és prou explícit.



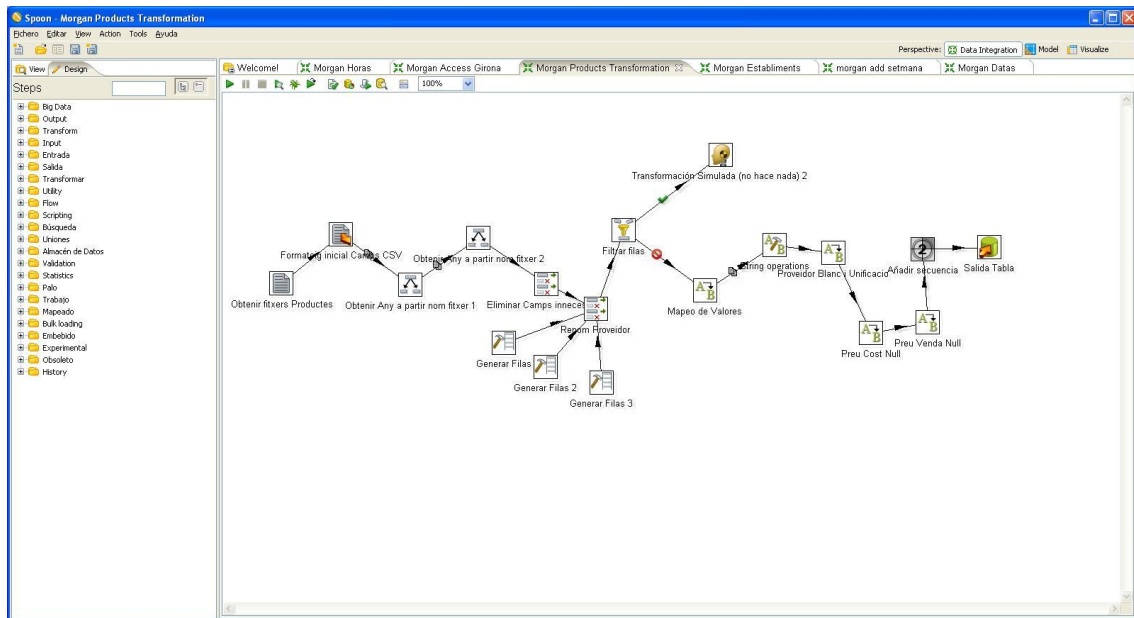
Il·lustració 17 Establiments - Transformació

Ara hem vist les 3 transformacions més senzilles. En el cas de les dates i les hores, degut a que precisament el fitxer Excel l'hem generat nosaltres, i per tant ja hem triat el format que ens ha sigut més còmode pel procés posterior.

I en el cas dels establiments, es podria dir el mateix, doncs del fitxer que ens ha enviat el client, l'hem manipulat per incorporar la informació que mancava, i adaptat a un format compatible amb el programa PDI.

De seguida es nota, que la transformació dels productes és més complicada. A més es carreguen alhora tots els fitxers de productes que es trobin al directori, siguin del any que siguin.

En la primera pantalla veiem tot l'esquema de la transformació, i en la següent, com fem ús dels comodins que ens facilita el programa, perquè carregui tots els fitxers de productes que trobi al nostre directori de carrega ("E:\UOC\Fitxer\"). Han de ser anomenats *Productes XXXX.csv* on XXXX correspon a l'any.



Il·lustració 18 Productes - Transformació

Fichero/Directorio	Comodin	Exclude wildcard	Required	Include subfolders
1	E:\UOC\Fitxer	([P]Productes [0-9]+).csv	N	N

Il·lustració 19 Productes - Càrrega fitxers

Aquí veiem com li donem un format inicial als camps, i cal fer notar com recuperem i passem endavant els noms dels arxius, doncs d'ells extraiem l'any.

CSV Input

Step name:

The filename field (data from previous steps):

Include the filename in the output?

Delimiter:

Enclosure:

NIO buffer size:

Lazy conversion?

Header row present?

Add filename to result

The row number field name (optional):

Running in parallel?

New line possible in fields?

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Id_producto	Integer	#	15		€	,	.	ninguno
2	Nom	String		72		€	,	.	ambos
3	Preu_venta	Number	#.#	15	2	€	,	.	ninguno
4	Proveedor	String		30		€	,	.	ambos
5	Tipologia	Integer	#	15		€	,	.	ninguno
6	IVA	Integer	#	15		€	,	.	ninguno
7	Preu_cost	Number	#.#	15	2	€	,	.	ninguno
8	Codi_barres	String	#	15		€	,	.	ambos
9	Venda_impuls	String				€	,	.	ambos

II-lustració 20 Productes - Format Camps

Partir campo

Nombre paso:

Campo a partir:

Separador:

Campos

#	Nuevo campo	ID	¿Eliminar ID?	Tipo	Longitud	Precisión	Formato	Grupo	Decimal	Moneda	NuloSi	Default	Trim type
1	nombre		N	String	40								ambos
2	Año_temp		N	String	8								ambos

II-lustració 21 Productes - Obtenir any

Ara s'adjunten una sèrie de pantalles on es veu les transformacions que patiran les dades que hem rebut, per tal d'adequar-les a les nostres necessitats. En concret, s'eliminen aquells registres que no tenen identificador de producte, es posa el preu cost i preu venda a zero, en cas de no estar informat, i es valida el format del camp Venda impuls.

Més interessant resulta el pas de Proveïdor Blanc i unificació, amb aquest pas es normalitza el nom del proveïdor, doncs s'havien detectat molts errors. Això es deu a que no és un camp normalitzat, no existeix cap taula mestre de proveïdors, i són els propis usuaris els que producte a producte han entrat aquests valors.

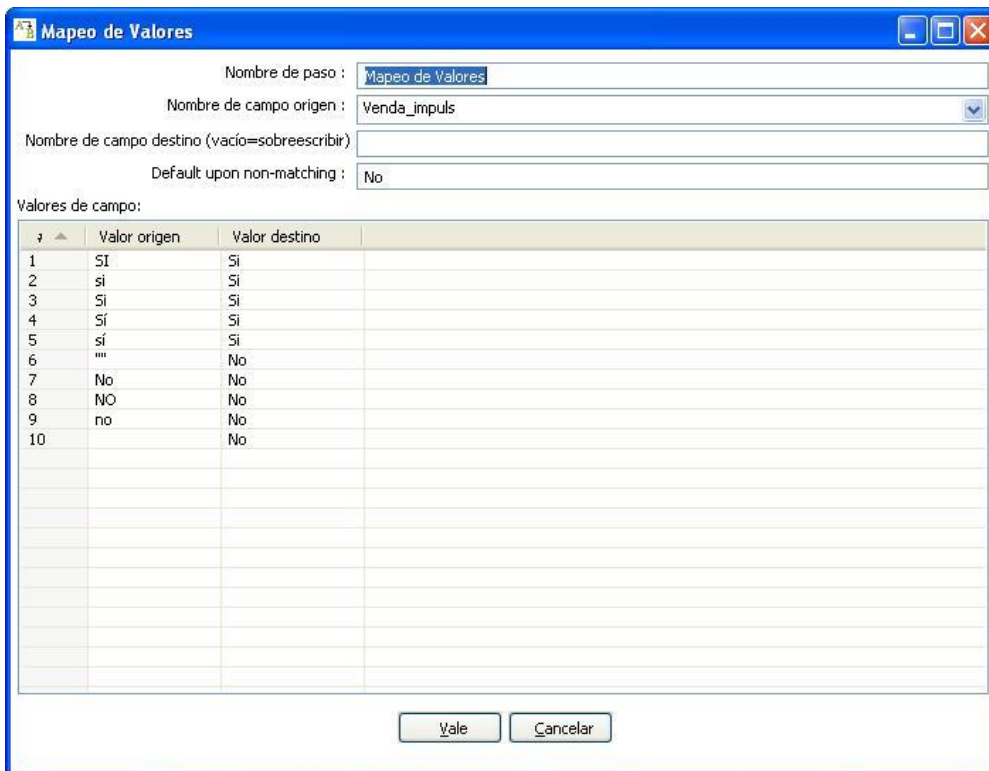
Per tant és normal que hi hagi forces errors, que el nom s'hagi escrit de forma diferent.

S'ha seguit el següent procediment:

- S'ha extret una llista de proveïdors ordenada alfabèticament, i amb el nombre de productes que fan referència a cadascun.
- S'ha procedit a relacionar aquells proveïdors que semblen ser el mateix. I d'aquests, s'ha seleccionat aquell al qui els productes feien major nombre de referències.
- I en el pas de la transformació, es canvia el nom a aquest últim.
- En cas de dubte, no s'ha modificat.
- Es valida amb el client.



II-lustració 22 Productes - Filtrar registros



II-lustració 23 Productes - Mapa Venda_impuls

Nombre de paso : Proveedor Blanc i Unificació

Nombre de campo origen : Proveedor

Nombre de campo destino (vacío=sobreescribir) :

Default upon non-matching :

Valores de campo:

#	Valor origen	Valor destino
1		Sense Proveedor
2	NULL	Sense Proveedor
3	" "	Sense Proveedor
4	AGRO NEC	AGRO-NEC
5	ANEX PLANT	ANEXPLANT
6	CERAS ROURA	CERES ROURA
7	CERAS ROURA S.A.	CERES ROURA
8	CNZP BORDET	CNCP BORDET
9	COLONI	COLONY
10	GREENDEL S.L.	GREENDEL
11	MAGISA	MAGHISA
12	MAIOL GERMANS	MAIOL
13	GERMANS MAIOL	MAIOL
14	ORIENTAL CRAFATS, S.A.	ORIENTAL CRAFTS
15	ORIENTAL CRAFTL	ORIENTAL CRAFTS
16	ORIENTAL CRAFTS S.A.	ORIENTAL CRAFTS
17	ORIENTAL CRAFTS	ORIENTAL CRAFTS
18	ORIENTAL CRAFTS	ORIENTAL CRAFTS
19	OVERSAT S. L.	OVERSAT
20	PRO SANDIMAS	PROSANDIMAS
21	SERMAG DIS.	SERMAG DISTRIBUCIONS

Vale Cancelar

II-lustració 24 Productes - Homogeneïtzació Proveïdors

Nombre de paso : Preu Cost Null

Nombre de campo origen : Preu_cost

Nombre de campo destino (vacío=sobreescribir) :

Default upon non-matching :

Valores de campo:

#	Valor origen	Valor destino
1	NULL	0
2		0

Vale Cancelar

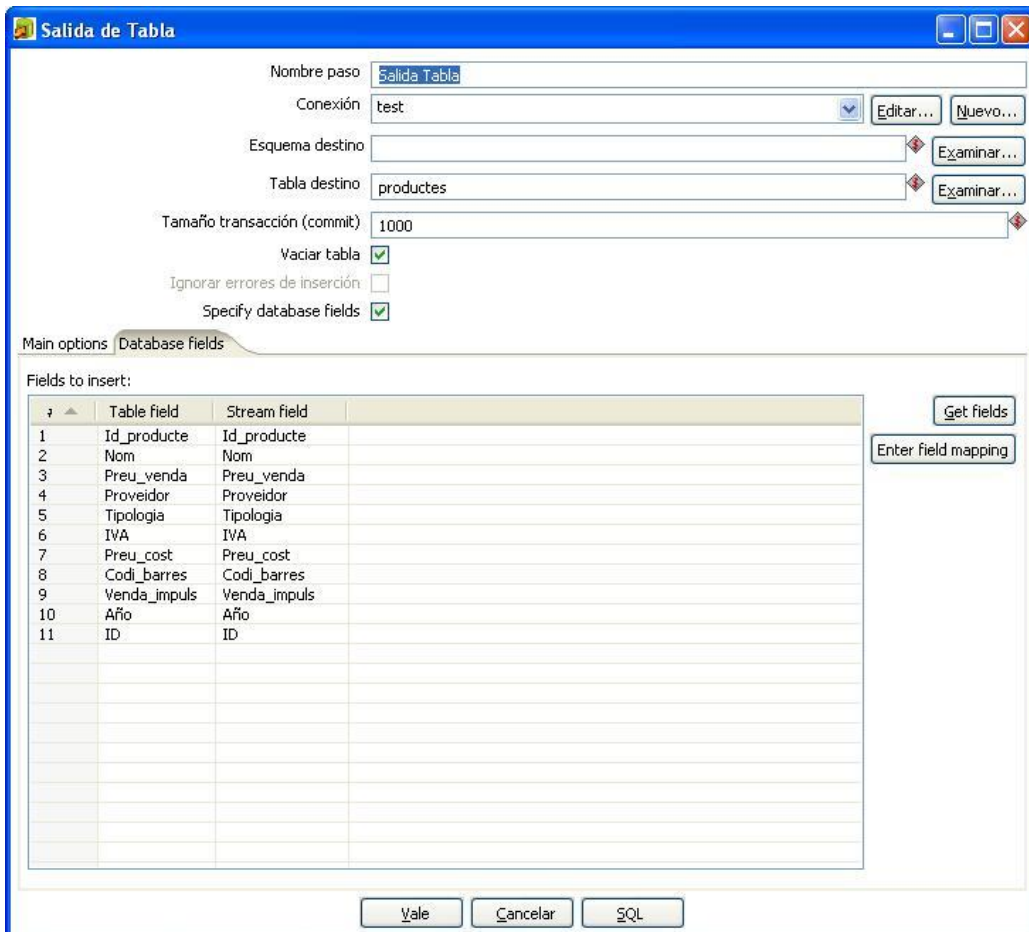
II-lustració 25 Productes - Preu Cost Nul

Aquí s'afegeix la clau substituïda.



II-lustració 26 Productes - Afegir clau

Per últim, es grava a la taula **PRODUCTES**. Hi ha marcat l'indicador de buidar taula. Amb lo que en aquest pas, buida la taula e incorpora els nostres registres. Aquest indicador és important alhora de realitzar el manteniment o actualitzacions posteriors.



II-lustració 27 Productes - Sortida Taula

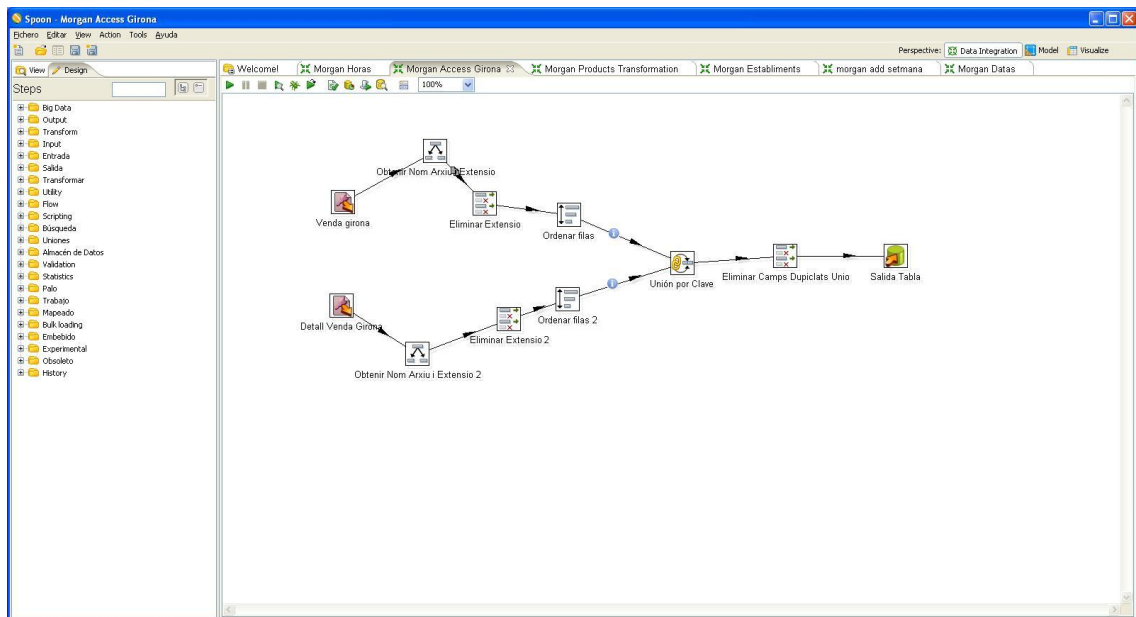
Finalment s'ha creat una transformació per carregar les vendes de cadascun dels establiments. Totes elles són idèntiques, per lo que tan sols n'examinarem una.

Hi ha però, un únic detall de diferència entra el primer establiment del que carreguem les vendes i la resta. Es tracta de l'indicador que acabem de veure de buidar taula. En el primer establiment està marcat, i en tota la resta, òbviament, no.

L'esquema tal com veiem és més senzill que en el cas de productes. Es veu com inicialment hi ha dos fluxos d'execució, un correspon a la capçalera de les vendes i l'altre al detall de les mateixes.

Pentaho ens ofereix la possibilitat de carregar directament des dels fitxers de la base de dades Microsoft Access. Carreguem les dades amb els connectors que ens facilita.

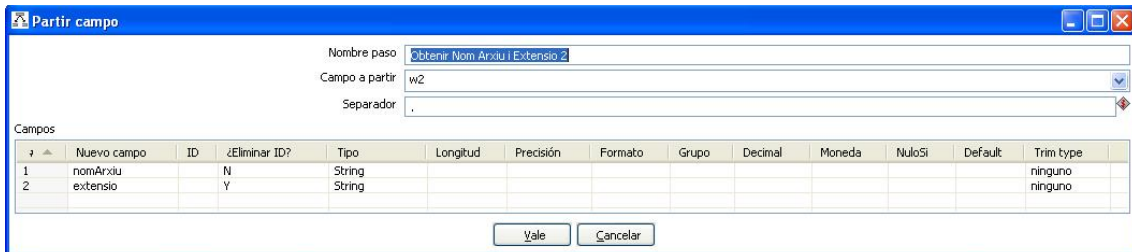
De forma semblant a com passava amb el fitxer de productes, passem el nom del fitxer als passos posteriors, per tal d'obtenir de forma automàtica el nom del establiment.



II-lustració 28 Vendes - Transformació

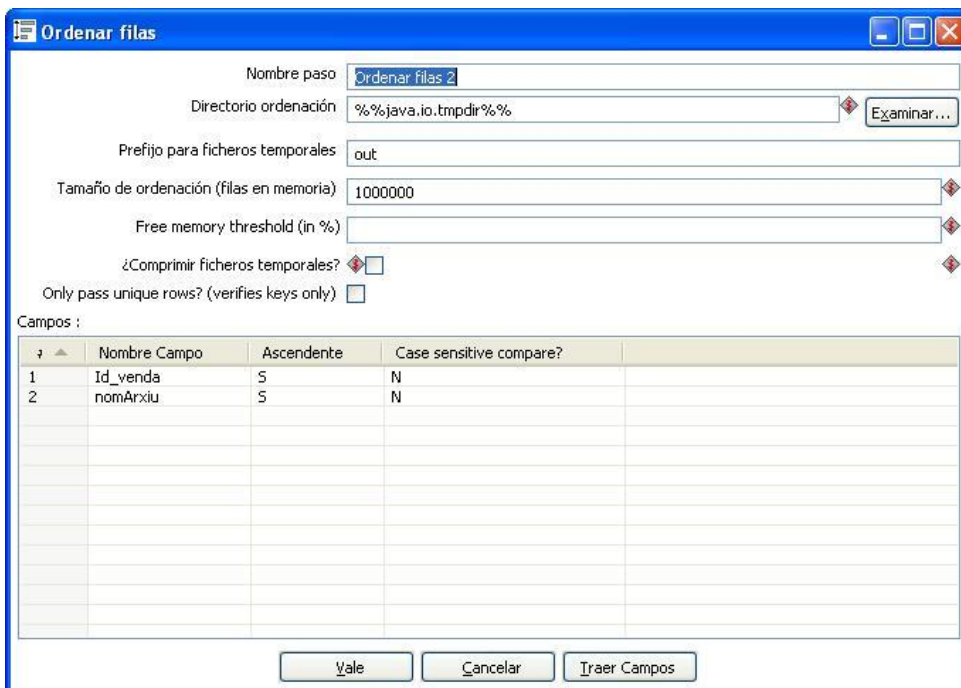
Field Name	Value
Short filename field	w2
Extension field	
Path field	
Size field	
Is hidden field	
Last modification field	
Uri field	
Root uri field	

II-lustració 29 Vendes - Nom del fitxer

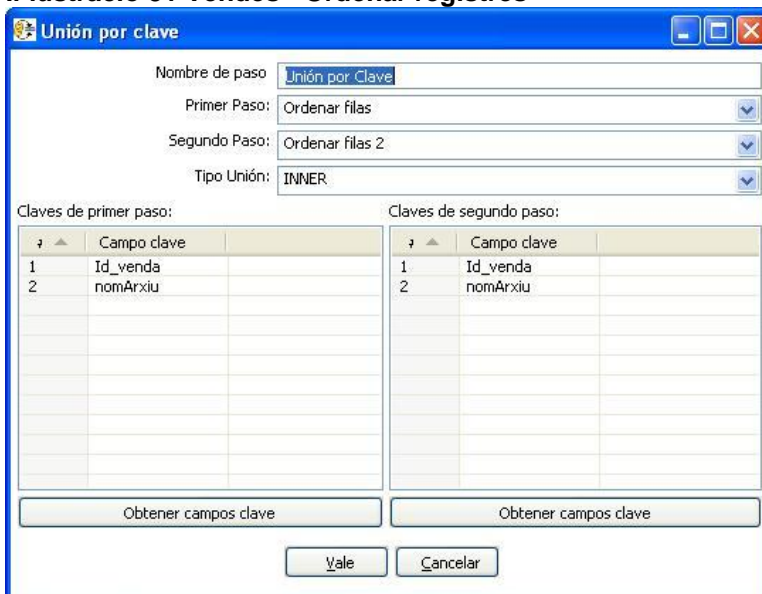


II-lustració 30 Vendes - Nom del establiment

S'ordenen els registres per l'identificador de venda i pel nom de l'arxiu, per tal de poder fer servir el pas "Unión por clave" que requereix aquesta ordenació. D'aquesta manera unim les dues taules, el detall amb la seva corresponent capçalera.

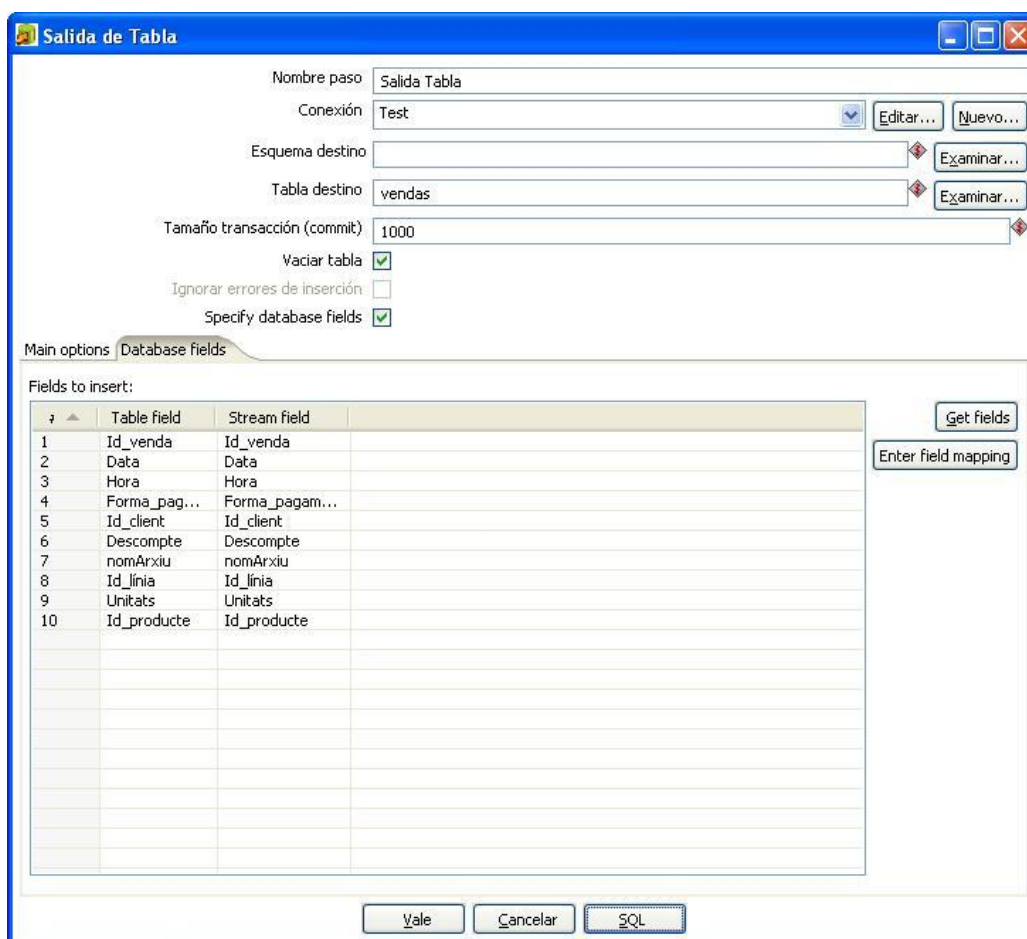


II-lustració 31 Vendes - Ordenar registres



II-lustració 32 Vendes - Unir Capçalera i detall

Per últim gravem les dades a la taula temporal Vendes.



II-lustració 33 Vendes - Sortida Taula

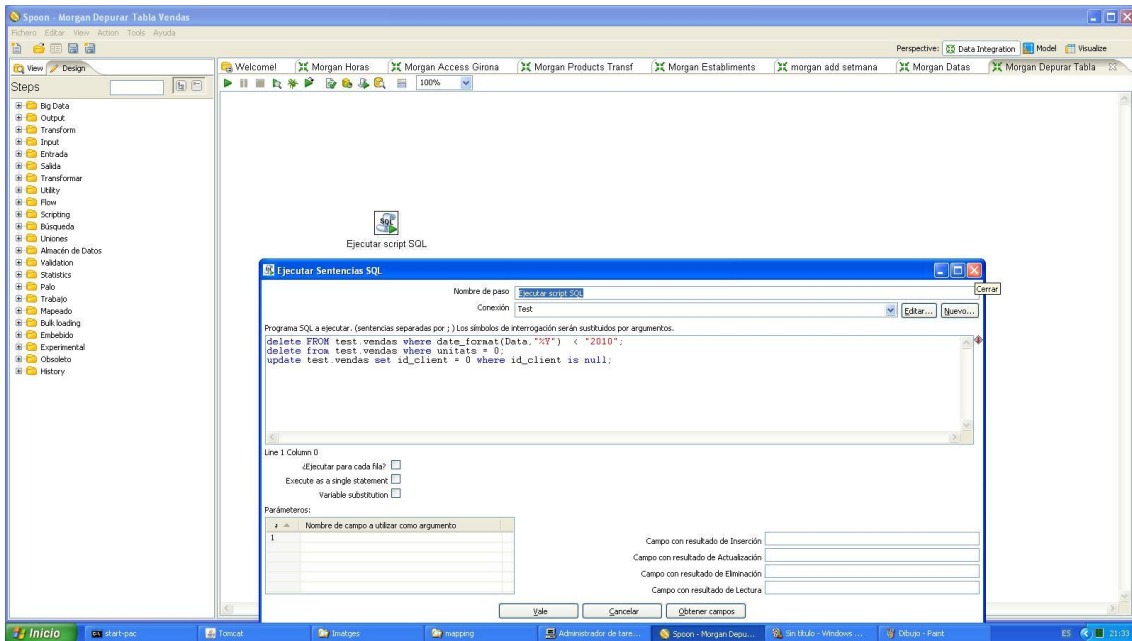
Amb això finalitza la càrrega temporal. Ja tenim totes les dades, amb els seus orígens heterogenis, en una única base de dades, MySQL en el nostre cas.

4.2.3 Càrrega Final

En aquest pas passem de la nostra taula de vendes temporal, a la nostra taula de fets, calculant tots els camps, mesures, que hem determinat que ens serien útils (marge, venda amb iva, sense, totals, descomptes). A més substituïm a la taula de fets cada referència a una dimensió, per la ID d'aquesta, la clau substituïda que hem creat amb anterioritat.

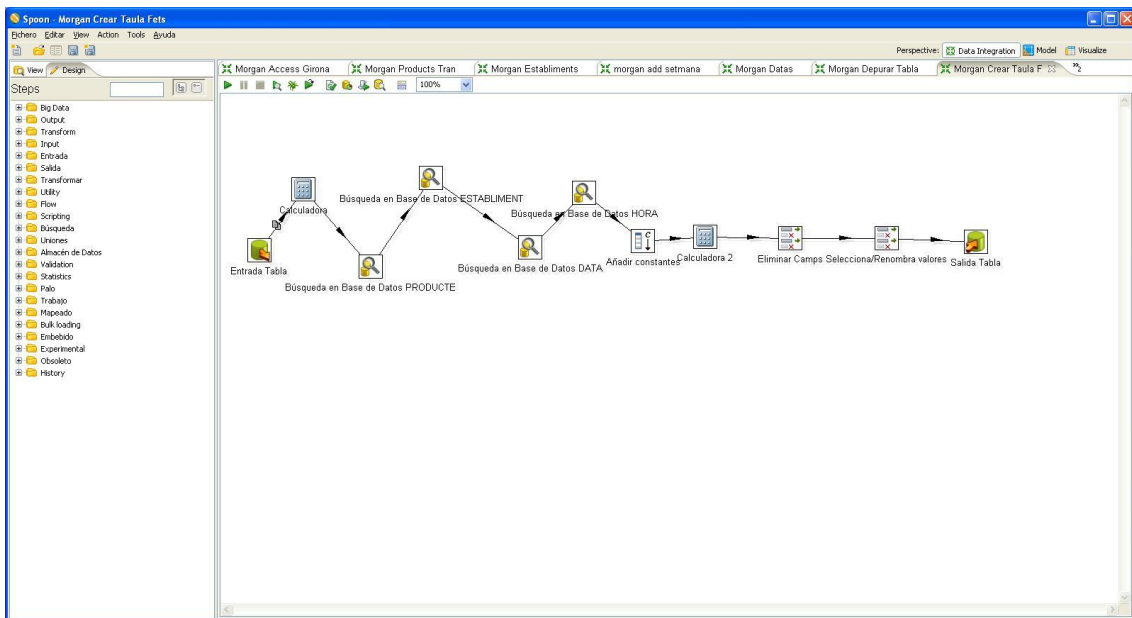
A nivell de base de dades es generen a la taula de fets claus foranies amb cada una de les taules de dimensions.

El primer pas és depurar la taula de vendes temporal. S'ha vist que existien registres anteriors al 2010, vendes sense unitats, i s'actualitza la taula perquè el camp client sigui zero i no un valor nul.



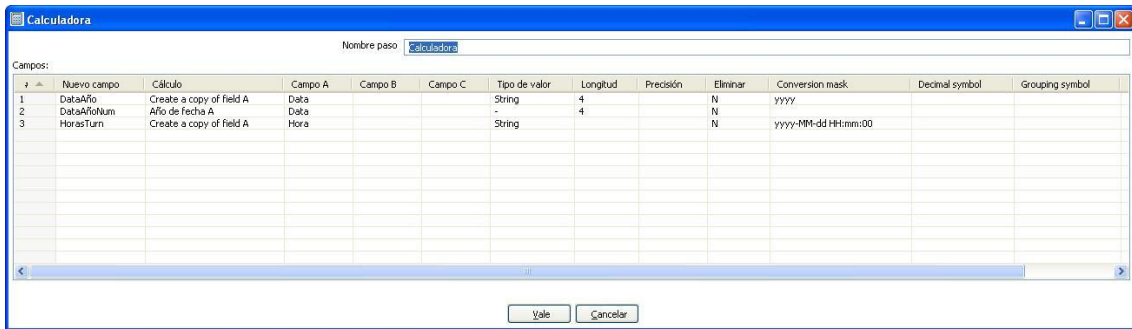
Il·lustració 34 Depuració registres

Tenint com entrada aquesta taula temporal, depurada, anem a realitzar l'última transformació.



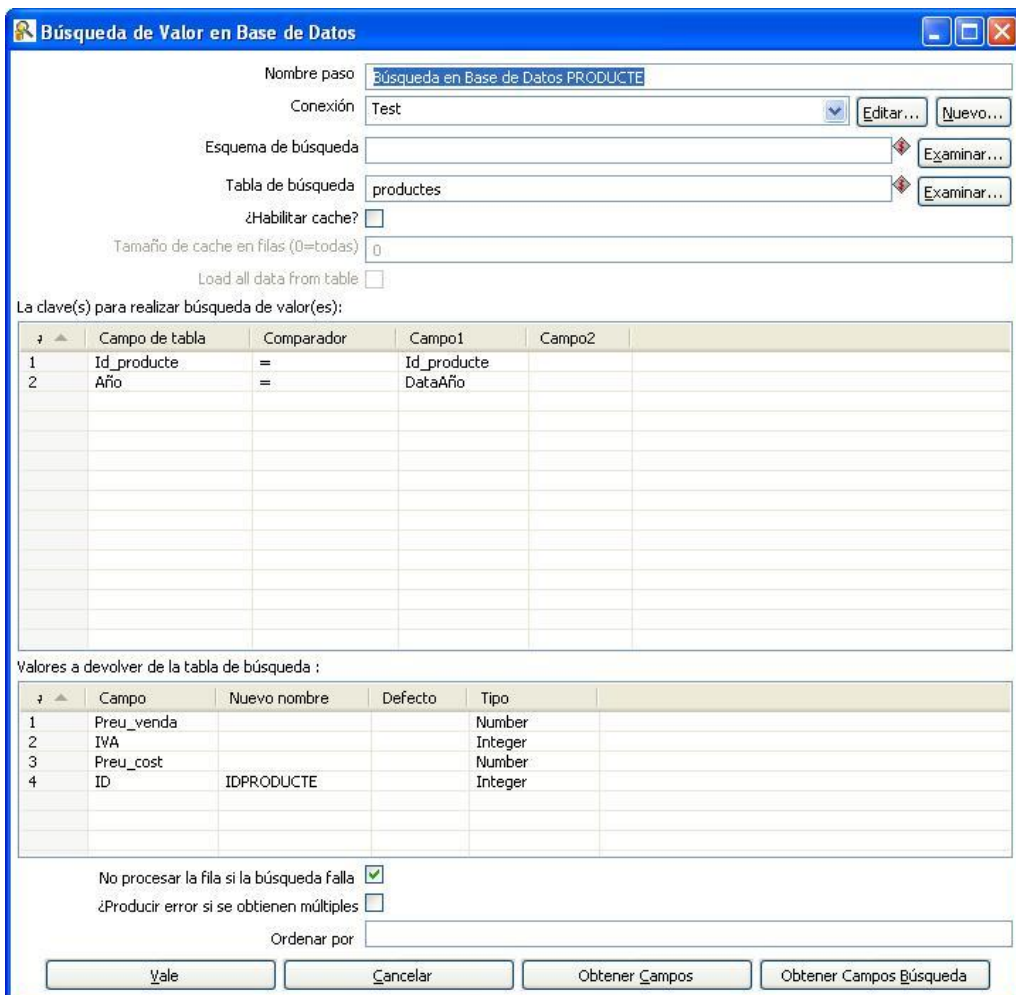
Il·lustració 35 Taula de Fets - Transformació

Calculem els camps i els hi donem el format que volem.



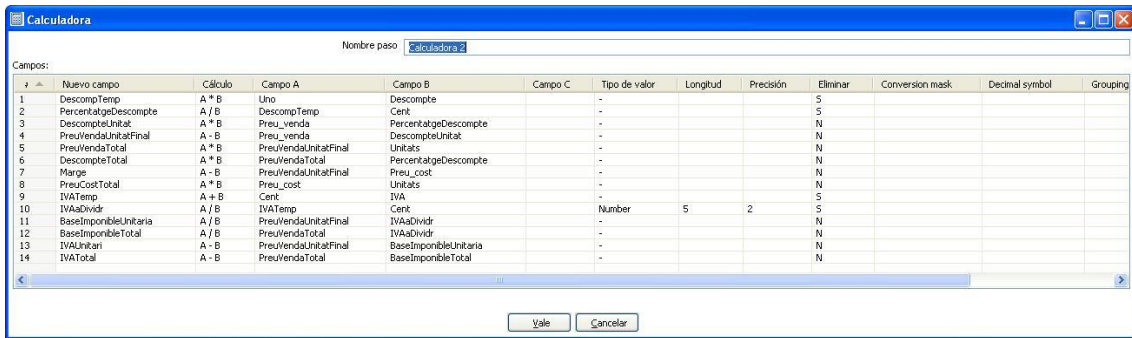
II-lustració 36 Taula de Fets – Càlculs

I fem la substitució del producte, establiment, data i hora per la seva clau substituta. Per exemple en el cas del producte. A més de fer aquesta substitució, recuperem de la taula producte, els preus, que farem servir per les nostres mesures.



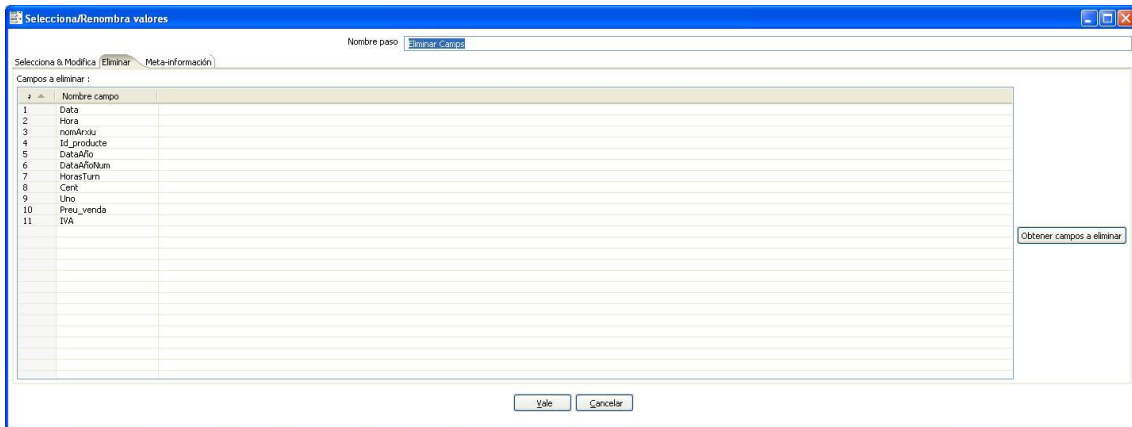
II-lustració 37 Taula de Fets - Recerca producte

Calculem totes les mesures que són càlculs d'altres.



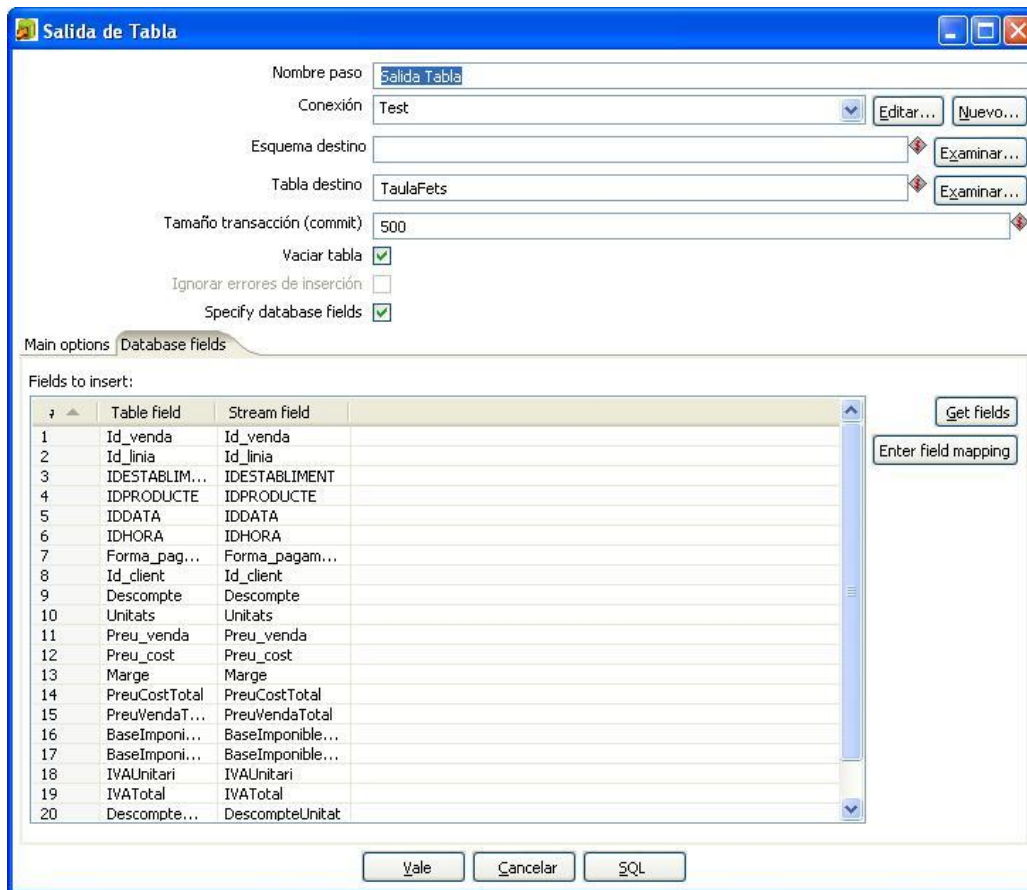
II-lustració 38 Taula de Fets - Calcular Mesures

I eliminem els camps que ja no necessitem, com per exemple el nom de l'establiment, la data, l'hora, el producte, ja tenim la nova clau.



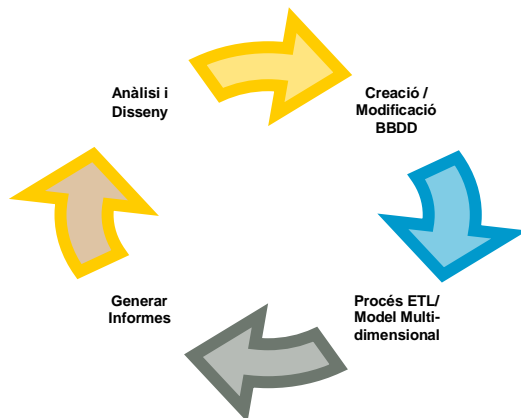
II-lustració 39 Taula de Fets - Netejar Camps

I finalment la sortida cap a la nostre taula de fets **TaulaFets**.



II-lustració 40 Taula de Fets - Sortida Taula

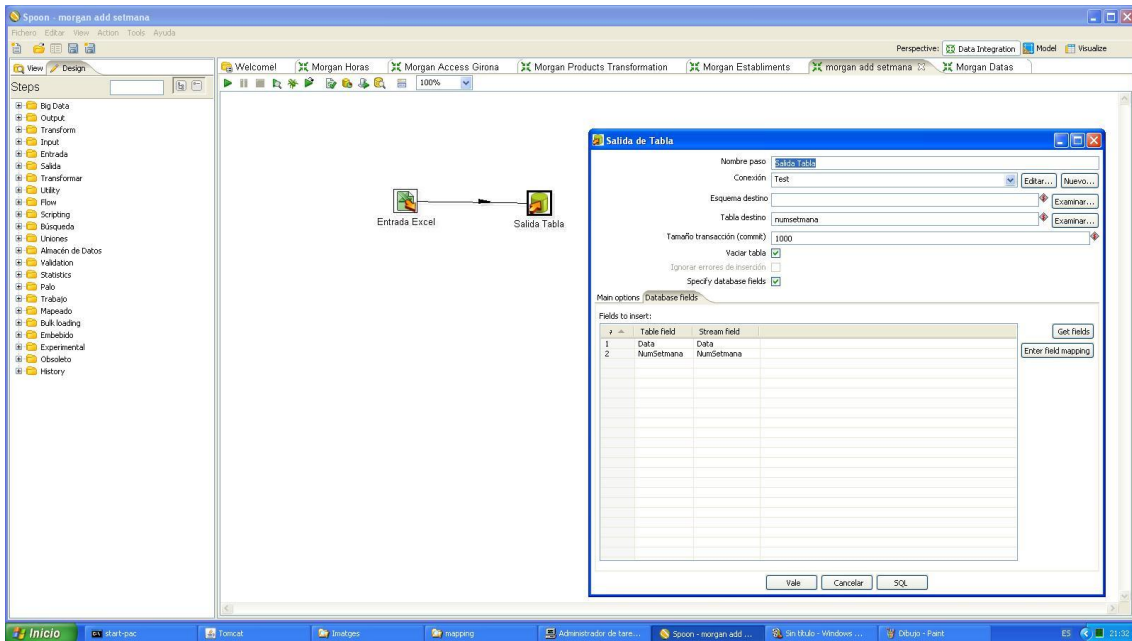
Aquí ha acabat la càrrega. Però el projecte és un problema de refinament. Sempre sorgeixen o bé problemes que no es tenien previstos o bé pensem en noves millores.



II-lustració 41 Refinament

I així ens adonem quan estem generant els informes, que ens faria falta en la taula de dimensions de les dates, el nombre de setmana, doncs en un informe se'ns demana. Aquesta dada poder es podria calcular per l'informe, però per evitar un pitjor rendiment, es decideix incorporar-la a la taula de dimensions.

Es fa recolzant-nos en un fitxer Excel, per aprofitar que aquest programa ja té una funció que ens indica el numero de setmana d'una data concreta. I un cop carregada, via MySQL afegim-lo a la nostre taula de dimensions.



II-Ilustració 42 Afegir Num.Setmana


Finalment, les nostres taules queden així:

Table Name:

Columns

Column Name	Datatype	PK	NN	UQ	BIN	UN	ZF	AI	Default
Id_venta	INT(11)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Id_linia	INT(11)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
IDESTABLIMENT	BIGINT(20)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
IDPRODUCTE	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
IDDATA	BIGINT(20)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
IDHORA	BIGINT(20)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Forma_pagament	VARCHAR(85)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Id_client	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Descompte	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Unitats	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Preu_venta	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Preu_cost	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Marge	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
PreuCostTotal	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
PreuVendaTotal	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
BaseImponibleUnitaria	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
BaseImponibleTotal	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
IVAUnitari	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
IVATotal	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
DescompteUnitat	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
DescompteTotal	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
DIACLIENT	VARCHAR(1)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL

II-Ilustració 43 BD Taula de Fets


 Table Name:


Collation:

Comments:

Columns

Column Name	Datatype	PK	NN	UQ	BIN	UN	ZF	AI	Default
IDENTIFICADOR	BIGINT(20)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
Hora	DATETIME	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Torn	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL


II-lustració 44 BD Hores


 Table Name:

Columns

Column Name	Datatype	PK	NN	UQ	BIN	UN	ZF	AI	Default
IDENTIFICADOR	BIGINT(20)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
Data	DATETIME	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
DiaSetmanaNum	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
DiaSetmana	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Estacio	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Mes	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
AnyData	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
MesNum	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Dia	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
MesAny	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Vacances	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
CapSetmana	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Equinocci Primavera	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Equinocci Tardor	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Solstici Estiu	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
SolsticiHivern	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
NumSetmana	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL

II-lustració 45 BD Data


 Table Name:

Columns

Column Name	Datatype	PK	NN	UQ	BIN	UN	ZF	AI	Default
Id_producte	BIGINT(20)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Nom	VARCHAR(72)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Preu_venda	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Tipologia	BIGINT(20)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
IVA	BIGINT(20)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Preu_cost	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Codi_barres	VARCHAR(15)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Venda_impuls	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Year	VARCHAR(4)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
ID	INT(11)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Proveidor	VARCHAR(30)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
productescol	VARCHAR(45)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL

II-lustració 46 BD Productes

Table Name: **establiments**

Collation: utf8 - default collation

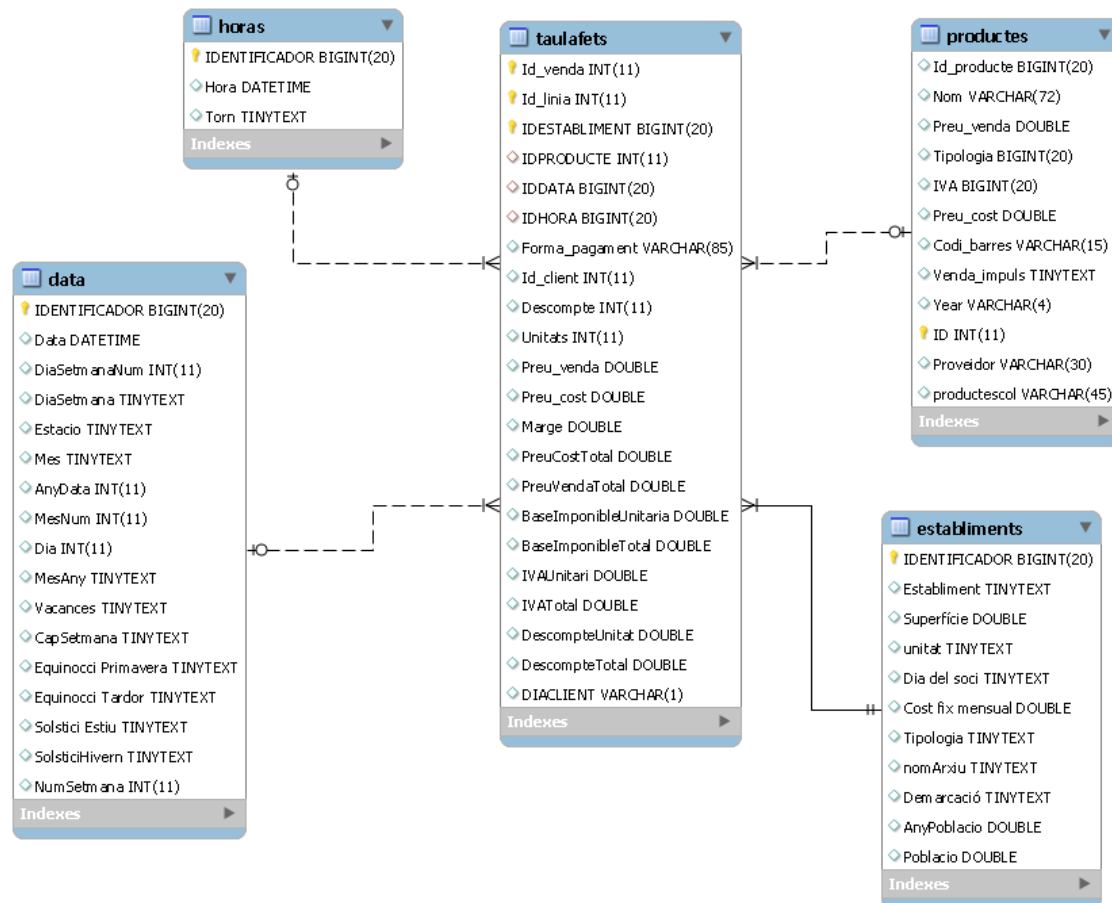
Comments:

Columns

Column Name	Datatype	PK	NN	UQ	BIN	UN	ZF	AI	Default
IDENTIFICADOR	BIGINT(20)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
Establiment	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Superfície	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
unitat	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Dia del soci	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Cost fix mensual	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Tipologia	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
nomArxiu	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Demarcació	TINYTEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
AnyPoblacio	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Poblacio	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL

II-lustració 47 BD Establiments

I el model:



II-lustració 48 Model

4.2.4 Treballs

Després de haver creat les nostres transformacions, el programa ens permet crear treballs, de manera que agrupem diferents transformacions, podem programar-les, i realitzar un control d'errors més complet.

Així també, serà molt més senzill fer el procés ETL, al tenir-lo automatitzat. O programar les actualitzacions.

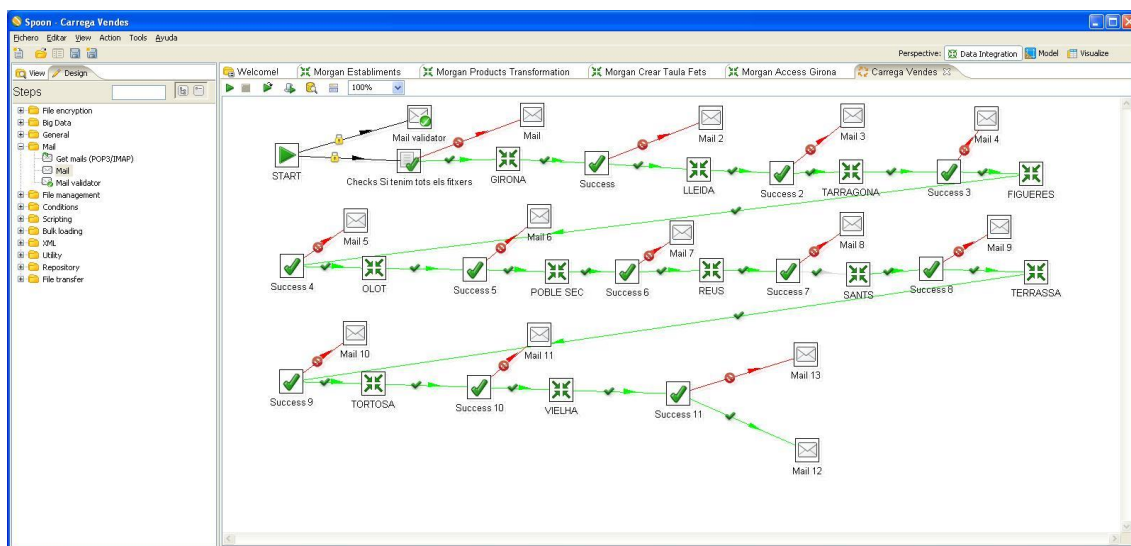
Veiem un exemple de la càrrega de la taula temporal de vendes. El primer element de la transformació ens permetrà programar-lo en cas que ho desitgem. Després controlem l'existència de tots els fitxers, de forma, que si falta algun, enviem un correu al responsable i avortem el procés.

Després anem realitzant les diferents transformacions de càrrega que hem vist abans i en cada una d'elles guardem un log, que s'enviarà amb el correu en cas de produir-se algun error.

Finalment si acaba tot el procés correctament també s'envia un mail de finalització al responsable.

Es podria també crear una treball que concatenés els diferents treballs.

Actualment falta parametritzar l'acció, per enviar correus amb el servidor SMTP que ens faciliti el client.



Il·lustració 49 Treball

4.2.5 Observacions

La càrrega en general ha anat força bé, en algun moment ha hagut algun rendiment baix en el tractament del fitxer de vendes unificat, però al ser processos que s'executen un únic cop, no ha hagut més problema.

En l'apartat de millores del document parlarem d'alguna millora que es pot introduir.

A nivell de dades tenir en compte alguns aspectes:

- En el calendari (data) s'ha carregat fins el 2013, per tal de tenir-ho preparat pel any vinent. Això fa que al buscar els anys possibles en alguns informes aparegui. No hi ha més problema, si el client considera que és incorrecte es pot esborrar aquesta informació i carregar-la conjuntament amb la resta de dades de l'any.
- S'han trobat dades anteriors a l'any 2010, com no es disposa d'inventari per aquests anys, no sabem a quin producte fa referència la venda, ni si falten més dades. Per tal de no tenir al nostre sistema informació incompleta s'han eliminat aquests registres.
- Les dades de població a les demarcacions corresponents al any 2012, s'han calculat com la població que ens facilita IDESCAT per l'any 2012, no com mitjana d'aquesta i la produïda a 1 de gener del 2013. Al no estar encara publicada aquesta última informació.
- A la base de dades de Vendes de Figueres no existeix cap correspondència entre capçalera i línies, la informació està corrupte, per lo que no es carrega al servidor. Es dona avís al client.
- S'ha realitzat un procés d'homologació de Proveïdors. Hem trobat que en la base de dades, al no haver mestre de dades de proveïdors, els usuaris han comés errades que fa que el mateix proveïdor aparegui escrit de forma diferent. S'ha triat com forma correcte aquella que tingui major nombre de registres. En algun cas, com hi ha la possibilitat que hagi dos proveïdors amb noms diferents, però ambdós possibles, s'ha deixat els dos noms. S'ha avisat al client per tal de confirmar l'homologació realitzada.

4.3 Model Multidimensional

Fent servir l'eina *Schema Workbench*, comencem a dissenyar el cub amb el que treballarem. Per poder usar l'eina, prèviament copiarem les llibreries que utilitzarà per connectar-se a la base de dades.

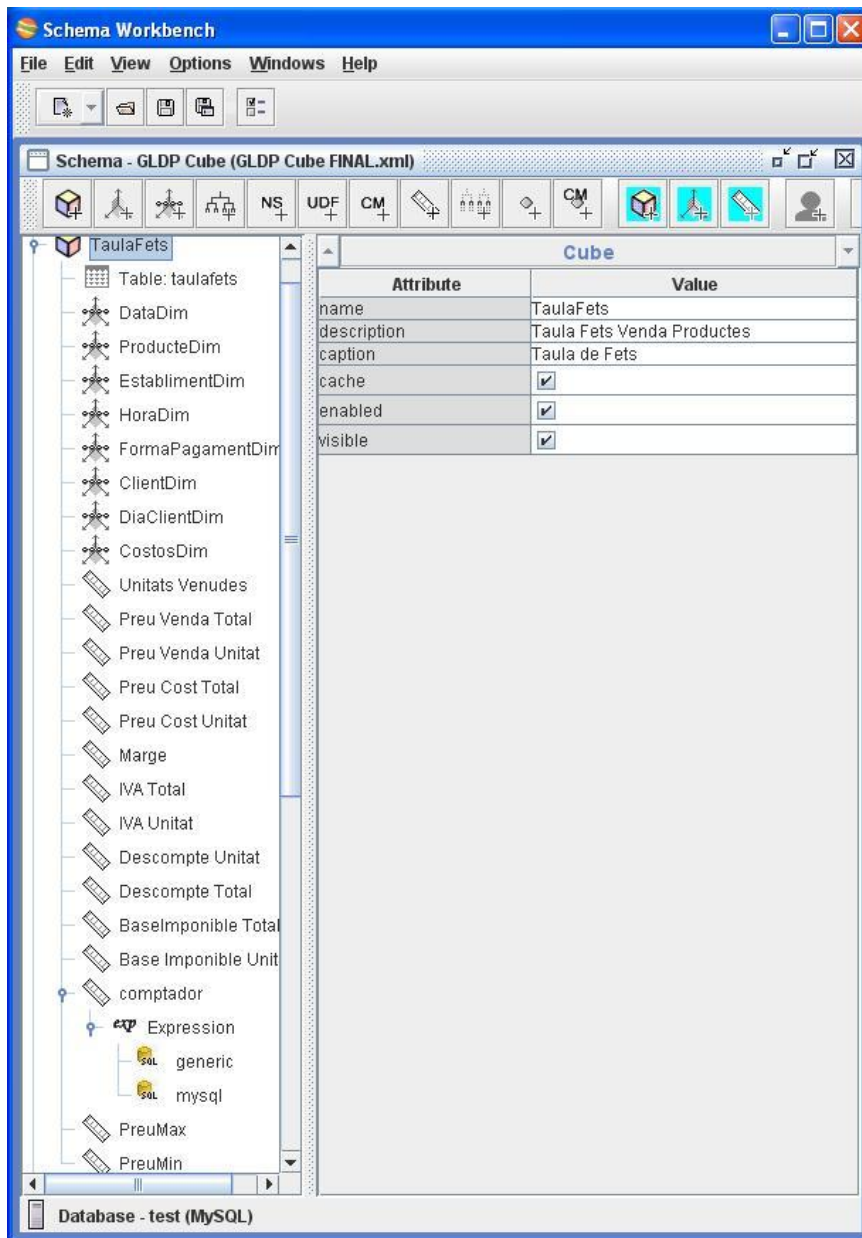
El cub és diu TaulaFets. En aquest esquema indiquem les diferents dimensions, i amb quina taula que es recolzen. A cada dimensió se li assigna una o més jerarquies, amb els seus diferents nivells.

Per últim s'assignen al cub.



Il·lustració 50 Cub - Dimensions

En el cub també es defineixen les diferents mesures que s'empraran i amb quin criteri s'agrupen, el més habitual és que es sumin.



II-lustració 51 Cub - Mesures

Cal notar que el nostre disseny té dimensions degenerades, aquelles que no tenen taula pròpia, i estan a la pròpia taula de fets. Aquestes dimensions no es poden modelar directament amb el Schema Workbench. Però si, es poden afegir modificant el fitxer XML i després continuar treballant amb l'eina sense problemes.

Les dimensions degenerades són l'identificador de client, la forma de pagament i el dia del client.

A més cal destacar que s'ha afegit una nova mesura calculada, que compta el nombre de registres de la taula de fets. Per no associar-ho a cap camp, deixem aquesta propietat en blanc, i afegim a la mesura dues expressions, una pel motor genèric i l'altre pel MySQL, el contingut és un simple asterisc, perquè compti totes les files.

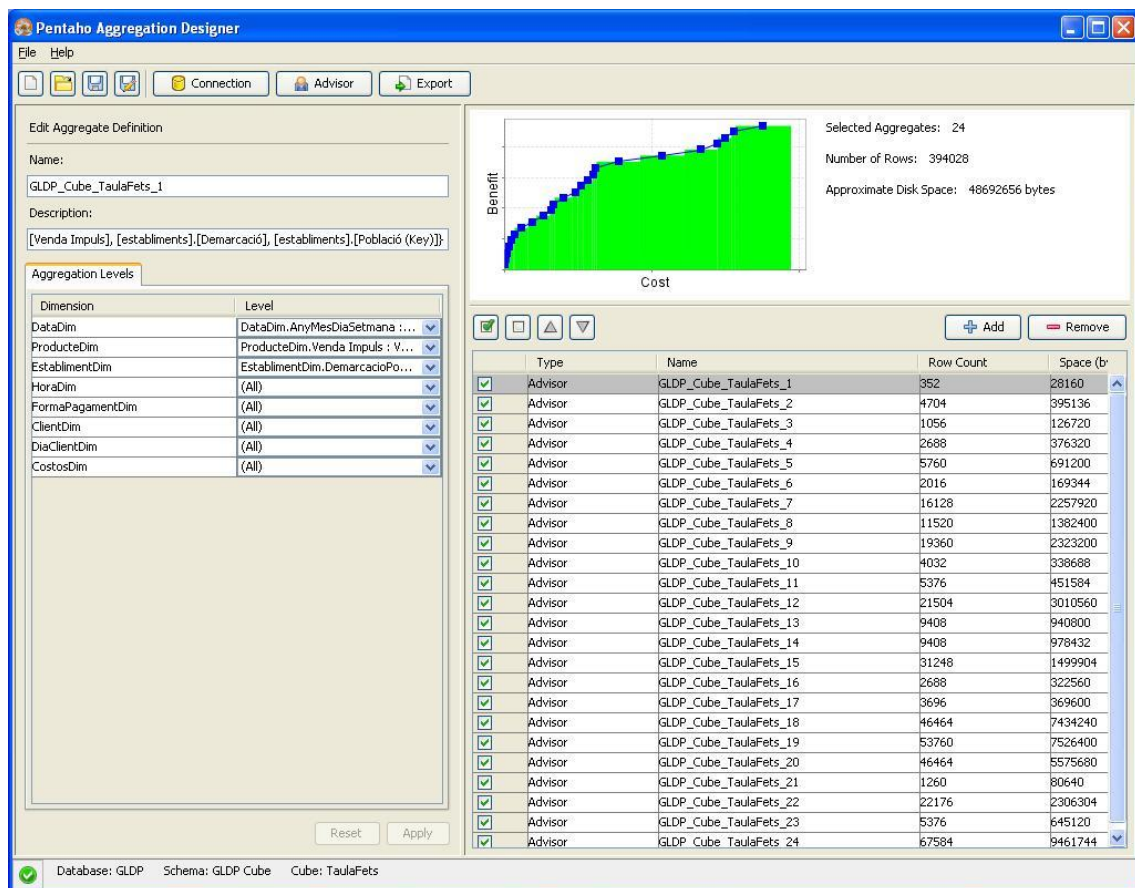
Un cop finalitzat el modelatge el pujarem al servidor per poder emprar-lo des del portal. Per poder realitzar aquesta tasca (i després de forma anàloga en la publicació dels

informes) cal editar un fitxer de configuració per informar la contrasenya que es farà servir per tal de poder publicar.

Un cop al servidor creem un carpeta GLDP on publicarem el nostre cub.

Un altre aspecte interessant en el nostre model multidimensional és l'ús de taules agregades. Per això ens baixem el *Pentaho Aggregation Designer* (PAD). Cal al utilitzar-ho per primer cop copiar, com hem fet abans amb el Schema Workbench, les llibreries per accedir a la base de dades. El seu ús és molt senzill, al executar-lo i triar el nostre cub li podem indicar quin nombre de taules agregades volem utilitzar, i deixem que analitzi el nostre sistema i ens faci les seves propostes. Òbviament podem modificar o incorporar les nostres.

Això ens generarà/executarà 3 punts, un script per crear les taules agregades, un altre per calcular-les i omplir-les. I finalment l'esquema del cub modificat que incorpora aquestes taules, per tal de publicar-ho.



Il·lustració 52 PAD - Agregacions

Si s'intenta fer automàticament, en el nostre cas falla. Però com podem obtenir el script a executar, ho executem sobre la base de dades mitjançant el MySQL Workbench. Allà ens adonem que un dels problemes és la grandària per defecte en la memòria intermèdia del MySQL, la qual ampliem de 50K a 250K.

4.4 Consultes

Fins aquí la major part de tasques han estat transparents per l'usuari, ara cal desenvolupar els diferents informes que ens ha demanat.

Primer cal crear un entorn de treball, en el nostre cas serà el portal. Hi ha dos accessos un és per la consola d'administració i l'altre pel portal pròpiament dit, on treballaran els usuaris.

Als dos es troben a la mateixa màquina, canvia el port amb el que s'accedeix. A la consola d'administració s'entra pel port 8099, i al portal pel 8080.

Per fer ús del portal, i les seves eines afegim un nou *Datasource*, és a dir li indiquem com connectar amb la nostre base de dades. L'afegim a la consola d'administració, i després l'afegim també en el fitxer de configuració *default.properties* en el directori *sample-jndi*.

Un cop fetes les configuracions bàsiques cal dissenyar els informes.

Bàsicament disposem d'entrada de 3 eines per realitzar els informes:

- Un *report wizard* que disposa el portal, per realitzar de forma gràfica informes molt senzills.
- També a través del portal la possibilitat de realitzar un anàlisi OLAP utilitzant el programa JPIVOT.
- Realitzar informes des de el programa Pentaho Report Designer, i després publicar-los pel seu ús des del portal.

Tots els informes s'han realitzat a través d'aquestes dues últimes eines, usant una connexió JNDI al nostre cub, i mitjançant l'ús de consultes MDX per extreure la informació que ens interessa.

Quan parlem de OLAP (*Online Analytical Processing*) ens referim al motor per analitzar els sistemes multidimensionals com el nostre. En el nostre cas fem servir el motor *Mondrian* que és tracta d'un ROLAP, és a dir un *Relational OLAP*, lo qual vol dir que es recolza en una base dades relacional.

Aquests motors són molt flexibles, i senzills, doncs es recolzen en bases de dades que poden ja existir, i en canvi ens permeten realitzar consultes MDX sobre elles, i gaudir de tots els avantatges del model multidimensional.

El peatge a pagar és el rendiment, que sol ser més baix que en el cas del MOLAP (*Multidimensional OLAP*).

Tornant als informes, molts cops durant el desenvolupament dels informes ens adonem que ens fa falta una nova mesura (com pot ser la mesura comptador que hem explicat en el punt anterior), o una nova jerarquia o nivell en una dimensió.

Examinem ara cada informe lliurat al client.

4.4.1 Informes i Anàlisi

Començarem examinant els informes estàtics. Tots ells admeten paràmetres per tal de seleccionar l'any o l'establiment que es vol consultar.

Aquesta decisió es deu al rendiment de la màquina virtual, per tal de fer ús de les facilitats de *slicing* del model multidimensional i obtenir un rendiment millor. El *slicing* consisteix en seccionar part del cub, per tal de tan sols treballar amb un tros. Amb el model relacional té molta semblança amb l'ús de les clàusules *where*, tan en sintaxi com en el resultat.

Sobre aquest tema cal destacar que Mondrian, el motor ROLAP que fem servir, fa ús d'una cache amb lo que el primer cop que executem la consulta és quan triga més.

En cada informe s'hi ha afegit l'emblema de l'empresa en la capçalera.

❖ Rànquing d'establiments

Veiem aquest informe ja penjat del portal. De fet s'han realitzat dos informes, un que mostra el rànquing d'establiments per unitats venudes i l'altre per volum de vendes (euros).

GLDP Ranking Venues per Establiment

Establiment	Vendes
GLDP TARRAGONA	348.809.721,84
GLDP TORTOSA	433.971.755,69
GLDP TERRASSA	412.362.599,97
GLDP REUS	408.007.043,87
GLDP VIELHA	402.330.204,34
GLDP LLEIDA	306.872.056,31
GLDP GIRONA	310.160.196,82
TOTAL	2.891.393.641,44

Il·lustració 53 Informe - Rànquing d'establiments I

GLDP Ranking Units Venues per Establiment

Establiment	Unitats Venudes
GLDP TARRAGONA	9.412.896
GLDP TORTOSA	7.440.911
GLDP TERRASSA	7.194.176
GLDP REUS	6.989.454
GLDP VIELHA	6.918.236
GLDP LLEIDA	6.679.296
GLDP GIRONA	5.372.681
TOTAL	50.013.650

Il·lustració 54 Informe - Rànquing d'establiments II

❖ Top Ten de Productes

En aquest cas també tenim el rànquing tan per unitats com per vendes. En ambdós categories el producte més destacat són les “Galletas Maria Dorada” amb el *packaging* de 4 unitats de 200gr.

Adjuntem una captura dins el *Pentaho Report Designer* per veure el seu aspecte. I on a la dreta es pot veure una finestra, amb les consultes que s’han creat i amb els elements que apareixen en l’informe.

Notar que hi ha dos consultes. L’explicació és que s’ha creat una consulta per tal que el desplegable que deixa triar l’any estigui informat amb tots els valors permesos. A tal efecte, s’ha creat una consulta que ens retorna els elements de la dimensió Data, a nivell d’any.

Producto	Unitats Vendues
Galletas Maria Dorada, 4x200 G.	49.606.429,000
Estroppo Plastico Lacteos, 2 Und.	36.309,00
Embalada De Atun Milanesa 150 G. P.N. F...	21.218,00
Estroppo Fibra Blanca Con Esponja	16.511,00
Esponja Limpiaclizado Caracas	12.198,00
Insecticida Concentrado Hojar Y Plantas 400.	11.763,00
Escobillero We	9.111,00
Kit Crema Calzado Antibrilante+Esponja ..	7.994,00
Esponja Ovalada N°4	7.871,00
Kit Crema Calzado Antibrilante+Esponja ..	7.440,00

II-lustració 55 Informe - Top Ten de productos

Structure Data

- Data Sets
 - Pentaho Analysis: GLDP: cub E: IUOCT: bend
 - Query 2
 - Query 1
 - [ProductDim.NomProducto] [ANSI]
 - [ProductDim.NomProducto] [BOMF]
 - [Measures] [Unitats Vendues] [OBJE]
- Environment
 - env: hostNamePort (String)
 - env: pentahoBaseURL (String)
 - env: requestURL (String)
 - env: roles (String)
 - env: roles-array (String[])
 - env: serverBaseURL (String)
 - env: solutionRoot (String)

II-lustració 56 Informe - Top Ten de productos II

Producto	Vendes
Galletas Maria Dorada, 4x200 G.	2.884.482.933,51
Insecticida Concentrado Hogar Y Planta 400.	1.488.679,72
Horchata Pet 1 L.	551.567,60
Harina Especial Fibra De Pescao 1 K.	248.357,16
Estopajo Plástico Licitos, 2 Unid.	203.245,47
Paleta Cocina Placa 5 Kg. Aprox.	181.615,80
Leche Desmullidora Sensation, 250 ML.	159.040,00
Crema Mano Nutritiva N°4 Sensation 150 ML.	143.857,10
Hierba Provenidas, 12 G.	129.415,65
Guandillo Molida, 32 G.	109.891,30

II-il·lustració 57 Informe - Top Ten de productes III

❖ Import mitjà de compra per soci i establiment

Per aquest informe també tenim l'establiment com paràmetre, cosa que té sentit, més enllà dels temes de rendiment, si recordem que els clients són únics per cada establiment.

Al igual que en el cas de l'any, s'ha creat una consulta per informar el desplegable amb tots els establiments existents.

ID CLIENT	Compra Mitja
0	21.124,93
3	18,26
4	5.451,34
8	35.351,30
35	59.625,45
43	3,20
52	73.627,10
123	1.307.580,00
125	323.950,00
126	99.740,18

II-il·lustració 58 Informe - Import mitjà de compra per soci i establiment

❖ Total vendes i marge net del grup

Mostrem per cada establiment el seu volum de vendes, i el seu cost fix, així com el marge resultant. Després mostrem el gran total del grup.

The screenshot shows the Pentaho BI console interface. The report title is 'GLDP Total Vendes i marge net del Grup'. The report displays a table with the following data:

Establiment	Vendes	Cost Fix	Marge Net
GLDP GIRONA	310.160.196,82	142000	310.015.196,82
GLDP LLEIDA	386.872.039,31	122000	386.749.039,31
GLDP REUS	404.087.043,87	104000	403.981.043,87
GLDP TARRAGONA	543.609.721,44	136000	543.473.721,44
GLDP TERREASSA	412.382.599,97	114800	412.247.799,97
GLDP TORTOSA	431.971.755,69	92400	431.879.355,69
GLDP VIELHA	402.330.284,34	68000	402.262.284,34
TOTAL	2.891.293.641,44	785.200	2.890.608.441,44

II-il·lustració 59 Informe - Total vendes i marge net del grup

❖ Percentatge de vendes de marques blanques per habitants

Entenem per marques blanques aquelles que el seu proveïdor és la pròpia cadena, GLDP. En aquest cas creem un informe a nivell de demarcació, on mostrem les unitats venudes de marques blanques, i el nombre d'habitants d'aquella demarcació. Després es calcula el rati.

The screenshot shows the Pentaho BI console interface. The report title is 'GLDP Marques Blanques'. The report displays a table with the following data:

Demarcació	Habitants	Unitats Venudes	Vendes GLDP per habitant
Balears	5520123	5.523,00	0,1%
Girona	754928	3.232,00	0,4%
Lleida	441038	6.676,00	1,5%
Tarragona	809911	12.921,00	1,6%

II-il·lustració 60 Informe - Marques Blanques

❖ Categorització de Clients A/B/C

Una forma molt habitual de classificar els clients és emprar la categorització ABC. Consisteix en agrupar els clients pel volum de vendes que generen. Aquest volum pot

variar, per exemple podríem agrupar-los amb un criteri de 70/15/15, de forma que els clients A serien aquells que suposen el 70% del volum, els B el 15% i els C la resta.

Així el mètode per fer aquesta classificació, és ordenar els clients per volum de vendes i anar agafant-los fins que suposen el 70%, tots ells són de categoria A, després es mira el 15%, i per últim el 15% restant.

D'aquesta manera l'empresa té molt clar quins son els seus gran clients, i quins ha de potenciar més.

Aquesta classificació es basa en el principi de Pareto, que ens diu, aplicat al nostre exemple, que el 20% dels clients són els causants del 80% de les nostres vendes.

Per aquest informe l'usuari pot triar l'any i establiment que vol consultar, i els percentatges amb els que es realitzarà la classificació. S'adjunten 3 captures jugant amb aquests valors.

Durant la elaboració dels diferents informes s'han fet servir forces planes de referència del llenguatge MDX i sobretots els fòrums de les comunitats *Pentaho*, i *stackoverflow*. Per realitzar aquest informe s'ha mirat especialment una entrada al de Microsoft Analysis Server.

Any a visualitzar: 2010
Establiment: GLDP GIRONA
Percentage Clients A: 85
Percentage Clients B: 10

ID CLIENT	Vendes	Vendes Acumulades	% Acumulat	Categoria
341	2.073.340,00	2.073.340,00	31,48	A
342	2.073.340,00	4.146.680,00	62,96	A
79	824.638,50	4.971.318,50	75,48	A
269	505.742,22	5.477.060,72	83,16	A
211	259.160,00	5.736.220,72	87,09	B

II-Il·lustració 61 Informe - Categorització Clients ABC I

ID CLIENT	Vendes	Vendes Acumulades	% Acumulat	Categoria
341	2.073.340,00	2.073.340,00	31,48	A
342	2.073.340,00	4.146.680,00	62,96	A
79	824.638,50	4.971.318,50	75,48	A
269	505.742,22	5.477.060,72	83,16	B
211	259.160,00	5.736.220,72	87,09	B

II-lustració 62 Informe - Categorització Clients ABC II

ID CLIENT	Vendes	Vendes Acumulades	% Acumulat	Categoria
341	2.073.340,00	2.073.340,00	31,48	A
342	2.073.340,00	4.146.680,00	62,96	A
79	824.638,50	4.971.318,50	75,48	B
269	505.742,22	5.477.060,72	83,16	B
211	259.160,00	5.736.220,72	87,09	C
336	206.130,00	5.942.350,72	90,22	C
346	192.146,60	6.134.517,32	93,14	C
50	88.857,96	6.223.375,28	94,49	C
319	70.707,75	6.294.083,03	95,56	C
337	64.839,35	6.358.942,38	96,55	C
325	59.043,25	6.417.985,63	97,45	C

II-lustració 63 Informe - Categorització Clients ABC III

Ara veurem els informes OLAP que s'han realitzat. Aquests difereixen dels que hem vist fins ara, en que són dinàmics. És a dir, mentre els que acabem de veure són uns informes estàtics, amb els informes OLAP mitjançant les tècniques de *slicing*, *drill-down* i *drill-up*, podem anar canviant de nivell, per veure la informació amb més o menys detall, depenent de les jerarquies que haguem definit prèviament.

Del *slicing* ja hem parlat en aquest mateix apartat, el cas del *drill down* és anar de la informació més resumida a més detallada (passar de anys a mesos per exemple), i *drill up* és anar en sentit contrari.

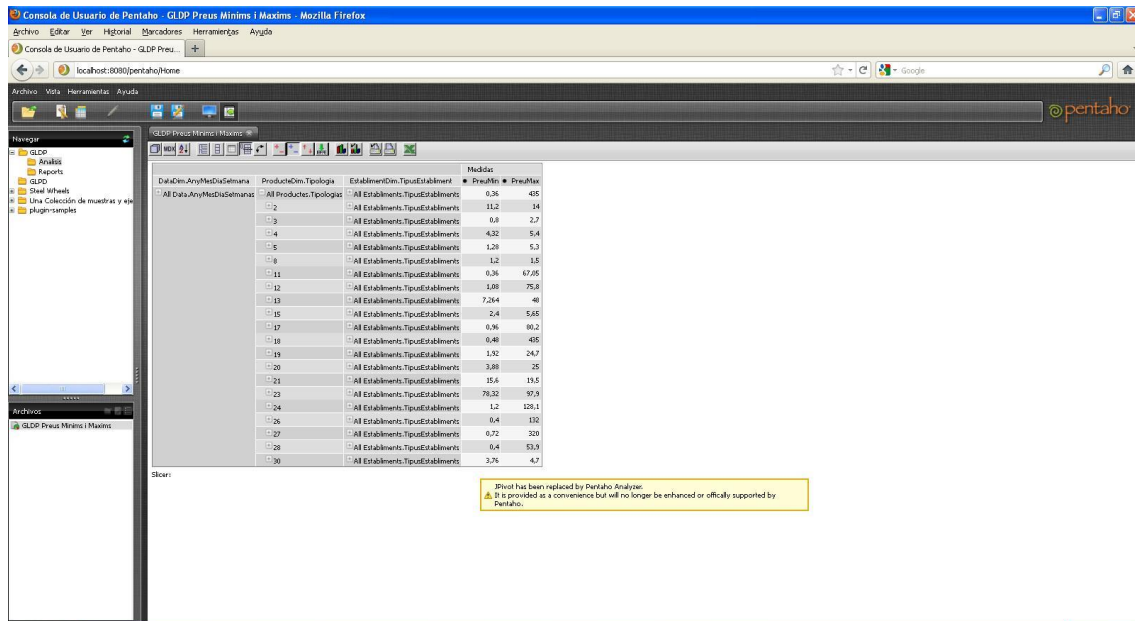
A més amb el JPivot se'ns permet moltes més funcionalitats, com afegir una nova mesura, o una nova dimensió al nostre anàlisi.

- ❖ Preus màxims i mínims per tipologia d'establiment i de producte

En aquest cas veiem un llistat amb les diferents tipologies de producte i d'establiment. En les tres dimensions que apareixen, la temporal, representada per la dimensió DataDim i la jerarquia AnyMesDiaSetmana, EstablimentDim amb la jerarquia TipusEstabliment, i la de ProducteDim amb Tipologia, en el seu costat esquerre apareix el símbol + que ens permetrà navegar per la jerarquia, en aquest cas, com estem en el element més resumit, realitzar un *drill-down*.

Mitjançant la botonera del JPIVOT es pot arribar a la mateixa funcionalitat.

Per exemple en el cas dels productes, si premem el símbol + ens mostraria els productes que són de cada tipologia, amb la seva informació producte a producte.

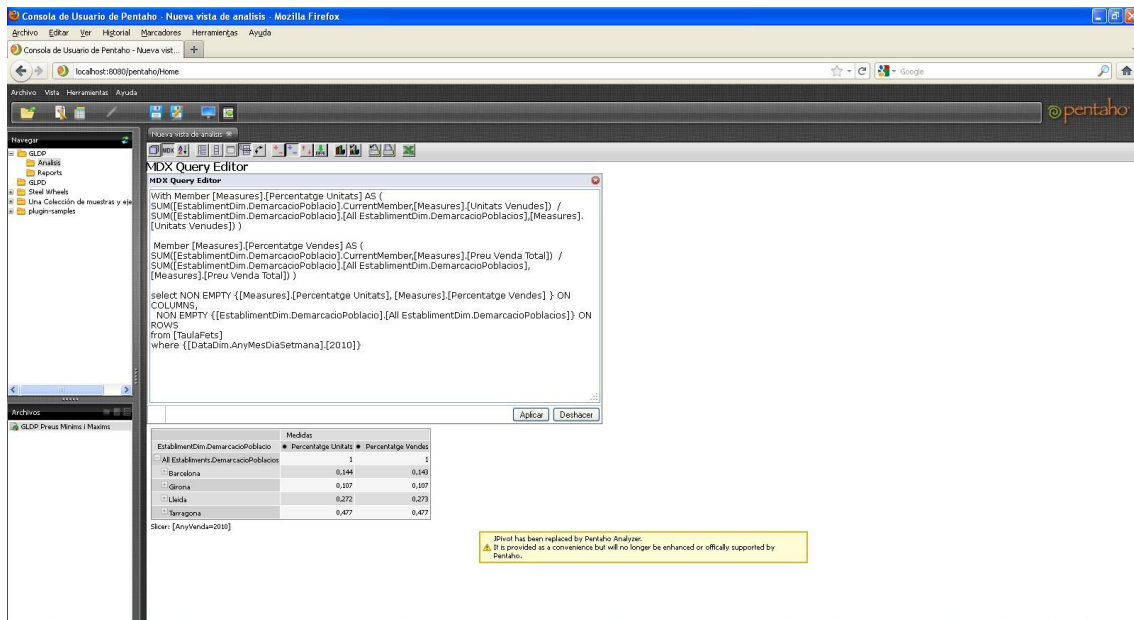


II-Il·lustració 64 Anàlisi - Preus Màxims i Mínims

- ❖ Percentatge de vendes respecte el total del grup (per establiment i per demarcació)

Per aquest informe mostrem la demarcació i el seu percentatge tan les unitats venudes, com pel volum de vendes.

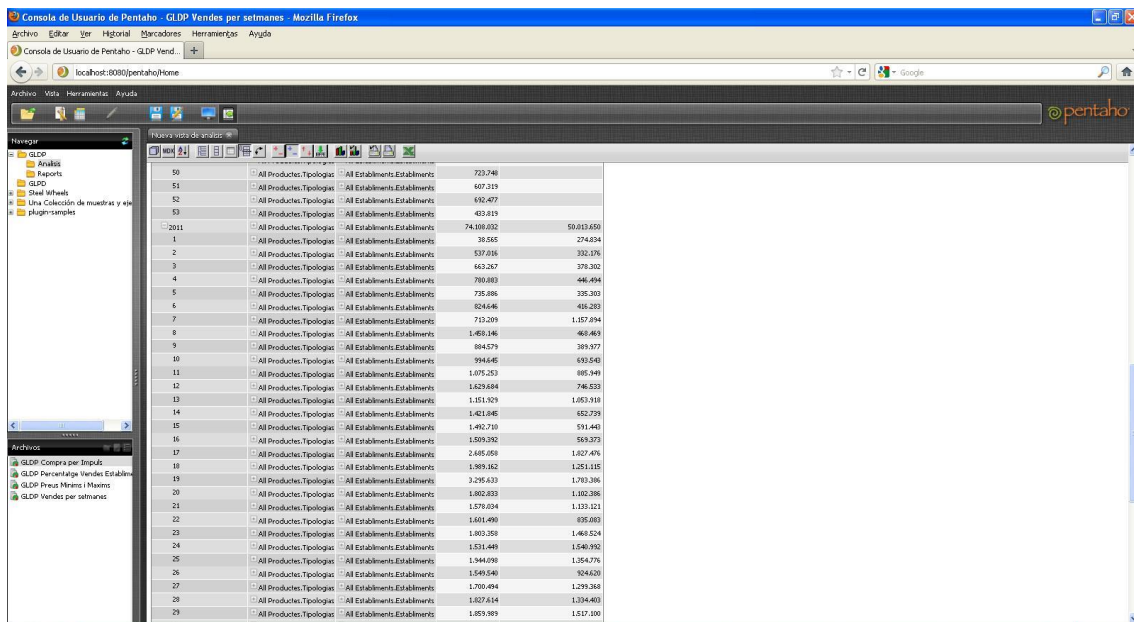
Aquí veiem una altra funcionalitat del JPIVOT que ens permet veure el codi MDX i editar-lo. En ell podem veure com es fa ús dels membres calculats per tal d'aconseguir mostrar la informació. És important en aquests emprar la propietat *CurrentMember* per tal de garantir la navegabilitat.



II-Il·lustració 65 Anàlisi - Percentatge VenDES

❖ Distribució setmanal i estacionalitat de vendes

Aquest informe ha provocat que incorporéssim a la dimensió data, el número de setmana. Veiem a més la comparació amb l'any anterior, és per això que en el cas de les dades del 2010, la segona columna apareix en blanc, doncs no hi ha dades del any anterior.

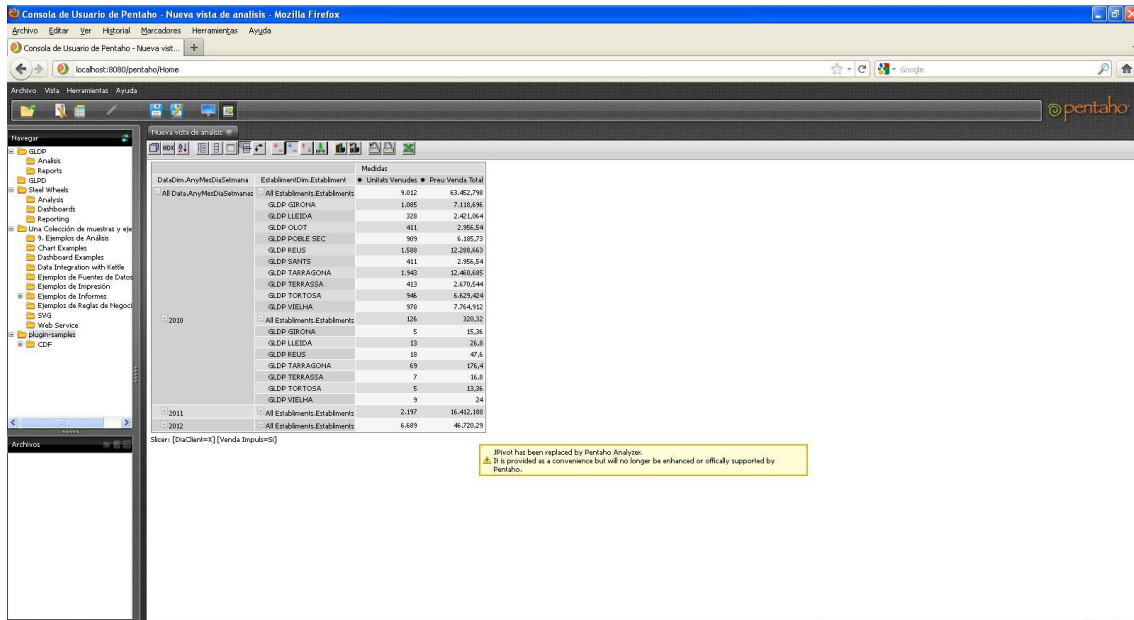


II-Il·lustració 66 Anàlisi - Distribució VenDES

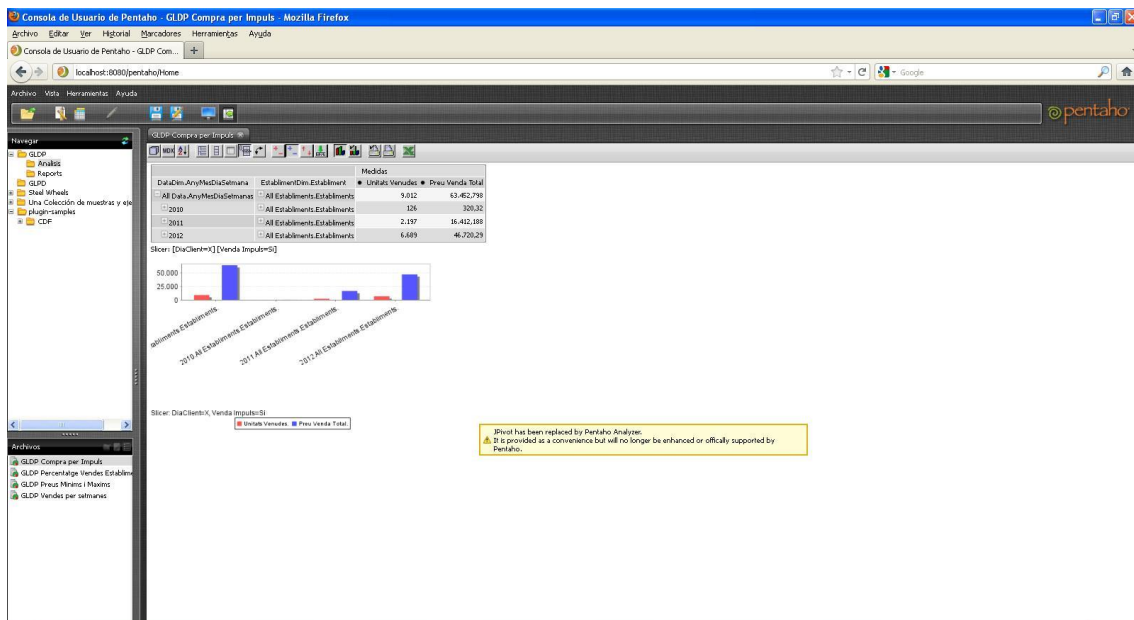
❖ Anàlisi de compra per impuls

Aquí mostrem les dades en unitats i en import dels productes que es col·loquen al costat de la caixa registradora, en promoció el dia del client.

Per veure-ho més clar, fem un *drill-up*, i llavors provem a utilitzar el gràfic.



II-lustració 67 Anàlisi - Compra per impuls I



II-lustració 68 Anàlisi - Compra per impuls II

5 Treball Futur

- ETL: Preparació de transformacions i treballs per l'actualització de dades.

Preparar un treball i modificar les transformacions que sigui necessari per tal d'automatitzar les actualitzacions de dades. Està clar que és possible carregar tot des de zero, però no és gens pràctic pel temps invertit quan hi ha tan sols un nou mes, per exemple.

- Informes: Ús de *Dashboards*.

Els *dashboards* són panells de control, on l'usuari té tota la informació més important, o enllaços per arribar a ella.

- Informes: Incorporar Gràfics als informes

De fet en algun informe ja s'ha fet. Com es veu en el cas dels informes dinàmics, es tractaria tan sols de prémer l'opció corresponent i ajustar els paràmetres desitjats. Aquesta tasca la pot realitzar el propi usuari.

En canvi en els informes estàtics realitzats amb el programa PRD (Pentaho Report Designer), el gràfic s'ha de incloure en el propi disseny de l'informe i no és possible fer-ho des de el portal.

- Informes: Millorar el seu disseny.

Seguint les recomanacions del client, fer més atractius els informes.

- Informes: Major ús de l'eina de Anàlisi.

Incorporar més informes dinàmics. Aquests informes permeten que l'usuari els manipuli per tant d'extreure més informació, descobrir noves. Es tracta en definitiva de dotar al sistema de més intel·ligència.

- General: Implementació rols i seguretat específiques.

Ser més estricte en la política de seguretat i els rols assignats. Crear clarament un rol d'administrador i limitar als usuaris per tal que no puguin veure ni manipular el que no els hi pertoca. És en aquest punt on definiríem també el rol d'usuari avançat.

6 Conclusions

Hem vist que avui en dia per empreses de certa grandària els magatzems de dades són imprescindibles. Els hi permeten obtenir nou coneixement a partir de les dades que inicialment tenen repartides entre diferents sistemes i diferents ubicacions.

I després centrat en aquest treball final de carrera, l'etapa d'estudi i disseny sembla la més important pel correcte desenvolupament del projecte.

En aquest cas, cal afegir el fet d'aprendre un nou paquet, en totes les fases del procés, des de la carrega amb el PDI, la creació del cub amb el Schema Workbench, i el PAD, com sobretot la generació d'informes amb el PRD i el JPIVOT, amb el llenguatge MDX.

Per tant, els objectius inicials que es plantejaven, a saber:

- 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.
- Perfeccionar la gestió i seguiment de projectes.
- Entregar un producte final on sigui mesurable aquest aprenentatge, i que compleixi els requeriments i funcionalitats demanats pel client.

Han estat assolits en tot el seu conjunt.

Queda un llarg camí per endavant per tal d'ampliar aquests coneixements, amb la relació de l'intel·ligència de negoci amb la mineria de dades, i amb la Big Data.

7 Llista de referències

Kimball , Ralph; Ross, Margy. *The Data Warehouse Toolkit (2nd Edition)*. John Wiley and Sons, 2002, Inc. ISBN 0-471-20024-7, pp 58-62.

Fòrum Microsoft SQL Server Analysis Services, 2008.

[<http://social.msdn.microsoft.com/forums/en-US/sqlanalysisservices/thread/141efac3-1129-46f3-be9f-9aabaadd2181/>, 28 de desembre del 2012]

8 Annexos

8.1 Annex A1 - Bibliografia

Kimball , Ralph; Ross, Margy (2002). *The Data Warehouse Toolkit (2nd Edition)*. John Wiley and Sons, Inc. ISBN 0-471-20024-7

Caserta, Joe; Kimball , Ralph (2004). *The Data Warehouse ETL Toolkit*. Wiley Publishing, Inc. ISBN 0-7645-7923 -1

Inmon, W.H. (2002). *Building the Data Warehouse (3rd Edition)*. John Wiley and Sons, Inc. ISBN 0-471-08130-2

Carina Roldán, María (2010). *Pentaho 3.2 Data Integration Beginner's Guide*. Packt Publishing ISBN 978-1-847199-54-6

Bouman, Roland; van Dongen, Jos (2009). *Pentaho Solutions Business Intelligence and Data Warehousing with Pentaho and MySQL*. Wiley Publishing, Inc. ISBN: 978-0-470-48432-6

Whitehorn, Mark; Zare, Robert; Pasumansky, Mosh (2002). *Fast Track to MDX*. Springer. ISBN: 1-85233-681-1

8.1.1 Enllaços d'interès a Internet

- <http://www.inmoncif.com>
- <http://www.rkimball.com>
- <http://www.dwreview.com>
- <http://www.1keydata.com/datawarehousing/datawarehouse.html>
- <http://www.intranetjournal.com/features/datawarehousing.html>
- <http://community.pentaho.com>
- <http://dev.mysql.com>
- http://es.wikipedia.org/wiki/Principio_de_Pareto
- <http://stackoverflow.com/questions/515114/the-best-mdx-references-on-the-net>
- <http://social.msdn.microsoft.com/forums/en-US/sqlanalysiservices/thread/141efac3-1129-46f3-be9f-9aabaadd2181/>