



# Anàlisi de vendes d'una companyia

**Joaquín Carlos Rodríguez De Torres Jaramillo**  
Màster d'Enginyeria Informàtica

**Víctor Ruiz Marques**  
Consultor

3 de Juliol de 2.015



Aquesta obra està subjecta a una llicència de [Reconeixement-NoComercial-SenseObraDerivada 3.0 Espanya de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FITXA DEL TREBALL FINAL

<b>Títol del treball:</b>	<i>Anàlisi de vendes d'una companyia</i>
<b>Nom de l'autor:</b>	<i>Joaquín Carlos Rodríguez de Torres Jaramillo</i>
<b>Nom del consultor:</b>	Víctor Ruiz Marques
<b>Data de lliurament (mm/aaaa):</b>	<i>07/2015</i>
<b>Àrea del Treball Final:</b>	<i>Business Intelligence</i>
<b>Titulació:</b>	<i>Màster d'Enginyeria Informàtica</i>
<b>Resum del Treball:</b>	
<p>Es presenta un anàlisi de dades de vendes d'una companyia. Mitjançant una metodologia de disseny i implementació de magatzems de dades (Datawarehousing) es descriu cada pas, des de l'anàlisi de les fonts de dades inicials en fitxers de text fins a l'obtenció d'informes que tracten de respondre a preguntes com si ha hagut un decrement de vendes als darrers anys o quina delegació és la ha venut més articles, entre d'altres, tot utilitzant programari lliure.</p>	
<b>Abstract:</b>	
<p>The following is a job about analytics of sales data of an organization. By means of a methodology of design and implementation of datawarehouses, it's described all steps on making one, from structure analysis of the initial data in text files to reports that try to answer usually questions related to organization's business as if there is a reduction of selling in past years or what's the division that sells more items inside the company, among others. Open source software was used for this purpose.</p>	
<b>Paraules clau:</b>	
datawarehouse, datamart, business intelligence, BI, open source, reporting, analytics	

## Índex

1. Introducció.....	1
1.1 Context i justificació del Treball.....	1
1.2 Objectius del Treball.....	1
1.3 Enfocament i mètode seguit.....	3
1.4 Planificació del Treball.....	3
1.5 Breu resumari de productes obtinguts.....	6
1.6 Breu descripció dels altres capítols de la memòria.....	6
2. Plataformes de BI.....	7
2.1 Comparativa Pentaho, SpagoBI, Eclipse BIRT.....	7
2.2 Elecció de la plataforma.....	8
3. Anàlisi de les dades inicials.....	9
4. Disseny del Data Mart.....	12
4.1 Disseny conceptual.....	12
4.2 Disseny lògic.....	14
5. Implementació del Data Mart.....	16
5.1 Elecció del SGDB.....	16
5.2 Disseny físic.....	16
6. Disseny i implementació dels processos ETL.....	18
6.1 Staging Area.....	18
6.2 Extracció, Transformació, Càrrega.....	21
7. Explotació del Data Mart.....	25
7.1 Informe-Evolució de les vendes.....	29
7.2 Informe-Productes i famílies de productes més venuts.....	34
7.3 Informe-Clients que els hi compren més (top 10).....	39
7.4 Informe-Distribució de vendes segons la zona del client.....	42
7.5 Informe-Evolució de les vendes en el temps.....	46
7.6 Informe-Comissions als comercials.....	52
7.7 Informe-Marge anual.....	55
8. Conclusions.....	58
9. Glossari.....	59
10. Bibliografia.....	60
11. Annexos.....	61
11.1 Notes d'instal·lació i configuració de l'entorn usat per al TFM.....	61

## Llista de taules

Taula 1: Relació entre fases del PMBOK i les PAC lliurades.....	3
Taula 2: Breu comparativa de tres plataformes de BI.....	7
Taula 3: Fitxers de text amb les dades originals.....	9
Taula 4: Anàlisi de les dades originals del fitxer Almacenes.csv.....	10
Taula 5: Anàlisi de les dades originals del fitxer Articulos.csv.....	11
Taula 6: Anàlisi de les dades originals del fitxer Clientes.csv.....	11

## Llista de consultes

Consulta 1: Importació de dades originals al Staging Area CSV.....	20
Consulta 2: Càrrega de la taula de D_MAGATZEMS.....	21
Consulta 3: Càrrega de valors no definits.....	21
Consulta 4: Càrrega de valor de client desconegut.....	22
Consulta 5: Càrrega de la taula D_TEMPS per a l'any 2009.....	22
Consulta 6: Càrrega de la taula de fets F_VENDES.....	24
Consulta 7: Informe a, evolució de les vendes per anys i delegacions.....	29
Consulta 8: Informe b, productes més venuts.....	34
Consulta 9: Informe b, famílies de productes més venudes.....	34
Consulta 10: Informe c, clients top.....	39
Consulta 11: Informe d, distribució de vendes segons zona del client.....	42
Consulta 12: Informe e, evolució temporal de les vendes.....	46
Consulta 13: Informe f, comissions.....	52
Consulta 14: Informe g, marges anuals.....	55

## Llista de figures

Il·lustració 1: Diagrama de Gantt per a la planificació del Treball.....	4
Il·lustració 2: Fitxer de text Almacenes.csv.....	10
Il·lustració 3: Diagrama d'entitats del model conceptual del Data Mart.....	13
Il·lustració 4: Diagrama del disseny lògic del Data Mart.....	15
Il·lustració 5: Diagrama del model físic del Data Mart.....	17
Il·lustració 6: Processos ETL i esquemes usats.....	18
Il·lustració 7: Staging area CSV.....	19
Il·lustració 8: Aspecte d'un nou informe en blanc obert a BIRT.....	25
Il·lustració 9: Seleccionant el tipus de font de dades.....	26
Il·lustració 10: Configuració de la connexió a la base de dades.....	26
Il·lustració 11: Diàleg selecció de font de dades.....	27
Il·lustració 12: Definit la consulta a la base de dades.....	28
Il·lustració 13: Previsualitzant les dades del Data Mart.....	28
Il·lustració 14: Estructura informe a.....	30
Il·lustració 15: Definint un Data Cube (1).....	31
Il·lustració 16: Definint un Data Cube (2).....	31
Il·lustració 17: Data Explorer mostrant el cub de dades creat.....	32
Il·lustració 18: Format d'eixida a l'informe.....	32
Il·lustració 19: informe a: evolució de les vendes per any i delegació.....	33
Il·lustració 20: Disseny de l'informe de productes més venuts.....	35
Il·lustració 21: Visualització d'informe de productes més venuts.....	35
Il·lustració 22: Paràmetres als informes.....	36
Il·lustració 23: Definició del paràmetre Empresa.....	36
Il·lustració 24: Definició del paràmetre Any.....	36
Il·lustració 25: Disseny de l'informe dels articles més venuts.....	37
Il·lustració 26: Diàleg de selecció de paràmetres.....	37

Il·lustració 27: Informe b: articles i famílies més venuts.....	38
Il·lustració 28: Disseny de l'informe dels clients que més compren.....	39
Il·lustració 29: Definint una agregació.....	40
Il·lustració 30: Definint una agregació (2).....	40
Il·lustració 31: Definint el filtre.....	41
Il·lustració 32: Definint el filtre (2).....	41
Il·lustració 33: Informe c: top 10 clients més fidels.....	41
Il·lustració 34: Disseny de l'informe de les zones amb més i menys vendes.....	43
Il·lustració 35: Informe d1: zones amb més vendes.....	44
Il·lustració 36: Informe d2: zones amb menys vendes.....	45
Il·lustració 37: Disseny de l'informe de l'evolució temporal de les vendes.....	47
Il·lustració 38: Disseny de l'informe de l'evolució temporal de les vendes (2).....	48
Il·lustració 39: Cub per a l'informe e.....	48
Il·lustració 40: Informe e1: evolució de les vendes per anys.....	49
Il·lustració 41: Informe e2: evolució de les vendes per trimestres (any 2013).....	50
Il·lustració 42: Informe e3: evolució de les vendes per mesos (any 2013).....	51
Il·lustració 43: Estructura de l'informe de les comissions.....	52
Il·lustració 44: Disseny de l'informe de les comissions (2).....	53
Il·lustració 45: Cub de dades per a l'informe de les comissions.....	53
Il·lustració 46: Informe f1: comissions percebudes per anys.....	54
Il·lustració 47: Informe f2: comissions percebudes per mesos.....	54
Il·lustració 48: Disseny de l'informe del marges anuals.....	56
Il·lustració 49: Informe dels marges anuals.....	56

# 1. Introducció

## 1.1 Context i justificació del Treball

Les empreses tenen avui dia necessitats més enllà de cobrir una demanda mitjançant l'oferiment d'un producte a un client: algunes d'aquestes necessitats són, per exemple, saber si pot millorar el producte, si és l'hora de canviar de mercat o si es continuarà a curt termini produint beneficis.

Amb l'emmagatzematge de les dades de tot tipus no n'hi ha prou. Tecnològicament és possible ara fer-ho en un volum de dades enguany inimaginable fa una dècada. A més, els avanços en d'altres tipus de tecnologies (sobretot sensorials relatives a dispositius portàtils i *wearables*) incrementen aquest volum encara més.

D'altra banda, la jerarquia a les empreses estableix quasi sempre un grup de persones que, en última instància, són les que prenen les decisions més importants quan a la direcció estratègica de la companyia (un comitè, un grup de directius, etc). Aquestes persones han de saber llegir, d'alguna manera, la immensa quantitat de dades mitjançant processos d'extracció i de creació d'informació que ajudi a la interpretació del rumb del negoci.

És per tant necessari disposar de tecnologies que facin aquesta feina més accessible. Aquestes s'engloben dins del paradigma de la Business Analytics (analítica de negoci), on s'apliquen disciplines com Disseny de Bases de Dades, Estadística o, fins i tot, Intel·ligència Artificial.

En aquest context, TOTSALLES, una empresa de vendes d'articles d'àmbit nacional, necessita saber si la sensació que tenen els seus responsables de que les vendes cada cop van baixant més o de que el cost afecta al marge de beneficis negativament, és alguna cosa més tangible i evident. Aquest Treball, doncs, intentarà que l'empresa tingui justificacions quan a l'hora de prendre decisions importants pel que fa al rumb del seu negoci.

## 1.2 Objectius del Treball

L'objectiu d'aquest treball és fer l'anàlisi de vendes del sistema de la empresa que es planteja, TOTSALLES. A partir d'un conjunt de fitxers amb dades de vendes (que simularan les dades recollides en un sistema ERP de l'empresa), s'haurà de dissenyar, implementar, carregar i explotar un Data Mart (visió parcial d'un Data Warehouse o magatzem de dades) per donar sortida a les necessitats analítiques d'aquest sistema.

Per tant, el treball tindrà els següents objectius:

1. Analitzar 3 de les diferents plataformes de BI disponibles en el mercat i fer la tria d'una de elles per fer l'explotació de la informació.
2. Fer el disseny i la implementació del Data Mart.

3. A partir del conjunt de dades facilitat s'haurà de fer la càrrega d'aquestes en el Data Mart a través dels processos ETL (Extracció, transformació i càrrega de dades).
4. Una vegada carregades les dades en el Data Mart s'haurà d'explotar aquesta informació per tal de poder extreure certs indicadors clau per tal de donar als responsables de la empresa TOTSALLES una visió analítica del seu sistema.

Les preguntes que es volen resoldre amb el Data Mart són les següents:

- a) **S'ha produït un decrement de les vendes en les seves delegacions respecte a les vendes d'un any enrere?** Necessiten un anàlisi que els mostri com han evolucionat les vendes en cadascuna de les seves delegacions, tant en unitats venudes com en import.
- b) **Quins són els productes i famílies de productes més venuts?** Detectar quins són els productes i famílies de productes amb més èxit de vendes (top 10 productes) i quins menys, segons la delegació i any tractat. Voldrien potenciar tant els productes/famílies que es venen menys (potser descatalogar-los finalment i/o substituir-los per altres), com els que es venen més.
- c) **Quins són els clients top?** Volen fidelitzar aquells clients que els hi compren més (top 10 clients).
- d) **Quina és la distribució de les vendes segons la zona del client?** Volen conèixer des de quines zones els hi compren més i des de quines menys. L'objectiu que es plantegen és moure els comercials allà on faci falta.
- e) **Quina és la l'evolució de les vendes en funció del temps?** Necessiten tenir una jerarquia a la dimensió temps per poder analitzar les vendes en funció dels mesos, trimestres i anys.
- f) **Quines són les comissions que s'han de liquidar als comercials?** Volen conèixer les comissions que els comercials han de rebre per la seva feina. Ho necessiten tant en format mensual com anual.
- g) **Quin és el marge anual obtingut?** Es vol saber el marge anual per les famílies de productes, entès com la relació entre el import de la venda i el cost de la mateixa.

Altres objectius que s'han considerats secundaris són:

- i. Iniciar-se a l'aprenentatge de l'anàlisi de negoci, nou del tot per a l'alumne.
- ii. Usar només solucions de programari lliure (entorn, eines de suport, etc.) sempre que sigui possible.



### 1.3 Enfocament i mètode seguit

D'acord amb els materials de l'assignatura del TFM, s'ha seguit una metodologia de gestió de projectes informàtics, donat la semblança que hi ha entre el Treball i un projecte TIC. En aquest sentit, la classificació del PMBOK, que divideix el projecte en cinc etapes, s'ha fet correspondre amb les tasques de cada PAC. Així:

<b>Fase PMBOK</b>	<b>Descripció</b>	<b>PAC lliurada</b>
<i>Iniciació</i>	Elecció del tema de treball (consultor)	PAC1
<i>Planificació</i>	Pla de treball (definició d'objectius, abast, elecció de plataforma de BI)	PAC1
<i>Execució</i>	Desenvolupament del treball: lliurables parcials (anàlisi de dades, disseny i implementació del Data Mart i dels processos ETL)	PAC2, PAC3, PAC4
<i>Seguiment i control</i>	Desenvolupament del treball: lliurables parcials (correcció d'errors trobats a PACs anteriors i/o notificats pel consultor)	PAC2, PAC3, PAC4
<i>Tancament</i>	Lliurables finals (producte, memòria, presentació)	PAC4

*Taula 1: Relació entre fases del PMBOK i les PAC lliurades*

A més, es tindrà en compte altres treballs fets per alumnes de la UOC i que són disponibles al repositori institucional de la UOC, O2<sup>1</sup>. Una aproximació d'aquest tipus permet reutilitzar idees, conceptes i maneres de treballar que poden estalviar temps d'investigació, a més d'aportar claus per a la consecució dels objectius primaris.

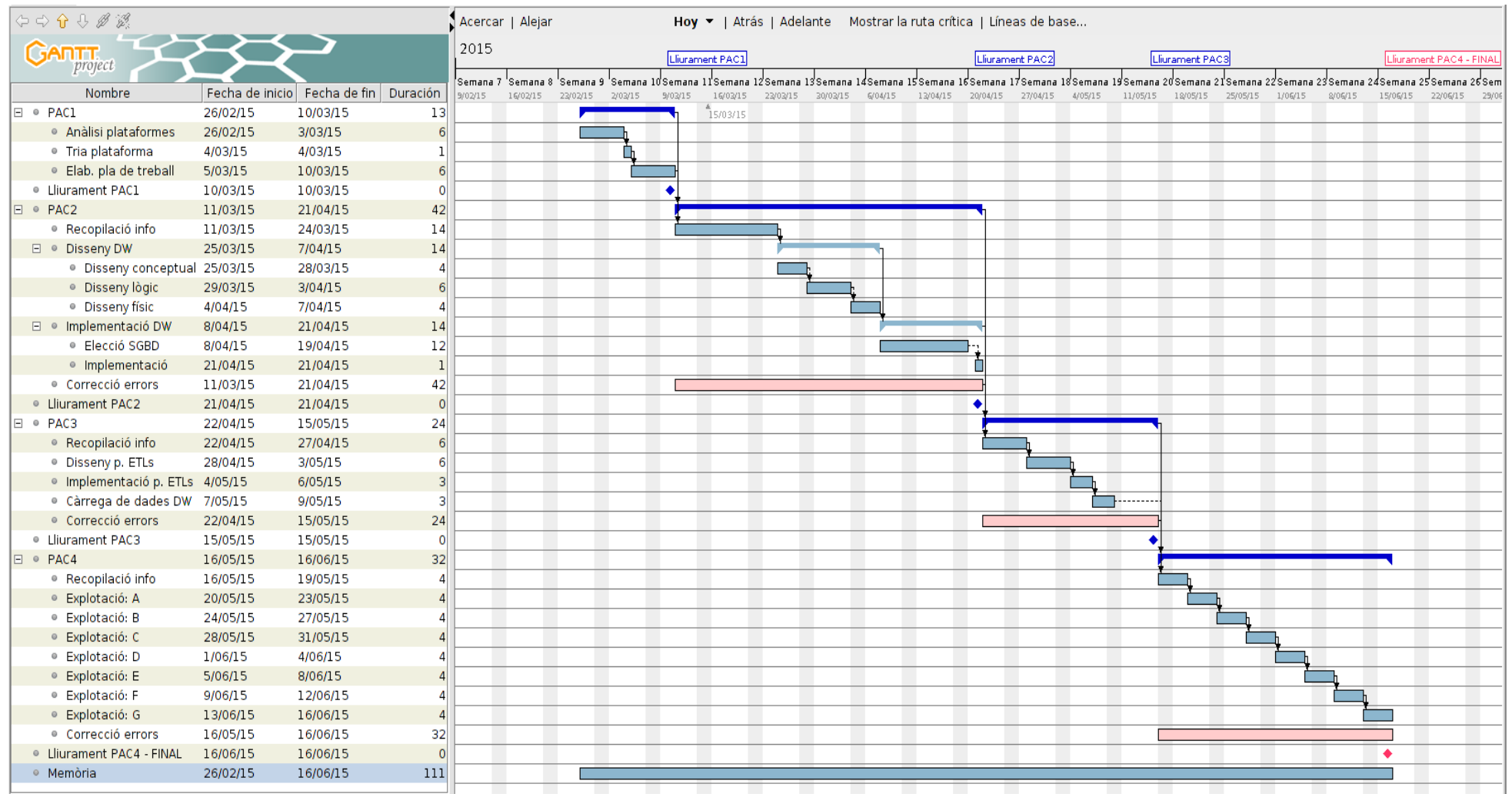
### 1.4 Planificació del Treball

Per escometre aquest projecte es disposa dels següents recursos:

- Ordinador amb processador de 4 nuclis, memòria RAM de 8 GB i 1 TB de disc dur.
- Sistema operatiu Linux Mint de 64 bits (basat en Ubuntu Linux 14.04)
- Connexió a Internet
- Recursos electrònics i de suport proporcionats pel consultor i disponibles a l'UOC.

A continuació es detalla una taula amb una divisió aproximada del TFM en tasques, i el diagrama de Gantt corresponent:

1 Accessible des de la Biblioteca de la UOC o mitjançant l'enllaç <http://openaccess.uoc.edu/webapps/o2/>



Il·lustració 1: Diagrama de Gantt per a la planificació del Treball

(captura de GanttProject per a Ubuntu Linux: <http://www.ganttproject.biz/>)

Al diagrama, les barres blaves corresponen a tasques i subtasques, i les barres roses a correcció d'errors, una tasca que en general es fa en paral·lel amb les altres, segons es van veient i/o els comunica el consultor quan corregeix les PAC.

Pel que fa a l'esforç i dedicació, la planificació reflexa 111 de dies de durada per a aquest projecte. S'ha considerat una dedicació diària d'entre 2 i 3 hores, tenint en compte circumstàncies personals i familiars. A més, s'ha considerat que els caps de setmana poden servir per recuperar el temps no dedicat als altres dies. L'estimació a la dedicació global, per a una dedicació diària de 2 hores, és de:

$$111 \text{ dies} \times 2 \text{ hores} = 222 \text{ hores totals}$$

Quan als riscos del projecte, s'han considerat els següents:

R1: Complexitat del projecte desborda les previsions inicials. Si bé es disposa de recursos molt valuosos com els que s'ha mencionat abans, pot ser que alguna fase requereixi més esforç i dedicació que les altres, ocasionant endarreriments.

- Probabilitat: Alta
- Impacte: Alt
- Pla de contingència: Sol·licitar ajuda al consultor, ús del fòrum.

R2: Eines de programari lliure no responen adequadament.

- Probabilitat: Baixa
- Impacte: Mitjà
- Pla de contingència: Per a cada eina, buscar una alternativa possible i funcionalment equivalent. Virtualitzar un segon sistema operatiu.

R3: Pèrdua parcial o total de la part del treball desenvolupat, ja sigui del producte (Data Warehouse i plataforma BI) o de la memòria final del treball.

- Probabilitat: Mitjana
- Impacte: Alt
- Pla de contingència: Bolcat regular a un núvol d'emmagatzematge (s'utilitzarà la plataforma *cloud* lliure Dropbox i GoogleDrive<sup>2</sup>)

---

2 Això pot ser útil per un projecte final com aquest, amb dades fictícies, però pot no ser viable en un projecte real, ja que la llei de protecció de dades obliga a garantir alguns aspectes que sistemes del núvol, com per exemple Dropbox, pot no garantir en funció del país on s'allotgin les dades. (Veure [http://www.agpd.es/portalwebAGPD/revista\\_prensa/revista\\_prensa/2012/notas\\_prensa/new\\_s/2012\\_07\\_01-ides-idphp.php](http://www.agpd.es/portalwebAGPD/revista_prensa/revista_prensa/2012/notas_prensa/new_s/2012_07_01-ides-idphp.php) o <http://www.multilegalauditblog.com/es-compatible-google-drive-dropbox-con-la-lopd/> )

## 1.5 Breu sumari de productes obtinguts

D'acord amb el pla d'estudis del TFM es lliura el següent conjunt de productes:

- Aquesta memòria, guia que recull tot el procés de desenvolupament del TFM, en format PDF.
- Carpeta comprimida amb els conjunts de fitxers resultants del producte: diagrames, informes, i scripts, en format ZIP.
- Presentació, vídeo a l'eina del Campus Present@.

## 1.6 Breu descripció dels altres capítols de la memòria

A continuació es descriu breument el contingut de cada capítol.

- Capítol 2: Plataformes de BI, on s'analitzen tres possibles solucions i s'escull la que servirà per explotar les dades, Eclipse BIRT.
- Capítol 3: Anàlisi de les dades inicials, on es fa un breu anàlisi de les dades de partida, i del format en que hi són.
- Capítol 4: Disseny del Data Mart, on s'expliquen els dissenys conceptual i lògic proposats, previs a la implementació física.
- Capítol 5: Implementació del Data Mart, on s'explica el disseny físic (la base de dades) sobre un sistema gestor escollit.
- Capítol 6: Disseny i implementació dels processos ETL, on s'explica com es carreguen les dades al emmagatzemament.
- Capítol 7: Explotació del Data Mart, on s'explica com s'han desenvolupat els informes.
- Capítol 8: Conclusions, on es recullen les impressions que l'alumne ha tingut del desenvolupament del projecte i les línies futures d'ampliació del mateix.

## 2. Plataformes de BI

Es tracta d'analitzar algunes comparatives i de triar una d'elles. Amb l'objectiu de desenvolupar aquest TFM amb eines de programari lliure, s'han analitzat tres plataformes englobades dintre d'aquest paradigma. No s'ha fet un anàlisi exhaustiu ja que aquest no és pas un dels objectius d'aquest Treball. Per obtenir més informació es pot consultar [1] i [2] <sup>3</sup>

### 2.1 Comparativa Pentaho, SpagoBI, Eclipse BIRT

Dintre de l'univers de plataformes de BI, s'han proposat tres. Es descriuen molt breument algunes de les característiques. Per a més informació, es pot consultar [4].

Plataforma	Tipus	Funcions principals
Pentaho	Suite	"Pentaho Data Integration" pels ETL (Extracció, Transformació i càrrega de dades), "Pentaho Workbench" per elaborar els cubs OLAP i "Pentaho Dashboard" i "Pentaho Report Designer" per la elaboració de informes (explotació de dades i quadres de comandament)
SpagoBI	Suite	OLAP, generació d'informes, Dashboards i Scorecards (reuneix tots els components BI).
BIRT	Conjunt de plugins per a Eclipse	Generació d'informes, creació de Dashboards i Scorecards.

*Taula 2: Breu comparativa de tres plataformes de BI*

En referència al document mencionat abans, les tres plataformes es descriuen força bé a l'apartat 9.1.17. Pentaho Business Analytics (pàgina 73), l'apartat 9.1.30. SpagoBI (pàgina 80) i l'apartat 9.1.19. Eclipse Birt (pàgina 77). Les funcionalitats principals van ser estretes d'aquí per a la comparativa de la taula anterior. A continuació els avantatges i inconvenients per a cadascuna:

#### **Pentaho**

Avantatges: Gran comunitat online. Complementos de programari lliure. Fàcil d'usar. Compatibilitat amb eines de programari de codi obert (Open Source)

Inconvenients: Complementos amb llicència privada (a més de la versió gratuïta (Community Edition), existeix la versió Enterprise (Enterprise Edition) que incorpora serveis i funcionalitats que la versió gratuïta no incorpora).

<sup>3</sup> En general no és fàcil trobar anàlisis molt complets sense haver de pagar a l'organització que els fa, ja que triar una bona plataforma de BI pot repercutir ostensiblement en els beneficis de l'empresa que vol fer l'analítica del seu negoci.

## **SpagoBI**

Avantatges: Es pot usar per a qualsevol solució BI. Programari lliure amb llicència MPLv2. Inclou processos ETL. Visions dels cubs OLAP configurables. Interpreta qualsevol tipus de dades d'entrada. Actualitzacions permanents. Suport gratuït. Etc.

Inconvenients: Cal registrar-se per obtenir-ho. Per a explotar tot el seu potencial, es necessiten d'usuaris experts.

## **Eclipse BIRT**

Avantatges: Pot cobrir tots els tipus de solucions BI. N'hi ha exemples de tot tipus, tutorials, molta informació,... Llicència EPL. Fàcil integració amb d'altres eines, etc.

Inconvenients: Per a aprofitar el seu potencial es requereix d'usuaris experts que siguin capaços de programar.

## 2.2 Elecció de la plataforma

Els criteris d'elecció van ser:

1. Pes. La instal·lació de les suites Pentaho i SpagoBI ocupen gairebé 1'5 GB d'espai d'emmagatzematge cadascuna, mentre que BIRT «només» ocupa uns 300 MB. Tot i que no n'hi ha problemes d'emmagatzematge, s'espera que una instal·lació amb menys pes sigui més senzilla. És important senyalar que el que ocupi menys no significa necessàriament que sigui més senzilla d'usar, però en tenir menys pes segurament tindrà menys funcionalitat implementada que pot ser no calgui per explotar el magatzem de dades.
2. Documentació i tutorials a Internet. Les tres plataformes disposen de molta informació a través de les seves comunitats, sobretot per les llicències de programari de codi obert amb que s'ofereixen.
3. Familiaritat. L'alumne ja ha treballat amb la plataforma Eclipse i amb el llenguatge de programació Java i espera, en aquest sentit, que la corba d'aprenentatge de l'entorn BI sigui suau.

Per tant, d'acord amb aquestes línies, la plataforma escollida va ser Eclipse BIRT.

### 3. Anàlisi de les dades inicials

Es tracta de fitxers en format de text pla, procedents del bolcat d'un ERP. Els camps estan limitats per un separador, un punt i coma<sup>4</sup>, on la primera línia indica una capçalera que denota cadascuna de les columnes.

Cada fitxer guarda dades d'un aspecte del sistema:

Nom del fitxer	Dades que guarda
Almacenes.csv	Magatzems
Articulos.csv	Articles
Clientes.csv	Clients (personal)
ClientesZona.csv	Clients i zones a les que pertanyen
Comerciales.csv	Comercials (personal)
Conceptos.csv	Tipus de conceptes (no articles)
Divisa.csv	Tipus de divises
Empresas.csv	Delegacions (5 en total)
FamiliasArticulos.csv	Articles i famílies a les que pertanyen
FormasPago.csv	Formes de pagament
TipoDato.csv	Tipus de dades (real, pressupost o previsió)
TipoLinea.csv	Tipus de línia (si es d'article o de concepte)
Ventas.csv	Informació principal de les vendes

*Taula 3: Fitxers de text amb les dades originals*

Com exemple, l'aspecte intern d'un aquests fitxers és:

---

4 El format CSV és un dels anomenats DSV, Delimiter-Separated Values (valors separats per un delimitador), en referència al caràcter que marca on acaba i on comença cada camp. Segons sigui aquest caràcter, el fitxer es pot trobar en un format CSV (Coma), SSV (de Semicolon, punt i coma, el nostre), PSV (de Pipe, canonada), etc. (Font: [http://en.wikipedia.org/wiki/Delimiter-separated\\_values](http://en.wikipedia.org/wiki/Delimiter-separated_values))

```

Almacenes.csv x
GEN_ID;NAME;PARENT (PARENT_GEN_ID)
+;Todos;
AIBE001;Almacén centro;+
AIBE002;Almacén norte;+
AIBE003;Almacén sur;+
PCIN01;Almacén secundario centro;+
PCIN02;Almacén secundario norte;+
PCIN03;Almacén secundario sur;+
-;Sin Almacén;
--;Sin Almacén;-|

```

Il·lustració 2: Fitxer de text Almacenes.csv

L'anàlisi del mateix passa per descriure cada camp/columna, identificar el tipus de dades que conté cada columna i si conté informació errònia o no rellevant, i que posteriorment s'haurà de treure o de modificar als processos ETL (veure el Capítol 6). Així:

Posició	Camp Original	Significat	Tipus de dades	Noms alternatius
1	GEN_ID	Codi del magatzem	Cadena	ID_ALMACEN, COD_ALMACEN
2	NAME	Nom del magatzem	Cadena	NOM, DESC
3	PARENT (PARENT_GEN_ID)	Referència exportada	Cadena	
<b>Comentari</b>		Tercera columna sense informació rellevant.		

Taula 4: Anàlisi de les dades originals del fitxer Almacenes.csv



Per al fitxer Articulos.csv:

Posició	Camp Original	Significat	Tipus de dades	Noms alternatius
1	GEN_ID	Codi de l'article	Cadena	ID_ARTICLE, COD_ARTICLE
2	NAME	Nom del magatzem	Cadena	NOM, DESC
3		Classe de l'article	Cadena	ID_CLASS, COD_CLASS
<b>Comentari</b>		Darrers registres sense informació rellevant.		

*Taula 5: Anàlisi de les dades originals del fitxer Articulos.csv*

Cientes.csv:

Posició	Camp Original	Significat	Tipus de dades	Noms alternatius
1	GEN_ID	DNI del client	Cadena	DNI, ID_CLIENT
2	NAME	Nom del client	Cadena	NOM
3	PARENT (PARENT_GEN_ID)	Sector laboral	Cadena	SECTOR, ID_SECTOR
<b>Comentari</b>		Es combina la informació dels clients amb el sector al que pertanyen. Darrers registres sense informació rellevant.		

*Taula 6: Anàlisi de les dades originals del fitxer Cientes.csv*

I així amb la resta de fitxers que, per comoditat, no s'inclouen a aquest document.

## 4. Disseny del Data Mart

### 4.1 Disseny conceptual

Per dur a terme el disseny conceptual, s'haurà de tenir en compte les necessitats del DW i la manera en que ajudarà a respondre les preguntes objectiu que es van plantejar. La idea es obtenir una organització de futures taules a la Base de Dades del magatzem que optimitzarà les consultes. Donat que aquest no és el disseny físic, no es parla ara de taules pròpiament, però sí d'entitats<sup>5</sup>.

Aquesta organització o arquitectura, de manera general, es compon per una o varies entitats principals denominades «de fets» (facts, en anglès), cadascuna representant un procés de negoci. Per al cas que ens ocupa, només es tindrà una: vendes<sup>6</sup>. Al voltant es tindrà les «dimensions» o entitats que representen «maneres» o «punts de vista» en que es pot referir també la informació de les entitats de fets.

Aquesta és l'arquitectura en forma d'estrella: entitats de fets al centre i entitats de dimensions al voltant i relacionades amb elles. En cas de que alguna entitat-dimensió es divideixi en «subentitats», l'arquitectura passaria a denominar-se «en flocs de neu» (snowflows, en anglés). Aquesta subdivisió es tradueix més endavant en una normalització de taules, i respon a la necessitat de representar informació que no pot recollir-se en una sola taula. Per exemple, si es té la dimensió Temps, es subdividirà en Anys, Trimestres i Mesos, o amb una entitat Productes, es subdividiria en Categories i Famílies.

Així doncs, per a cadascuna de les preguntes plantejades, i la resposta la qual és un informe (report), s'identifiquen les dimensions relacionades:

**a. S'ha produït un decrement de les vendes en les seves delegacions**

**respecte a les vendes d'un any enrere?** Necessiten un anàlisi que els mostri com han evolucionat les vendes en cadascuna de les seves delegacions, tant en unitats venudes com en import.

Dimensions: Empreses, Temps → Anys

**b. Quins són els productes i famílies de productes més venuts?** Detectar quins són els productes i famílies de productes amb més èxit de vendes (top 10 productes) i quins menys, segons la delegació i any tractat.

Dimensions: Empreses, Articles → Famílies, Temps → Anys

**c. Quins són els clients top?** Volen fidelitzar aquells clients que els hi compren més (top 10 clients).

Dimensions: Clients

**d. Quina és la distribució de les vendes segons la zona del client?** Volen conèixer des de quines zones els hi compren més i des de quines menys.

Dimensions: Clients → Zones

---

5 L'objectiu d'aquest disseny és el d'obtenir un diagrama d'entitat-relació (ERD).

6 És per això que en comptes de tenir un Datawarehouse, la denominació més exacta seria la de Datamart, ja que només s'analitzarà una activitat de l'empresa: les vendes. Un Datamart és per tant un subconjunt (molt més manejable) d'un DW.

**e. Quina és la l'evolució de les vendes en funció del temps?** Necessiten tenir una jerarquia a la dimensió temps per poder analitzar les vendes en funció dels mesos, trimestres i anys.

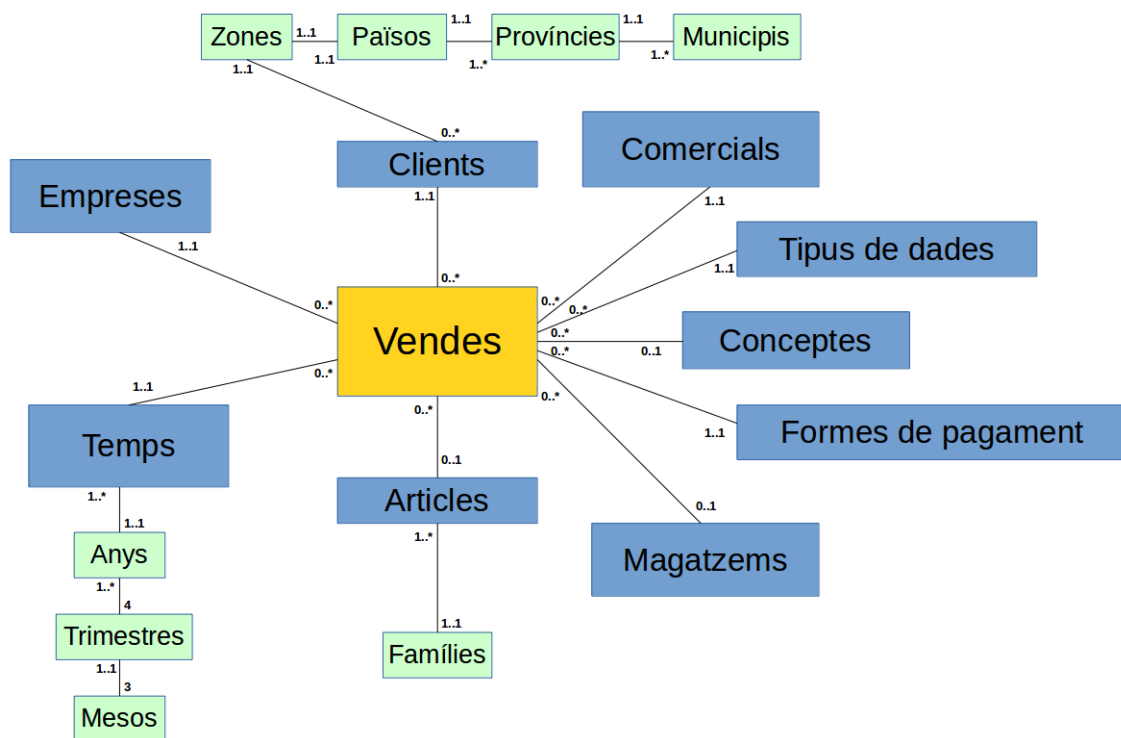
Dimensions: Temps → Anys → Trimestres → Mesos

**f. Quines són les comissions que s'han de liquidar als comercials?** Volen conèixer les comissions que els comercials han de rebre per la seva feina. Ho necessiten tant en format mensual com anual.

Dimensions: Comercials, Temps → Anys → Mesos

**g. Quin és el marge anual obtingut?** Es vol saber el marge anual per les famílies de productes, entès com la relació entre el import de la venda i el cost de la mateixa.

Dimensions: Articles → Famílies



*Il·lustració 3: Diagrama d'entitats del model conceptual del Data Mart*

El bloc central es correspon amb l'entitat de fets (activitat). Els blocs de color blau representen dimensions (punts de vista de l'activitat, el «per ...» de les preguntes). Els blocs de color verd representen jeràrquies dins d'algunes dimensions. És per tant un esquema en floccs de neu.

Respecte a la cardinalitat, una relació de «uns a molts» indica que una instància d'una entitat pot relacionar-se amb varies instàncies d'una altra. Per exemple, una venda es realitza per un (i només un) comercial, però un comercial pot fer cap o moltes vendes. Això es tindrà en compte a l'hora de fer el disseny lògic ja que les jeràrquies, en cas de que presentin una cardinalitat

«de un a molts» es podran incloure a les dimensions com si fossin atributs<sup>7</sup>. La cardinalitat de 0..1 per a les entitats/dimensions Articles, Magatzems i Conceptes respon al fet de que una venda ho és d'un article o d'un concepte. Anàlogament, per a una venda d'un concepte, no té sentit que aparegui el codi del magatzem, ja que no es refereix a un article<sup>8</sup>.

## 4.2 Disseny lògic

En aquest pas intermedi entre el nivell conceptual i el nivell físic (en el que es defineix la base de dades final) s'estableixen les relacions entre dimensions i fets mitjançant claus, i es defineixen els atributs més importants per a cada entitat, així com les mètriques. És per tant un nivell «pont», els elements dels quals poden referir-se com a blocs lògics o com a taules.

Les decisions més importants d'aquest disseny van ser:

1. Als blocs/taules Empreses i Magatzems es podrien tenir com a claus els propis codis, però s'ha optat per una implementació alternativa. En realitat no s'ha trobat un avantatge clar fent usar una o altra.
2. Als blocs/taules Clients i Comercials, el DNI de la persona es podria fer usar com a clau, però s'ha considerat l'opció de minimitzar el repartiment «gratuit» a la base de dades d'una dada de caràcter personal per qüestions de privacitat (si, per exemple, la taula de fets es veu compromesa en la seva seguretat, no es posa en risc la divulgació d'aquesta dada).
3. Les jerarquies dels blocs Zones, Temps i Famílies han restat englobades a les respectives entitats/dimensions origen. Tal i com es va comentar de passada en dissenyar el model conceptual, si un element no té suficient informació per ser considerada com una dimensió apart, es pot considerar incloure'l a la dimensió mare. Aquesta és una decisió que dependent del volum de dades que s'està tractant, pot influir finalment al rendiment a l'hora de fer consultes, ja que s'haurà d'accedir a més o menys taules físiques.

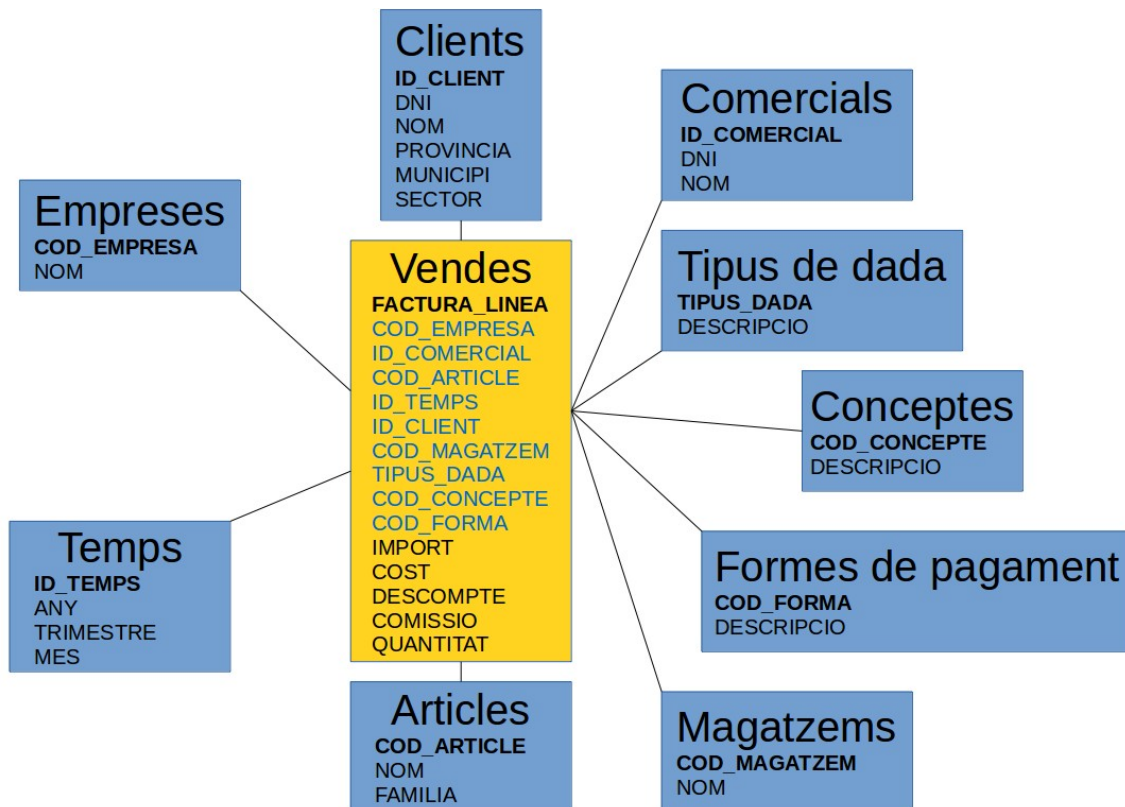
Es mostra finalment el diagrama del model lògic. Al bloc de fets (vendes), els camps de color blau corresponen a claus foranies cap a les taules-dimensions. La clau primària estarà formada per la concatenació de el nombre de factura i el nombre de la línia. D'aquesta forma no es perd aquesta informació i es pot rastrejar els registres a la font original. La resta de camps són mètriques que ajudaran a presentar la informació requerida als informes.

A les dimensions (resta de blocs) els camps en negreta representen les claus primàries.

---

7 La primera "gran dificultat" consisteix en si considerar un determinat element una dimensió apart o bé un atribut més.

8 Estrictament parlant, es podria veure que un concepte està relacionat amb un article, però aquí es suposa que són elements independents.



*Il·lustració 4: Diagrama del disseny lògic del Data Mart*

Així doncs, les preguntes objectiu són clau per a determinar les dimensions, ja que cadascuna representa un punt de vista del sistema. Així, si tenim vendes, «coses» que voldríem saber al respecte serien «qui ha fet la venda» (clients), «quan es va fer» (temps), «per quin comercial» (comercials), «en què consisteix la venda» (article o concepte), etc.

Respecte a la composició de cada dimensió, és a dir, els seus atributs, estan reflectits a l'esquema lògic i sempre es tracta d'informació relacionada. Per exemple, d'un client es recull el nom i la zona on viu, d'un comercial el seu nom i la comissió, etc.

## 5. Implementació del Data Mart

### 5.1 Elecció del SGDB

Per a escometre el disseny físic, és a dir, la implementació del model lògic d'abans, s'ha triat com a gestor de bases de dades Apache Derby<sup>9</sup> a través del connector per a Eclipse. Es tracta d'una implementació en Java molt fàcil de configurar a l'entorn i que demostra tenir uns més que acceptables paràmetres de rendiment i de requeriments<sup>10</sup>.

Com ja es va mencionar anteriorment, el SGDB del Datamart serà Apache Derby (Cloudscape o JavaDB segons implementacions de IBM i de Oracle respectivament). Tot i així, el model es perfectament transportable a un altre SGDB com PostgreSQL, més robust i pot ser més indicat per a un magatzem de dades (ben és cert que la instal·lació és més pesada).

### 5.2 Disseny físic

Les decisions de disseny més importants van ser:

Notació de les taules: les dimensions del sistema de BI es denoten pel prefixe «D\_» als noms de les taules. La taula de fets corresponent a les vendes és «F\_VENDES».

Tipus de les claus primàries: les claus de les taules Clients, Comercials i Temps són numèriques (tipus INTEGER). Les claus de Articles, Magatzems, Formes de pagament, Tipus de dada, Empreses i de la taula de vendes (factura i número de línia) són del tipus «cadena» (VARCHAR).

Mètriques: el tipus de les mètriques Import, Cost, Descompte i Comissió és DECIMAL. El de la Quantitat és INTEGER.

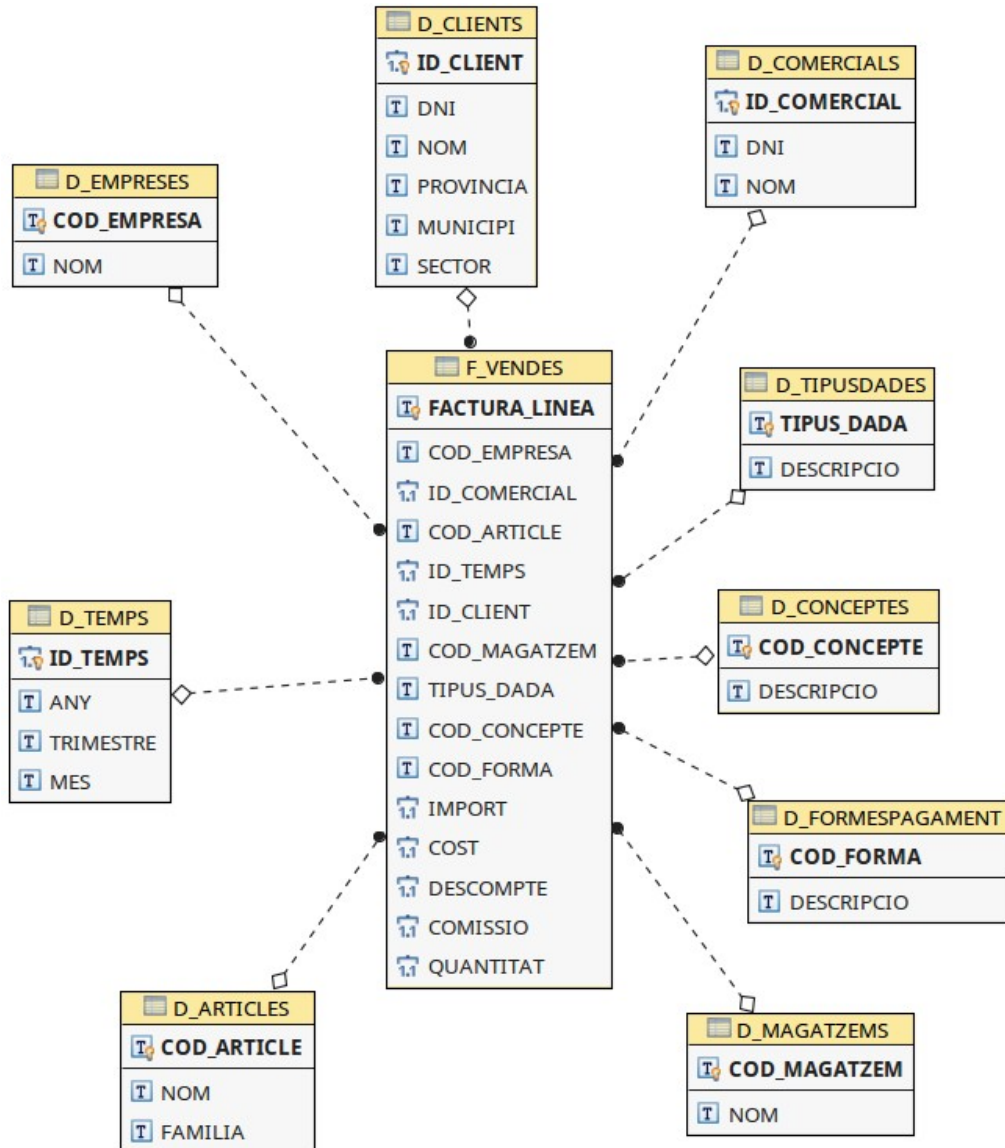
Es mostra a continuació la proposta de implementació. Les taules s'han creat usant l'eina Dbeaver<sup>11</sup> per a definir els camps i les relacions entre taules. A continuació, s'ha obtingut el diagrama i l'script en SQL.

---

9 <http://db.apache.org/derby/>

10 Veure <http://www.vogella.com/tutorials/EclipseDataToolsPlatform/article.html> quan a la configuració d'aquest SGDB sobre la plataforma Eclipse.

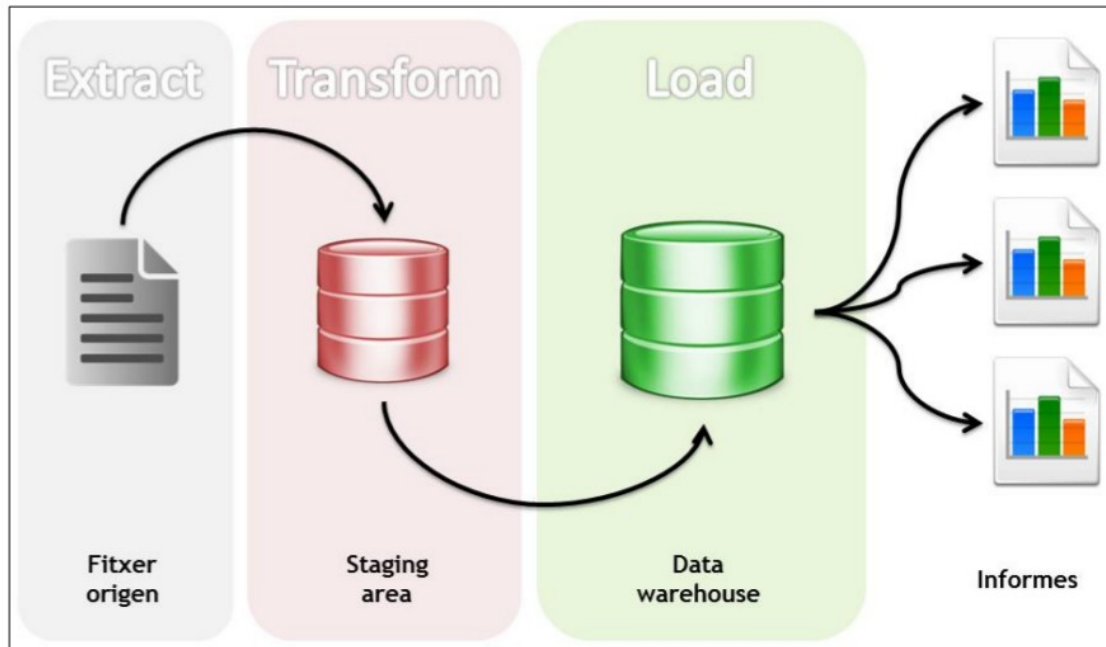
11 <http://dbeaver.jkiss.org/>







Il·lustració 5: Diagrama del model físic del Data Mart

## 6. Disseny i implementació dels processos ETL

Per processos ETL s'entén una sèrie de passos que cal fer a les dades de partida abans de que la informació pugui ser explotada. De manera esquemàtica es mostra en què consisteixen.



	<b>Fitxer origen</b>	Conjunt de fitxers amb les dades de l'empresa en text pla.
	<b>Staging area</b>	(CSV) Magatzem de dades on es realitzaran les transformacions temporals prèvies a la càrrega.
	<b>Data warehouse</b>	(TFM) Magatzem de dades corporatiu amb l'estructura adequada per a la seva explotació.
	<b>Informes</b>	Documents d'anàlisi que ajudaran a la pressa de decisions.

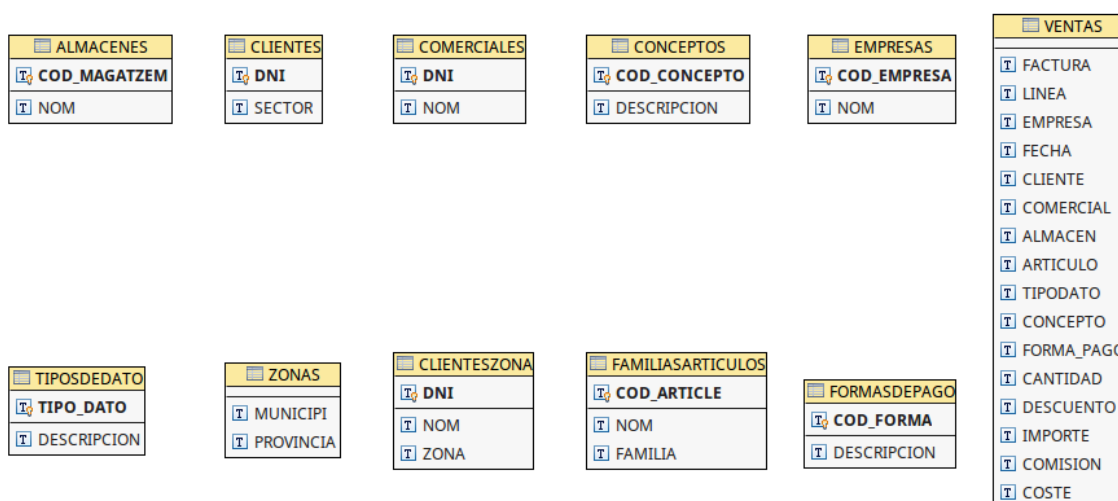
*Il·lustració 6: Processos ETL i esquemes usats*  
(imatge extreta de [3] amb permís de l'autor)

### 6.1 Staging Area

Com ja s'ha vist, les dades són a fitxers de text en format CSV i l'objectiu és passar-les a l'esquema del magatzem de dades. Per fer-ho, s'ha considerat la creació d'un altre esquema intermedi, o conjunt de taules, anomenat *Staging area*, que reflecteixi exactament aquesta mateixa informació i que aconseguixi



posteriorment que els processos ETL siguin més ràpids<sup>12</sup>. Aquest «procés previ» s'anomena EXTRACCIÓ i es farà a l'esquema CSV. L'esquema físic de taules és el següent:



Il·lustració 7: Staging area CSV

Els camps de cada taula s'han anomenat atenent a l'anàlisi de les dades que es va fer anteriorment, i són els més rellevants. No és necessari establir relacions de claus foranes, ja que això es farà més endavant al magatzem pròpiament dit (esquema TFM). A més a més, s'ha dividit l'arxiu dels clients en dos: un amb la relació entre clients i la zona a la que pertanyen (importat a la taula CLIENTESZONA) i un altre amb la relació d'un municipi amb la província, als registres finals (importat a la taula ZONAS), ja que ambdós tipus d'informació estaven barrejats<sup>13</sup>.

Un cop creat aquest esquema intermedi CSV, es tracta d'importar cada fitxer per separat a cada taula. S'ha usat el programari Squirrel<sup>14</sup> perquè ha possibilitat la importació automàtica de la majoria de registres de tots els fitxers sense donar cap error, i per la flexibilitat a l'hora de «renombrar» les columnes<sup>15</sup> i de seleccionar al codificació de caràcters<sup>16</sup>, a més de proporcionar una interfície gràfica que permet visualitzar còmodament els resultats de les extraccions.

12 L'objectiu és no treballar directament amb les fonts de dades (en aquest cas, amb els fitxers de text), sinó fer una «instantània» o còpia a una base de dades des d'on aplicar els processos ETL. Les fonts originals podran actualitzar-se periòdicament i un procés (normalment en segon pla) faria la còpia cap a aquesta *Staging area*.

13 Bàsicament el que s'ha fet és anar al fitxer ClientesZona.csv i crear un altre fitxer Zonas.csv amb la informació municipis-províncies del final. S'ha considerat que la mida del primer fitxer és del tot manejable i per això no ha hagut problemes d'usabilitat en fer la divisió manualment.

14 <http://squirrel-sql.sourceforge.net/>

15 Una altra bona eina és SQL Workbench/J (<http://www.sql-workbench.net/>). La plataforma DTP per a Eclipse també permet importar però té menys flexibilitat.

16 La codificació que s'ha usat és ISO-8859-1, europea. Amb UTF-8 (per defecte a Derby) les cadenes de text no es mostrarien bé i s'hauria de fer un tractament posterior.

No obstant això, si el que es vol és que aquesta importació/extracció cap a la àrea intermèdia es pugui fer de manera desatesa amb l'ajuda d'un planificador de tasques (com la comanda `cron` de Unix o el Task Scheduler de Windows) llavors s'ha de crear un fitxer de comandes, anomenat script de lots o *batch file*.

Gairebé tots els SGBD tenen una eina per ser executada a la línia de comandes (sqlplus a Oracle, psql a PostgreSQL, etc). A Derby, aquesta eina és `ij`. La manera de programar aquest traspàs automàtic és:

1. Crear un fitxer `Extract.sql`, per exemple, amb el contingut:

```
connect 'jdbc:derby://localhost:1527/MyDB_NetworkServer;create=true';
set current schema csv;
delete from almacenes;
delete from clientes;
delete from clienteszona;
delete from comerciales;
delete from conceptos;
delete from empresas;
delete from familiasarticulos;
delete from formasdepago;
delete from tiposdedato;
delete from ventas;
CALL SYCS_UTIL.SYCS_IMPORT_TABLE ('CSV','ALMACENES','Almacenes.csv',
';',null, 'ISO-8859-1', 0);
CALL SYCS_UTIL.SYCS_IMPORT_TABLE ('CSV','CLIENTES','Clientes.csv',
';',null, 'ISO-8859-1', 0);
CALL SYCS_UTIL.SYCS_IMPORT_TABLE ('CSV','CLIENTESZONA','ClientesZona.csv',
';',null, 'ISO-8859-1', 0);
CALL SYCS_UTIL.SYCS_IMPORT_TABLE ('CSV','COMERCIALES','Comerciales.csv',
';',null, 'ISO-8859-1', 0);
CALL SYCS_UTIL.SYCS_IMPORT_TABLE ('CSV','CONCEPTOS','Conceptos.csv',
';',null, 'ISO-8859-1', 0);
CALL SYCS_UTIL.SYCS_IMPORT_TABLE ('CSV','EMPRESAS','Empresas.csv',
';',null, 'ISO-8859-1', 0);
CALL SYCS_UTIL.SYCS_IMPORT_TABLE
('CSV','FAMILIASARTICULOS','FamiliasArticulos.csv', ';',null, 'ISO-8859-1',
0);
CALL SYCS_UTIL.SYCS_IMPORT_TABLE ('CSV','FORMASDEPAGO','FormasPago.csv',
';',null, 'ISO-8859-1', 0);
CALL SYCS_UTIL.SYCS_IMPORT_TABLE ('CSV','TIPODEDATO','TipoDato.csv',
';',null, 'ISO-8859-1', 0);
CALL SYCS_UTIL.SYCS_IMPORT_TABLE ('CSV','VENTAS','Ventas.csv', ';',null,
'ISO-8859-1', 0);
```

*Consulta 1: Importació de dades originals al Staging Area CSV*

2. Construir la comanda que crida a `ij` (cal tenir accés a `derbyrun.jar` de la distribució binària de Derby al `CLASSPATH` de Java):

```
java org.apache.derby.tools.ij Extract.sql
```

3. Finalment, fer que el planificador executi la comanda anterior, per exemple, a principi de mes. Les dades que més canviaran seran les relatives als clients, comercials, articles i vendes sobretot.

## 6.2 Extracció, Transformació, Càrrega

A continuació es descriuen els processos d'extracció, transformació i càrrega<sup>17</sup> que ompliran l'esquema del Datamart. En general consisteixen en agafar les dades originals (que ja es tenen a l'esquema CSV), fer transformacions prèvies, càlculs o mapatges entre columnes (si escau) i finalment inserir les noves versions d'aquestes dades a l'esquema TFM). Per tant, al menys un procés d'aquest tipus (per a cada taula al magatzem) implica una *selecció* de dades (corresponent-se amb una instrucció SELECT al SGBD) i una *inserció* (corresponent-se amb una instrucció INSERT).

La manera d'omplir les Taules D\_MAGATZEMS, D\_EMPRESES, D\_COMERCIALS, D\_ARTICLES, D\_TIPUSDADES, D\_CONCEPTES i D\_FORMESPAGAMENT és molt semblant. Per exemple, per al cas de la taula D\_MAGATZEMS, es fa una selecció dels camps COD\_MAGATZEM i NOM de la taula CSV.ALMACENES, filtrant (condició WHERE) els que no tinguin al primer camp (el codi) el valor '+' ja que no correspon a cap registre vàlid, i a més a més, els que tinguin un guió, '-' al mig, ja que aquests registres tampoc són vàlids. Un cop feta aquesta selecció, simplement s'insereix a la taula TFM.D\_MAGATZEMS, als corresponents camps.

```
insert into TFM.D_MAGATZEMS (COD_MAGATZEM, NOM)
select COD_MAGATZEM, NOM
from CSV.ALMACENES
where COD_MAGATZEM <> '+' and COD_MAGATZEM not like '%-%';
```

*Consulta 2: Càrrega de la taula de D\_MAGATZEMS*

A algunes taules-dimensions és necessari fer una inserció més posterior. És el cas de D\_MAGATZEMS, D\_ARTICLES, D\_CONCEPTES, D\_FORMESPAGAMENT, per al cas de que hi hagi vendes que no tinguin els codis associats:

```
insert into TFM.D_MAGATZEMS (COD_MAGATZEM, NOM) values ('NO_AL', 'Sense
magatzem');
insert into TFM.D_ARTICLES (COD_ARTICLE, NOM) values ('NO_ART', 'Sense
article');
insert into TFM.D_CONCEPTES (COD_CONCEPTE, DESCRIPCIO) values ('NO_C',
'Sense concepte');
insert into TFM.D_FORMESPAGAMENT (COD_FORMA, DESCRIPCIO) VALUES ('NO_F',
'Sense forma p.');
```

*Consulta 3: Càrrega de valors no definits*

Per a omplir la taula-dimensió dels clients, D\_CLIENTS, s'haurà de consultar les taules de l'esquema CSV CLIENTES (per als camps Dni, Nom i Sector), CLIENTESZONA (per al camp Municipi) i ZONAS (per al camp Província). La manera de fer-ho, tot enllaçant els registres sense que es perdi informació, es usant instruccions JOIN internes, és a dir, que relacionin sempre dues registres de dues taules per un camp amb igual valor no nul. Així, el Dni d'un client a la taula CLIENTES haurà d'existir a la taula CLIENTESZONA per poder referir els altres camps. En cas de que a l'hora de carregar la taula de

---

<sup>17</sup> En anglès, Extract, Transform, Load

vendes, no existeixi un determinat client referit a CLIENTESZONA, el registre a F\_VENDES tindrà com a codi de client 1 (ú)<sup>18</sup>.

```
insert into TFM.D_CLIENTS (NOM) values ('Client desconegut'); -- ID_CLIENT=1
```

*Consulta 4: Càrrega de valor de client desconegut*

Per a omplir la taula-dimensió de temps, D\_TEMPS, tenint en compte que només tenim dades de 5 anys, el més pràctic és fer la inserció sense selecció prèvia, amb valors constants<sup>19</sup>. Aquesta taula genera una columna autoincrementada que serà la que després sigui referenciada a la taula F\_VENDES com a clau forana. Es mostra un exemple de com seria per un any qualsevol:

```
insert into TFM.D_TEMPS ("ANY", TRIMESTRE, MES) VALUES ('2009', '1', '01');
-- Gener
insert into TFM.D_TEMPS ("ANY", TRIMESTRE, MES) VALUES ('2009', '1', '02');
-- Febrer
insert into TFM.D_TEMPS ("ANY", TRIMESTRE, MES) VALUES ('2009', '1', '03');
insert into TFM.D_TEMPS ("ANY", TRIMESTRE, MES) VALUES ('2009', '2', '04');
insert into TFM.D_TEMPS ("ANY", TRIMESTRE, MES) VALUES ('2009', '2', '05');
insert into TFM.D_TEMPS ("ANY", TRIMESTRE, MES) VALUES ('2009', '2', '06');
insert into TFM.D_TEMPS ("ANY", TRIMESTRE, MES) VALUES ('2009', '3', '07');
insert into TFM.D_TEMPS ("ANY", TRIMESTRE, MES) VALUES ('2009', '3', '08');
insert into TFM.D_TEMPS ("ANY", TRIMESTRE, MES) VALUES ('2009', '3', '09');
insert into TFM.D_TEMPS ("ANY", TRIMESTRE, MES) VALUES ('2009', '4', '10');
insert into TFM.D_TEMPS ("ANY", TRIMESTRE, MES) VALUES ('2009', '4', '11');
insert into TFM.D_TEMPS ("ANY", TRIMESTRE, MES) VALUES ('2009', '4', '12');
```

*Consulta 5: Càrrega de la taula D\_TEMPS per a l'any 2009*

Un cop són omplertes les taules-dimensions, és el torn de la taula de fets o taula principal, F\_VENDES. Es carreguen tots els registres per a no perdre informació. Si algun codi de client, concepte, etc. no és informat, la inclusió d'aquestes «vendes» també pot indicar dades "incorrectes" registrades directament en el sistema transaccional (l'ERP). I és a partir d'aquesta informació on els responsables de TOTSALES poden acabar decidint que han de millorar els seus sistemes.

Si un article no està identificat, pot ser que la venda es refereixi a un concepte. Si és així, llavors es pot esperar que el magatzem tampoc estigui identificat. La primera part de la instrucció de càrrega té en compte això i, en cas de que es retorni un codi d'article, de concepte o de magatzem nuls, modifica els valors perquè siguin NO\_ART, NO\_C o NO\_AL respectivament. De manera semblant passa amb un codi de client o de forma de pagament.

A continuació s'especifiquen les taules-dimensions a que es referiran i com estaran enllaçades amb instruccions JOIN externes, ja que no es vol descartar els registres, tot i que no tinguin camps informats<sup>20</sup>.

<sup>18</sup> El codi és indiferent, podria haver seguit el 0 perfectament.

<sup>19</sup> Derby no permet PL/SQL com Oracle o T-SQL com SQL Server, i per tant no es pot considerar una solució programàtica amb bucles, per exemple. S'hauria de programar una classe en Java i cridar-la des d'un procediment emmagatzemat.

<sup>20</sup> Les estructures CASE-WHEN permeten la selecció o no de determinats camps segons aquests siguin NULLs o no a la consulta.

Per al cas de la taula D\_TEMP, es fa l'ús de la expressió de conversió de tipus de dades CAST, de la funció d'extracció de sub-cadenes SUBSTR i de la funció de arrodoniment a l'alça (CEIL), que serveixen per a calcular a quin trimestre pertany una determinada data de venda.

```

insert into TFM.F_VENDES (FACTURA_LINEA, COD_EMPRESA, ID_COMERCIAL, COD_ARTICLE, ID_TEMPS, ID_CLIENT, COD_MAGATZEM, TIPUS_DADA, COD_CONCEPTE, COD_FORMA, QUANTITAT,
DESCOMPTE, IMPORT, COMISSIO, COST)
select * from (
select FACTURA || ' ' || LINEA as FACTURA LINEA,
TFM.D_EMPRESAS.COD_EMPRESA, TFM.D_COMERCIALS.ID_COMERCIAL,
CASE TFM.D_ARTICLES.COD_ARTICLE
WHEN is null THEN 'NO_ART'
else TFM.D_ARTICLES.COD_ARTICLE
end
as ARTICLE,
TFM.D_TEMPS.ID_TEMPS,
CASE TFM.D_CLIENTS.ID_CLIENT
WHEN is null then 1
else TFM.D_CLIENTS.ID_CLIENT
end
as CLIENT,
CASE TFM.D_MAGATZEMS.COD_MAGATZEM
WHEN is null THEN 'NO_AL'
else TFM.D_MAGATZEMS.COD_MAGATZEM
end
as MAGATZEM,
TFM.D_TIPUSDADES.TIPUS_DADA,
CASE TFM.D_CONCEPTES.COD_CONCEPTE
WHEN is null THEN 'NO_C'
else TFM.D_CONCEPTES.COD_CONCEPTE
end
as CONCEPTE,
CASE TFM.D_FORMESPAGAMENT.COD_FORMA
WHEN is null THEN 'NO_F'
else TFM.D_FORMESPAGAMENT.COD_FORMA
end
as FORMAPAGAMENT,
CAST(CSV.VENTAS.CANTIDAD as SMALLINT) as QUANTITAT, CAST(CSV.VENTAS.DESCUENTO as DECIMAL(5, 2)) as DESCOMPTE, CAST(CSV.VENTAS.IMPORTE as DECIMAL(7, 2))
as IMPORT, CAST(CSV.VENTAS.COMISION as DECIMAL(5, 2)) as COMISSIO, CAST(CSV.VENTAS.COSTE as DECIMAL(7, 2)) as COST
from CSV.VENTAS
left outer join TFM.D_EMPRESAS on CSV.VENTAS.EMPRESA = TFM.D_EMPRESAS.COD_EMPRESA
left outer join TFM.D_COMERCIALS on CSV.VENTAS.COMERCIAL = TFM.D_COMERCIALS.DNI
left outer join TFM.D_ARTICLES on CSV.VENTAS.ARTICULO = TFM.D_ARTICLES.COD_ARTICLE
left outer join TFM.D_CONCEPTES on CSV.VENTAS.CONCEPTO = TFM.D_CONCEPTES.COD_CONCEPTE
left outer join TFM.D_TEMPS on (SUBSTR(CSV.VENTAS.FECHA, 7, 4) = TFM.D_TEMPS."ANY" and SUBSTR(CSV.VENTAS.FECHA, 4, 2) = TFM.D_TEMPS.MES and
CAST(CAST(CAST(CEIL(CAST(SUBSTR(CSV.VENTAS.FECHA, 4, 2) AS INTEGER) / 3.0) AS DOUBLE) AS INTEGER) as CHAR) = TFM.D_TEMPS.TRIMESTRE)
left outer join TFM.D_CLIENTS on CSV.VENTAS.CLIENTE = TFM.D_CLIENTS.DNI
left outer join TFM.D_MAGATZEMS on CSV.VENTAS.ALMACEN = TFM.D_MAGATZEMS.COD_MAGATZEM
left outer join TFM.D_TIPUSDADES on CSV.VENTAS.TIPODATO = TFM.D_TIPUSDADES.TIPUS_DADA
left outer join TFM.D_FORMESPAGAMENT on CSV.VENTAS.FORMA_PAGO = TFM.D_FORMESPAGAMENT.COD_FORMA) as T;

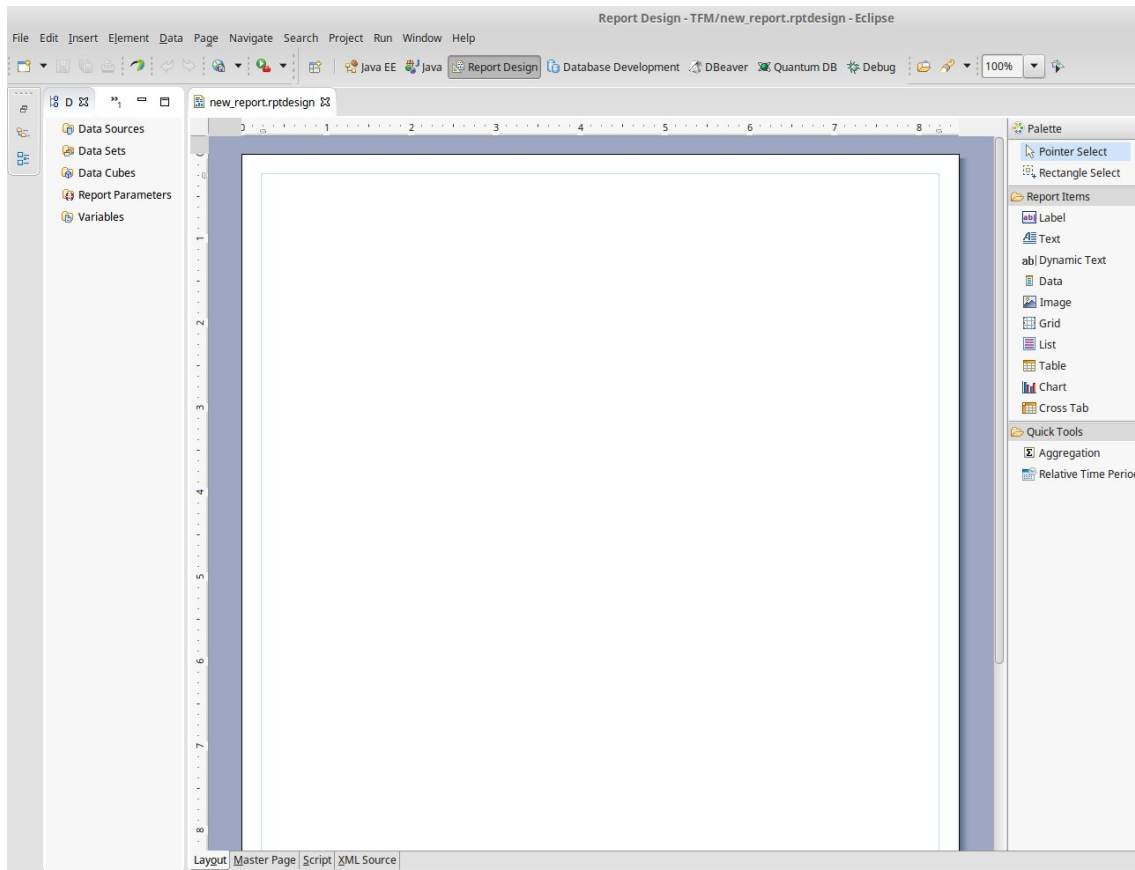
```

*Consulta 6: Càrrega de la taula de fets F\_VENDES*

## 7. Explotació del Data Mart

Un cop es disposa de la base de dades del Data Mart amb les dades carregades, és el torn de treure'n informació per respondre les preguntes plantejades als objectius d'aquest Treball, és a dir, és l'hora de l'explotació del sistema.

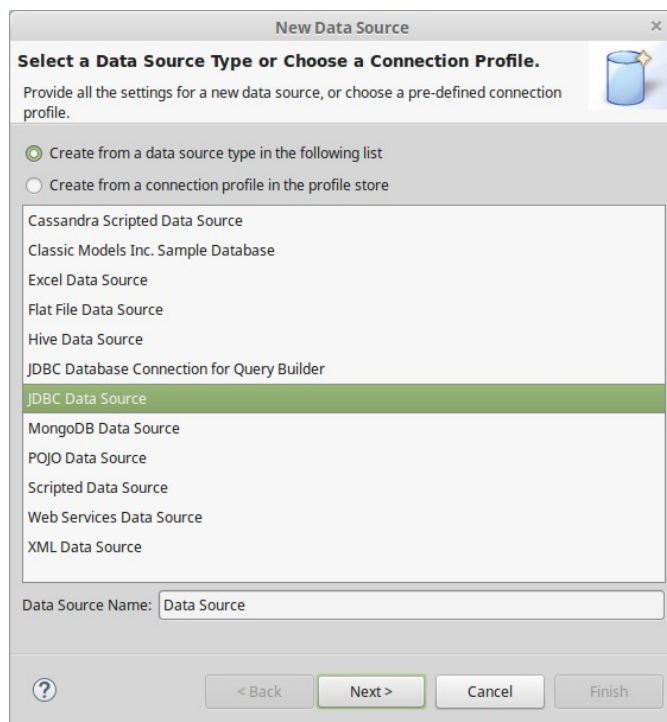
El disseny de l'informe comença a Eclipse BIRT amb File → New → Report. El nom del fitxer tindrà l'extensió .rptdesign.



*Il·lustració 8: Aspecte d'un nou informe en blanc obert a BIRT*

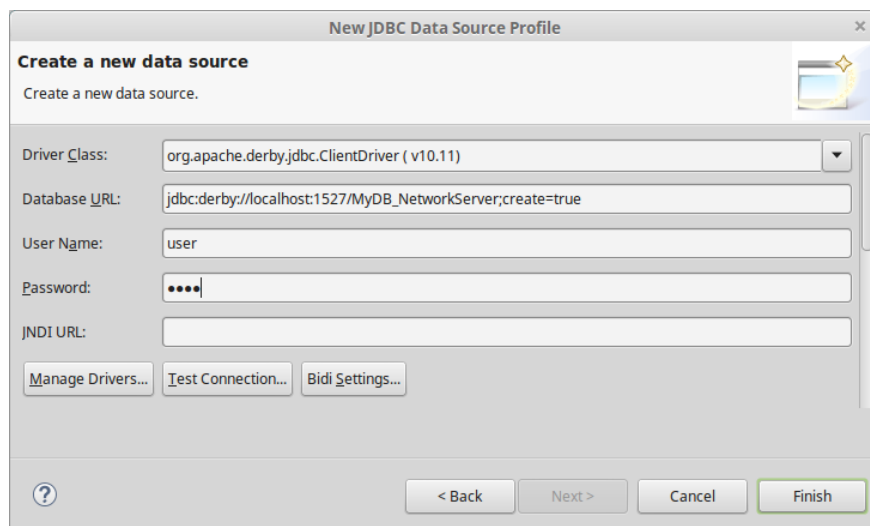
A continuació es defineix l'origen de dades (Data Source) que serà evidentment el Datamart recentment omplert. Per fer això, botó dret del mouse a Data Source i «New Data Source». Es selecciona «JDBC Data Source» com a tipus de font de dades<sup>21</sup>:

21 Es pot seleccionar també «JDBC Database Connection for Query Builder» on la consulta que es farà per al Data Set (subconjunt de dades) es pot definir de manera gràfica.



*Il·lustració 9: Selecció de tipus de font de dades*

A continuació s'especifiquen el *driver*<sup>22</sup>, la URL de la base de dades i les credencials necessàries:

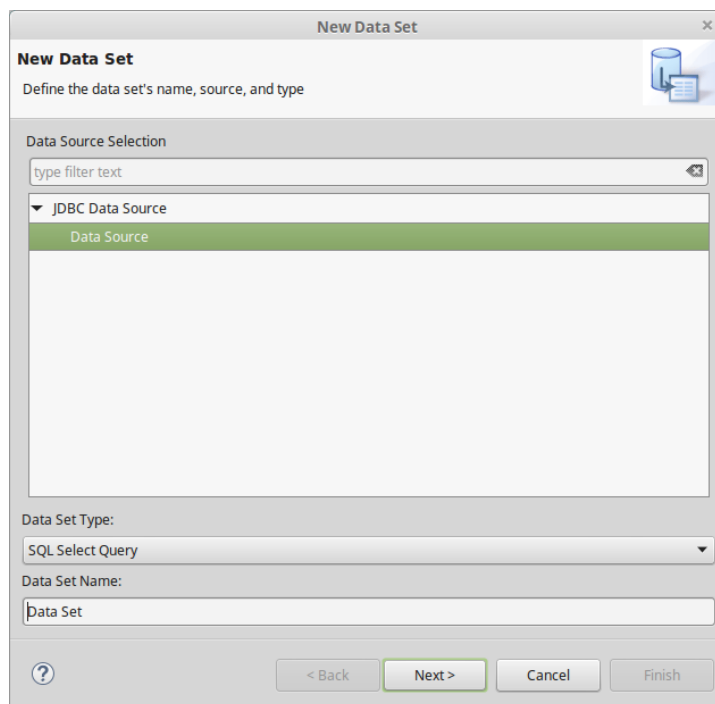


*Il·lustració 10: Configuració de la connexió a la base de dades*

<sup>22</sup> A Derby es poden utilitzar dos tipus de *drivers*. L'*Embedded Driver* només permet una instància de la JVM i és la manera habitual quan és distribuïda una aplicació que usa una base de dades, ja que va amb ella. L'altre tipus, *Network Client Driver*, permet més d'una instància i és la manera recomanable quan s'està desenvolupant.

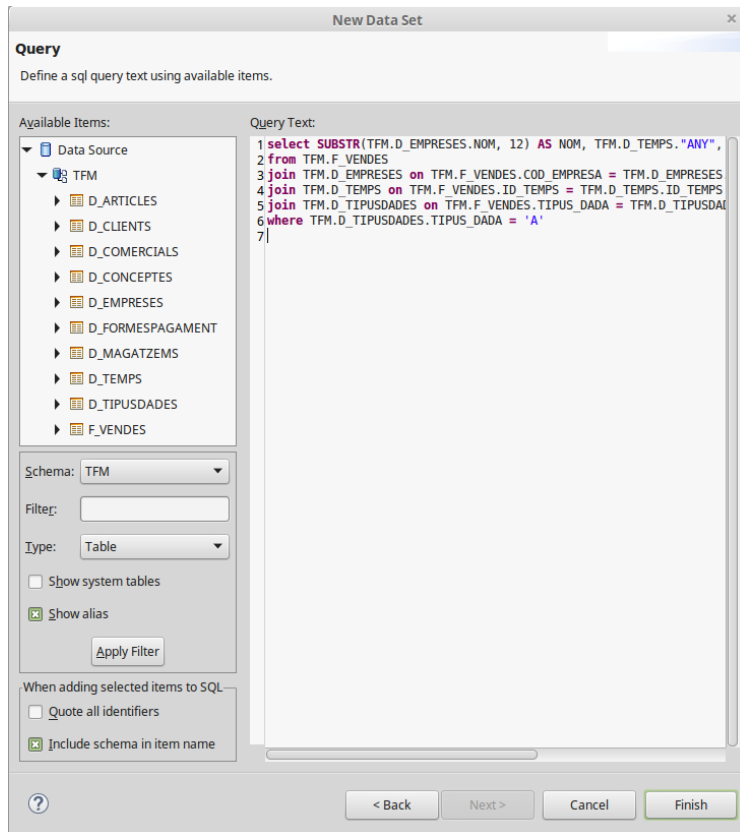


A continuació es crea el contenidor amb el subconjunt de dades que farà falta per a l'informe. Botó dret sobre Data Sets i «New Data Set». La font de dades per defecte apareix seleccionada, així com el tipus:



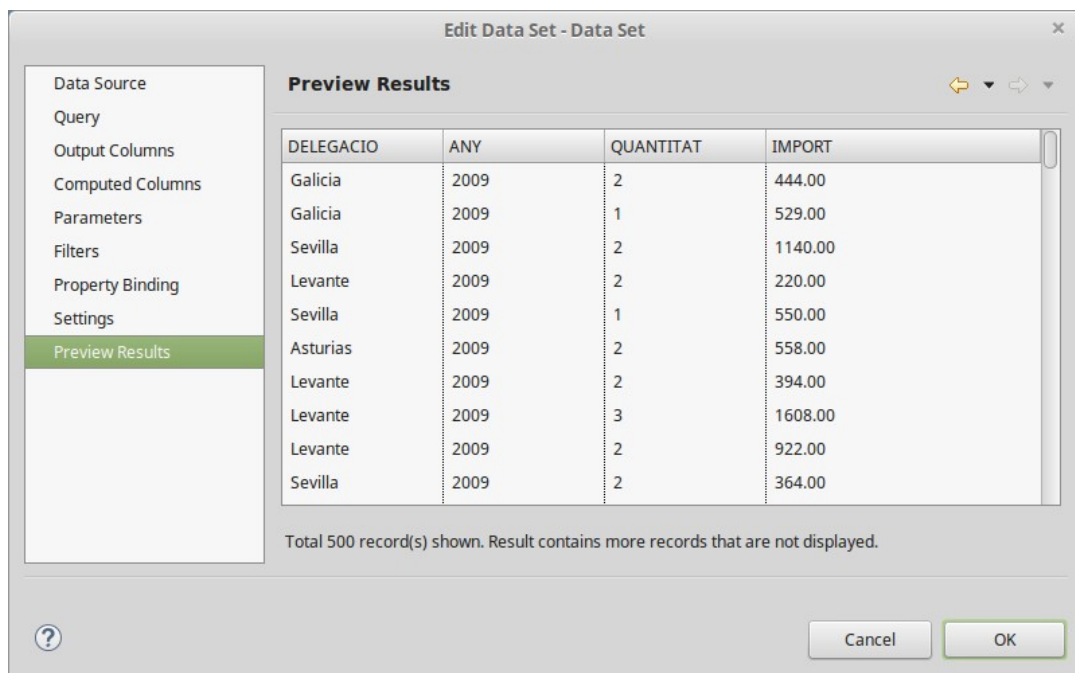
*Il·lustració 11: Diàleg selecció de font de dades*

Depenent del tipus d'informe, la consulta que proporcionarà les dades serà diferent ja que, per a cada pregunta, s'ha d'identificar les dimensions (taules) i les mètriques/mesures/indicadors que es necessiten per a cada informe. Per exemple:



*Il·lustració 12: Definit la consulta a la base de dades*

Un cop definida la consulta, es pot previsualitzar una mostra per veure com serien les dades:



*Il·lustració 13: Previsualitzant les dades del Data Mart*

A la part esquerra, el Data Explorer sempre mostrarà les estructures dels conjunts de dades creats i, opcionalment, els cubs de dades (una manera d'organització que pot resultar més efectiva a l'hora de definir els informes).

Als següents apartats es descriu el disseny i la implementació d'aquests informes. Per a cadascú s'especifica la pregunta que sol·licita aquest informe, el nom/s dels informes generats, les taules implicades del Data Mart, les mètriques, la consulta en SQL, l'estructura de l'informe, una visualització i finalment una conclusió sobre l'anàlisi.

## 7.1 Informe-Evolució de les vendes

**a. S'ha produït un decrement de les vendes en les seves delegacions respecte a les vendes d'un any enrere?** Necessiten un anàlisi que els mostri com han evolucionat les vendes en cadascuna de les seves delegacions, tant en unitats venudes com en import.

Informe: Nombre de vendes i imports totals per delegacions i anys (TFM\_Pregunta\_a.rptdesign).

Taules: Temps, Empreses, Tipus de dades (reals, A), Vendes

Mètriques: Quantitat, Import

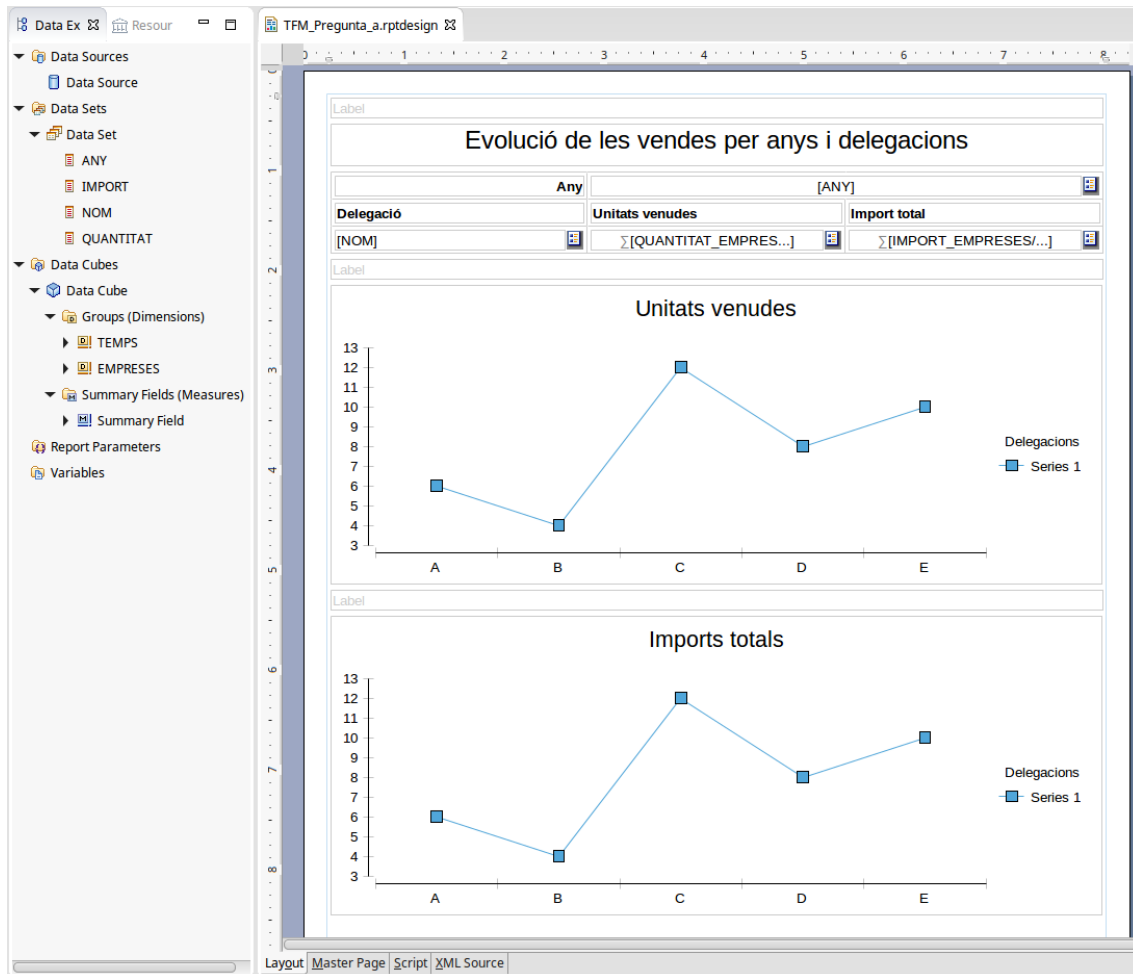
Consulta:

```
select SUBSTR(TFM.D_EMPRESSES.NOM, 12) AS NOM, TFM.D_TEMPS."ANY",
TFM.F_VENDES.QUANTITAT, TFM.F_VENDES.IMPORT
from TFM.F_VENDES
join TFM.D_EMPRESSES on TFM.F_VENDES.COD_EMPRESA = TFM.D_EMPRESSES.COD_EMPRESA
join TFM.D_TEMPS on TFM.F_VENDES.ID_TEMPS = TFM.D_TEMPS.ID_TEMPS
join TFM.D_TIPUSDADES on TFM.F_VENDES.TIPUS_DADA =
TFM.D_TIPUSDADES.TIPUS_DADA
where TFM.D_TIPUSDADES.TIPUS_DADA = 'A'
and CAST(TFM.D_TEMPS.MES as INTEGER) <= 5
```

*Consulta 7: Informe a, evolució de les vendes per anys i delegacions*

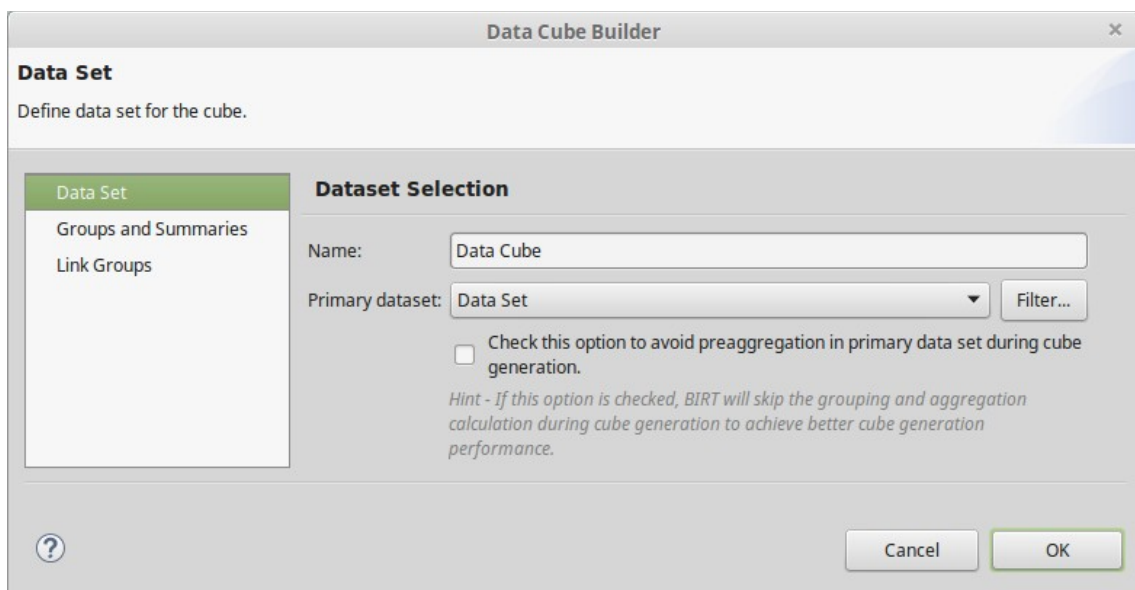
Només es compararan els 5 primers mesos de cada any, ja que al 2013 no tenim dades reals després de maig, per això el 5 a la clàusula WHERE. També hagués estat correcte fer l'anàlisi deixant de banda el 2013, veient l'evolució de les vendes segons el comportament dels anys anteriors sencers.

Estructura:



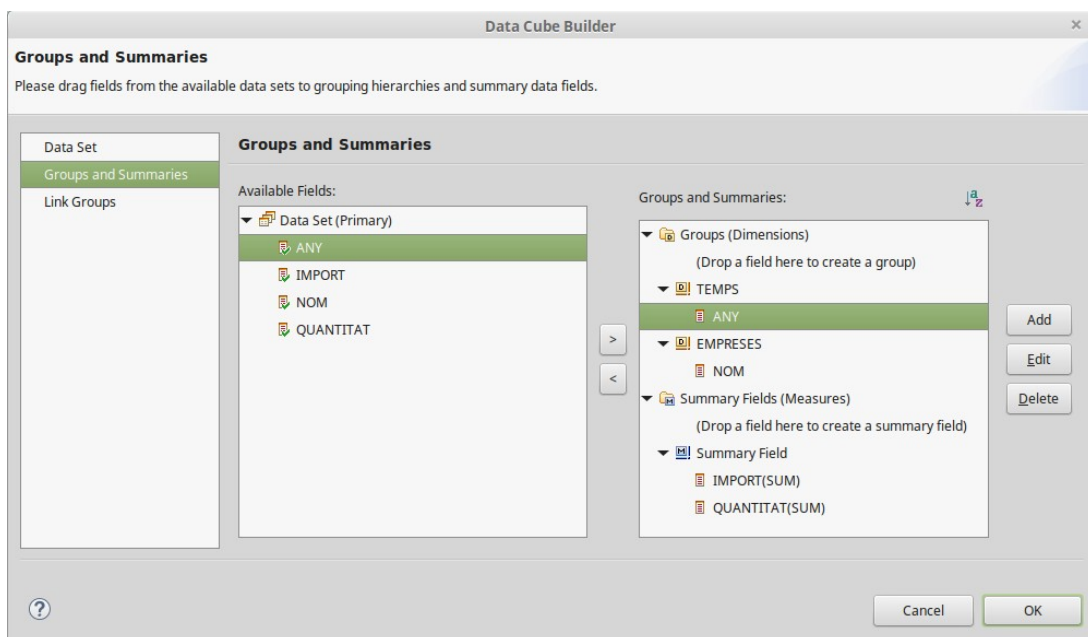
II-Iustració 14: Estructura informe a

L'element que mostrarà les dades tabulades és un CrossTab, molt indicat per mostrar dades en files i columnes, juntament amb les mètriques. El seu ús implica la creació d'un Data Cube (al Data Explorer de la esquerra, clic amb el botó dret del ratolí → New Data Cube):



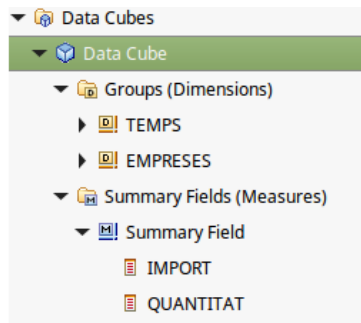
*Il·lustració 15: Definint un Data Cube (1)*

Es selecciona el Data Set relacionat i es constitueixen els grups o dimensions arrossegant els camps del Data Set seleccionat abans. Si tinguéssim més d'un Data Set es podrien relacionar-los a «Link Groups»:



*Il·lustració 16: Definint un Data Cube (2)*

Al Data Explorer es mostra com s'ha creat finalment el cub. Aquests camps són els que s'usaran al disseny de l'informe, arrossegant-los sobre ell:

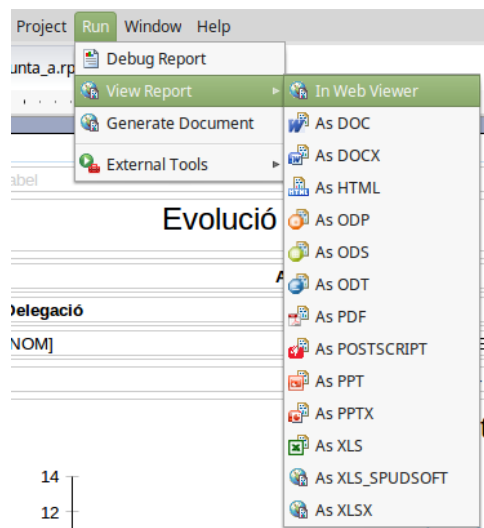


*Il·lustració 17: Data Explorer mostrant el cub de dades creat*

Per a les gràfiques s'ha usat l'element Chart, també de la secció «Report Items» a la paleta de la dreta.

#### Visualització:

BIRT permet visualitzar l'informe en diferents formats (PDF, DOC, etc.). Per defecte es mostra al navegador (Web Report):



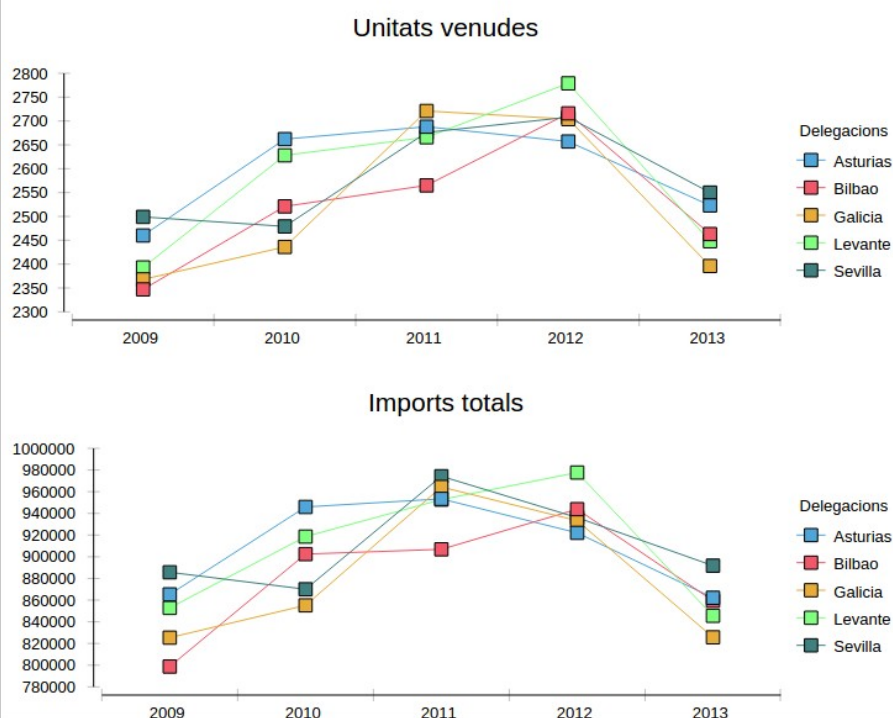
*Il·lustració 18: Format d'eixida a l'informe*

# TOTSALES

.selling you know.

## Evolució de les vendes per anys i delegacions

Any	2009		2010		2011		2012		2013	
Delegació	Unitats venudes	Import total	Unitats venudes	Import total	Unitats venudes	Import total	Unitats venudes	Import total	Unitats venudes	Import total
Asturias	2.460	865.295 €	2.662	945.932 €	2.688	953.382 €	2.657	922.058 €	2.523	862.165 €
Bilbao	2.347	798.751 €	2.521	902.449 €	2.565	906.883 €	2.716	943.885 €	2.463	859.429 €
Gàlcia	2.368	825.378 €	2.436	855.181 €	2.721	964.343 €	2.704	933.287 €	2.396	825.685 €
Levante	2.393	852.836 €	2.628	918.607 €	2.666	952.724 €	2.779	977.680 €	2.448	845.557 €
Sevilla	2.499	885.676 €	2.479	870.025 €	2.676	974.273 €	2.708	936.119 €	2.550	891.771 €



Il·lustració 19: informe a: evolució de les vendes per any i delegació

### Conclusió:

Es comprova amb aquest anàlisi de les dades reals (no de previsions o de pressuposts) que l'any 2013 va ser especialment dolent amb les vendes a les cinc delegacions, amb un apreciable decaïment al nombre d'unitats venudes, cosa que repercuteix directament als imports totals respectius. Per no disposar de dades més enllà de maig al 2013, no es pot dir si va acabar amb dèficit o amb superàvit en comparació amb els darrers anys per al mateix període.

## 7.2 Informe-Productes i famílies de productes més venuts

**b. Quins són els productes i famílies de productes més venuts? Detectar quins són els productes i famílies de productes amb més èxit de vendes (top 10 productes) i quins menys, segons la delegació i any tractat.**

Informe1: Nombre de vendes per article, per delegació i per any

Informe2: Nombre de vendes per famílies, per delegació i per any

Reports: TFM\_Pregunta\_b.rptdesign i TFM\_Pregunta\_b\_alt.rptdesign

Taules: Articles, Empreses, Temps, Vendes

Mètriques: Quantitat

Consulta: Per als articles més venuts:

```
select TFM.D_EMPRESSES.COD_EMPRESA, SUBSTR(TFM.D_EMPRESSES.NOM, 12) as
NOMEMPRESA, TFM.D_ARTICLES.NOM as NOMARTICLE, TFM.D_TEMPS."ANY",
SUM(TFM.F_VENDES.QUANTITAT) as QUANTITAT_TOTAL
  from TFM.F_VENDES
    join TFM.D_ARTICLES on TFM.D_ARTICLES.COD_ARTICLE =
TFM.F_VENDES.COD_ARTICLE
    join TFM.D_EMPRESSES on TFM.D_EMPRESSES.COD_EMPRESA =
TFM.F_VENDES.COD_EMPRESA
    join TFM.D_TEMPS on TFM.D_TEMPS.ID_TEMPS = TFM.F_VENDES.ID_TEMPS
    join TFM.D_TIPUSDADES on TFM.D_TIPUSDADES.TIPUS_DADA =
TFM.F_VENDES.TIPUS_DADA
  where TFM.D_ARTICLES.COD_ARTICLE <> 'NO_ART' and
TFM.D_TIPUSDADES.TIPUS_DADA = 'A'
    and CAST(TFM.D_TEMPS.MES as INTEGER) <= 5
  group by TFM.D_EMPRESSES.COD_EMPRESA, TFM.D_EMPRESSES.NOM,
TFM.D_ARTICLES.NOM, TFM.D_TEMPS."ANY"
  order by QUANTITAT_TOTAL desc
```

*Consulta 8: Informe b, productes més venuts*

Per a les famílies més venudes, la consulta és pràcticament igual:

```
select TFM.D_EMPRESSES.COD_EMPRESA, SUBSTR(TFM.D_EMPRESSES.NOM, 12) as
NOMEMPRESA, TFM.D_ARTICLES.FAMILIA as FAMILIA, TFM.D_TEMPS."ANY",
SUM(TFM.F_VENDES.QUANTITAT) as QUANTITAT_TOTAL
  from TFM.F_VENDES
    join TFM.D_ARTICLES on TFM.D_ARTICLES.COD_ARTICLE =
TFM.F_VENDES.COD_ARTICLE
    join TFM.D_EMPRESSES on TFM.D_EMPRESSES.COD_EMPRESA =
TFM.F_VENDES.COD_EMPRESA
    join TFM.D_TEMPS on TFM.D_TEMPS.ID_TEMPS = TFM.F_VENDES.ID_TEMPS
    join TFM.D_TIPUSDADES on TFM.D_TIPUSDADES.TIPUS_DADA =
TFM.F_VENDES.TIPUS_DADA
  where TFM.D_ARTICLES.COD_ARTICLE <> 'NO_ART' and
TFM.D_TIPUSDADES.TIPUS_DADA = 'A'
    and CAST(TFM.D_TEMPS.MES as INTEGER) <= 5
  group by TFM.D_EMPRESSES.COD_EMPRESA, TFM.D_EMPRESSES.NOM,
TFM.D_ARTICLES.FAMILIA, TFM.D_TEMPS."ANY"
  order by QUANTITAT_TOTAL desc
```

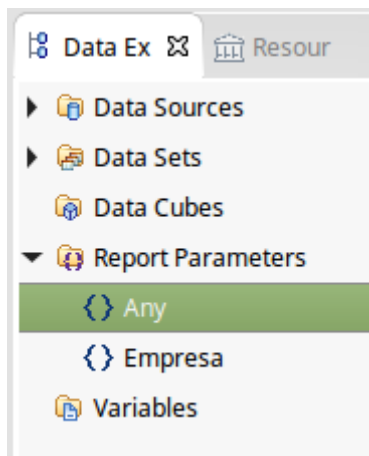
*Consulta 9: Informe b, famílies de productes més venudes*

Estructura:

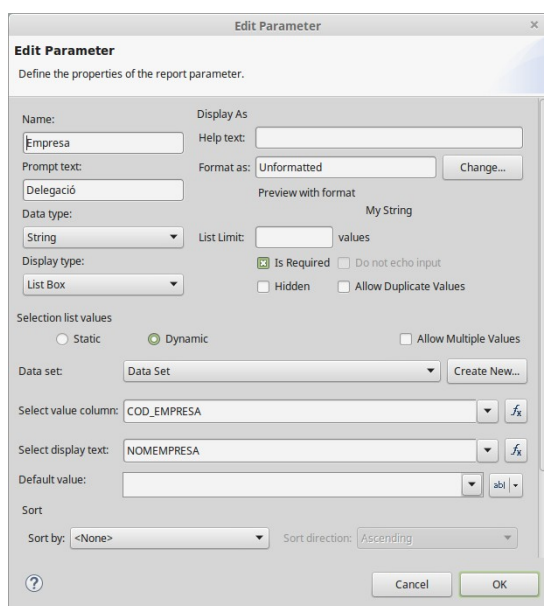
Es poden adoptar dos tipus d'estructures: la primera opció és la de presentar tota la informació per a cada any i delegació, de manera semblant a l'informe anterior:



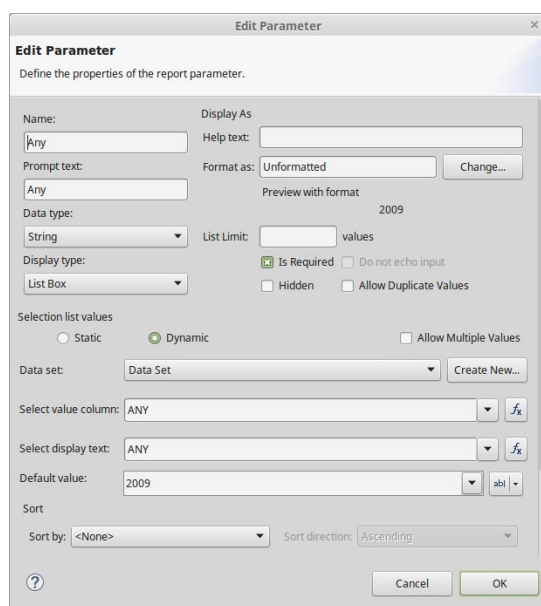




*Il·lustració 22:  
Paràmetres als informes*

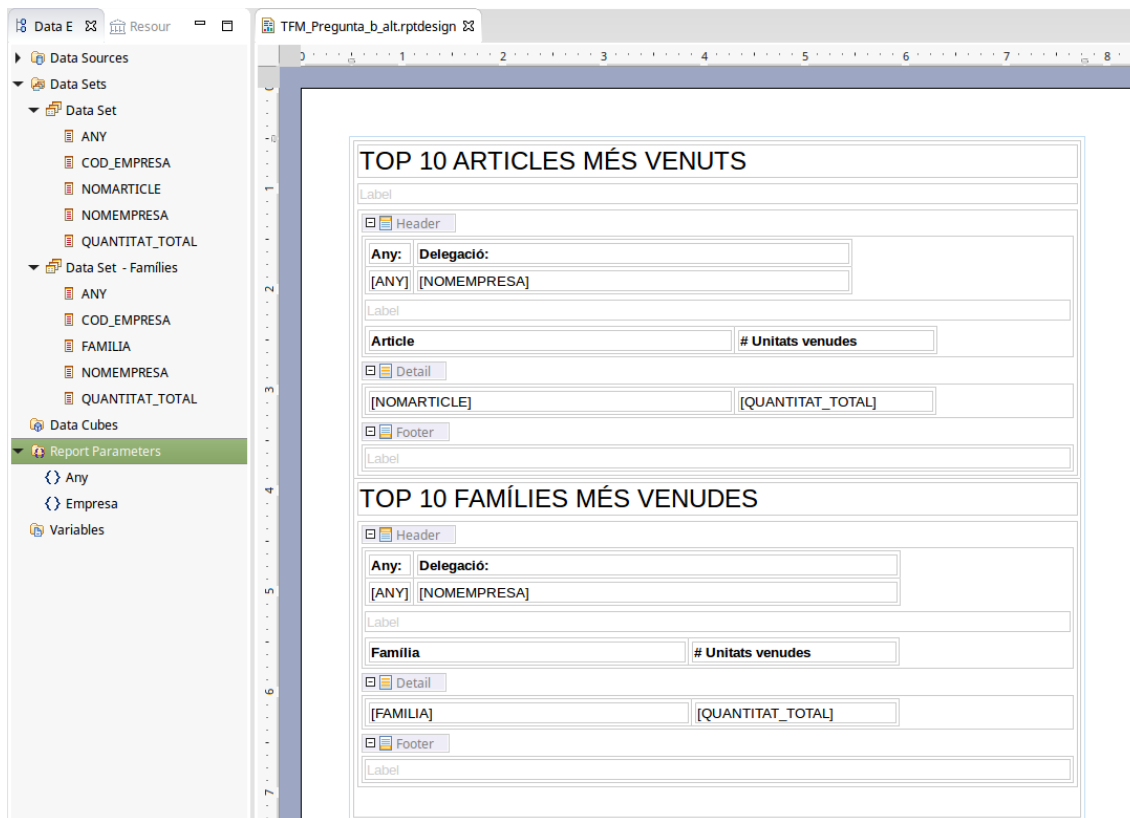


*Il·lustració 23: Definició del paràmetre  
Empresa*



*Il·lustració 24: Definició del paràmetre  
Any*

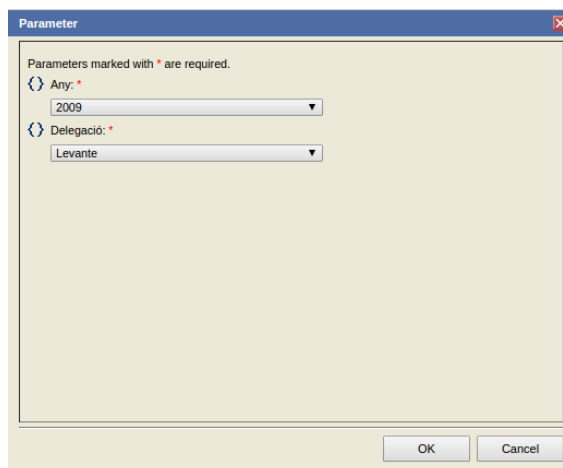
L'estructura alternativa seria:



*Il·lustració 25: Disseny de l'informe dels articles més venuts*

Visualització:

Abans de mostrar els resultats, s'han de proporcionar els valors de l'any i la delegació:



*Il·lustració 26: Diàleg de selecció de paràmetres*

BIRT Report Viewer

Showing page 1 of 1

Go to page:

# TOTSALES

.selling you know.

## TOP 10 ARTICLES MÉS VENUTS

Any: 2009 Delegació: Levante

Article	# Unitats venudes
Aspiradora ufesa AC 6230 Bogge Potencia	76
Frogorífic Balay 3KF17861	46
Càmera digital Nikon 1 J1 Rojo 10mm	40
Càmera digital Nikon D3200 Cuerpo	39
Aspiradora de mano Electrolux ZB2908W	37
Barbacoa elèctrica Philips HD 4418/20	35
Cafetera Fagor CCA 15 R	34
Bateria BEEEM F3001.110	34
Barbacoa Severin PG 8523	34
Barbacoa Severin PG 8521	33

## TOP 10 FAMÍLIES MÉS VENUEDES

Any: 2009 Delegació: Levante

Família	# Unitats venudes
ELECTRODOMESTICOS	1610
INFORMATICA	728
ELECTRONICA CONSUMO	551
CASA / JARDIN	500
FOTOGRAFIA / VIDEO / ÓPTICA	486
CONSOLAS Y VIDEOJUEGOS	218

(C) TOTSALES 2015

*Il·lustració 27: Informe b: articles i famílies més venuts*

### Conclusió:

Les vendes d'electrodomèstics pràcticament han dominat el mercat a totes les delegacions, inclús al 2013, on el nombre de vendes va ser molt inferior respecte a la resta d'anys. Dintre d'aquesta categoria, aspiradores i frigorífics van ser els articles més venuts. En aquest sentit, potser TOTSALES hauria de centrar-se en aquesta categoria per restaurar els beneficis a nivells passats.

### 7.3 Informe-Clients que els hi compren més (top 10)

**c. Quins són els clients top?** Volen fidelitzar aquells clients que els hi compren més (top 10 clients).

Informe: Nombre de vendes per client

Report: TFM\_Pregunta\_c.rptdesign

Taules: Articles, Empreses, Temps, Vendes

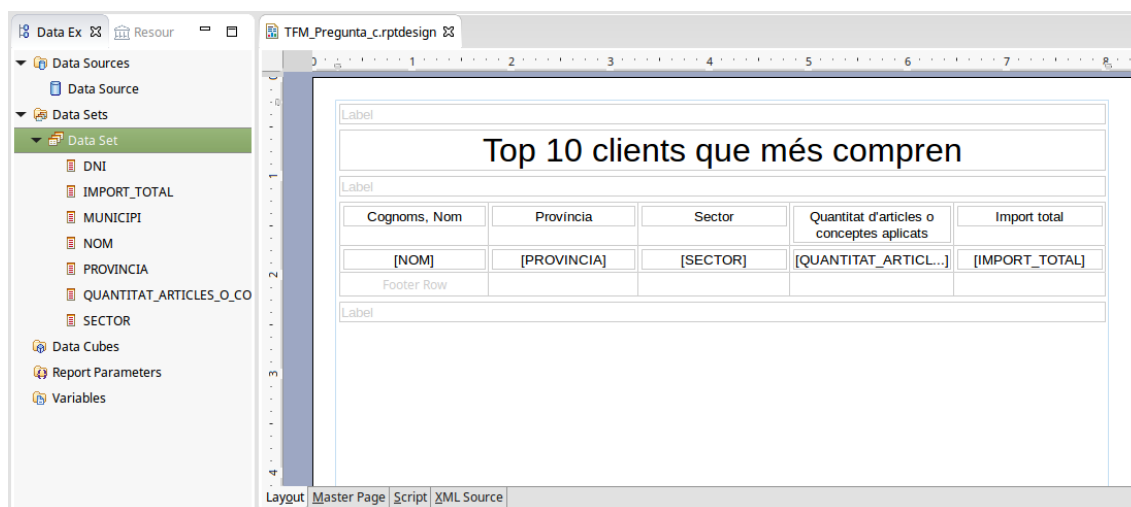
Mètriques: Quantitat, Import

Consulta:

```
select Clients.DNI, Clients.NOM, Clients.PROVINCIA, Clients.MUNICIPI,
Clients.SECTOR, SUM(Vendes.QUANTITAT) AS QUANTITAT_ARTICLES_0_CONCEPTES,
SUM(Vendes.IMPORT) AS IMPORT_TOTAL
from TFM.F_VENDES as Vendes
join TFM.D_CLIENTS as Clients on Vendes.ID_CLIENT = Clients.ID_CLIENT
join TFM.D_TIPUSDADES as TipusDades on Vendes.TIPUS_DADA =
TipusDades.TIPUS_DADA
where Vendes.TIPUS_DADA = 'A'
and CAST(TFM.D_TEMPS.MES as INTEGER) <= 5
group by Clients.DNI, Clients.NOM, Clients.PROVINCIA, Clients.MUNICIPI,
Clients.SECTOR
order by QUANTITAT_ARTICLES_0_CONCEPTES DESC
```

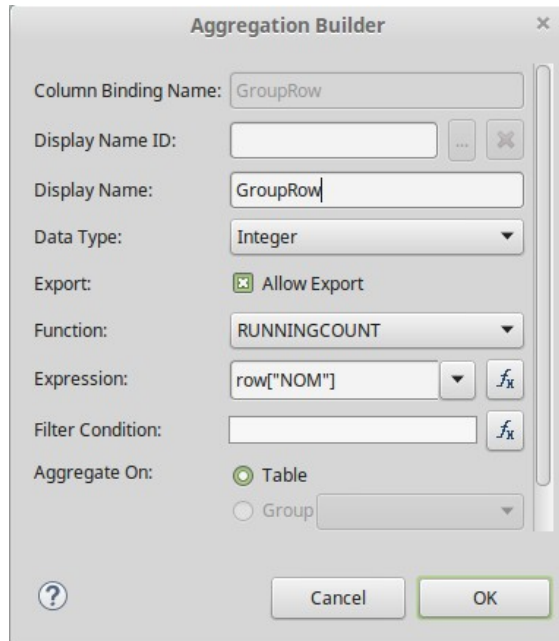
*Consulta 10: Informe c, clients top*

Estructura:

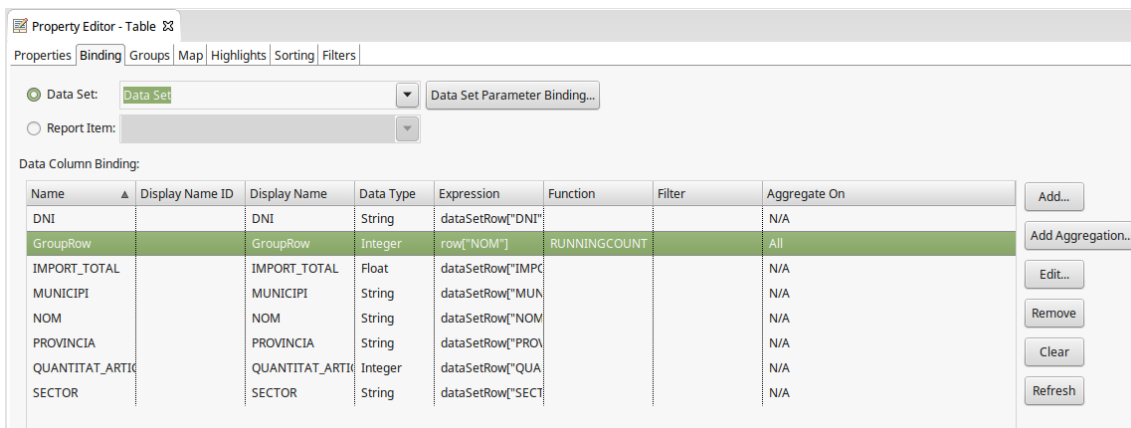


*Il·lustració 28: Disseny de l'informe dels clients que més compren*

Per obtenir un TOP N (com als informes anteriors) s'ha d'obtenir primer tots els registres i seguidament fer el truncament. Una manera de fer-ho és, a BIRT, a l'editor de propietats de l'element contenidor que mostrarà les dades (en aquest cas, una TABLE). A la pestanya Binding, es defineix una Aggregation amb la funció RUNNINGCOUNT amb el nom, per exemple, de GroupRow:

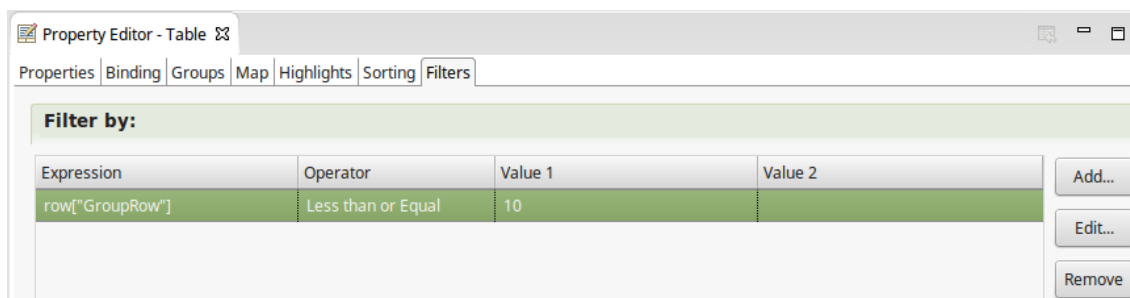


Il·lustració 29: Definint una agregació

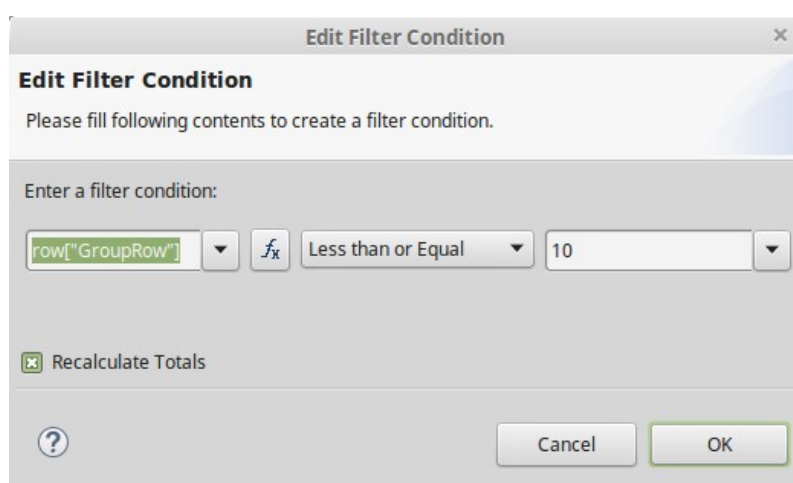


Il·lustració 30: Definint una agregació (2)

Per últim es defineix el filtre amb la condició «menys que o igual a»:



Il·lustració 31: Definint el filtre



Il·lustració 32: Definint el filtre (2)

Visualització:

**TOTSALES**  
..selling you know.

### Top 10 clients que més compren

Cognoms, Nom	Província	Sector	Quantitat d'articles o conceptes aplicats	Import total
MIRA MORENO, ISABEL	Cáceres	PEQUEÑO COMERCIO	2.183	772.129,00 €
RIBE PIE, JORDI	Avila	PEQUEÑO COMERCIO	1.959	686.618,00 €
CUEVAS- CAMACHO, M.JOSE	Barcelona	HOTELES	1.903	654.454,00 €
Nuria Torres Suárez	Cuenca	PEQUEÑO COMERCIO	1.885	656.406,00 €
GASCO CORRAL, CARLOS JAVIER	Tarragona	DISTRIBUIDORES	1.854	645.398,00 €
SAEZ PELAYO, JOSEFA	Vizcaya	DISTRIBUIDORES	1.839	645.564,00 €
HIGUERAS RODRIGUEZ, ANGUS	Teruel	MAYORISTAS	1.802	627.333,00 €
ROY BELLO, LORENZO	Navarra	SERVICIOS	1.787	613.186,00 €
VELASCO MARTIN, JUANA	Huelva	PEQUEÑO COMERCIO	1.768	601.266,00 €
ALTES BALANYA, MIREIA	Guadalajara	PEQUEÑO COMERCIO	1.753	613.370,00 €

(C) TOTSALES 2015

Il·lustració 33: Informe c: top 10 clients més fidels

## Conclusió:

Els petits comerços són els clients que més quantitats d'articles o conceptes han demanat durant el període 2009-2013 (dades reals) amb el 50% de presència, amb una lleugera penetració del sector hotelier. Són aquests clients els que més i millor tractament haurien de rebre per part de TOTSALLES.

### 7.4 Informe-Distribució de vendes segons la zona del client

**d. Quina és la distribució de les vendes segons la zona del client?** Volen conèixer des de quines zones els hi compren més i des de quines menys.

Informe1: Nombre de vendes per zona de client (més, províncies i municipis)

Informe2: Nombre de vendes per zona de client (menys, províncies i municipis)

Report: TFM\_Pregunta\_d.rptdesign

Taules: Articles, Empreses, Temps, Vendes

Mètriques: Quantitat, Import

Consulta:

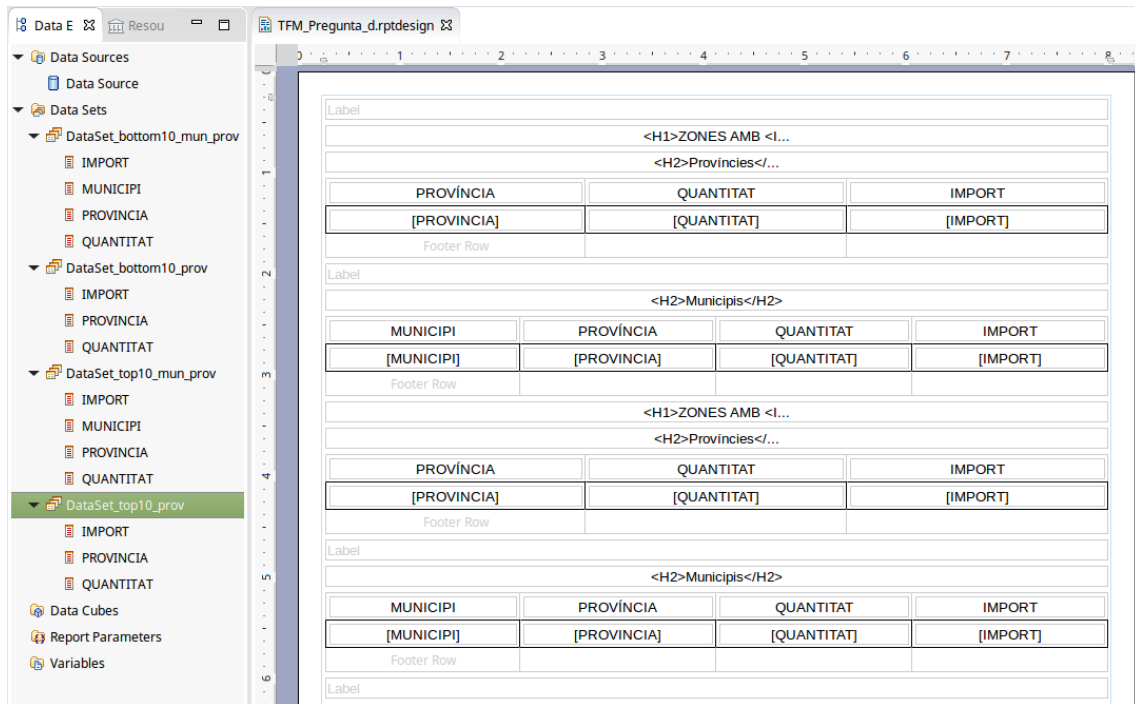
```
select Clients.PROVINCIA, SUM(Vendes.QUANTITAT) as QUANTITAT,
SUM(Vendes.IMPORT) as IMPORT
from TFM.F_VENDES as Vendes
join TFM.D_CLIENTS as Clients on Vendes.ID_CLIENT = Clients.ID_CLIENT
join TFM.D_TIPUSDADES as TipusDades on Vendes.TIPUS_DADA =
TipusDades.TIPUS_DADA
where Vendes.TIPUS_DADA = 'A'
group by Clients.PROVINCIA
order by QUANTITAT desc
```

*Consulta 11: Informe d, distribució de vendes segons zona del client*

La resta de consultes és molt similar. L'ordre serà descendent si es busca els que més han venut i ascendent si es busca els que menys.

Estructura:





*Il·lustració 34: Disseny de l'informe de les zones amb més i menys vendes*

Visualització:

Showing page 1 of 2 Go to page:

# TOTSALLES

.selling you know.

## ZONES AMB MÉS VENDES (període 2009-2013)

### Províncies

PROVÍNCIA	QUANTITAT	IMPORT
Girona	8.623	3.001.702,00 €
Cáceres	7.857	2.767.468,00 €
Rioja (La)	7.298	2.554.775,00 €
Zaragoza	6.721	2.353.677,00 €
Barcelona	6.548	2.299.463,00 €
Burgos	5.995	2.102.164,00 €
Guadalajara	5.923	2.071.196,00 €
Valencia/València	5.869	2.064.746,00 €
Zamora	5.595	1.957.765,00 €
Valladolid	5.388	1.882.966,00 €

### Municipis

MUNICIPI	PROVÍNCIA	QUANTITAT	IMPORT
Solana de Rioalmar	Avila	3.308	1.164.195,00 €
Bellaguarda	Lleida	3.226	1.108.740,00 €
San Vicente de la Sonsierra	Rioja (La)	3.167	1.108.238,00 €
Badia del Vallès	Barcelona	3.146	1.090.295,00 €
Tremedal de Tormes	Salamanca	3.134	1.110.271,00 €
Torrubia	Guadalajara	3.065	1.068.803,00 €
Herrera de Alcántara	Cáceres	3.063	1.083.207,00 €
Montaverner	Valencia/València	3.061	1.079.066,00 €
Espolla	Girona	3.019	1.058.984,00 €
Gallardos (Los)	Almeria	2.886	1.002.670,00 €

(C) TOTSALLES 2015

*Il·lustració 35: Informe d1: zones amb més vendes*

La visualització Web permet la paginació de diferents zones de l'informe. Així, es pot separa l'informe de les zones amb més vendes de les que menys:

TOTSALLES .selling you know.			
ZONES AMB MENYS VENDES (període 2009-2013)			
Províncies			
PROVÍNCIA	QUANTITAT	IMPORT	
Castellón/Castelló	1.242	440.986,00 €	
Córdoba	1.296	461.148,00 €	
Cantabria	1.402	495.663,00 €	
Soria provincia	1.500	509.965,00 €	
Huesca	1.503	526.356,00 €	
Balears (Illes)	1.531	527.714,00 €	
Sevilla	1.606	562.795,00 €	
Jaén Provincia	1.648	568.960,00 €	
Toledo	1.662	588.907,00 €	
Asturias	1.683	586.573,00 €	
Municipis			
MUNICIPI	PROVÍNCIA	QUANTITAT	IMPORT
Monesterio	Badajoz	975	330.319,00 €
Valgañón	Rioja (La)	1.126	405.317,00 €
Navarredonda y San Mamés	Madrid	1.140	404.940,00 €
Valdestillas	Valladolid	1.162	403.442,00 €
Fresno de Caracena	Soria	1.211	421.495,00 €
Aranda de Moncayo	Zaragoza	1.215	436.102,00 €
Lumpiaque	Zaragoza	1.240	430.285,00 €
Pina de Montalgrao	Castellón/Castelló	1.242	440.986,00 €
Alcubilla de las Peñas	Soria	1.255	441.290,00 €
Boñar	León	1.265	440.634,00 €

(C) TOTSALLES 2015

*Il·lustració 36: Informe d2: zones amb menys vendes*Conclusió:

Es veu que n'hi ha municipis que concentren un volum de vendes alt respecte a la província a la que pertanyen. Anàlogament passa amb els municipis amb menys vendes. Per tant, s'hauria d'optimitzar les rutes de lliuraments de comandes cap a aquestes zones per a assegurar el proveïment.

## 7.5 Informe-Evolució de les vendes en el temps

**e. Quina és la l'evolució de les vendes en funció del temps?** Necessiten tenir una jerarquia a la dimensió temps per poder analitzar les vendes en funció dels mesos, trimestres i anys.

Informe1: Dades de vendes per anys

Informe2: Dades de vendes per trimestres

Informe3: Dades de vendes per mesos

Report: TFM\_Pregunta\_e.rptdesign

Taules: Temps, Vendes

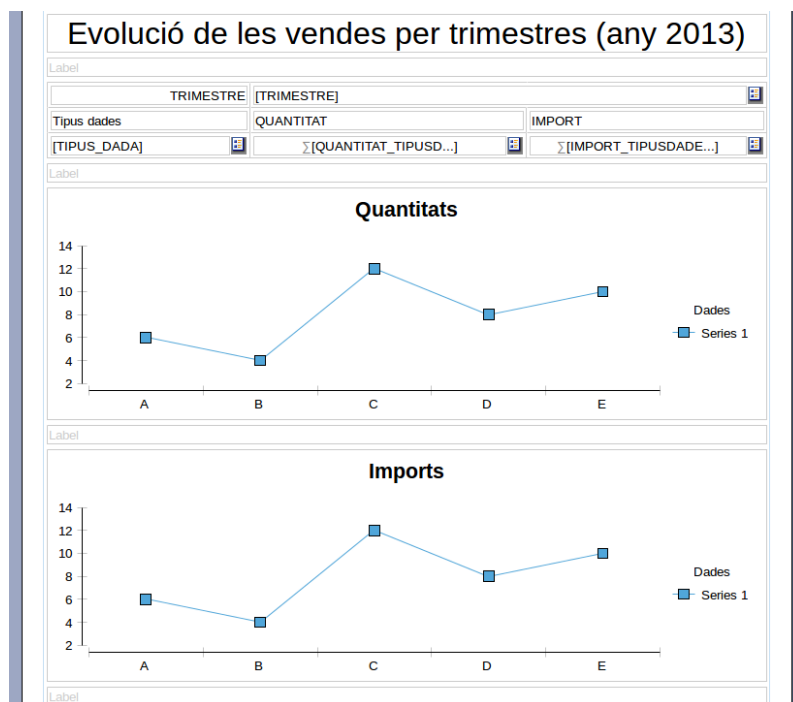
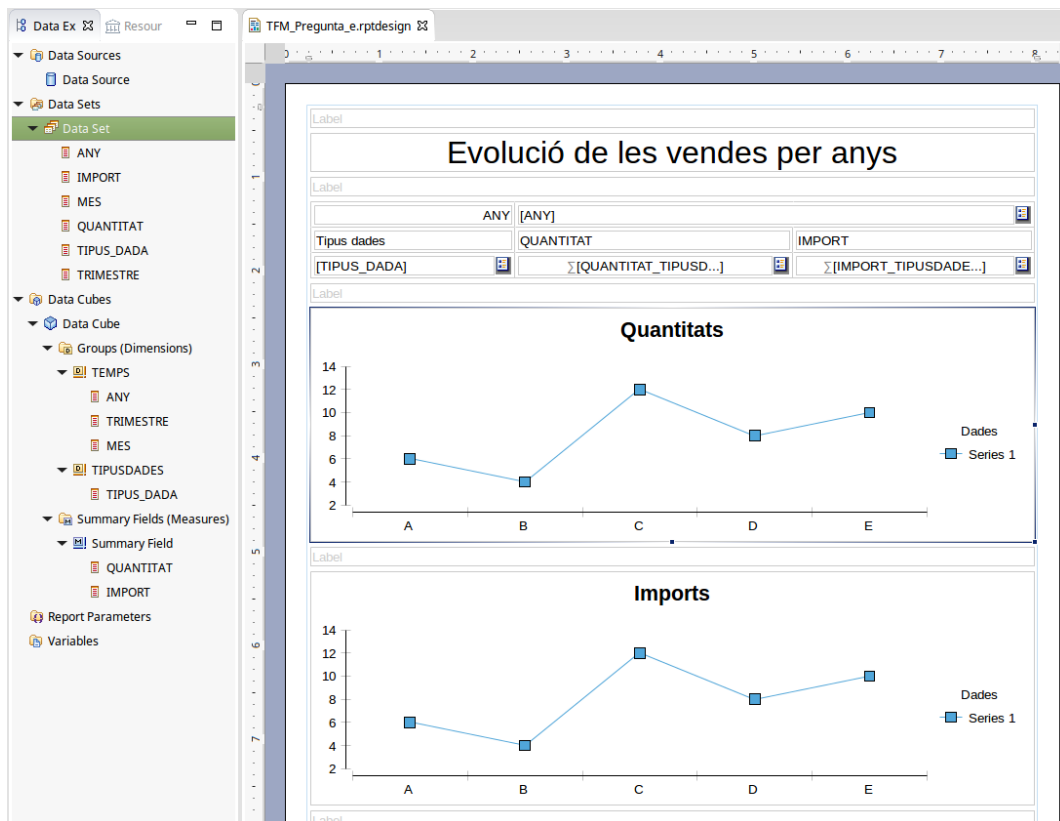
Mètriques: Quantitat, Import

Consulta:

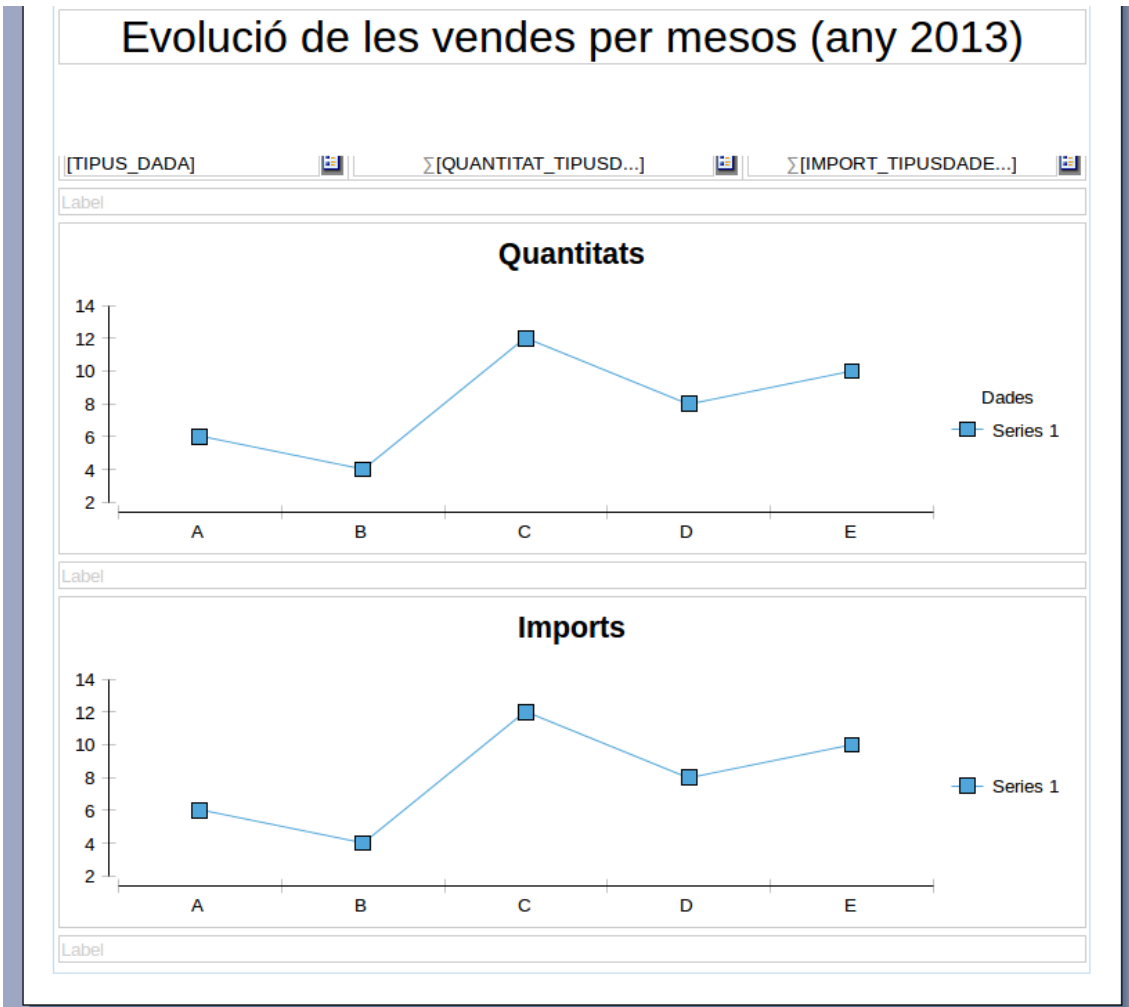
```
select
  CASE TFM.D_TIPUSDADES.TIPUS_DADA
    WHEN 'A' then 'Reals'
    WHEN 'B' then 'Pressupost'
    WHEN 'F' then 'Previsió'
  end
  as TIPUS_DADA,
  TFM.D_TEMPS."ANY",
  CASE TFM.D_TEMPS.TRIMESTRE
    WHEN '1' then '1 (Gen-Mar)'
    WHEN '2' then '2 (Abr-Jun)'
    WHEN '3' then '3 (Jul-Set)'
    WHEN '4' then '4 (Oct-Des)'
  end
  as TRIMESTRE,
  CASE TFM.D_TEMPS.MES
    WHEN '01' then '01-Gen'
    WHEN '02' then '02-Feb'
    WHEN '03' then '03-Mar'
    WHEN '04' then '04-Abr'
    WHEN '05' then '05-Mai'
    WHEN '06' then '06-Jun'
    WHEN '07' then '07-Jul'
    WHEN '08' then '08-Ago'
    WHEN '09' then '09-Set'
    WHEN '10' then '10-Oct'
    WHEN '11' then '11-Nov'
    WHEN '12' then '12-Des'
  end
  as MES,
  TFM.F_VENDES.QUANTITAT, TFM.F_VENDES.IMPORT
from TFM.F_VENDES
join TFM.D_TEMPS on TFM.F_VENDES.ID_TEMPS = TFM.D_TEMPS.ID_TEMPS
join TFM.D_TIPUSDADES on TFM.F_VENDES.TIPUS_DADA =
TFM.D_TIPUSDADES.TIPUS_DADA
where CAST(TFM.D_TEMPS.MES as INTEGER) <= 5
order by TFM.D_TEMPS.TRIMESTRE
```

*Consulta 12: Informe e, evolució temporal de les vendes*

Estructura:

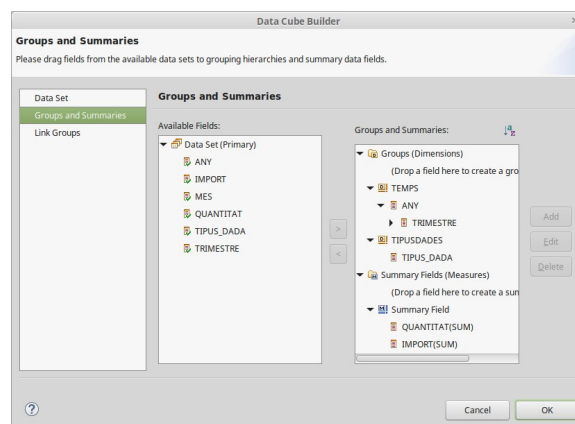


*Il·lustració 37: Disseny de l'informe de l'evolució temporal de les vendes*



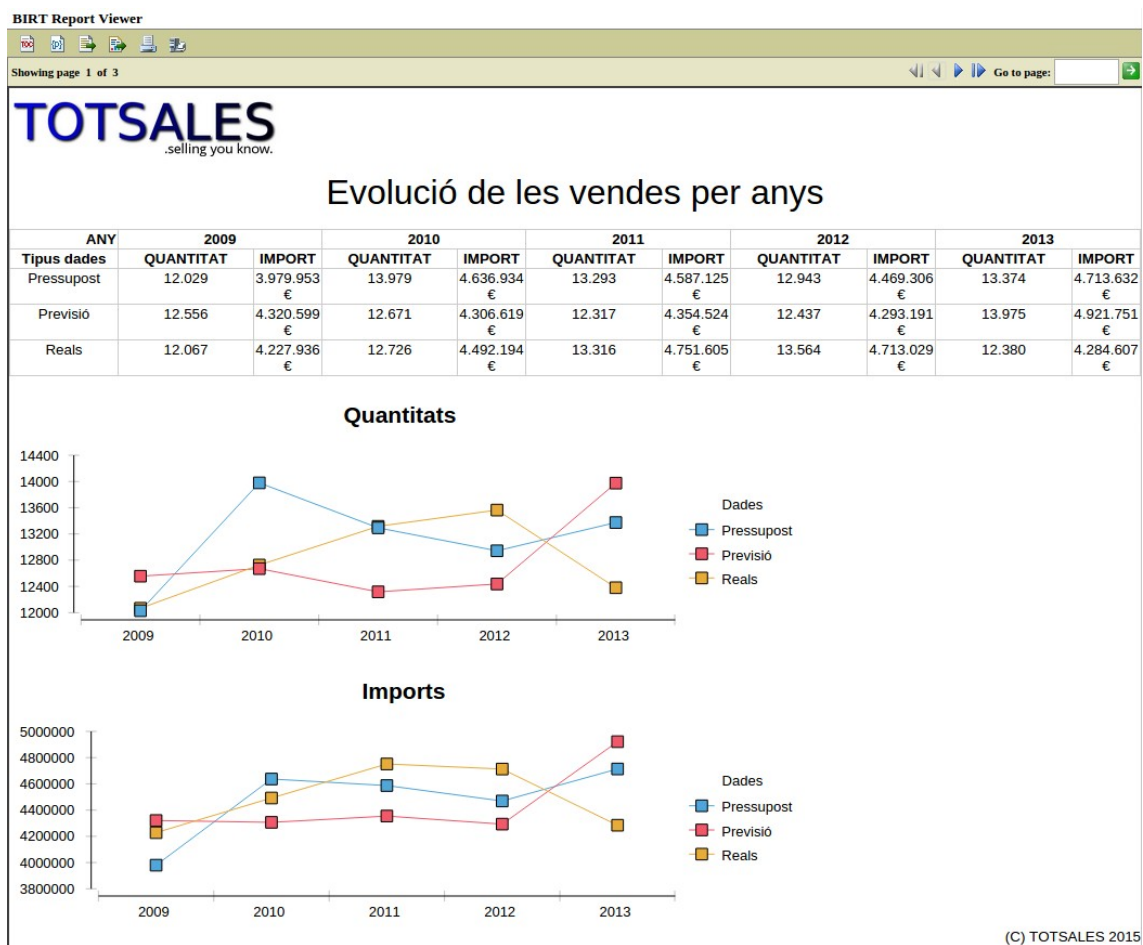
*II-lustració 38: Disseny de l'informe de l'evolució temporal de les vendes (2)*

S'ha usat de nou l'element Crosstab per representar les dades, el qual, com ja s'ha dit, implica la creació d'un cub de dades:



*II-lustració 39: Cub per a l'informe e*

## Visualització:



Il·lustració 40: Informe e1: evolució de les vendes per anys

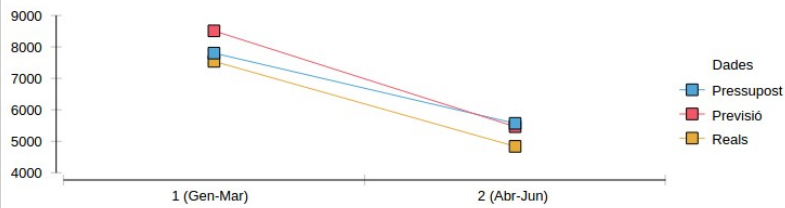
# TOTSALES

.selling you know.

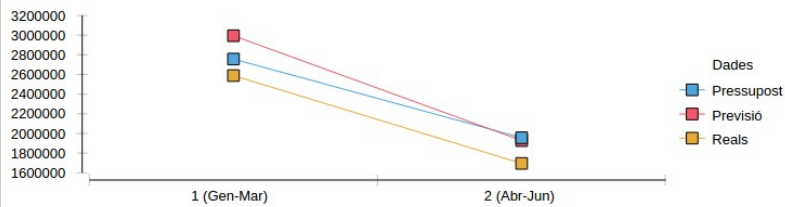
## Evolució de les vendes per trimestres (any 2013)

TRIMESTRE	1 (Gen-Mar)		2 (Abr-Jun)	
	QUANTITAT	IMPORT	QUANTITAT	IMPORT
Pressupost	7.803	2.756.833 €	5.571	1.956.799 €
Previsió	8.510	2.994.130 €	5.465	1.927.621 €
Reals	7.541	2.588.959 €	4.839	1.695.648 €

### Quantitats



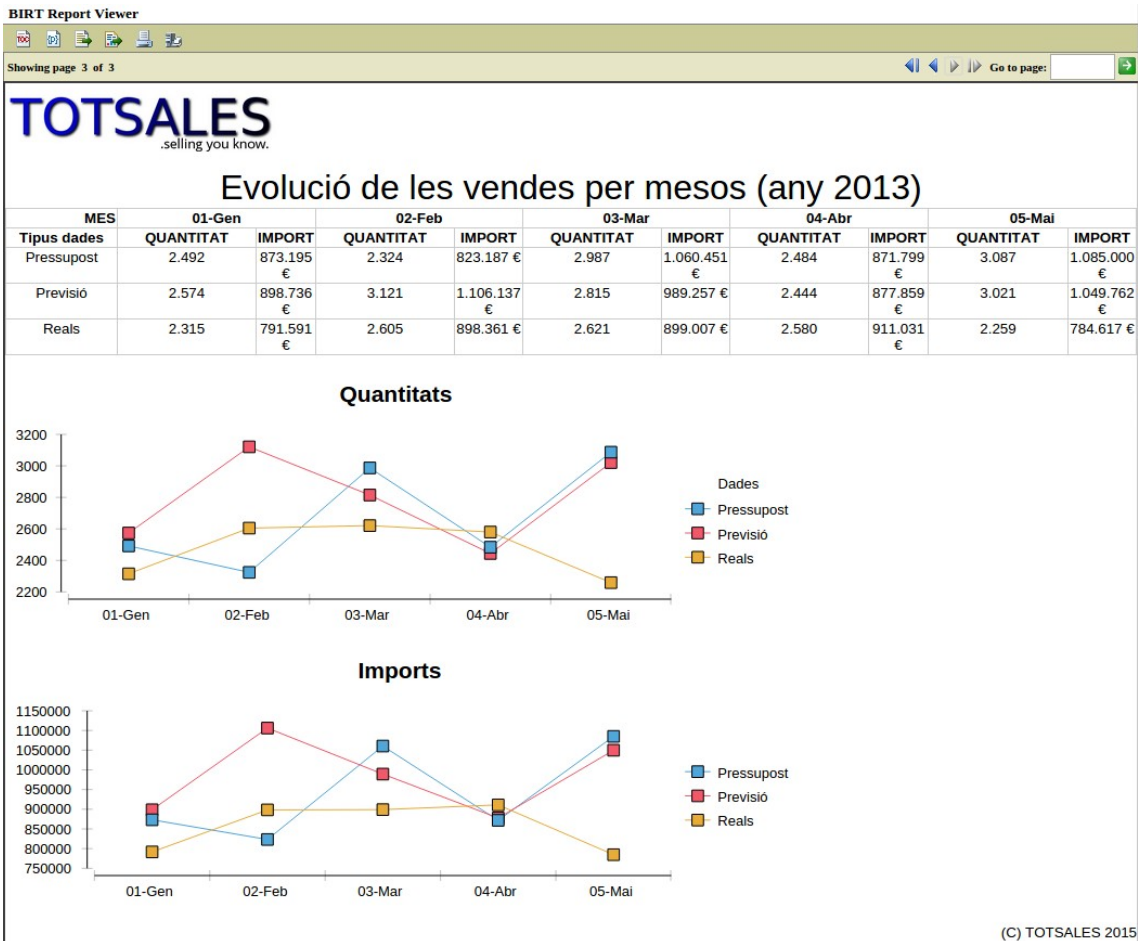
### Imports



(C) TOTSALES 2015

Il·lustració 41: Informe e2: evolució de les vendes per trimestres (any 2013)





Il·lustració 42: Informe e3: evolució de les vendes per mesos (any 2013)

### Conclusió:

Les dades reals (només disponibles fins a maig del 2013) mostren que la disparitat de les quantitats venudes i dels imports respecte a les dades de previsió i/o de pressupost començà gairebé a principis de l'any 2013. Ja a febrer la previsió que es va fer va ser molt diferent de la realitat. En aquest sentit, s'haurien de revisar els processos d'estimació de vendes. D'altra banda, es confirma que al mes de maig les vendes comencen a caure.

## 7.6 Informe-Comissions als comercials

**f. Quines són les comissions que s'han de liquidar als comercials? Volen conèixer les comissions que els comercials han de rebre per la seva feina. Ho necessiten tant en format mensual com anual.**

Informe1: Comissions segons el comercial per anys

Informe2: Comissions segons el comercial a l'any 2013 per mesos

Report: TFM\_Pregunta\_f.rptdesign

Taules: Comercials, Tipus de dades, Vendes

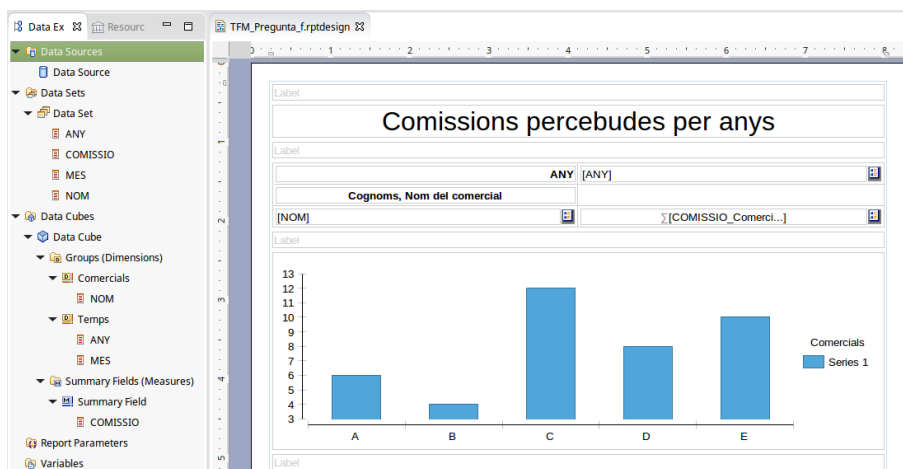
Mètriques: Comissió

Consulta:

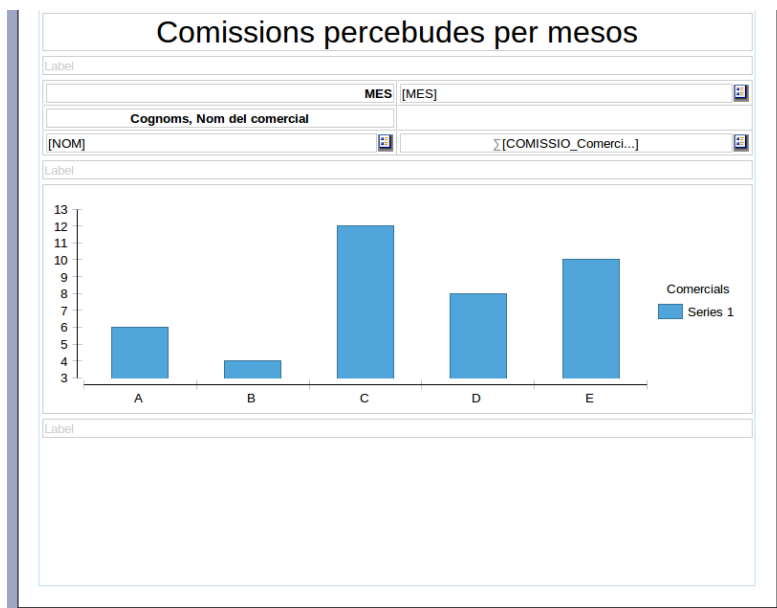
```
select TFM.D_COMERCIALS.NOM, TFM.D_TEMPS."ANY",
CASE TFM.D_TEMPS.MES
WHEN '01' then '01-Gen'
WHEN '02' then '02-Feb'
WHEN '03' then '03-Mar'
WHEN '04' then '04-Abr'
WHEN '05' then '05-Mai'
WHEN '06' then '06-Jun'
WHEN '07' then '07-Jul'
WHEN '08' then '08-Ago'
WHEN '09' then '09-Set'
WHEN '10' then '10-Oct'
WHEN '11' then '11-Nov'
WHEN '12' then '12-Des'
end
as MES,
TFM.F_VENDES.COMISSIO
from TFM.F_VENDES
join TFM.D_TEMPS on TFM.D_TEMPS.ID_TEMPS = TFM.F_VENDES.ID_TEMPS
join TFM.D_COMERCIALS on TFM.D_COMERCIALS.ID_COMERCIAL =
TFM.F_VENDES.ID_COMERCIAL
join TFM.D_TIPUSDADES on TFM.D_TIPUSDADES.TIPUS_DADA =
TFM.F_VENDES.TIPUS_DADA
WHERE TFM.F_VENDES.TIPUS_DADA = 'A'
and CAST(TFM.D_TEMPS.MES as INTEGER) <= 5
```

Consulta 13: Informe f, comissions

Estructura:

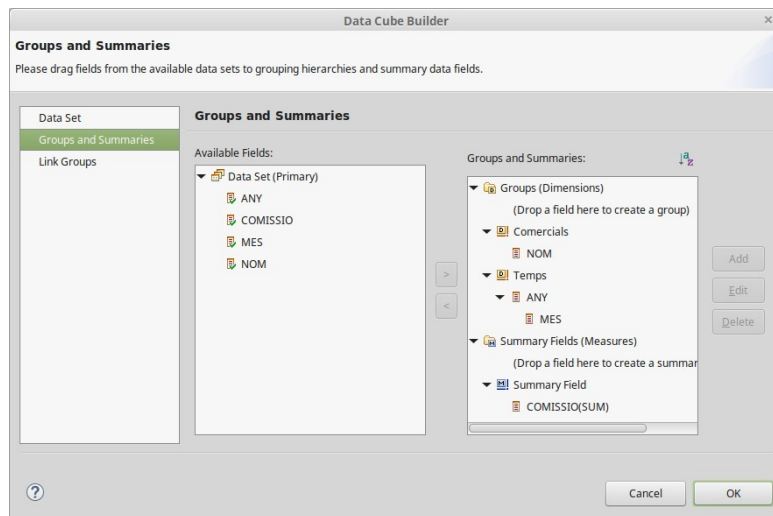


Il·lustració 43: Estructura de l'informe de les comissions



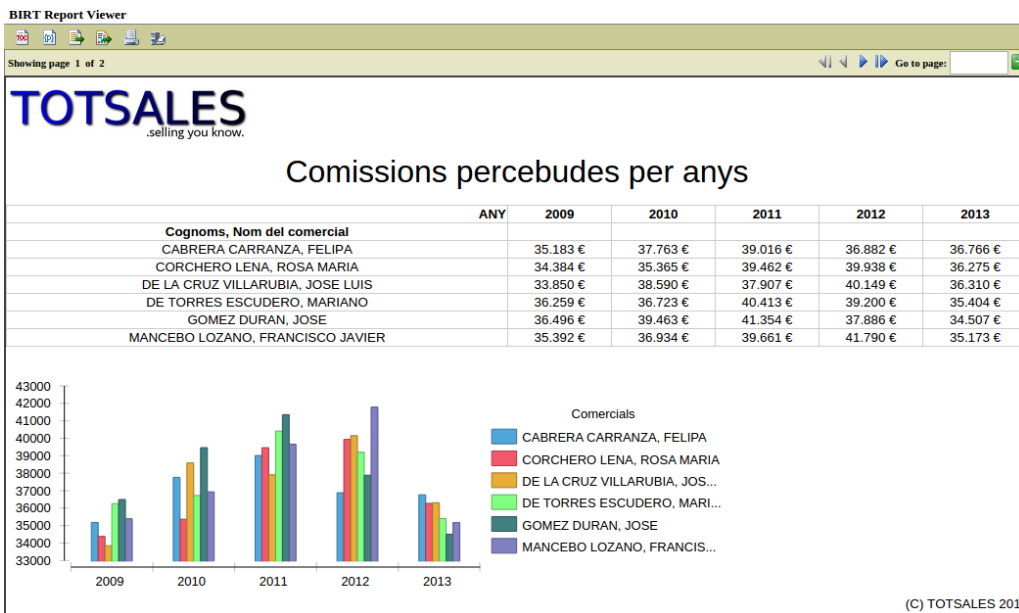
*Il·lustració 44: Disseny de l'informe de les comissions (2)*

El Data Cube creat és:

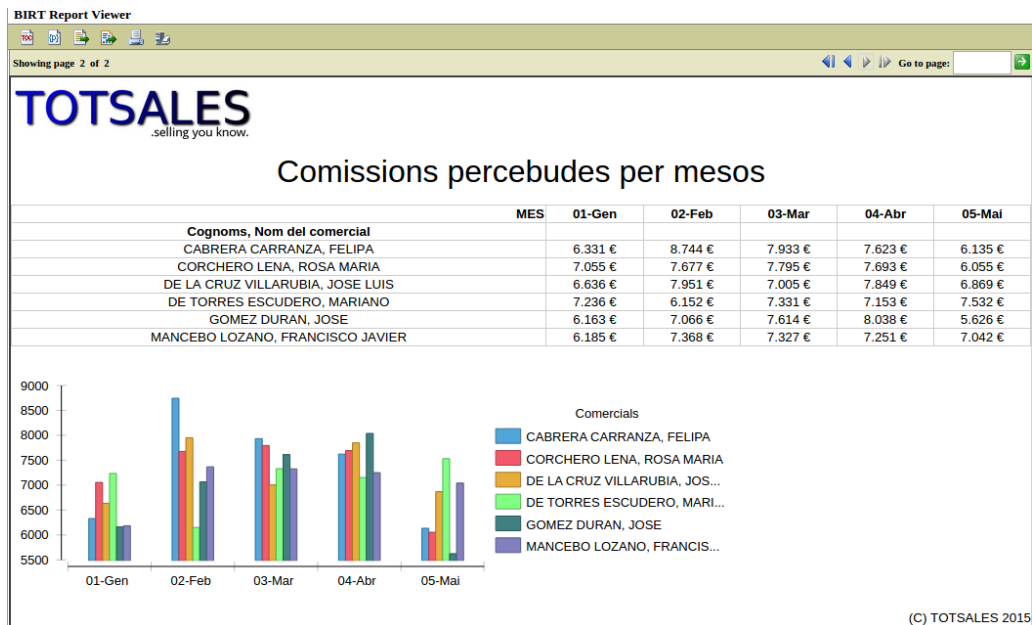


*Il·lustració 45: Cub de dades per a l'informe de les comissions*

Visualització:



Il·lustració 46: Informe f1: comissions percebudes per anys



Il·lustració 47: Informe f2: comissions percebudes per mesos

Conclusió:

En concordància amb els altres informes, les comissions disminueixen cap a l'any 2013. A aquest any, amb dades reals (fins a maig), es veuen que són molt desiguals, probablement per un desajust quan a la càrrega de feina durant els dos primers trimestres.

## 7.7 Informe-Marge anual

**g. Quin és el marge anual obtingut?** Es vol saber el marge anual per les famílies de productes, entès com la relació entre el import de la venda i el cost de la mateixa.

Informe: Diferències entre imports i costos de les vendes, per famílies de productes (articles)

Report: TFM\_Pregunta\_g.rptdesign

Taules: Articles, Temps, Tipus de dades, Vendes

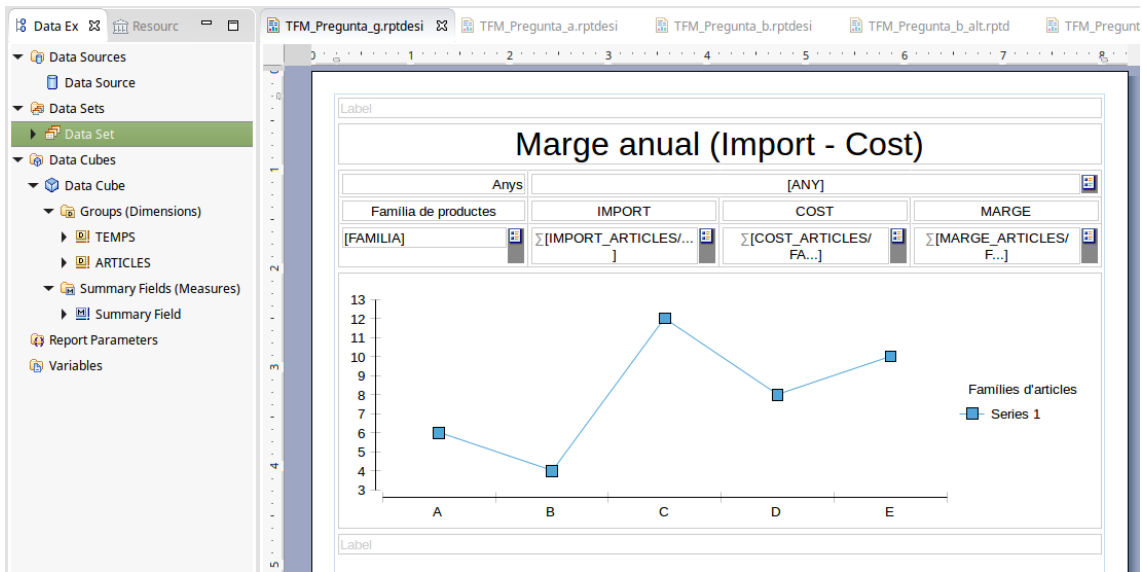
Mètriques: Import, Cost, Marge (Import - Cost)

Consulta:

```
select TFM.D_TEMPS."ANY",
       CASE TFM.D_ARTICLES.FAMILIA
       WHEN is null then 'Sense Família'
       ELSE TFM.D_ARTICLES.FAMILIA
       END
       as FAMILIA,
       TFM.F_VENDES.IMPORT, TFM.F_VENDES.COST
from TFM.F_VENDES
join TFM.D_ARTICLES on TFM.F_VENDES.COD_ARTICLE =
TFM.D_ARTICLES.COD_ARTICLE
join TFM.D_TEMPS on TFM.F_VENDES.ID_TEMPS = TFM.D_TEMPS.ID_TEMPS
join TFM.D_TIPUSDADES on TFM.F_VENDES.TIPUS_DADA =
TFM.D_TIPUSDADES.TIPUS_DADA
where TFM.D_TIPUSDADES.TIPUS_DADA = 'A'
and CAST(TFM.D_TEMPS.MES as INTEGER) <= 5
```

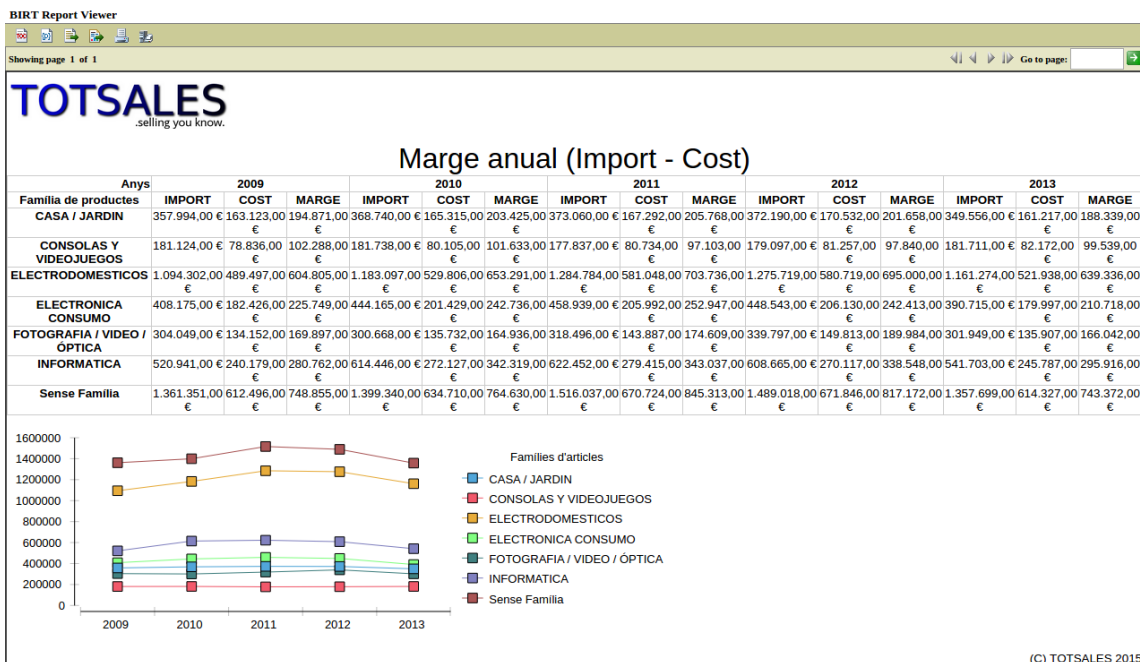
*Consulta 14: Informe g, marges anuals*

Estructura:



Il·lustració 48: Disseny de l'informe del marges anuals

Visualització:



Il·lustració 49: Informe dels marges anuals

Conclusió:

El marge és més gran a la família d'electrodomèstics, i per tant és més gran el benefici derivat. Les dades «sense família» probablement es refereixen a conceptes. Altres famílies com el de Consoles i videojocs no tenen un benefici tan clar com la primera. Tot i així, no s'aprecien marges negatius (tots són sobre zero).

## 8. Conclusions

S'ha presentat un conjunt d'informes que permetran als directius de TOTSALLES treure conclusions esclaridores sobre el que passa amb el rumb del negoci, i que poden facilitar l'adopció d'una o varies estratègies que permetin aconseguir més beneficis, ja que, com es demostra, les vendes a maig del 2013 estan baixant. L'obtenció d'aquests informes van constituir els objectius aquest Treball.

La realització d'aquest Treball ha suposat una bona base de coneixement per iniciar-se en l'analítica de negoci. S'han presentat eines informàtiques i s'han usat d'altres que permeten iniciar-se en aquest paradigma.

Menció especial pel que fa a l'ús del programari de codi obert, oferint nivells de qualitat bastant alts, tot i que, l'entorn Eclipse no s'hagi portat bé amb la targeta gràfica de l'estació de treball en algunes ocasions.

Pel que fa a la planificació, las tasques definides des del principi han respectat la temporització en general, si bé la quarta i darrera PAC, la més completa de totes, va ser també la que més improvisacions ha tingut degut a que s'havia de lliurar l'explotació dels informes juntament amb la memòria, sense poder esperar correccions del consultor respecte als primers.

Són dues les línies de treball futur considerades:

1. Definir nous informes, referents als conceptes i a les formes de pagament. Per exemple, formes de pagament per anys i zones o top 10 conceptes més aplicats, també per anys i zones.
2. Implementar el Data Mart en una altra plataforma de BI, com SpagoBI o Pentaho.



## 9. Glossari

**BI**, Business Intelligence, Intel·ligència o Analítica de Negoci. Conjunt de processos, metodologies i eines usat per una companyia amb la finalitat última de millorar el seu negoci (definició pròpia).

**BIRT**, Business Intelligence Reporting Tool, nom de la plataforma de Eclipse per a l'analítica de negoci.

**DM**, DataMart, subconjunt d'un DataWarehouse, només referit a una sola activitat del negoci d'una empresa. Aquest Treball tracta amb les Vendes de TOTSALLES.

**DW**, DataWarehouse (en català, Magatzem de Dades), base de dades dissenyada i implementada per donar suport generalment a una gran quantitat de dades procedents d'activitats de negoci (no només de les Vendes) d'una empresa (definició pròpia).

**ERD**, Entity-Relation Diagram (en català, Diagrama d'Entitats i Relacions, un tipus de diagrama de blocs usat al disseny del model conceptual d'un sistema, generalment una base de dades.

**ETL**, Extract-Transform-Load (en català, Extracció-Transformació-Càrrega), grup de processos previs a l'explotació d'un sistema de BI i que preparen les dades per a ser analitzades.

**PMBOK**, Project Management Body Of Knowledge, (en català, cos de coneixement per a l'administració de projectes), es tracta d'un conjunt de guies estandarditzades per a l'administració de projectes de tot tipus, no només informàtics.

**SGBD**, Sistema Gestor de Base de Dades, nom genèric de tota plataforma o programari que implementa una Base de Dades.

**SQL**, Structured Query Language, (en català, Llenguatge estructurat de consultes), és el llenguatge estàndard que s'usa per fer consultes a una base de dades.

**TIC**, Tecnologies de la Informació i Comunicació (en anglès IT), grup de tecnologies informàtiques

## 10. Bibliografia

[1] CompareBI.com, Febrer 2015

<http://www.comparebi.com/vendor-reviews/>

[2] TodoBI, Febrer 2015

<http://todobi.blogspot.com.es/>

[3] TFC – Construcció i explotació d'un magatzem de dades per a l'anàlisi d'informació immobiliària, Àlex Ullate Antón, 2012

[4] Trabajo Final de Grado - Análisis de herramientas BI en el mercado actual, Antonio Carlos Vera García, 2015

<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/40088/9/averagarTFG0115memoria.pdf>

[5] Recurs UOC – Sportsinsight.com: comercialización de información deportiva vía web , José Ángel Martín Carballo , PID\_00206519

# 11. Annexos

## 11.1 Notes d'instal·lació i configuració de l'entorn usat per al TFM.

Per a desenvolupar el nucli del Treball, s'ha usat i configurat el següent programari:

### - Java

- Descarregar l'última versió de Java: <https://www.java.com/en/>

### - Eclipse BIRT.

- Descarregar de <http://www.eclipse.org/birt/>. Descomprimir l'arxiu .ZIP a qualsevol ubicació, per exemple a /home/usuari en un entorn Linux, [C:\](#) a un entorn Windows. L'executable s'anomena eclipse o eclipse.exe.
- Dins de la plataforma, s'ha d'instal·lar la plataforma d'accés a base de dades DTP. Es troba a <https://eclipse.org/datatools/>, o mitjançant l'opció «Install New Software» (veure tutorial a <http://www.vogella.com/tutorials/EclipseDataToolsPlatform/article.html>).
- Després d'instal·lar el SGBD, afegir dintre de les Preferències de projecte el jar derbyclient.jar, o modificar el CLASSPATH per a que apunti a ell.

### - Apache Derby.

- Descarregar de [http://db.apache.org/derby/derby\\_downloads.html](http://db.apache.org/derby/derby_downloads.html). Seleccionar la versió lib que només conté els arxius .jar amb els binaris i descomprimir l'arxiu .ZIP o .tar.gz a qualsevol ubicació, igual que amb l'eclipse.
- Per a engegar el servidor, a la carpeta bin, executar startNetworkServer.

```
Terminal
joa@Vulcano ~ $ cd db-derby-10.11.1.1-bin/bin
joa@Vulcano ~/db-derby-10.11.1.1-bin/bin $ ./startNetworkServer &
[1] 3265
joa@Vulcano ~/db-derby-10.11.1.1-bin/bin $ Tue Jun 16 07:12:49 CEST 2015 : Se ha
instalado el gestor de seguridad con la política de seguridad de servidor básic
a.
Tue Jun 16 07:12:49 CEST 2015 : Servidor de red Apache Derby: Se ha iniciado 10.
11.1.1 - (1616546) y está listo para aceptar las conexiones en el puerto 1527
```

*El terminal de L'Spock*