

3274 Integració d'un business intelligence en un ERP

Memòria del Projecte Fi de Carrera

d'Enginyeria en Informàtica

realitzat per

Diego Galera Sánchez

i dirigit per

Maria Vanrell Martorell

Bellaterra, Setembre de 2011

El sotasignat, David Tomàs Rubinat
de l'empresa, Xarxes, Sistemes i Serveis de Catalunya, SLU

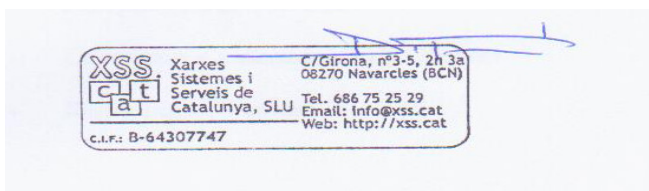
CERTIFICA

Que el treball a què correspon aquesta memòria ha estat realitzat en l'empresa sota la seva supervisió mitjançant conveni

Signat el 14 d'octubre de 2010 amb Registre núm. 6097

firmat amb la Universitat Autònoma de Barcelona.

Així mateix, l'empresa en té coneixement i dóna el vist-i-plau al contingut que es detalla en aquesta memòria.



Signat: David Tomàs Rubinat
Divendres, 9 de Setembre del 2011

La sotasignat, Maria Vanrell Martorell

Professor/a de l'Escola Tècnica Superior d'Enginyeria de la UAB,

CERTIFICA:

Que el treball a què correspon aquesta memòria ha estat realitzat sota la seva direcció per en Diego Galera Sánchez

I per tal que consti firma la present.

Signat: Maria Vanrell Martorell

Bellaterra, 13 de setembre de 2011

Als meus pares, germans i parella, per haver-me ajudat a tenir l'oportunitat d'arribar fins aquí.

Agraïments

Especialment a David Tomàs Rubinat, Gerent de l'empresa Xarxes, Sistemes i Serveis de Catalunya, SL (e-consultic.com), per haver-me donat aquesta oportunitat de realitzar aquest projecte d'empresa, i del seu suport.

Al Grup Alimentari Llobet, per haver cedit les dades dels anys 2008, 2009 i 2010 per la realització d'aquest projecte.

A Maria Vanrell Martorell per haver-me ajudat a realitzar aquest projecte, sobretot al inici, on vaig tenir tot el seu suport i donar-me l'oportunitat de presentar aquest projecte.

A Isabel Burdó Expósito, per haver-me aguantat tot aquest temps, que he estat amb el projecte de final de carrera i el seu suport incondicional.

Índex

Capítol 1	1
Introducció	1
1.1 Presentació del Projecte	1
1.2 Objectius del Projecte	2
1.3 Estudi de la situació actual	3
1.4 Anàlisi de viabilitat tècnica	4
1.5 Planificació Temporal	6
1.5.1 Planificació temporal realitzada	8
1.6 Estructura de la memòria	10
Capítol 2	11
Conceptes bàsics	11
2.1 Business Intelligence	11
2.2 ETL	12
2.3 Datawarehouse	12
2.4 Cubs OLAP	12
2.5 ERP	13
2.6 TPV	13
2.7 CRM	13
Capítol 3	15
Recollida i tractament de les dades	15
3.1 Recollida de les dades	15
3.2 Estudi de les dades	17
3.3 ETL (Extracció, Transformació i Càrrega)	22

3.3.1	Extracció.....	23
3.3.2	Transformació.....	24
3.3.3	Càrrega	25
Capítol 4	27
Datawarehouse: Base de dades de l'aplicació	27
4.1	Recollida de les dades	27
4.2	1a Versió del Datawarehouse: dades en cru	28
4.3	2a Versió del Datawarehouse: particionat vendes-any.	30
4.3.1	Particionament de taules	30
4.3.1.1	Tipus de particionament de taules amb MySQL.....	32
4.3.2	Particionament aplicat a les taules del datawarehouse	33
4.4	Cubs OLAP.....	34
4.5	3a Versió del Datawarehouse: particionat de vendes-Any-establiment.....	36
Capítol 5	39
Sistema Web de Generació de Gràfics.....	39
5.1	Patró de creació de gràfiques.....	39
5.2	Generació de gràfics amb php: jpgraph.....	41
5.3	Sistema Login.....	41
5.4	Generació de gràfics implementats	42
5.4.1	Gràfiques per establiment.....	43
5.4.1.1	Volum de facturació entre anys.....	43
5.4.1.2	Volum de productes venuts-anys.....	44
5.4.1.3	Volum de productes venuts-anys (3a versió del DW)	45
5.4.1.4	Comparació de temps d'execució entre 2a i 3a versió DW.....	45
5.4.1.5	Mitjana de vendes de productes.....	47
5.4.2	Gràfiques per producte	48
5.4.2.1	Volum de vendes-anys.	48
5.4.2.2	Vendes per establiments-anys.....	49
5.4.2.3	Mitjana de vendes totals.....	51
5.4.2.4	Volum de vendes entre productes-establiment-any.....	52
Capítol 6	55
Conclusions	55
6.1	Desenvolupament del projecte.....	55

6.2 Treball futur	57
Bibliografia.....	59

Índex de figures

Taula 1: Volum de dades que treballem 1	5
Figura 1.1 Planificiació inicial	7
Figura 1.2 Planificiació final.....	9
Figura 3.1 Programa Spoon de Pentaho.	17
Figura 3.2. Base de dades generada en format MySQL	22
Figura 3.3 Esquema bàsic d'entrada de dades, procés ETL i càrrega del Datawarehouse	23
Figura 4.1 Esquema de disseny del Datawarehouse versió 1	29
Figura 4.2 Mostra el sistema de particionament d'una taula amb MySQL.....	32
Figura 4.3 Exemple de cub OLAP del nostre Datawarehouse en la taula vendes-any.....	36
Figura 4.4 Sub-Cubs de OLAP del Datawarehouse.....	37
Figura 5.1 Arbre de decisió per a la creació de les gràfiques.....	40
Figura 5.2 Esquema base de la generació de gràfics.....	40
Figura 5.3 Sistema Login d'entrada al sistema Business Intelligence	42
Figura 5.4 Menú principal de l'aplicació Business Intelligence	43
Figura 5.5 Volum de facturació de l'establiment Jaume 1 de Manresa (2008-2009)	44
Figura 5.6 Productes venuts per l'establiment Drets de Manresa entre anys.....	45
Taula 5.1: Temps d'execució amb segona versió del Datawarehouse.....	46
Taula 5.2: Temps d'execució amb tercera versió del Datawarehouse	46
Figura 5.7 Mitjana de vendes del producte en l'establiment Numància de Manresa.....	48
Figura 5.8 Vendes del producte "Berberecho Baymar Pequeño" entre anys.....	49
Figura 5.9 Vendes del producte "Sardinas CUCA OLIVA 120G" en 2008	50
Figura 5.10 Vendes producte "Pate fines h. Tarradelles en 2009 entre establiments	51
Figura 5.11 Mitjana de vendes d'aigua Viladrau 1,5l per tot el grup.....	52
Figura 5.12 Vendes d'aigua de 1.5l entre Font Vella i Viladrau en Numancia(2008).....	53
Figura 5.13 Vendes d'aigua de 1.5l entre Font Vella i Viladrau en Economat(2008).....	53

CAPÍTOL 1

INTRODUCCIÓ

Aquest capítol, recull l'anàlisi de viabilitat del projecte de final de carrera Business Intelligence en un ERP, els objectius del projecte, anàlisi i diagrama de temps del projecte.

1.1 PRESENTACIÓ DEL PROJECTE

La competència en el món empresarial, és cada dia més dura, i s'ha d'anar innovant i prenent noves i bones decisions, pel benefici de l'empresa. En aquest punt, és de vital importància la idea del Business Intelligence (BI), ja que és pitjor no tenir informació disponible, que tenir molta informació i no saber que fer amb ella.

La intel·ligència en els negocis o Business Intelligence, dóna la solució al problema plantejat anteriorment, ja que podem generar escenaris, pronòstics, informes que puguin ajudar a la presa de decisions en l'empresa, accessible des de qualsevol dispositiu mòbil amb accés a entorn web.

En l'actualitat, hi ha una gran varietat de software de Business Intelligence, o aplicacions similars que poden ser utilitzades en les diferents àrees d'una empresa, com poden ser els departaments de vendes, màrqueting, finances, etc. El handicap més gran, és que la gran

majoria, són eines del tipus software privatiu que tenen un cost afegit, com ara les llicències, etc.

Cada dia que passa, hi ha més empreses que s'estan beneficiant de la implantació d'un sistema de BI en el seu negoci. I cada vegada, és més important poder utilitzar aquestes eines. Segons estudis [14-15], amb el pas del temps, es convertirà en una necessitat per a totes les empreses, el fet de disposar d'algun sistema de Business Intelligence, per la presa de decisions.

1.2 OBJECTIUS DEL PROJECTE.

Aquest projecte, es centra en la creació d'un prototip de software de Business Intelligence, destinat al grup de distribució alimentària Llobet, amb diferents línies de negoci en supermercats. L'estudi de tota la informació generada per les diferents línies de negoci, faran que el nostre software doni informació útil per a que sigui més fàcil la presa de decisions. Aquest prototip, respon a la demanda que el grup Llobet, de distribució alimentaria va realitzar a l'empresa e-consultic, petita empresa dedicada en el desenvolupament web, LOPD i consultoria TIC en la qual formava part i que es troba ubicada en la Catalunya Central, formada inicialment per 4 treballadors fixes, que en l'actualitat ha reduït el seu personal per falta de liquiditat, afectant el meu lloc de treball.

L'objectiu principal d'aquest projecte, és **el desenvolupament d'un prototip d'aplicació amb una interfície amigable, que permeti l'extracció d'informació útil per la presa de decisions.** Per tal d'assolir aquest objectiu, he dividit el treball en 6 tasques que alhora ens ajudaran a conformar el projecte, trobant la principal dificultat per realitzar-lo en el gran volum de dades que hem de gestionar i degut a la manca de finançament, que ens obliga a desenvolupar l'aplicació amb eines open source.

Tasca 1:

Anàlisi de requeriments: per poder desenvolupar aquest anàlisi, s'han de realitzar diverses reunions amb els departaments del grup alimentari, que estan implicats en el projecte.

Tasca 2:

Estudi de les eines: per poder realitzar els requeriments del projecte, es necessari fer un estudi de les eines de codi lliure, que podem trobar en el mercat amb cost 0 pel projecte.

Tasca 3:

Estudi de les dades: per poder realitzar l'agregació de les informacions obtingudes en les diferents fonts, es necessari fer un estudi de les dades de què es disposen.

Tasca 4:

Transformació de dades: per poder crear les taules, necessitem abans fer una normalització de les dades. Aquesta, es pot dur a terme fent transformacions, sobre aquestes que permetin obtenir dades de qualitat.

Tasca 5:

Definició de taules: s'haurà de realitzar la distribució de la informació en diferents taules, i s'hauran de crear taules amb pre-càlculs, per ajudar a realitzar la recerca de la informació en un temps curt, que permeti la realització de consultes per part de les eines front-end.

Tasca 6:

Interfícies d'usuari: per poder realitzar consultes al datawarehouse, es crearan interfícies d'usuari per realitzar les consultes, i mostrar la informació obtinguda en un format útil i de fàcil interpretació, utilitzant diferents interfícies gràfiques. Per donar resposta als indicadors més importants pel client, com són els d'ofertes, compres i botigues o supermercats.

1.3 ESTUDI DE LA SITUACIÓ ACTUAL

Actualment, el camp de Business Intelligence és un camp que s'està estudiant i treballant dia a dia, fent que cada vegada més, un major nombre d'empreses, estiguin interessades en poder implementar un sistema BI en la seva estructura d'empresa.

S'han trobat moltes empreses que treballen amb aquest tema, i molts productes que ofereixen solucions, però gairebé totes són del tipus software privatiu, com són:

- Alterian (www.alterian.es)
- ApeSoft (www.apesoft.com)
- BusinessObjects (www.iberica.businessobjects.com)
- Cognos (www.cognos.com/es)
- Microsoft Business Intelligence (www.microsoft.es/bi)
- PowerData (www.powerdataib.com)
- QlikView (www.qlikviewspain.com)

Aplicacions BI Open Source:

- Pentaho (www.pentaho.com)

Com que no existeix una aplicació total del tipus Open Source, ja que l'aplicació pentaho, té parts obertes, i altres en els quals s'han d'adquirir una llicència. Per aquest projecte, s'ha decidit utilitzar algunes de les eines de pentaho que són del tipus Open Source, que faciliten la feina.

1.4 ANÀLISI DE VIABILITAT TÈCNICA

Abans de realitzar el projecte, es va contemplar la possibilitat de què el projecte fos viable, tant a nivell econòmic, temporal i acadèmic. Cal doncs, estudiar sí en el mercat podem trobar tot el material necessari, pel desenvolupament del projecte amb cost 0.

Com s'ha vist anteriorment, el projecte consta de varies parts diferenciades que són l'anàlisi de requeriments, estudi de les eines, estudi de les dades, transformació de les dades, creació de les taules i de les interfícies d'usuari, per poder realitzar les consultes, i mostrar la informació de forma gràfica i útil.

La dificultat del projecte, no es centra en la definició de les consultes per part de l'usuari, sinó en el volum de dades amb les què es treballen, ja que el temps de resposta de l'aplicació ha de ser curt. Per tal de veure el volum de dades amb les què s'han treballat, ho podem observar a través de la següent taula:

	Volum de dades
Número de registres de vendes en 2008, 2009, 2010	+ de 21.000.000 de línies de registres
Número de productes	+ de 27.000 productes diferents
Número d'establiments	76 establiments

TAULA 1: VOLUM DE DADES QUE TREBALLEM 1

Per poder desenvolupar el projecte, es va necessitar obtenir les dades per a la realització de l'estudi del projecte. Aquestes dades, han estat cedides pel grup alimentari Llobet a l'empresa e-consultic, i en conseqüència per la realització d'aquest projecte de final de carrera.

La transformació de les dades és un pas important pel projecte. Aquest problema, que pot arribar a ser crític és també conegut com ETL (extracció, transformació i càrrega), és el que permet poder tenir informació vàlida i de qualitat. Per poder fer aquest estudi i transformar les dades, utilitzarem l'eina gràfica kettle o Spoon de Pentaho.

Per poder tenir les dades disponibles en taules en una base de dades, s'han avaluat diferents tipus de solucions disponibles en el mercat. Per la limitació de cost, s'ha decidit utilitzar un servidor web Apache i el servidor SQL de MySQL.

Per la creació de les interfícies d'usuari utilitzarem el llenguatge php, per la facilitat de treballar en entorn web, i per la facilitat d'integrar-ho en qualsevol plataforma d'accés, ja que l'objectiu del prototip és la creació de gràfiques utilitzant les dades històriques.

Inicialment, es va decidir treballar les dades obtingudes, amb la col·lecció d'eines i algoritmes del paquet Weka (llenguatge Java), per poder treballar amb models predictius, degut a l'alteració de la planificació inicial, aquest punt no es va poder desenvolupar, i es va convertir en un apartat per treballar en el futur.

1.5 PLANIFICACIÓ TEMPORAL

Els diferents subobjectius, que formen aquest projecte es troben relacionats amb el desenvolupament de les tasques. Així doncs, es va desenvolupar l'aplicació, utilitzant un model de construcció en prototip, arribant a desenvolupar en tres fases de generacions de prototips, fent que en cada fase s'introdueixi l'anàlisi d'un nou indicador (ofertes, compres i establiments).

El diagrama de Gantt de la planificació inicial es troba en la següent figura.

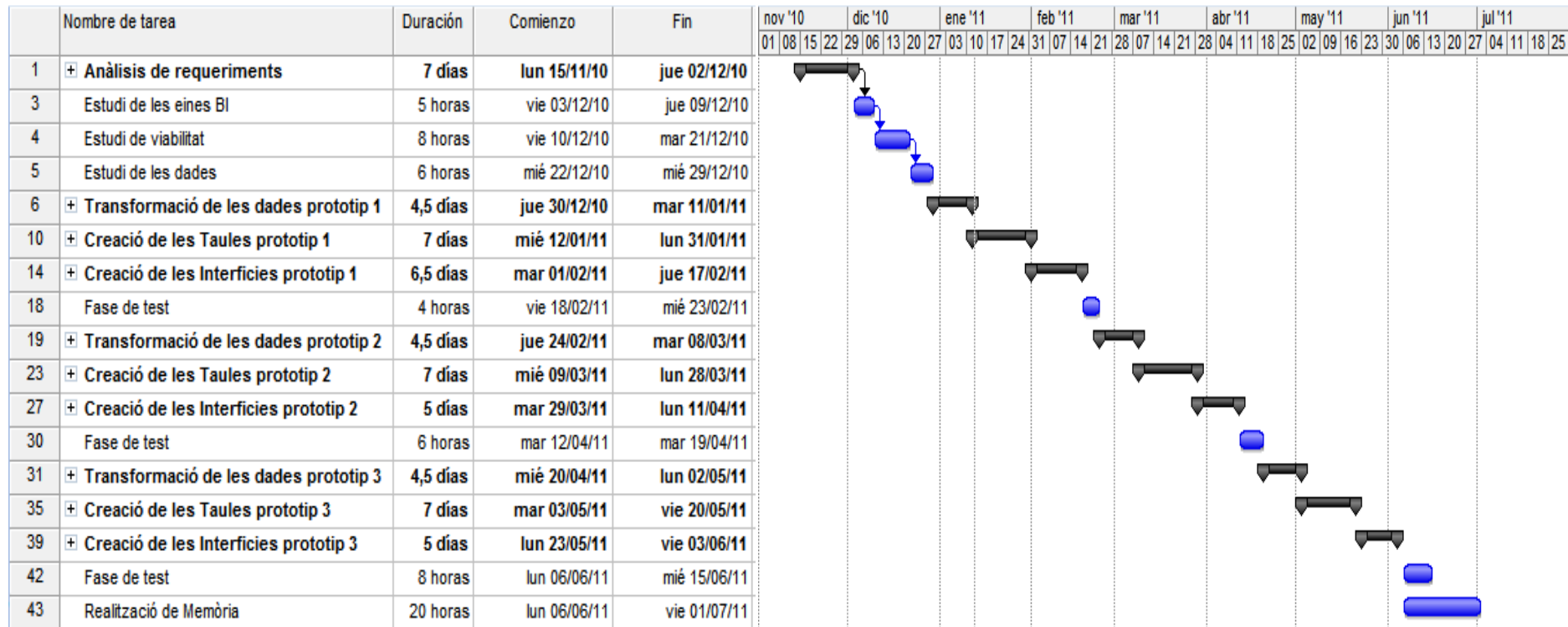


Figura 1.1 Planificació inicial

El rang de dades planificades respecte les evolucions dels prototips eren les següents:

Prototip 1 - Gener - Febrer

Prototip 2 - Març – Abril

Prototip 3 - Maig – Juny

1.5.1 PLANIFICACIÓ TEMPORAL REALITZADA

Aquesta planificació es va veure alterada, per diversos problemes que s'han descrit en la memòria, aquest fet va modificar la planificació de tasques inicials com veurem en el diagrama gannt següent, augmentant el temps que es va compatibilitzar per la creació de les taules i càrrega de les dades, ampliant-se respecte a la planificació inicial en un mes. Un fet important, va ser que dins de l'horari laboral, no es va poder compaginar el desenvolupament del BI, amb les tasques laborals del lloc de treball.

CAPÍTOL 1: ANÀLISI DE VIABILITAT

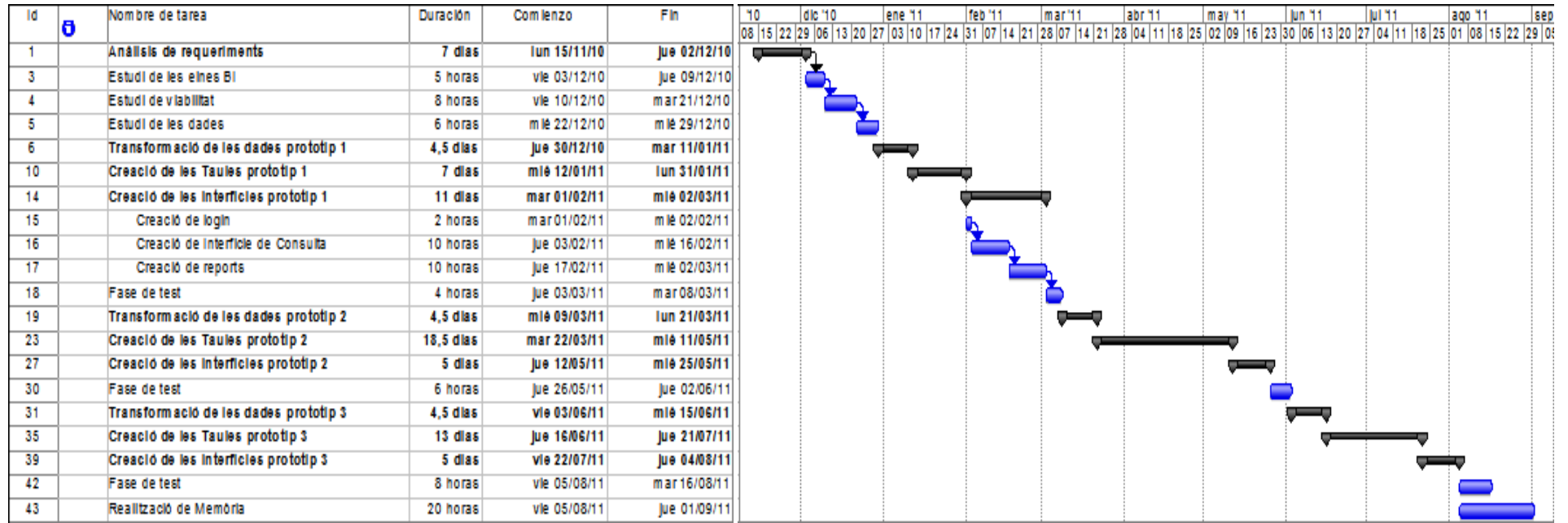


Figura 1.2 Planificació final

1.6 ESTRUCTURA DE LA MEMÒRIA

Aquesta memòria es troba estructurada en 6 capítols: introducció, conceptes bàsics, recollida i tractament de les dades, Datawarehouse, sistema web de generació de gràfics i conclusió.

1. El capítol que esteu llegint és la introducció, on s'ha explicat quins són els objectius del projecte, d'on va sorgir la idea de realitzar-lo, estudi de viabilitat, planificació inicial i la real i així l'estructura de la memòria.
2. El segon capítol, és un recull dels conceptes bàsics necessaris per entendre la resta de projecte, donant definicions de les diferents eines, que podem trobar en un Business Intelligence i amb quines eines han de conviure i comunicar-se.
3. El tercer capítol, va de la recollida i tractament inicial de les dades, que ens van cedir el grup alimentari per la realització d'aquest projecte, de com s'ha realitzat tot l'estudi de les dades, transformació i càrrega en la nova Base de dades.
4. El quart capítol, està dedicat a quina estructura ha de tenir la base de dades central de l'aplicació business intelligence (Datawarehouse).
5. El cinquè capítol, es dedica a la transformació de les dades en gràfiques, estudiant quines són les que necessitem per mostrar de manera més optima, explicant quines podem generar amb l'aplicació.
6. L'últim capítol, està reservat a les conclusions, on podrem veure un resum del què ha estat el desenvolupament del projecte, possibles millores i línies de treball pel futur.

CAPÍTOL 2

CONCEPTES BÀSICS

Aquest capítol, és un recull dels conceptes bàsics que es desenvoluparan en la resta de capítols. El capítol comença amb un repàs de terminologia sobre el món empresarial i de màrqueting. Seguidament s'exposa que és un Business Intelligence, i els passos principals per realitzar-lo. Per últim, comentarem la importància i necessitat de tenir dades rellevants per qualsevol negoci.

2.1 BUSINESS INTELLIGENCE

Business Intelligence (BI) és l'habilitat de transformar les dades en informació, i la informació en coneixement. Amb aquesta transformació, aconseguim que el coneixement generat pugui optimitzar el procés, que implica la presa de decisions dins dels negocis.

A l'hora de prendre les decisions sempre involucra acceptar un risc, el BI té com objectiu minimitzar en tot lo possible aquesta incertesa en la presa de decisions, analitzant totes les dades generades per l'empresa i que es tenen guardades en bases de dades històriques, i així ajudar-nos en la decisió final.

El concepte Business Intelligence, va néixer amb la necessitat de poder comptar amb informació rellevant de qualsevol departament dins de l'empresa, podent provenir les dades de múltiples llocs, fent que la informació rellevant i important sigui accessible de manera ràpida, oportuna i fàcil, per tal d'obtenir avantatges competitius.

El Business Intelligence, va destinat als usuaris responsables dins la direcció de l'empresa o organització, que puguin influir o decidir en la presa de decisions.

2.2 ETL

Les sigles ETL provenen de l'anglès de Extreure, Transformar i Carregar (*Extract, Transform and Load*). Aquest és el procés d'organitzar, moure dades de diferents fonts, donar-li un format predeterminat, netejar els camps sense informació rellevant, per després carregar-los en la base de dades on treballarà el BI, omplint les dades elaborades directament en el Datawarehouse.

2.3 DATAWAREHOUSE

Datawarehouse, és la base de dades de l'empresa o organització, on tenim la informació centralitzada i organitzada en format històric, que es caracteritzada per integrar i depurar la informació d'una o més fonts, per després processar-la i permetre amb una velocitat de resposta alta, l'anàlisi de les dades amb diferents punts de vista. El Datawarehouse és la base de dades amb la què treballa el Business Intelligence, per consultar les peticions del client, i mostrar les dades històriques de l'empresa.

2.4 CUBS OLAP

Els cubs OLAP són les sigles de "*Online Analytic Processing*". Es tracta de subconjunts de dades del DataWarehouse, de forma organitzada i estructurada en multi dimensions. Els cubs, permeten la mecanització de recerca de dades de forma ràpida, amb un temps de resposta estable. Les dimensions dels cubs, són factors del negoci, que podem seleccionar per tenir les dades organitzades en 3 dimensions.

2.5 ERP

Les sigles ERP provenen de l'anglès "Enterprise Resource Planning", és a dir, sistema que permet realitzar una gestió integral de tots els recursos d'una empresa o organització. Específicament dissenyat, per poder automatitzar i modelar la gran majoria de processos dins de l'empresa, com són les àrees de finances, comercial, producció, logística, màrqueting, etc.

L'ERP, planifica els recursos dins de l'empresa o organització, i la relació de totes les àrees, integrant tota la informació generada al llarg de la vida de la mateixa, eliminant enllaços entre els diferents departaments empresarials i de negoci, optimitzant els processos empresarials, centralitzant tota la informació en un mateix lloc, permetent l'accés de tota la informació i mantenint la integritat de les dades.

Gràcies als ERP s'eliminen molts dels programes aïllats, que s'utilitzen en la gestió de cada una dels departaments d'una empresa.

2.6 TPV

Les sigles TPV signifiquen Terminal Punt de Venda. És un hardware amb una aplicació informàtica, que facilita les tasques de gestió d'un establiment comercial d'un punt de venda al públic. El mateix sistema, és l'encarregat de la creació i impressió dels tiquets de venda, mitjançant les referències de productes, i que al mateix temps ens pot ajudar a gestionar l'estoc, utilitzant les bases de dades on disposem de tota la informació dels productes de l'establiment.

2.7 CRM

Les sigles CRM, que en anglès volen dir "Customer Relationship Management", administració de la relació amb els clients. Els CRM són programes informàtics, que donen suport en la gestió de les relacions amb els clients en la venda i en les campanyes de màrqueting.

Els sistemes CRM, fan que la estratègia de negoci aplicada sigui centrada en el client, intentant avançar-se a les peticions dels consumidors, fent que els futurs compradors tinguin tot allò que necessiten en el moment que ho necessiten.

CAPÍTOL 3

RECOLLIDA I TRACTAMENT DE LES DADES

En aquest capítol, s'exposa el procés que hi ha hagut en la recollida de la informació, per començar a estudiar les dades, que han estat cedides pel client i per realitzar aquest projecte. Els tractaments que s'han aplicat són per treballar les dades i poder-les incorporar en el Datawarehouse del Business Intelligence.

3.1 RECOLLIDA DE LES DADES

Com el projecte de Business Intelligence és d'empresa, el grup alimentari Llobet va cedir a Xarxes, Sistemes i Serveis de Catalunya, SL (e-consultic.com), empresa amb la què es realitza el BI, algunes de les taules que pertanyen a l'ERP que tenen implementat en la seva empresa, donant les taules de vendes i productes dels històrics 2008, 2009 i 2010.

Els fitxers que ens han cedit per realitzar el projecte són:

- Almacen.dbf
- Desefarg.dbf
- Desegra_articulos.dbf

- Tiendas.dbf
- Ubica.dbf
- li2008_2009.dbf
- li2010.dbf

El primer problema que ens trobem és en els fitxers de bases de dades que ens han cedit, estan en format dbf. El format dbf és un format de dBASE, el primer Sistema de Gestió de base de dades que es va començar a utilitzar massivament en els ordinadors de l'època, on la versió III Plus va ser un dels software més venuts a finals dels 80 i principis dels 90.

Així el primer pas, és la transformació dels fitxers dbf a un sistema més actual de base de dades, en aquest cas s'ha escollit el sistema de base de dades MySQL, per la seva potència i pel seu cost 0, ja que MySQL és de codi obert, sent una de les bases de dades més utilitzades actualment en el desenvolupament d'aplicacions.

Després d'avaluar diferents opcions, per la conversió d'un fitxer amb format dbf a un altre format compatible amb MySQL, es va escollir per a realitzar aquesta transformació, el programa Spoon del paquet Pentaho. Aquest programa, ens ofereix una interfície gràfica i intuïtiva, per generar la transformació dels fitxers en cru, i que en aquest cas són de taules xbase (dbf) a un fitxer sql.

A la figura 3.1, mostrem una captura en el moment de transformació d'un dels fitxers cedits pel grup alimentari a un format compatible amb mysql.

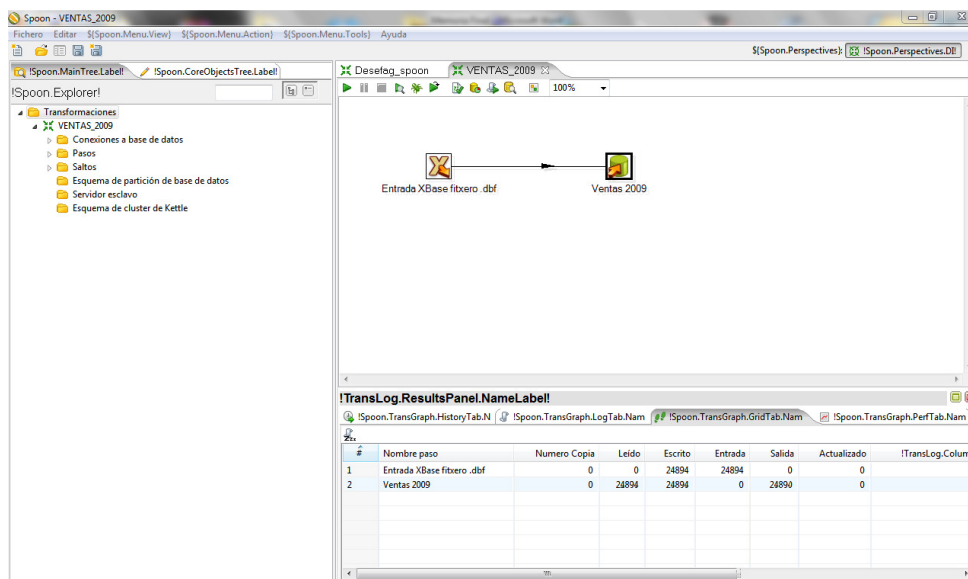


Figura 3.1 Programa Spoon de Pentaho.

3.2 ESTUDI DE LES DADES

Al final de tota la transformació de les dades dbf a format mysql, podem observar tots els camps que hi han en cada una de les taules, per poder realitzar l'estudi i quins camps volem que tingui el nostre Datawarehouse, i sobre quines transformacions necessitem realitzar.

La taula número 1 "Almacen" que obtenim després de la transformació a MySQL, podem observar que té els següents camps:

Taula 1: "Almacen"

- | | |
|------------|----------|
| REFE | ULTPRECO |
| CORRESPOND | ULTPREDI |
| DESEFAGR | ULTPROVE |
| DESCORTA | PRECOSTM |
| DESC | PREVENT |
| DESC1 | MARGESTA |
| MARCA | TIPOENVA |
| FORMATO | MARCAOFE |
| TIPO | EXISTINI |
| UNIDADCA | EXISTSAL |
| UNIDADDI | EXISTOTA |
| IVA | CANTPEND |
| STOCKMIN | ULTCANTD |
| STOCKMAX | EXISTSAA |
| FECHALTA | FECILTIN |

EXISTULT	BALAETIQ
PCMEDULT	BALAFETI
MARCAINV	BALATXT1
NETIQUET	BALATXT2
MARCAGRA	BALATXT3
MARCASIN	BALAING1
MARCAETI	BALAING2
DCTOEMP	BALAING3
PLU	ULTFCAD
HUECOPK	CENTDO
HUECOSV	CTRLCAJA
LARGO	PRECVMIN
ANCHO	CAJCAPA
ALTO	EMPR
PESO	REFEALTE
CAJPALET	ORDALTE
PRECVEN2	ENVASEUN
DCTO	ENVASECJ
MERMNATU	ENVASECG
CALCMARG	INTRASTAT
COMP	ENLAZADO
AVERIAS	COMPRADOR
PUNTOS	GRUPOTR
DIASCAD	TPALET
DESC2	ALCOHOL
DESCORTA2	CTARIC
PREMNETN	CODRESTI
ULTPRENE	PORRESTI

Com podem observar en la taula número 1 “Almacen”, tenim molts camps que ens poden donar una àmplia informació d’un producte, però ens trobem amb un problema, i és que la gran majoria dels camps de què disposem estan buits o sense informació rellevant. Aquest problema ens fa plantejar la definició, de quins camps agafarem de la taula número 1 “Almacen” pel nostre Datawarehouse, que més endavant explicarem.

Taula 2: DESEGRA ARTICULOS

En la taula número 2 “Desegra_articulos” tenim els següents camps:

DPTO
SECC
FAMI
GRUPO
FAMILIA

El camp “DPTO”, ens indica amb un número (de l’1 al 9), de quin departament pertany el grup de producte.

El camp “SECC”, ens indica amb dos xifres el identificador de la secció dins d’un departament.

El camp “FAMI”, ens indica també amb dos xifres el identificador de família dins d’una secció.

El camp "GRUPO", és un camp en format "integer" (de l'1 al 9), que ens indica el grup que pertany dins d'una família, secció i departament.

En el camp "FAMILIA", ens indica el nom de categoria de producte, amb la combinació dels altres camps.

Taula 3: Desefarg

En la taula número 3 "Desefarg", disposem de dos únics camps que són:

REFE
DESEFAGR

El camp "REFE", és el codi de referència del producte, amb el què relacionem els detalls del producte, amb el camp "DESEFAGR".

El camp "DESEFAGR" és la combinació dels camps "DPTO", "SECC", "FAMI" i "GRUPO" de la taula 2 "DESEGRA_ARTICULOS".

En aquesta taula número 3, tenim la combinació dels productes per referència amb la relació de quin grup pertanyen.

Taula 4: "Tiendas"

En la taula 4 "Tiendas", és on disposem de tota la informació dels establiments de venda que pertanyen al grup alimentari, cada un amb la seva informació.

Com veurem en la següent llista, existeixen molts camps que aporten informació dels establiments, però que no és informació important pel desenvolupament del projecte.

Els camps de la taula número 4 són:

CLAVE	TIPOTIENDA	SUCURSAL
NOMBRE	TARIFA	SSEMPRE
SOCIEDAD	DIRECTOR	ETIQUETAS
CALLE	ADJUNTO	ROLLSMEDIO
APARTADOC	JEFACA	ROLLSMAX
CODPOS	FINIFIES	ROLLSMIN
CIUDAD	FFINFIES	CIERRESEM
NIFDNI	SECCIONES	SERVICIO
TELEF1	IMPAVAL	CONTROLEX
TELEF2	FFINAVAL	RECDIS
FAX	CREDITOMAX	CLIEFACT
FECAPER	CODIGOBAN	TARIFAD
SUPERFICIE	BANCO	CTRLUBI
SCANNER	CLAVESICA	CTRLCAD
NUMTPV	CUENTA	COSVEN

PALECAPA	UBICAUTO	
HOJAPREP	GRUPOCHE	TIPOTEPO
CTRLEXI	RADIOFRE	CONMATRAZ
SERIES	ORDPICK	CONPACKIN
PROVFACT	ORDPASI	IDPASAL
EMPR	ORDCLAS	ARFSDIAS
GRPLANT	NLINDEXGL	ARFSGRPT
TIPOCEN	CHKARTI	ARFSOPER
DIVISA	CRITDIVI	TPETENT
EANEDI	AGEHDE	MODOREP
DCTOEMP	AGEHHA	CTRLLOTE
ZONAAUTO	AGEINT	SINIVA
ZONAPED1	AGEDES	RADIOFAB
ZONAPED2	AGEMUE	
ZONAPED3	CAJMAXPED	
ZONAPED4	REPMUELLE	
ZONAPED5	TIPOPREP	
ZONAPED6	AUTOFALP	
ZONAPED7		
ZONAPED8		
ZONAPED9		
ZONAPED10		
SEGVUELTA		
FORMETIEN		
CONMATRIC		
LOTEOBLIG		

Més endavant, s'explicarà quins camps de la taula ens quedarem per realitzar dins del nostre Datawarehouse, ja que moltes de les dades que disposem a la taula, estan buides o no són correctes. Per exemple, l'estoc d'un producte que té el valor del camp negatiu de la magnitud de -33240102 unitats.

Taula 5: "Ubica"

En la taula número 5 "Ubica", podem observar les dades sobre un producte, del qual podem trobar en qualsevol establiment, ja que al ser un grup alimentari compostat per 5 empreses, i que cadascuna té un objectiu diferent, i ens interessa saber quin producte tenim en cada establiment, per poder comparar resultats entre botigues que posseixin el mateix producte.

Els camps de la taula número 5 "Ubica", són els següents:

REFE	DPTOCOST
UBICÇ	STOCKMIN
DESEFAGR	STOCKMAX
EXISTACT	PREM
FECULTEN	RECDIS
EXULTINV	PROV
FECULTIN	FECSURTI
EXANTULT	EXISTCJ
PVENULT	EXULTINCJ
MARCAINV	EXANTULCJ

En aquesta taula, podem veure camps que poden ser de molta utilitat, però estudiant els valors que contenen les dades, veiem que molts dels registres estan en blanc o NULL. Més endavant veurem quin camps utilitzarem en el nostre Datawarehouse del BI.

Taules 6 i 7: li2008_2009 i li2010

En els fitxers li2008_2009 i li2010, veiem totes les línies de vendes generades, entre els períodes de temps dels anys 2008, 2009 i 2010 de tots els establiments del grup alimentari.

Cada línia representa un producte venut, en els supermercats, per exemple, sí en un tiquet de compra tenim 10 productes diferents, obtenim a la base de dades 10 línies de registres, una línia per cada producte venut.

En aquestes dues taules (6 i 7) tenim els mateixos camps, que s'han generat amb les compres, que podem observar en els següents camps:

NALB	IVA
NPED	TIPOIVA
REFE	PRECOSTM
CANT	PREC
PRECIO	VALIDADO
FECH	DCTO
TIENDA	DIVISA
TIPO	PNETON

En les línies de cada venda, tenim tota la informació generada en la venda d'un producte, com ara el preu, quantitat, data, etc.

Com podem veure, les dades de les què disposem en tots els fitxers .dbf, s'observen en la següent figura 3.2, que no mostra tot el llistat d'algunes taules, degut al seu gran número de camps. Per una altra banda, podem observar en diferents taules, la mateixa dada, però amb nom diferent. Per exemple, el valor "UBIC" de la taula "ubica", és el valor "CLAVE" de la taula de "tiendas", el mateix passa amb el camp "TIENDA" de les línies de vendes, amb el camp "CLAVE" de la taula "tiendas".

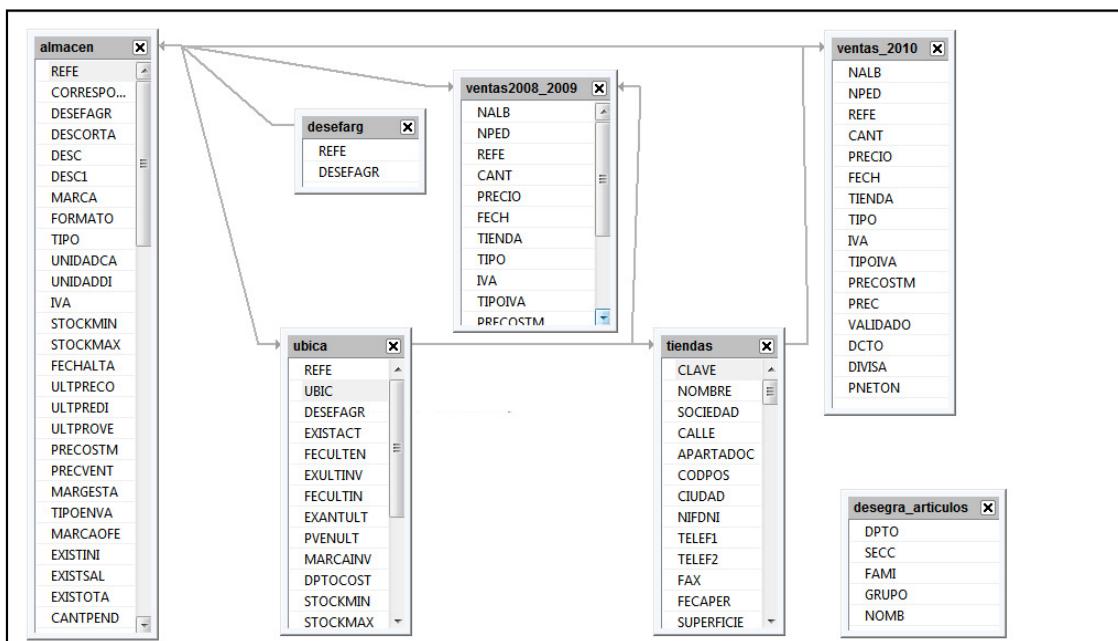


Figura 3.2. Base de dades generada en format MySQL

3.3 ETL (EXTRACIÓ, TRANSFORMACIÓ I CÀRREGA)

Les sigles ETL, signifiquen en anglès Extreure, Transformar i Carregar (Extract, Transform and Load). Aquest procés, és el que ens permet incloure i moure dades de múltiples fonts, donant-li un format determinat, netejant-los i carregant-los a una altra base de dades, que en aquest cas, la carregarem directament en el nostre Datawarehouse.

Els processos ETL, els podem utilitzar per la integració de sistemes heretats de la gestió del negoci, és a dir, d'aplicacions antigues existents en les empreses, que generen informació aïllada, de les quals integrem en el sistema BI. Un exemple d'aquestes aplicacions són els ERP's. En aquest cas, tenim les taules de diferents fonts com són els ERP i els TPV. En la següent figura 3.3, es pot veure tot el procés entre el recull de dades i l'alimentació del nostre Datawarehouse.

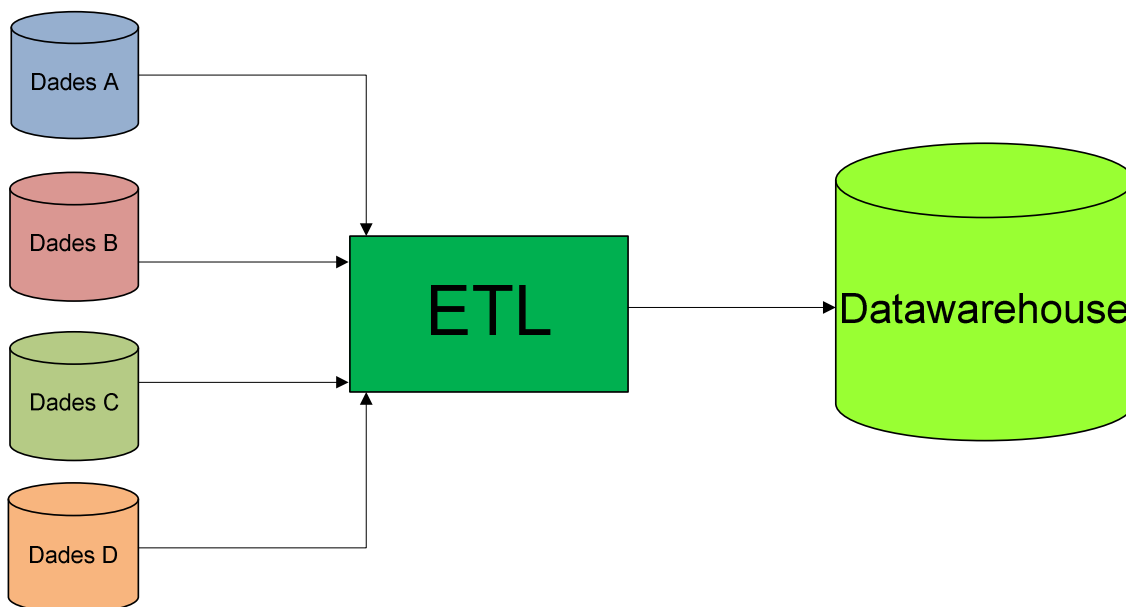


Figura 3.3 Esquema bàsic d'entrada de dades, procés ETL i càrrega del Datawarehouse

3.3.1 EXTRACCIÓ

La primera part del procés ETL, consisteix en extreure la informació dels sistemes d'origen, en el nostre cas dels fitxers .dbf, dels diferents sistemes de procedència de l'empresa, cada conjunt per separat pot arribar a contenir una organització de les dades o formats diferents.

Un exemple dels diferents formats que hem trobat, és que disposem de dades en els fitxers en format "date" i "datetime", referències als mateixos camps en diferents taules, tenim uns en format "varchar" i en una altra taula en format "integer".

Una part important d'aquest pas, és l'anàlisi d'aquestes dades per poder analitzar i verificar que el contingut d'aquestes que s'han extret, compleixin unes pautes, és a dir, que les dades siguin consistents. En el nostre cas, hem estudiat les dades dels camps obtinguts, després de transformar-les en accessibles en format MySQL, i comprovant que hi ha moltes dades amb camps amb valor NULL o erronis.

Un requeriment molt important en aquest procés, és que l'extracció de les dades causin un mínim impacte en els sistemes orígens (ERP, etc...), ja que sí les dades que s'han d'extreure tenen un volum de dades considerable, el sistema origen podria baixar de rendiment o col·lapsar-se, i per tant, obtenir en aquesta funció en sistema Business Intelligence, les dades

el més actualitzades possible, s'hauria de desenvolupar cada cert període temps. Per això, en sistemes grans les operacions són en horaris i dies, amb un impacte 0 o mínim, per exemple, en caps de setmana, hores nocturnes, etc).

En el nostre cas, no tenim accés a l'ERP, ja que estan en procés d'inserció d'un nou sistema en tot el grup i no ens han permès tenir accés a les taules del nou sistema, però sabem que el nou sistema ERP que incorpora el Grup Alimentari, les bases de dades que utilitzaran seran en SQL Server. Per tant, hauríem de realitzar scripts, per poder sincronitzar les dades amb el nostre Datawarehouse, comprovar quins camps són els guardats en les taules, fer la conversió i transformació d'aquests camps amb el format del nostre Datawarehouse.

3.3.2 TRANSFORMACIÓ

En aquest pas del procés ETL, consisteix en l'aplicació de regles o funcions per les dades que han estat extretes, per convertir-les en les que seran carregades. Algunes de les dades requereixen alguna transformació, per tal de tenir la informació en un format homogeni.

Algunes de les accions que s'han realitzat són:

- Selecció de determinades columnes per la seva càrrega, un típic exemple, és que les columnes amb valors nuls no es carregin en el nostre sistema.
- Obtenció de nous valors calculats, en el nostre cas s'ha determinat el valor mes, que s'ha extret de la data de venda, de les taules de vendes de cada any.
- Calcular valors totals de múltiples files de dades. (Disposar de les dades totals de vendes de cada un dels establiments, etc.)
- Aplicacions de validació de dades, consistents en l'aplicació en cada cas:
 - Dades correctes. Entra en la següent etapa que és la Càrrega.
 - Dades incorrectes. Tractar les dades, per intentar transformar o rebutjar el registre complet.
- Transformació de tipus de dades, d'un format a un altre.

3.3.3 CÀRREGA

Aquesta fase, és la que va després de la transformació i que procedeix a carregar les dades transformades en el nostre sistema destí, del nostre Datawarehouse.

En el nostre Datawarehouse, posseïrem de tot l'historial en registres de les vendes dels anys 2008, 2009 i 2010, de forma que es pugui fer una "auditoria" de les dades, per poder tenir tota la història d'un valor en un llarg període de temps.

Tenim dos formats bàsics a desenvolupar en el procés de càrrega:

- Rolling: Procés que s'aplica per poder mantenir diferents nivells de "granularitat". Per aquest motiu la informació és resumida en varis nivells, unitats de temps, etc.
- Acumulació simple: Consisteix en realitzar un resum de totes les transaccions compreses en un període, transportant el resultat com una única transacció al Datawarehouse, emmagatzemant el valor calculat, típicament en un sumatori, promig d'un valor, etc.
- Còpia predefinida: Còpia de les línies de registre de les taules ETL, amb la selecció de camps dels formats correctes del nostre Datawarehouse.

En aquesta fase de càrrega, s'interactua directament amb la base de dades del nostre Datawarehouse, el resultat de tot el procés, fa que les dades siguin de qualitat (valors únics, integritat referencial, rang de valor, cap dada amb valor null, etc). Això fa que hi hagi una garantia de qualitat de les dades en el procés.

CAPÍTOL 4

DATAWAREHOUSE: BASE DE DADES DE L'APLICACIÓ

En aquest capítol s'explica quines característiques té el nostre Datawarehouse, de quina estructura es compon, amb les seves característiques bàsiques i les versions desenvolupades de les taules, per tal d'ajudar a l'anàlisi de dades amb el menor temps de còmput.

4.1 RECOLLIDA DE LES DADES

El Datawarehouse és la base de dades corporativa d'una empresa, en el nostre cas, la base corporativa és la de tot un grup alimentari, que està compostat per 5 empreses.

En aquesta base de dades estan integrades i depurades les informacions de les diferents fonts d'entrada, amb l'objectiu de donar una alta velocitat de resposta, a les peticions de dades realitzades sobre la base d'informació del Datawarehouse.

Les característiques que ha de tenir un Datawarehouse, sobre les seves qualitats són les següents:

- **Integritat:** Les dades de les quals es compon el Datawarehouse, estan integrades dins d'una estructura consistent, tenint la informació estructurada amb diferents nivells de detall, fent que la informació sigui adaptada per a totes les necessitats i que puguin plantejar a tots els usuaris del Business Intelligence.
- **Temàtica:** Les dades necessàries pel procés de generació, per poder entendre el negoci, provenen d'un entorn operacional i han estat integrades en el nostre

Datawarehouse. Les dades estan organitzades per “temàtiques” per facilitar l'accés a elles, i així proporcionar l'enteniment per tots els usuaris. Per exemple, les dades que són d'establiment i per any estan ubicades en una única taula dins de la base de dades. D'aquesta manera, les dades es troben centrades en una única taula, de forma que la resposta a l'hora d'una demanda d'un producte en concret per part del client és més fàcil, ja que tota la informació es troba dins de la mateixa taula, per tal de poder reduir el temps de computació, com podrem observar més endavant.

- **Històrica.** La variable del temps, és una part implícita de la informació que tenim desada en el Datawarehouse. En el sistema operacionals, les dades reflecteixen l'estat d'activitat del negoci en el moment del present. En canvi, la informació que tenim emmagatzemada en el Datawarehouse, ens serveix entre altres coses per realitzar anàlisis de tendències, fet que implica disposar de la informació guardada i organitzada, tot depenent de les variables en el temps i així permetre comparacions entre dates.
- **No volàtil:** Tota la informació de què disposem dins el Datawarehouse, simplement són dades de lectura que mai podran ser modificades, és a dir, que quan s'actualitzen les taules, els valors s'afegeixen al final, però mai sense cap influència en les dades ja existents dins de la taula, perquè la informació que hi ha desada és històrica, i no es pot canviar en el transcurs del temps.

4.2 1A VERSIÓ DEL DATAWAREHOUSE: DADES EN CRU

La primera versió del nostre Datawarehouse, va ser la d'emmagatzemar les dades directament, en la què consistia en la reducció dels camps en les taules de l'inici, posteriorment la separació de la informació en anys, composta per 3 taules de vendes, cadascuna d'elles relacionades amb un any diferent entre el 2008 i 2010, amb una dimensió aproximada de cada taula de 7 milions de línies. A més també contindrà les taules de “tiendas”, “ubicación”, “productos”, user (taula on tenim emmagatzemada la informació dels usuaris, amb accés a les dades i a l'aplicació amb els camps Nom, Contrasenya, E-mail).

Així obtindrem com a resultat, la següent estructura de base de les dades del què es compondrà el nostre Datawarehouse, ho podem veure en la figura 4.1 que apareix a continuació, on es pot apreciar el dibuix en format estrella centrat per cada any.

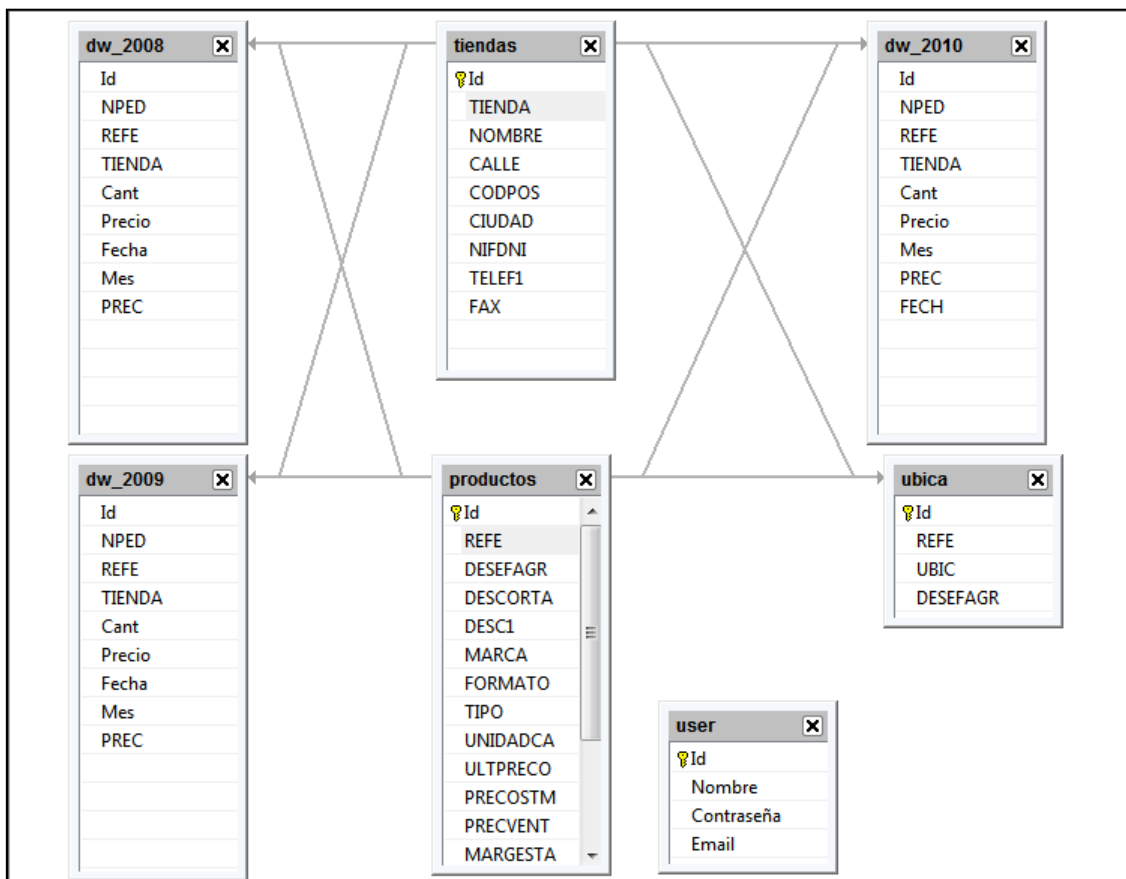


Figura 4.1 Esquema de disseny del Datawarehouse versió 1

Més endavant parlarem dels cubs d'OLAP, allà podrem veure el motiu perquè la primera versió del nostre Datawarehouse, va ser desestimada. El motiu del perquè no es va escollir la 1a versió va ser pel temps de càlcul, ja que per realitzar els nostres estudis de les dades i creacions de les gràfiques, es va obtenir un temps de resposta molt baix (més de 40 segons), amb pics fora dels límits d'execució dels scripts en php, que té configurat en el servidor on hem realitzat les proves i com a conseqüència també en l'accés al nostre servidor http, en aquest cas, de l'Apache, donant un temps de resposta superior al seu límit. Per aquests motius va fer que la versió 1 del nostre Datawarehouse fos una opció desestimada, ja que volem un temps de resposta alt i amb accés al nostre BI mitjançant la utilització d'Internet.

4.3 2A VERSIÓ DEL DATAWAREHOUSE: PARTICIONAT VENDES-ANY.

La idea en aquesta segona versió del Datawarehouse, va ser la de dividir la informació, per algun mecanisme, per ajudar al gestor de base de dades a l'hora de treballar amb els milions de registres que tenim de les taules de vendes dels anys 2008, 2009 i 2010.

Es va escollir el concepte de particionar (partitionim) la informació, continguda en les taules de vendes per mesos. Aquest particionament, s'ha de definir a l'hora de crear les taules, ja que és important la seva definició per tal de saber quines són les particions que es volen tenir en cada taula. En aquest cas les taules es van dividir en 13 particions, el perquè del número 13, ve degut a que tenim els 12 mesos de l'any, i l'últim camp de partició és per les dades que estan fora d'aquest rang, en cas d'una nova introducció de dades i que aquestes no entrin en el particionament.

Seguidament, s'explica com s'ha realitzat aquest particionament de les taules vendes.

4.3.1 PARTICIONAMENT DE TAULES

El motiu d'aquest particionament, és que un dels objectius que trobem dins el sistema Business Intelligence, és el factor velocitat a l'hora d'obtenir la informació. És a dir, que les consultes es realitzin en el menor temps possible, intentant evitar els mateixos problemes de rendiment que podríem trobar-nos en la versió anterior (dades en cru), o com podem trobar en els sistemes operacionals (ja que per generar un informe amb gràfiques poden tardar fins i tot hores o dies, tenint personal dedicat).

Per evitar aquest problema de velocitat i rendiment, a l'hora de dissenyar físicament el DW, utilitzarem el mètode de particionament. El motiu d'utilitzar aquesta iniciativa, és perquè en les taules del nostre DW disposem de milions de registres, que se situen tots ells en una única taula, provocant així que el gestor de base de dades hagi de moure totes aquestes dades en el moment de consultar una petició. La idea es basa en trossejar en petits segments la taula, fent que augmenti el rendiment i la velocitat del sistema.

El particionament, ens permet distribuir la informació en porcions dins d'una taula, aquestes dades s'agreguen dins d'una porció (sub-taula) depenent d'unes regles. Per poder realitzar aquest particionament, existeixen dues formes d'implementació:

- **Manual:** La partició es podria fer en el procés de càrrega en l'ETL, creant les taules fent una separació de les dades de vendes, per mesos i anys. Amb aquesta partició del nostre sistema Business Intelligence, hauria de ser capaç de gestionar aquest particionament lògic de les dades. En cada moment, hauríem de saber a quina taula necessitem consultar o inserir les dades generades. Aquest fet provoca una gestió molt bona en el tema de consultes a la base de dades, ja que hem de tenir la capacitat de poder controlar tota la gestió de les taules del nostre Datawarehouse. Aquesta opció causa una problemàtica, fent que la gestió del particionament sigui més complexa.
- **Automàtic:** Les diferents parts de les taules, podran ser emmagatzemades en diferents ubicacions del sistema de forma automàtica, segons ens permet el SGBDR (Relational Database Management System) que estem utilitzant. La gestió del particionament es realitza de forma automàtica i totalment independent de l'usuari, perquè des de fora es permeti veure la taula de vendes sencera sense particionament, que en realitat la informació està repartida en diverses sub-taules dins de la mateixa taula de vendes. Aquesta gestió la realitza de forma automàtica, el motor de la base de dades a l'hora d'introduir els registres dins la taula, com podem veure en la següent figura 4.2, que mostra un esquema de com queda la distribució de la informació.

Per exemple, sí en la sortida del procés ETL, tenim un registre que té el valor "X" dins el camp definit com a regla de partició dins la taula, el registre s'introdueix en la sub-taula determinada segons valor que apareix en el camp utilitzat pel particionament. Encara que l'usuari només veu la taula gran de vendes, la informació està organitzada en una estructura de sub-taules.

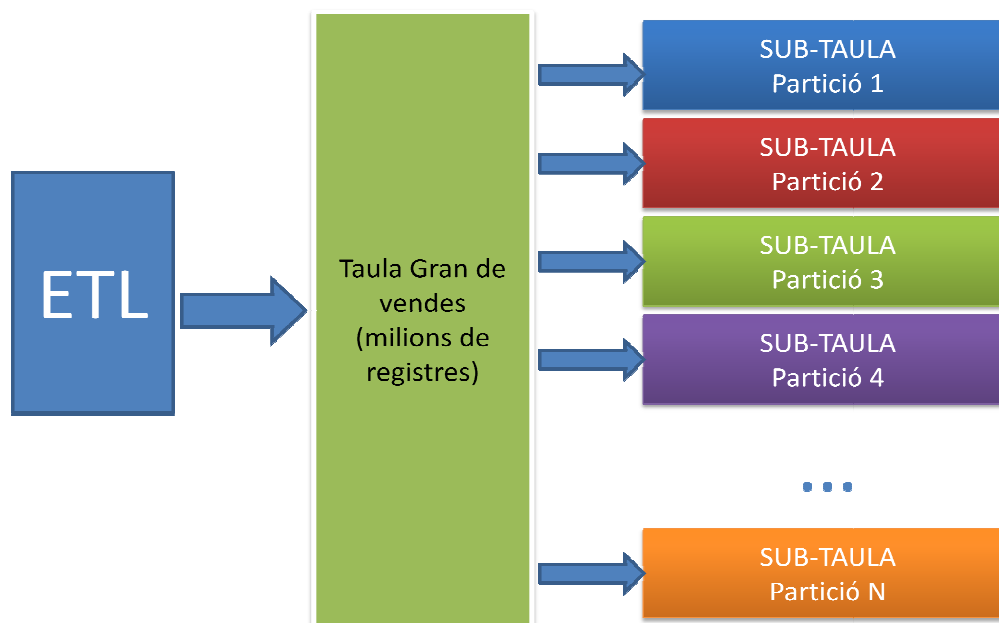


Figura 4.2 Mostra el sistema de particionament d'una taula amb MySQL

Aquest particionament depèn del tipus de Base de dades que s'utilitzi, ja que es pot realitzar aquesta partició en format horitzontal o vertical, és a dir, per fila o columna. En el nostre cas, el sistema és MySQL, que treballa amb el particionament horitzontal (per files).

4.3.1.1 TIPUS DE PARTICIONAMENT DE TAULES AMB MySQL.

MySQL té implementada la partició en horitzontal, on podem escollir entre els 4 tipus següents:

- **RANGE:** L'assignació dels registres dins de la taula, es realitza en particions segons un rang de valors definits sobre un determinat camp de la taula, indicant el nombre de particions a crear, i per a cada partició, el rang de valors que serà la condició. De manera que, quan es vulgui introduir un registre a la base de dades, ha de tenir el valor indicat dins del rang de la columna o expressió indicada, fent que el registre s'inclogui en aquesta partició.
- **LIST:** L'assignació de registres de la taula en múltiples particions, es realitza segons una llista de valors definida sobre una determinada columna de la taula, indicant el nombre de particions a crear, i per a cada partició. La llista de valors serà la condició

per introduir-se en la sub-taula, de forma que la llista de valors ha de contenir tots els valors dels què podem dividir.

- **HASH:** Aquesta partició, està pensada per repartir la informació de forma equitativa, dins de les múltiples particions. Mentre que “LIST” i “RANGE”, la divisió la decidim nosaltres, en la què podem tenir particions buides i d’altres amb una quantitat de registres alta.
- **KEY:** Aquesta opció és similar a “HASH”, la funció la proporciona el propi MySQL automàticament, un dels requisits ha de ser la clau primària de la taula o índex únic, com a camp de particionament.

A l’hora de crear el particionament, ens vam trobar amb un problema important a l’hora de particionar les taules del Datawarehouse, ja que aquesta propietat no es va implementar fins la versió de MySQL 5.1, aquesta circumstància va causar problemes a l’hora de la creació de les taules. Després de molts intents per trobar el perquè no funcionava i no es realitzaven les particions de forma correcta, vam trobar que la causant era que la versió del servidor on es realitzaven totes les proves, disposava d’una versió de MySQL inferior a 5.1 (5.0.27), i és per aquest motiu, que es veia afectada la correcta realització del particionament.

Aquest fet feia generar un error a l’hora de la creació de les taules amb particionament. Apart d’aquest tema, existeix una altra limitació en el particionament que és molt important, ja que només la partició es pot realitzar per camps amb valor numèric. Aquest fet, fa que la primera idea que es va tenir per utilitzar directament el camp “FECHA” de la taula de vendes, no es pogués fer servir, ja que el format “date” no es considera com un número, per això, es va crear un nou camp a la taula anomenat mes, on tenim una extracció del mes del camp “FECHA”.

4.3.2 PARTICIONAMENT APLICAT A LES TAULES DEL DATAWAREHOUSE

El particionament de les taules de vendes dels anys 2008, 2009 i 2010, s’han trossejat pel camp mes amb l’opció de format RANG, fent que hi hagi 13 particions (12 pels mesos, més un últim valor, per si hi hagués algun fora de límit). En el següent codi sql podem veure com s’ha utilitzat l’opció del particionament, en la creació de les taules de vendes del nostre Datawarehouse.

```

CREATE TABLE IF NOT EXISTS `dw_20xx` (
  `Id` bigint(20) NOT NULL,
  `NPED` varchar(14) DEFAULT NULL,
  `REFE` varchar(16) DEFAULT NULL,
  `TIENDA` varchar(11) DEFAULT NULL,
  `Cant` double NOT NULL,
  `Precio` double NOT NULL,
  `Fecha` date NOT NULL DEFAULT '9999-12-31',
  `Mes` int(11) NOT NULL,
  `PREC` double DEFAULT NULL
) ENGINE=MyISAM DEFAULT CHARSET=latin1

PARTITION BY RANGE (Mes)
(PARTITION p1 VALUES LESS THAN (1) ENGINE = MyISAM,
PARTITION p2 VALUES LESS THAN (2) ENGINE = MyISAM,
PARTITION p3 VALUES LESS THAN (3) ENGINE = MyISAM,
PARTITION p4 VALUES LESS THAN (4) ENGINE = MyISAM,
PARTITION p5 VALUES LESS THAN (5) ENGINE = MyISAM,
PARTITION p6 VALUES LESS THAN (6) ENGINE = MyISAM,
PARTITION p7 VALUES LESS THAN (7) ENGINE = MyISAM,
PARTITION p8 VALUES LESS THAN (8) ENGINE = MyISAM,
PARTITION p9 VALUES LESS THAN (9) ENGINE = MyISAM,
PARTITION p10 VALUES LESS THAN (10) ENGINE = MyISAM,
PARTITION p11 VALUES LESS THAN (11) ENGINE = MyISAM,
PARTITION p12 VALUES LESS THAN (12) ENGINE = MyISAM,
PARTITION p13 VALUES LESS THAN (13) ENGINE = MyISAM);

```

Fitxer SQL de creació de les taules de vendes.

4.4 CUBS OLAP

Els cubs OLAP que provenen de les sigles de “On-Line Analytic Processing”, són les eines en què se centren en la capacitat d’analitzar i explorar les dades. Aquest fet, ens permet canviar l’enfocament típic de què està passant?, mitjançant les eines de generació d’informes i de gràfics, per passar a perquè està passant?...

Per descobrir les incògnites, els usuaris que accedeixen al BI, han de navegar entre les dades, per analitzar amb detall i descobrir patrons de vendes que es troben dins les dades històriques. Als cubs OLAP, ens proporcionen un anàlisi interactiu de diferents dimensions de dades, per exemple en temps, productes, establiments, etc, diferenciant per diversos nivells de detall com anys, mesos, trimestres, etc.

Per realitzar l'anàlisi mitjançant els cubs d'Olap, hem de saber quines són les dimensions que disposem d'informació, fent que els directius de l'empresa puguin consultar les diferents dimensions dins del negoci. Aquests detalls estan estructurats per diferents nivells, en el nostre cas, la informació de vendes està per anys i per mesos.

Els sistemes OLAP, han de tenir els següents requeriments:

- Anàlisis complexos de les dades.
- Anàlisi de dades amb diferents punts de vista.
- Suportar anàlisis complexos de dades amb un gran volum d'informació.

OLAP és la solució pel tractament de dades, per tal d'agilitzar la consulta d'un gran volum de dades i les consultes amb l'existència d'informació redundant.

Aquesta funció, ens proporciona la capacitat de dividir la informació en parts més petites, perquè aquests siguin analitzats en diferents punts de vistes.

En la següent figura 4.3 podem observar un cub amb 3 dimensions: temps, productes, establiments. Depenent de la consulta que realitzem podrem visualitzar les vendes que s'han realitzat per un establiment, mes, any o producte, donant el resultat en trimestres, semestres, varis productes, etc. Com podem observar en la següent imatge, tenim les dimensions del nostre cub, amb accés de les dades del Datawarehouse.

Gener	Març		
	Establiment 1	Establiment 2	Establiment 3
Producte 1	212	534	254
Producte 2	21	46	33
Producte 3	310	321	200
Producte 4	120	234	131
Producte 5	43	78	55
Producte 6	12	32	21

Figura 4.3 Exemple de cub OLAP del nostre Datawarehouse en la taula vendes-any

4.5 3A VERSIÓ DEL DATAWAREHOUSE: PARTICIONAT DE VENDES-ANY-ESTABLIMENT

En aquesta tercera versió del Datawarehouse, l'idea és la de combinar els dos mètodes de particionament manual i automàtic, per tal de millorar el rendiment en algunes de les consultes de què disposem dins de l'aplicació Business Intelligence.

Aquesta fusió dels mètodes, ha consistit en crear manualment una taula per cada establiment i per any, amb el particionament automàticament per mes, fent que les dades de les quals disposem, siguin centrades en establiment i any.

La causa d'aquesta decisió seria la creació de múltiples taules de vendes, en aquest cas de 76 establiments x 3 anys (2008, 2009 i 2010) aconseguint un total de 228 taules. Aquesta solució s'ha implementat amb cubs d'Olap independents per cada establiment, fent l'accés a les dades a través de consultes amb recerca per establiment per poder reduir el temps considerablement.

Amb la tercera versió del Datawarehouse, disposem de les dades ben repartides, fent que les magnituds de registres de cada taula, es redueixin en un gran nombre, i així permeti l'ajuda a la reducció del temps de computació en l'accés a les dades. En la figura 4.4 podem observar que l'accés a la informació en cada taula pot tenir l'accés molt ràpid.

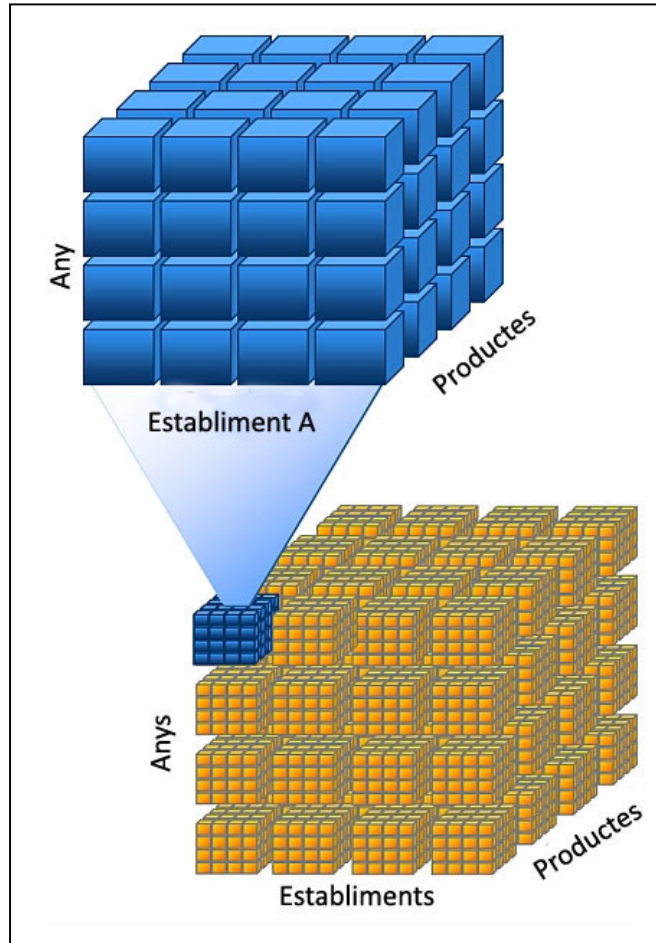


Figura 4.4 Sub-Cubs de OLAP del Datawarehouse

CAPÍTOL 5

SISTEMA WEB DE GENERACIÓ DE GRÀFICS

En aquest capítol s'explica quines característiques té la interfície d'usuari, que s'ha dissenyat en el prototip del BI, la selecció dels gràfics que s'han generat, quines opcions s'ofereixen al usuari, per la consulta de les dades que conté el Datawarehouse.

5.1 PATRÓ DE CREACIÓ DE GRÀFIQUES

Per implementar el sistema web de generació de gràfics, es va fer un estudi de quin tipus de gràfiques eren les més adequades per les peticions d'informació, fent referència a les necessitats inicials de generació de gràfiques demanades pel client. Amb aquest estudi es va trobar un patró de disseny dins de les gràfiques que necessitàvem, tenint aquestes 3 dimensions que segueixen un patró de disseny en format arbre, tenint el valor de l'eix x com a variable de temps en mesos, en l'eix y hi haurà les variables del càlcul de volum de vendes, volum de facturació, etc i en la dimensió de les dades, tindrem les variables que utilitzarem per a fer la comparació, com poden ser productes, establiments o mitjanes.

En la següent figura 5.1 tenim un esquema de l'arbre de generació de gràfiques, en el qual mostrem la distribució de les dades. I en la figura 5.2 tenim un exemple d'un gràfic base.

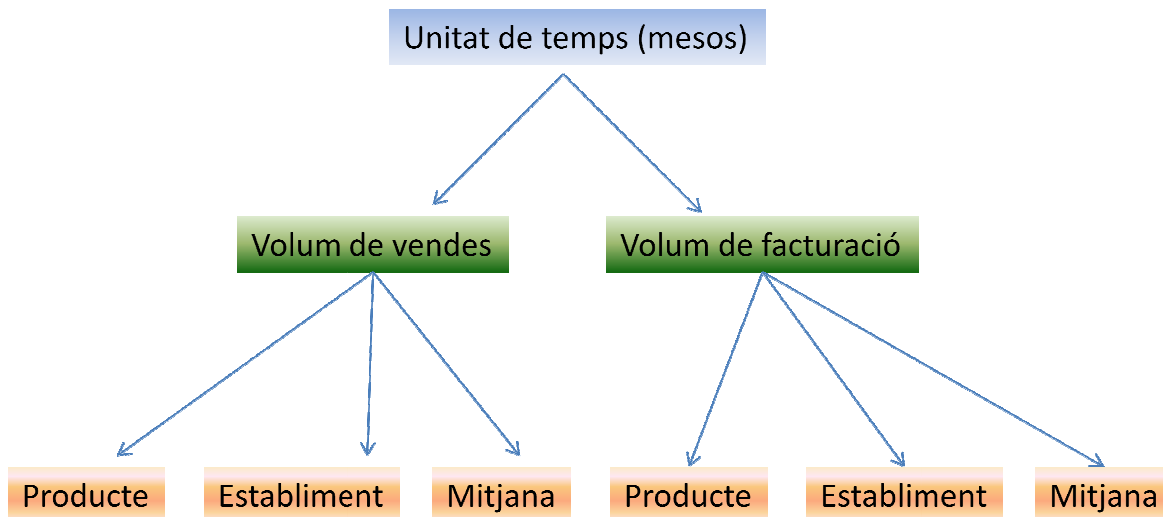


Figura 5.1 Arbre de decisió per a la creació de les gràfiques

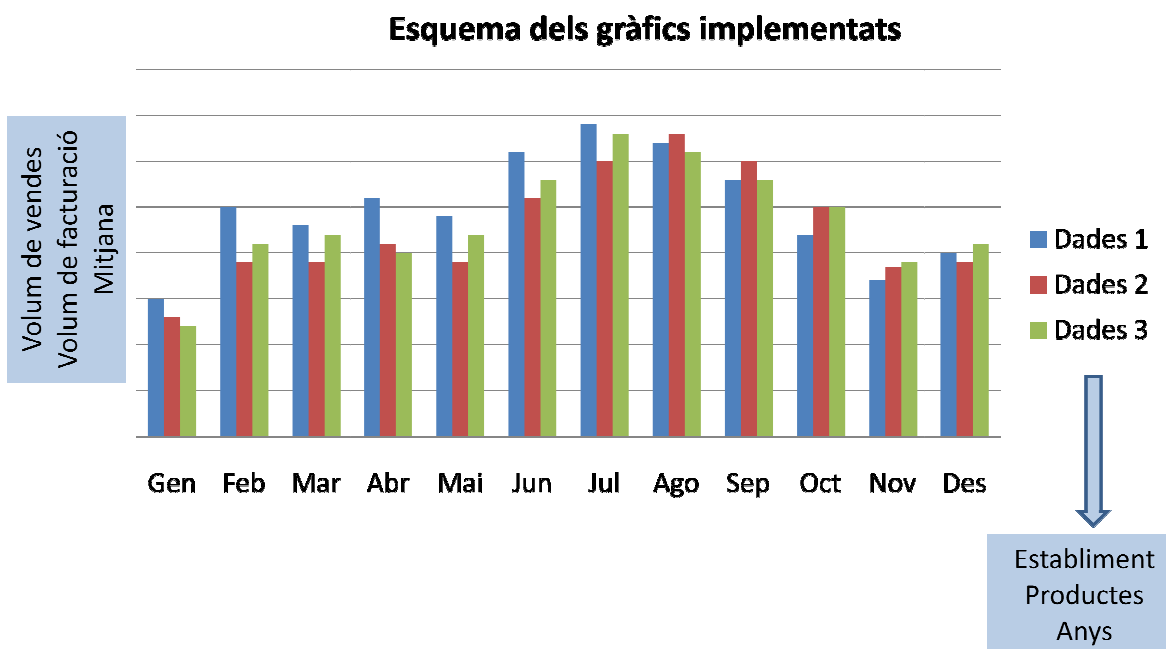


Figura 5.2 Esquema base de la generació de gràfics

5.2 GENERACIÓ DE GRÀFICS AMB PHP: JPGRAPH

Per poder generar les gràfiques desitjades per les consultes que es realitzen, s'ha utilitzat la llibreria gràfica jppgraph, donant la facilitat de crear diferents tipus de gràfiques amb llenguatge PHP i també la incorporació dels gràfics en una pàgina html.

S'ha utilitzat la versió no professional, ja que amb la varietat de gràfics que ens ofereix la llibreria en la versió free, té la suficient potència per mostrar les dades amb el format que desitgem, perquè la majoria de gràfiques tenen el format de barres, de línies, o la combinació de les dues.

Per aquest motiu totes les gràfiques que es generen en l'aplicació, han estat creades amb aquesta llibreria de gràfics de php.

5.3 SISTEMA LOGIN

S'ha implementat un sistema de login, perquè volem restringir l'accés a l'aplicació a usuaris que no estiguin identificats, ja que hi ha dades molt importants per l'empresa, i interessa establir una seguretat en l'entrada.

Aquest login podria ser opcional, ja que la primera idea, va ser la d'inserir l'aplicació Business Intelligence de creació de gràfiques dins d'un portal del grup alimentari, creat amb un CMS com Joomla, Drupal, Wordpress... Aquest portal, creat amb un sistema de gestió de contingut ja porta el seu propi sistema de gestió d'usuaris.

Per tant, el sistema de login que s'ha desenvolupat conté un formulari d'entrada, on tenim d'introduir el nom d'usuari i la contrasenya, tal i com es pot apreciar en la següent figura 5.3. El sistema comprova si existeix un usuari amb aquest nom, i verifica sí la contrasenya és correcta. Per tenir més seguretat, la contrasenya esta implementada amb l'algoritme sha1 d'encryptació que a l'hora de desar en la base de dades dins la taula user.

També hem implementat el sistema de canvi de contrasenya per part de l'usuari, fent que l'aplicació sigui més segura, al poder modificar la contrasenya en qualsevol moment.



Figura 5.3 Sistema Login d'entrada al sistema Business Intelligence

5.4 GENERACIÓ DE GRÀFICS IMPLEMENTATS

En el menú principal, podem trobar els enllaços amb accés als diferents formularis per la creació de les diferents gràfiques que s'han predefinit, fent que sigui un entorn de fàcil accés, tant per ordinador, com per qualsevol dispositiu amb accés a Internet.

En el menú, podem observar l'accés al llistat de gràfiques que podem realitzar, d'aquesta manera mostrem els resultats obtinguts de la consulta creada per l'usuari en l'aplicació. Les 8 gràfiques que s'han predefinit són:

- Establiment: Volum de Facturació entre anys i botiga.
- Establiment: Volum de productes venuts per establiment i anys.
- Establiment: Volum de productes venuts per anys amb el DW de 3a versió (experimentació)
- Establiment: Mitjana de vendes per producte.
- Producte: Mitjana de vendes de producte per tot el grup.
- Producte: Volum de vendes per anys i mesos.
- Producte: Volum de vendes de producte entre dos establiments.
- Producte: Volum de vendes entre productes per establiment i any.

Podem veure en la següent figura 5.4, una imatge de com es mostra el menú principal, quan és selecciona l'opció d'establiment, on és mostren les gràfiques predefinides que podem generar.



Figura 5.4 Menú principal de l'aplicació Business Intelligence

5.4.1 GRÀFIQUES PER ESTABLIMENT

5.4.1.1 VOLUM DE FACTURACIÓ ENTRE ANYS.

Amb l'entrada de menú dins d'establiment, podem realitzar una comparació del volum de vendes existents entre dos anys. Dins del mateix establiment i amb la informació del volum de facturació mensual i anual, podem veure l'evolució de la facturació entre anys. Amb aquesta informació podem realitzar la comparació i així tindrem una visió més general de les vendes de l'establiment. Ho podem veure en la figura 5.5., que segueix a continuació, on podem observar el volum de facturació de l'establiment Jaume 1 de Manresa, en els anys 2008 i 2009.

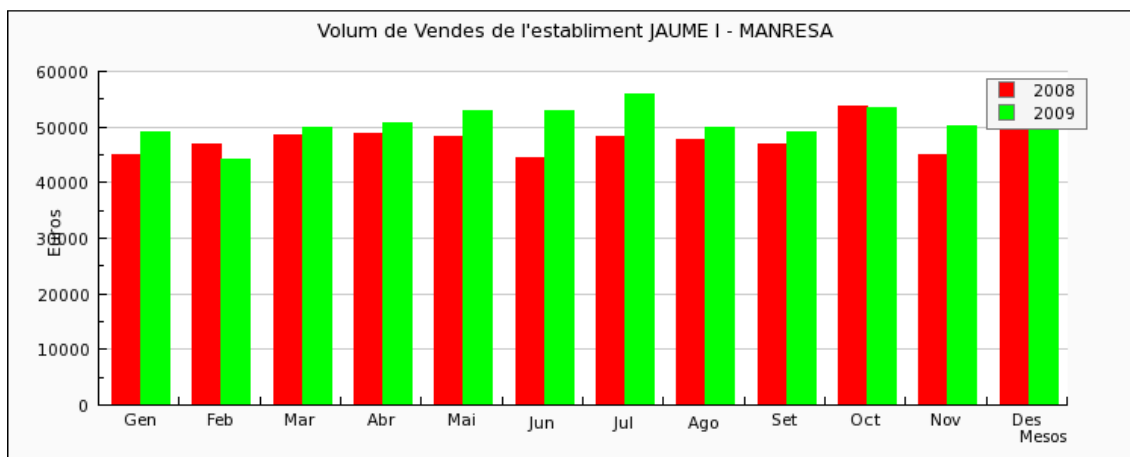


Figura 5.5 Volum de facturació de l'establiment Jaume 1 de Manresa (2008-2009)

5.4.1.2 VOLUM DE PRODUCTES VENUTS-ANYS

Aquesta entrada de menú, ens serveix per fer l'estudi del volum de productes que es mouen per botiga, en diferents períodes de temps que tenim emmagatzemats amb els històrics. Aquesta informació és útil pel departament de logística, ja que aquestes dades són importants per a ells, per saber quin volum de venda té cada botiga.

En la figura 5.6, es mostra la informació del volum de productes venuts de l'establiment "DRETS" de Manresa, en els anys 2008, 2009 i 2010 de forma mensual. Podem observar que els últims dos mesos de 2010 el valor és 0, això és degut a que no disposem de dades, ja que el llistat de vendes d'aquesta botiga, dels mesos de novembre i desembre de 2010 no van cedides per l'empresa en el seu dia. En la figura també podem apreciar que hi ha un augment de vendes pel què fa al 2010, respecte als altres anys que es mantenen més o menys al mateix nivell.

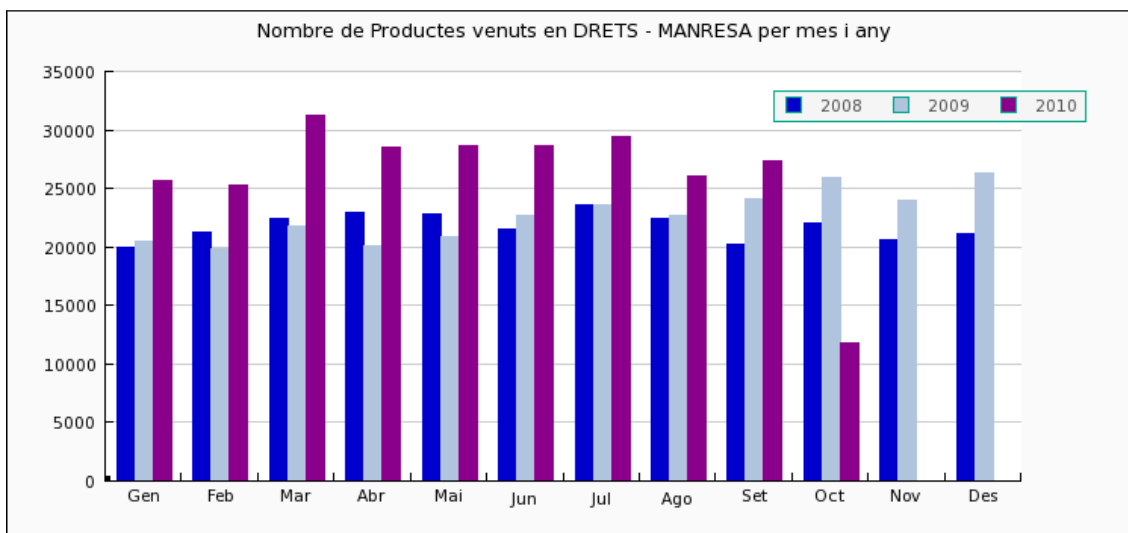


Figura 5.6 Productes venuts per l'establiment Drets de Manresa entre anys

5.4.1.3 VOLUM DE PRODUCTES VENUTS-ANYS (3A VERSIÓ DEL DW)

Aquesta entrada mostra la mateixa informació que l'apartat anterior, amb una diferència considerable, és que aquí tenim implementada l'entrada anterior amb la segona versió del Datawarehouse, distribuint la informació en 3 taules grans, una per cada any i particionada per mesos. En aquesta entrada, tenim distribuïda la informació en una taula independent per establiment i per any, implementada amb la 3a versió del Datawarehouse, podent accedir a la taula específica de l'establiment per any i particionada per mesos.

Aquesta entrada de menú, s'ha realitzat per fer l'estudi de la comparació, en temps d'execució del mateix nombre de còmput, entre la segona i tercera versió del DW, demostrant que el temps es redueix considerablement de la tercera versió envers a la segona versió implementada del Datawarehouse, com podrem veure seguidament.

5.4.1.4 COMPARACIÓ DE TEMPS D'EXECUCIÓ ENTRE 2A I 3A VERSIÓ DW

En aquest apartat, comparem el temps de resposta entre les dues versions del Datawarehouse, que s'han implementat fent una comparació de temps, en les mateixes condicions.

Per realitzar aquest càlcul, s'ha implementat una funció d'escriptura d'un fitxer log, per obtenir un registre del temps d'execució entre versions. S'han realitzat cinc proves de comparació de temps per a 8 establiments diferents, on tenim diferents nombres de vendes realitzades en cada establiment. En la següent taula número 5.1 veurem els temps obtinguts, el càlcul d'una mitjana de temps per botiga en la segona versió del Datawarehouse (OLAP de vendes, any, mesos).

	Prova1	Prova2	Prova3	Prova4	Prova5	Mitjana
T03	9 seg.	11 seg.	9 seg.	11 seg.	11 seg.	10,2 seg.
T04	12 seg.	11 seg.	12 seg.	12 seg.	12 seg.	11,8 seg.
T07	8 seg.	8 seg.	9 seg.	9 seg.	8 seg.	8,4 seg.
T09	10 seg.	10 seg.	9 seg.	10 seg.	9 seg.	9,6 seg.
T10	13 seg.	13 seg.	13 seg.	13 seg.	13 seg.	13 seg.
T101	9 seg.	9 seg.	9 seg.	9 seg.	9 seg.	9 seg.
T102	9 seg.	10 seg.	9 seg.	9 seg.	10 seg.	9,4 seg.
T103	11 seg.	12 seg.	12 seg.	11 seg.	11 seg.	11,4 seg.

Taula 5.1: Temps d'execució amb segona versió del Datawarehouse

En la següent taula 5.2, tenim els temps de les mateixes proves, que s'han realitzat per la 3a versió del Datawarehouse, on s'ha implementat taules amb OLAP vendes tenda, any i mesos.

	Prova1	Prova2	Prova3	Prova4	Prova5	Mitjana
T03	4 seg.	5 seg.	5 seg.	4 seg.	4 seg.	4,4 seg.
T04	9 seg.	6 seg.	5 seg.	5 seg.	5 seg.	6 seg.
T07	6 seg.	2 seg.	2 seg.	2 seg.	2 seg.	2,8 seg.
T09	7 seg.	3 seg.	3 seg.	3 seg.	3 seg.	3,8 seg.
T10	7 seg.	7 seg.	7 seg.	7 seg.	7 seg.	7 seg.
T101	7 seg.	3 seg.	3 seg.	3 seg.	2 seg.	3,6 seg.
T102	6 seg.	3 seg.	3 seg.	3 seg.	3 seg.	3,6 seg.
T103	8 seg.	8 seg.	4 seg.	5 seg.	5 seg.	6 seg.

Taula 5.2: Temps d'execució amb tercera versió del Datawarehouse

Després d'analitzar totes les proves i consultar la taula de temps, podem observar en les dades obtingudes, la mitjana de la tercera versió és sempre inferior respecte a la segona versió del Datawarehouse. Mirant els resultats obtinguts en les proves, assolim una mitjana de reducció de **5,72 segons**, per tant, el temps de còmput es redueix considerablement. En alguns casos, podem observar una reducció considerable del 60% de temps d'execució, demostrant que amb les dades distribuïdes en alguns dels casos, pot ser la solució per reduir el temps de consulta. En canvi, per a gràfiques on volem obtenir valors de tots els establiments del grup alimentari, es pot demostrar que tenir les dades centrades en una taula serà més efectiu.

5.4.1.5 MITJANA DE VENDES DE PRODUCTES

Amb aquesta opció, obtenim una mitjana de vendes d'un producte entre els anys 2008, 2009 i 2010. Aquesta informació és mostrada en la gràfica utilitzant el format de línia, donant informació de la mitjana de vendes que es mostrada en la gràfica, fent que això pugui ser informació útil, per realitzar més estudis per a futures prediccions de vendes amb series temporals, o altres algoritmes de pronosticacions de vendes que existeixen, ja que en molts dels estudis, es necessari conèixer la mitjana de vendes per mes i per establiment.

En la figura 5.7 es veuen les vendes del producte amb referència 001012 amb el nom "Berberecho Baymar pequeño OL-120" en els anys 2008, 2009 i 2010 per l'establiment Numancia de Manresa. Les columnes en forma de barra i la línia de color verd ens mostra la mitjana d'aquestes vendes per mes.

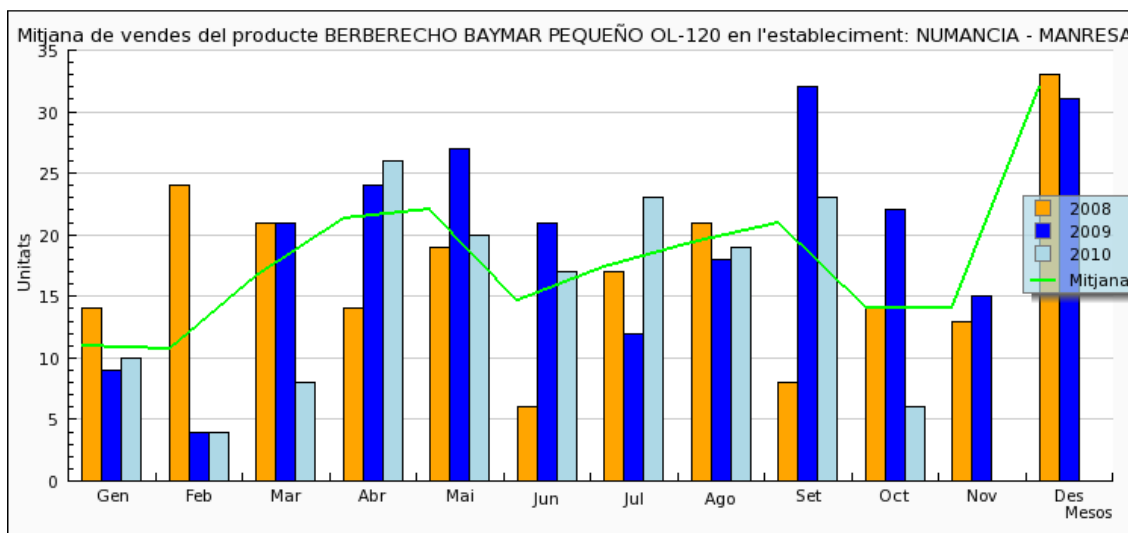


Figura 5.7 Mitjana de vendes del producte en l'establiment Numància de Manresa

5.4.2 GRÀFIQUES PER PRODUCTE

5.4.2.1 VOLUM DE VENDES-ANYS.

Amb aquesta entrada de menú, podem estudiar el volum de vendes que tenim d'un producte al llarg dels anys, on podríem detectar èpoques de l'any on les vendes del producte tendeixen a augmentar o disminuir, per després fer un estudi del motiu del perquè existeix aquesta tendència. Posteriorment es podria establir una estratègia de màrqueting per fer augmentar les vendes en aquestes èpoques en què el volum és més baix. Podem observar en la figura 5.8, on mostra la informació del producte "Berberecho Baymar Pequeño Ol-120", amb informació en mesos i anys, del número vendes d'aquest producte en concret. Podem observar que en març del 2008 va ser el mes que va generar més vendes, en canvi podem apreciar que en l'any 2009 el mes que es va vendre més aquest producte va ser l'Octubre. En definitiva, observant els dos anys en general podem apreciar en què l'any 2008 la venda de "Berberecho Baymar Pequeño Ol-120" més elevada va ser als principis d'any mentre que en l'any 2009, la venda va ser cap a finals d'any. També podem observar que a finals de l'any 2009 hi ha un augment considerable de la venda d'aquest producte respecte als últims mesos de l'any 2008. Es tractaria doncs, de fer un estudi del motiu d'aquest increment de vendes investigant-ne la causa, degut a algun factor en especial, ja sigui una campanya de màrqueting o la situació del producte dins del establiment, per alguna baixada de preu, oferta puntual del producte, etc.

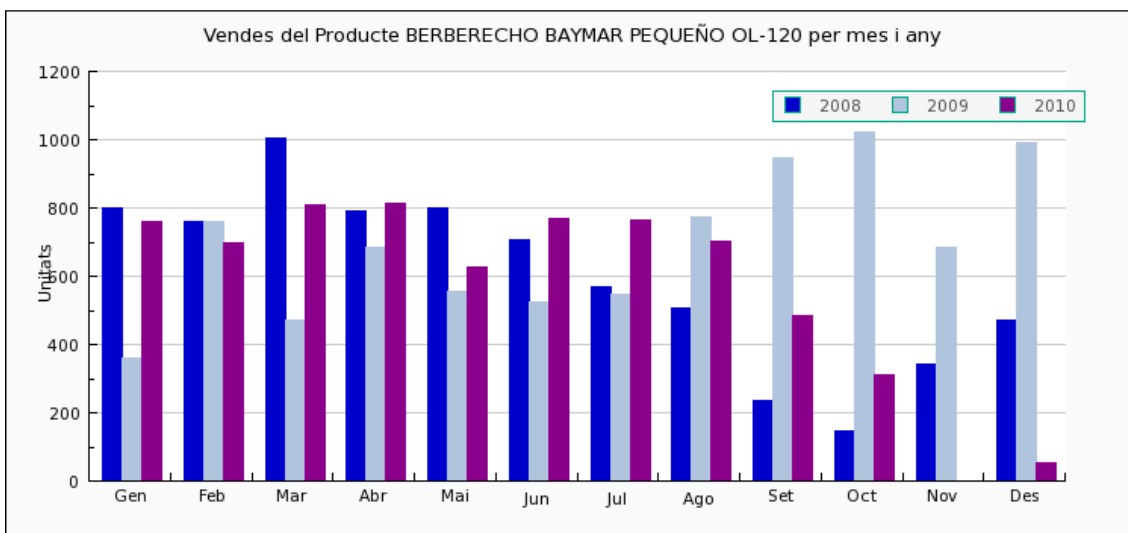


Figura 5.8 Vendes del producte “Berberecho Baymar Pequeño” entre anys

5.4.2.2 VENDES PER ESTABLIMENTS-ANYS

La comparació de productes entre dos establiments per any, és una opció que dóna molta utilitat per l'estudi de les dades als equips directius, ja que ens ofereix una eina de molta potència, per comparar un mateix producte entre dos establiments. Aquesta informació fa que sigui una font d'informació molt important, per realitzar estudis de vendes, per revisar molts aspectes que poden afectar en la compra d'un producte o un altre.

En la Figura 5.9 podem observar el resultat de comparar les vendes del producte “Sardina CUCA aceite oliva 120G” entre els establiments de Numancia i d'Economat en l'any 2008.

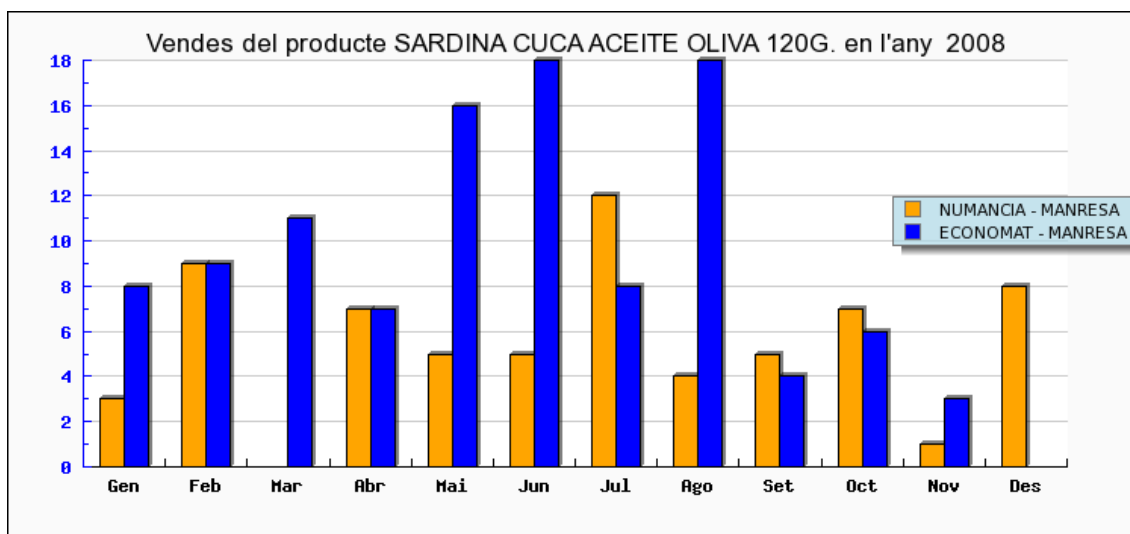


Figura 5.9 Vendes del producte "Sardinas CUCA OLIVA 120G" en 2008

Observant les vendes en l'any 2008, podem veure algunes diferències bastant importants entre un establiment i l'altre, entre els mesos de maig i juny on predomina la venda d'aquest producte és a l'Economat – Manresa, i el mateix passa en l'altre establiment de Numancia, on l'augment de venda és superior en els mesos de juliol, octubre i desembre.

Un altre exemple és la figura 5.10, que ens mostra la gràfica del producte "Pate Finas Hiervas de Terradelles de 135G" dels establiments "La Parada" i "Xup" de Manresa. Podem veure que les vendes del producte són molt més altes en l'establiment La Parada que en el establiment del Xup, tot i que les dues botigues estan en la mateixa ciutat, la distribució geogràfica pot ser un punt a estudiar a l'introduir un producte.

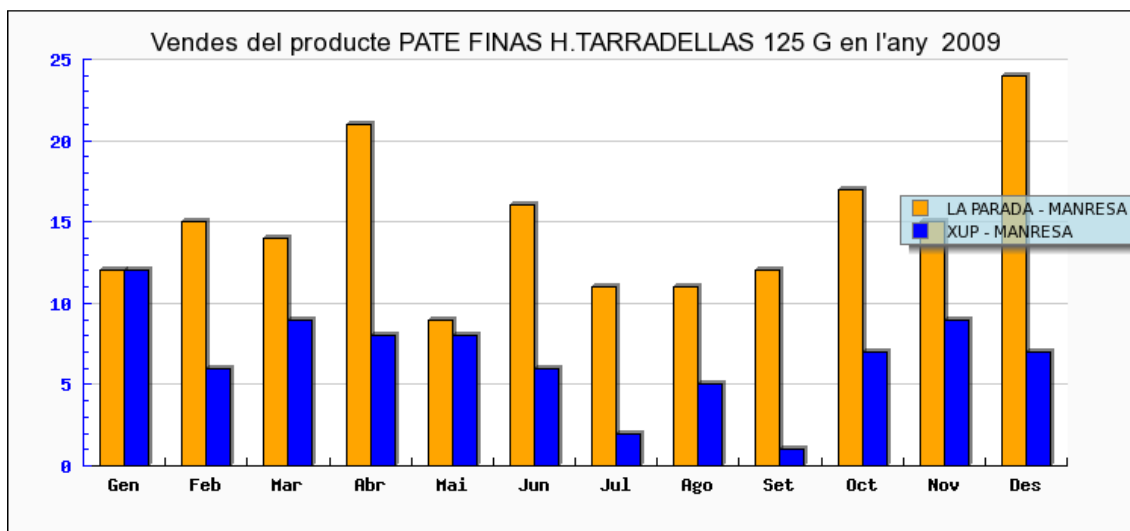


Figura 5.10 Vendes producte "Pate fines h. Tarradelles en 2009 entre establiments

5.4.2.3 MITJANA DE VENDES TOTALS

Amb aquesta opció del menú, obtindrem la mitjana de vendes d'un producte dins de tot el grup, entre els anys 2008, 2009 i 2010. Aquesta informació mostra les vendes del producte en forma de columna per cada any, i la mitjana de vendes en forma de línia. Aquesta informació ens serveix per comparar la venda de productes entre tots els establiments i així decidir alguna campanya per un producte a escala de tot el grup.

En la següent figura 5.11 podem veure la mitjana de vendes del producte Aigua de Viladrau de 1.5l.

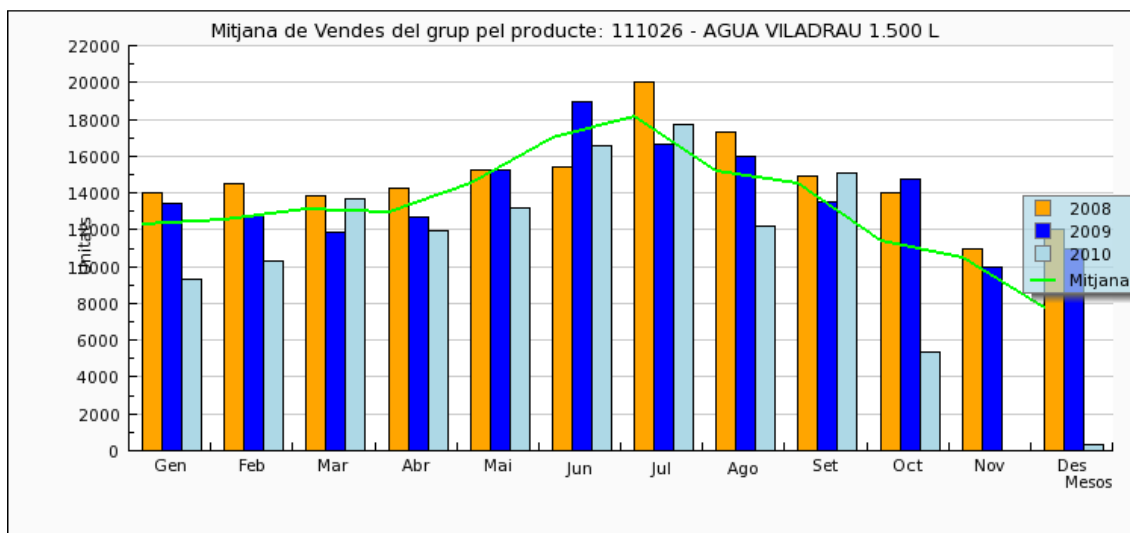


Figura 5.11 Mitjana de vendes d'aigua Viladrau 1,5l per tot el grup

Com podem observar en la gràfica, veiem un augment de les vendes de l'aigua Viladrau en els mesos d'estiu, fet que podem veure que es repeteix en el transcurs dels anys, demostrant que hi ha un patró de vendes, encara que tindríem de considerar diferents variants com per exemple el tipus d'estiu que ha fet en aquests darrers anys, amb el supòsit de si ha estat un any amb altes temperatures es pot haver venut més aquest producte o pel contrari un any de temperatures suaus en aquests mesos s'hauria pogut esdevenir menys vendes.

5.4.2.4 VOLUM DE VENDES ENTRE PRODUCTES-ESTABLIMENT-ANY

Amb aquesta opció del menú, ens serveix per poder comparar les vendes de dos productes en un establiment i per un any determinat. Aquesta informació és molt valuosa, ja que podem comparar productes de la mateixa categoria, i veure quin té més vendes d'on podem extreure informació important, si les vendes són per preu, qualitat, distribució del producte dins de l'establiment, sí afecten les ofertes, etc.

En la següent figura 5.12 podem veure la comparació del número d'ampolles de 1,5l venudes en l'establiment "Numancia" (Manresa) en l'any 2008. Analitzant la gràfica, podem observar que les ampolles de 1,5l de la marca Font Vella supera a les de Viladrau.

En la figura 5.13 podem veure la comparació dels mateixos productes que en la figura 5.10, però en aquest cas per l'establiment Economat de Manresa, i podem observar que en varis mesos, aquest cop Viladrau es va vendre més que Font Vella.

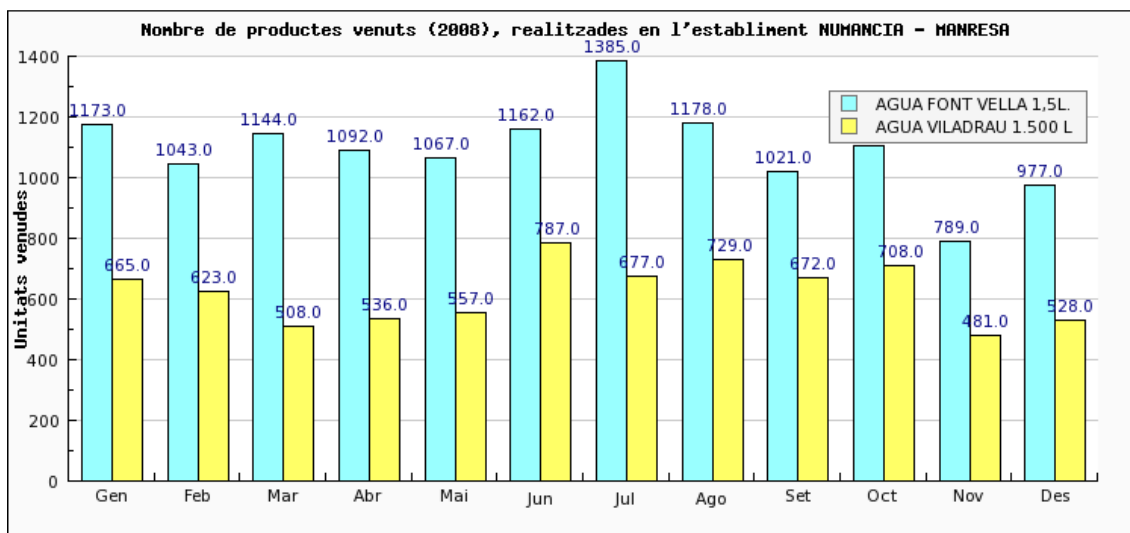


Figura 5.12 Vendes d'aigua de 1.5l entre Font Vella i Viladrau en Numancia(2008)

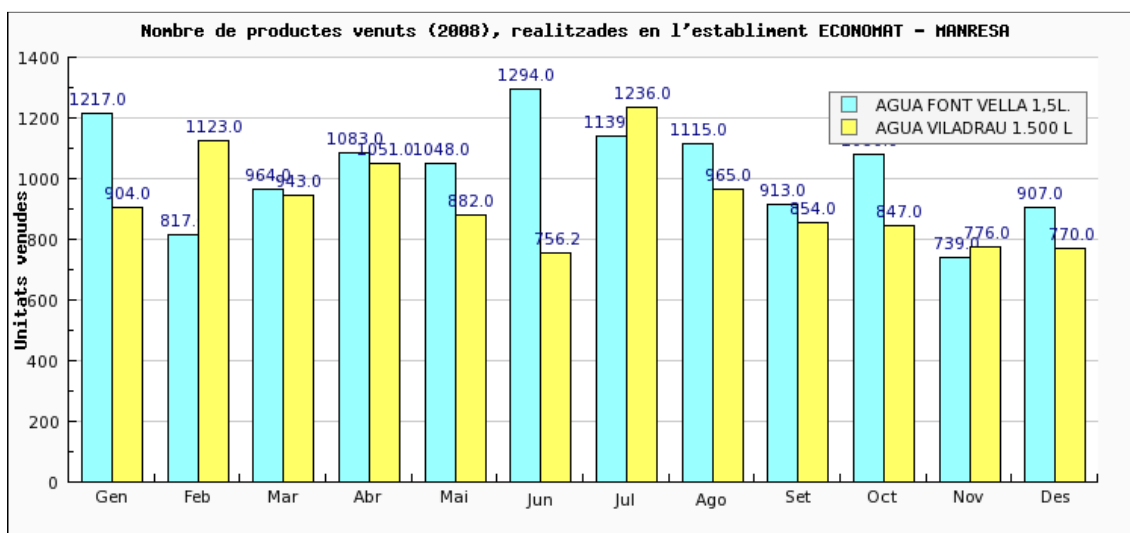


FIGURA 5.13 Vendes d'aigua de 1.5l entre Font Vella i Viladrau en Economat(2008)

CAPÍTOL 6

CONCLUSIONS

6.1 DESENVOLUPAMENT DEL PROJECTE

Aquest projecte, m'ha donat l'oportunitat de fer una introducció força completa al treballar amb grans bases de dades en el món real, i veure molts dels aspectes en què pot ajudar un Business Intelligence dins d'un negoci. Un dels aspectes, comença per tot el què afecta el tractament del gran volum d'informació que pot disposar una empresa, és important saber que encara que disposem de molta informació, això no implica coneixement, ja que sí tenim moltes dades però amb elles no tenim informació de qualitat, no serveix de gran cosa, en canvi sí aquest gran volum de dades, es tracten correctament podem obtenir informació de qualitat, i aquesta la podem transformar en coneixement.

La idea inicial del projecte era la creació un sistema complert de Business Intelligence, però de seguida em vaig adonar que l'aplicació era molt complexa, per realitzar en les hores estipulades en el PFC, per aquest motiu es va decidir realitzar un prototip, per arribar a la demanda del client d'obtenció d'informació gràfica, ja que per realitzar un sistema complex de Business Intelligence es necessitaria més hores de treball per portar-lo a terme fins al final.

Per poder arribar als objectius plantejats d'obtenció de gràfiques, es va necessitar més temps de l'estipulat del què en un principi s'havia previst en la planificació inicial, ja que al començar un projecte, la planificació ha de ser molt acurada i disposar d'un marge de temps per tal d'absorbir qualsevol alteració en l'organització, ja que un petit problema pot modificar tota la planificació inicial prevista.

L'objectiu bàsic d'obtenció de gràfiques, es va complementar amb èxit, fent el sistema web de generació de gràfiques dins del prototip de Business Intelligence. A falta de temps en la realització del projecte i arribar a l'objectiu de treballar les dades amb algoritmes de datamining, és causada per la dificultat d'estudiar amb algoritmes de previsió de series temporals, ja que és molt complexa.

He après que en qualsevol procés de ETL, és un pas bàsic per la realització d'un estudi de dades, detectant quines dades ens aporten informació important per l'empresa, i així poder treballar amb elles, i transformar-les en un format adequat per l'estudi en grans bases de dades. Amb la demostració de la importància de disposar d'una estructura equilibrada, dins del nostre Datawarehouse, podem obtenir un rendiment "òptim", en la recerca de dades, realització de consultes complexes, en les taules compostes per milions de registres i elaborant consultes de forma més optimitzada. D'aquesta manera, podrem millorar el temps de resposta de les dades quan estan tractades i preparades, envers un sistema relacional.

Per obtenir aquest rendiment, es va treballar amb diverses versions del nostre Datawarehouse del projecte, millorant així el temps de resposta en la creació de les gràfiques, per l'elaboració dels informes en temps d'execució.

Utilitzant aquestes metodologies i tecnologia, ajuden a treballar en la transformació d'informació d'un sistema transaccional de referència a una base de dades (Datawarehouse) de forma estructurada i optimitzada.

Així doncs, tenir un business Intelligence dins de l'empresa pot arribar a ser un avantatge competitiu molt gran, fent que l'empresa disposi d'informació addicional i privilegiada per resoldre els possibles problemes que poden sorgir en la gestió, ajudant a realitzar estudis de mercat, detecció de necessitats de negoci, estudis de màrqueting, etc, sent una eina molt important per augmentar la competitivitat de l'empresa dins del mercat empresarial.

6.2 TREBALL FUTUR

El projecte deixa moltes portes obertes, per la millora en la investigació i extracció d'informació del Datawarehouse:

- Millorar en la varietat de consultes, que podrien tenir a disposició els directius de l'empresa amb l'aplicació.
- Recomanar al grup alimentari, la creació d'una targeta client per part dels supermercats, per poder obtenir informació sobre diferents característiques de compres de grups de clients (per edat, sexe, nacionalitat, etc), per poder disposar d'una gran varietat de variables per mostrar i estudiar, observant les característiques per a cada grup de clients.
- Realitzar un estudi de dades amb algoritmes de predicció de vendes, per ajudar a estimar futures vendes en un període de temps i planificant noves estratègies de negoci.
- Crear una versió específica per dispositius tablets, encara que amb el format html i php, és accessible, es podria fer una interfície pròpia per l'accés a l'aplicació generada pel projecte.
- Crear una interfície gràfica, on el client pugui crear les seves gràfiques, escollint quines variables vol introduir en cada eix, fent la creació de les gràfiques de forma més dinàmica i atractiva.

BIBLIOGRAFIA

- [1] Business Intelligence fácil. <http://www.businessintelligence.info/>
- [2] El rincon del BI. <http://churriwifi.wordpress.com/category/business-intelligence/>
- [3] Jose Luis Cano. Business Intelligence: Competir con Informacion. ESADE
http://www.iwith.org/pdf/Libro_BI_Competer_con_Informacion.pdf
- [4] Particionado de tablas en MySQL
<http://blog.whynotsoluciones.com/particionado-de-tablas-en-mysql/>
- [5] Documentation MySQL – Chapter 17.Partitioning
<http://dev.mysql.com/doc/refman/5.1/en/partitioning.html> .
- [6] Carmen Wolff. Implementado un Datawarehouse. 2002
- [7] Teoria del Datewarehouse.
<http://en.juantxu.net/doku.php/dwh> . 2010
- [8] El data warehouse entró en nuestras vidas... para quedarse.
<http://eltamiz.com/elcedazo/2009/06/08/el-data-warehouse-entro-en-nuestras-vidas-para-quedarse/> . 2009
- [9] Business Intelligence para principiantes: ¿qué es y para qué sirve un DataWarehouse?
<http://www.materiabiz.com/mbz/ityoperaciones/nota.vsp?nid=33043>
- [10] Luke Welling / Laura Thomson. Desarrollo web con PHP y MySQL. Anaya Multimedia
2005.
- [11] Most powerful PHP-driven charts JpGraph. <http://jpgraph.net/>
- [12] Simple PHP chart drawing library. Libchart.
<http://naku.dohcrew.com/libchart/pages/introduction/>

BIBLIOGRAFIA

[13] Pchart Documentation.

<http://pchart.sourceforge.net/documentation.php?topic=datastructure>

[14] Roberto Clemente Navarrete Carrasco. Business Intelligence: La necesidad actual.

<http://www.gestiopolis.com/recursos/documentos/fulldocs/ger/binna.htm> 2002

[15] Chris Maxcer. ¿Qué hay detrás del auge de Business Intelligence?

<http://www.help400.com/asp/scripts/nwart.asp?Num=207&Pag=16&Tip=T> Oct 2010

A handwritten signature in blue ink, appearing to read 'Diego Galera Sánchez', is written over two horizontal lines. The signature is stylized and somewhat cursive.

Firmat: Diego Galera Sánchez
Bellaterra, setembre 2011

Resum

Aquest projecte mostra la importància que pot tenir un sistema basat en Business Intelligence, dins d'una empresa o organització, donant una eina per augmentar la competitivitat, treballant les dades que s'obtenen dels diferents sistemes de gestió que hi ha dins l'empresa. Aquest gran nombre de dades històriques les transformarem per formar una base de dades de qualitat, i les explorarem per tal d'extreure'n informació útil en format gràfic, per ajudar a la presa de decisions per part dels directius.

Resumen

Este proyecto muestra la importancia que puede tener un sistema basado Business Intelligence, dentro de una empresa u organización, dando una herramienta para aumentar la competitividad, trabajando los datos que se obtienen de los diferentes sistemas de gestión que hay dentro de la empresa. Este gran número de datos históricos las transformaremos para formar una base de datos de calidad, y las exploraremos para extraer información útil en formato grafico, para la ayuda en la toma de decisiones por parte de los directivos.

Abstract

This project shows the importance that a Business Intelligence based system may have inside a company or an organization, giving a tool to increase competitiveness, working the data obtained from the different systems from the company. This huge number of historical data is going to be transformed, to form a quality database, and explored to extract useful information in a graphic format, in order to help the management to take decisions.