Construcció i explotació d'un magatzem de dades

Joan Fajula i Castany

Universitat oberta de Catalunya
Enginyeria tècnica en informàtica de gestió
Àrea de magatzems de dades
Consultor: José Ángel Martín Carballo
11/01/2007

Dedicatòria i agraïments

Als meus pares, que no només m'han regalat la vida sinó que m'han ensenyat com afrontar-la i gaudir-ne.

Vull agrair també a la Rosana, tot el suport que m'ha donat al llarg de tota la carrera i la paciència mostrada en tots aquests caps de setmana en què ens hem quedat a casa perquè calia entregar una PAC...

Al José Ángel, gràcies pel suport i per la flexibilitat en els moments difícils.

També voldria recordar amb agraïment als creadors de la base de dades Access, aquest fou el primer sistema de gestió de bases de dades amb el que vaig tenir contacte i el que em despertà la curiositat per continuar investigant i aprenent més coses referents a aquests sistemes que al final han esdevingut el camp on es desenvolupa la meva activitat professional quotidiana.

Sense vosaltres aquest TFC no hauria estat possible. Gràcies de tot cor.

Resum

A partir de quatre sistemes operatius diferents que serveixen als departaments de Marketing, Vendes, Recursos Humans i Finances de l'empresa Globalhard i mitjançant un procés d'extracció, transformació i càrrega de les dades, construirem un magatzem de dades únic que respongui a les necessitats analítiques de l'empresa i doni suport al procés de presa de decisions.
El magatzem estarà organitzat en tres capes i les dades assoliran el seu estat definitiu a la capa BASE.
A través d'un programari client del magatzem de dades (*Oracle Discoverer*) modelarem jerarquies, derivarem informació de les dades creant atributs nous i organitzarem les dades en quatre àrees diferents preparant-les perquè els usuaris puguin elaborar informes *ad-hoc*. Aquests informes seran dinàmics i permetran als usuaris recórrer les jerarquies amunt i avall, veure nivells de detall diferents i/o crear-ne de nous. A més de les dades com a tals els informes també oferiran gràfiques. Tots aquests informes s'actualitzaran dinàmicament a mida que canviïn les dades del magatzem.

Índex de continguts

Dedicatòria i agraïments 2
Resum 3
Índex de continguts 4
Memòria 5
Introducció 5
Justificació del TFC i context en el qual es desenvolupa: punt de partida i aportació del TFC 5
Objectius del TFC 7
Enfocament i metodologia 8
Planificació i fases del projecte 10
Diagrama de Gannt 16
Llistat de productes obtinguts : 17
Anàlisi 18
Dimensions : 18
Diagrames del model conceptual 22
Disseny 25
Procés d'ETL (Extracció transformació i càrrega de dades al magatzem) 29
Diagrama de l'arquitectura programari 35
Diagrama de l'arquitectura maquinari 36
Disseny de la base de dades i diagrama del model físic 37
Disseny i descripció del informes creats 39
Captures de pantalla del modelatge de dades i jerarquies 45
Conclusions 47
Línies d'evolució futura 48
Glossari 49
Bibliografia 50

Memòria

Introducció

Justificació del TFC i context en el qual es desenvolupa: punt de partida i aportació del TFC

La coneguda companyia de venda al major de productes d'electrònica de consum, GlobalHard, disposa actualment d'un conjunt de sistemes heterogenis per gestionar la informació departamental sobre Vendes, Recursos Humans, Marketing, i Finances que asseguren el correcte funcionament diari de l'empresa. Actualment existeix la necessitat de que els responsables dels diferents departaments de la companyia disposin d'informació integrada, fiable i fàcilment accessible per gestionar correctament els recursos de la companyia, conèixer tendències de mercat i aconseguir, per tant, un avantatge competitiu respecte els seus principals competidors. Per això el gerent s'ha plantejat la possibilitat d'implantar un sistema d'informació basat en un magatzem de dades corporatiu. El gerent vol, a més, poder accedir a informació resumida sobre cada departament de la companyia.

Aquest TFC és una proposta per construir un magatzem de dades per consolidar dades en informació útil que permeti prendre decisions i establir noves funcionalitats de suport al procés de presa de decisions de GlobalHard.

Els executius sovint necessiten informació del passat, informació que descrigui la situació actual i informació que permeti anticipar el futur.

Per analitzar i transformar les dades en informació útil, els executius i el seu personal han de tornar a entrar la informació no integrada en els seus propis mini-sistemes d'una manera manual.

Construirem un magatzem de dades per tal de facilitar l'accés a la informació per a una major part d'usuaris que no tindran perquè està familiaritzats ni conèixer on es guarden les dades o com està codificada la informació en els diferents sistemes operatius.

Pel que fa als informes el problema principal és que els sistemes operatius actuals no estan dissenyats per donar suport a un procés analític i no s'està emprant cap mena d'eina *ad-hoc* que ofereixi funcionalitats analítiques com el *drill down* o el *slice & dice*, o permeti la construcció d'escenaris *i si...?*

Les dades enfocades a la presa de decisions o bé no estan integrades, o bé estan definides amb ambigüitat o bé són inexistents.

Quan diem que les dades no estan integrades ens referim que estan repartides en més d'un sistema operatiu. Així la informació emmagatzemada a un sistema no té perquè correspondre's amb la d'un altre. A més, l'entorn actual no dóna suport a les activitats que requereixen comparar, combinar, relacionar, etc. dades d'un sistema amb dades d'un altre sistema. La feina necessària per conglomerar la informació és massa gran i a vegades impossible.

Quan diem que les dades són inexistents és perquè les necessitats dels executius no es van tenir en compte en dissenyar els sistemes OLTP, pel què després no estan disponibles.

Quan diem que estan definides amb ambigüitat ens referim a que hi ha una manca d'enfocament unificat en la definició i classificació de les dades. Les necessitats de cadascun dels sistemes OLTP són diferents i per tant també ho són les seves definicions i classificacions de les dades.

Procés d'estratificació de dades : no hi ha una política que controli el procés pel qual la informació es recull i s'accedeix i és distribuïda en el marc de GlobalHard.

Aquest projecte dota GlobalHard d'un servei de magatzem de dades que donarà suport al procés de presa de decisions. Aquesta funcionalitat dotarà l'empresa de dades rellevants a les activitats de planificació amb independència de la seva procedència.
També ampliarà el ventall d'individus que puguin tenir accés a la informació i que no tindran perquè conèixer on està guardada la informació o la seva procedència.
El magatzem de dades contindrà dades consistents, formatades, transformades i sovint resumides de la resta de sistemes operatius. També consolidarà i organitzarà aquesta gran quantitat de dades transaccionals en un número raonable de taules i camps que es podran consultar i analitzar per donar suport a la presa de decisions.
Els usuaris accediran a la informació a través d'eines senzilles d'emprar que tindran una interfície còmoda i que els permetrà que la informació sigui fàcilment accessible.

Beneficis :

A diferència dels informes impresos que són habituals a GlobalHard, el magatzem de dades ens permetrà disposar d'una informació més flexible, més actual, i més adequada a les necessitats emergents. A la vegada, ja que en el procés d'extracció i transformació de les dades incorporarem regles de negoci i altres mesures de qualitat ens serà més fàcil de mantenir dades més consistents i més acurades.

Objectius del TFC

Amb els magatzem de dades s'ha de donar resposta a les necessitats bàsiques dels diferents departaments :

- Vendes: analitzar les vendes per productes (productes més venuts, vendes per regió, marges, etc.)
- Recursos humans : analitzar el número d'empleats de l'empresa, els sous mitjans per tipus d'empleat.
- Marketing : analitzar els productes venuts per tipus de client, canal de venda.
- Finances : analitzar els pressupostos de la companyia.

En permetre aquesta anàlisi donarem suport als diferents departaments en la presa de decisions.

Cada departament tindrà accés a les seves dades i en serà el propietari, de manera que si un departament vol accedir a les dades d'un altre haurà se l'haurà d'autoritzar prèviament.

Enfocament i metodologia

El projecte d'implementació del magatzem de dades és molt més que un projecte tècnic. Requereix un coneixement dels processos del dia a dia de GlobalHard així com de l'entorn on la informació es genera i és farà servir.

Directrius d'implementació del projecte :

- Desenvolupament interactiu ja que tenim una varietat força gran de necessitats d'informació.

- Habilitat de dotar el procés de presa de decisions d'informació executiva per diferents tipus d'anàlisi, des d'escenaris del tipus *i si...?* fins a qüestions com la distribució del benefici entre departaments, o la necessitat d'accedir a dades de clients en funció del seu volum de negoci, antiguitat, o dades geogràfiques.

- Tots aquests requeriments no es poden solucionar d'un dia per l'altre. Així, el magatzem de dades cal que es desenvolupi en un procés evolutiu del tipus, un pas cada vegada, començant amb un conjunt de funcionalitats clau, ben definides i on el consens de la seva utilitat és total. Aquestes funcionalitats s'aniran expandint a mida que passi el temps i s'adaptaran a les necessitats de cada moment.

- Control de les dades : Aquest projecte no només és un projecte d'extracció i carrega de dades de diversos sistemes operatius a una base de dades central. També implica unificar les dades, resolent conflictes que hi pugui haver entre sistemes. Ha donar resposta a la qüestió de l'accés a les dades, a la seva propietat així com al manteniment de les mateixes. Així la definició de polítiques, procediments i directrius d'administració i accés a les dades és un punt clau d'aquest projecte.

- Participació i col·laboració : La participació de diferents membres de GlobalHard (de diferents àrees o departaments) és un punt bàsic per el correcte desenvolupament d'aquest projecte.

- Implementació de la gestió de metadades centralitzada : La informació descriptiva de les dades del magatzem de dades és clau per tal que els usuaris entenguin el què conté. Les metadades ajuden a identificar els tipus d'informacions disponibles i en descriuen les fonts i les regles de transformació durant el procés de migració cap al magatzem de dades.

- Dotar el projecte d'una estructura de suport organitzativa : de la mateixa manera que els sistemes operatius actuals, el magatzem de dades és més que simple maquinari i programari. Necessita persones molt qualificades i amb experiència en les necessitats dels executius de GlobalHard així com en les capacitats de la varietat d'eines emprades en l'arquitectura del magatzem de dades. Aquest equip ha de monitorar les necessitats informatives de l'organització, promoure l'estandardització de les definicions de dades i les seves classificacions, coordinar l'establiment de polítiques d'accés a les dades, l'establiment d'estàndards i rols pel què fa a la gestió de les dades i el seu

accés. Han de donar suport a l'hora d'analitzar i interpretar les dades, donar suport en la utilització de les eines, la seva instal·lació i administrar-ne la seguretat i la configuració.

Planificació i fases del projecte

La interacció que s'empra en implementar l'arquitectura del magatzem de dades recomana una implementació en fases. Així dividirem el projecte en tres fases principals. Les primeres dues fases duraran 3 mesos, mentre que la tercera és de duració il·limitada ja que representa el futur creixement i evolució del magatzem de dades.

Com a qualsevol projecte d'implementació, hi ha diversos riscos associats a la implementació d'un magatzem de dades. La identificació primerenca i la planificació poden reduir significativament el impacte de qualsevol adversitat en el projecte.

Crear un magatzem de dades és dotar a la presa de decisions d'un accés més gran i de més qualitat a la informació. Això pot provocar un canvi cultural pel què fa a la manera com es valora la informació i la seva anàlisi.

La implementació d'un magatzem de dades és molt diferent de la compra i implementació d'un sistema ERP. En el fons, es basa en una nova tecnologia basada en pàgines web i un nou ventall de productes. El personal tècnic necessitarà competències analítiques, aprendre noves tecnologies de modelatge així com noves tècniques d'afinament del rendiment de bases de dades. Aquest risc es pot reduir si els dotem de l'aprenentatge necessari i els deixem el temps adequat per la implementació de les primeres fites.

L'èxit de la implementació del magatzem de dades depèn en gran part de la capacitat dels usuaris d'especificar les seves necessitats, avaluar i provar a fons les eines del *front-end*, fer proves de dades, i fer proves d'aplicació. Caldrà doncs fer un esforç per tal de reorganitzar les tasques diàries per tal de dotar-los del temps necessari per dur poder realitzar les aquí esmentades.

La fase de infraestructura :

En aquesta fase ens dotarem del programari, maquinari necessaris per dur a terme el projecte.
També contractarem personal, prioritzarem les necessitats de negoci i arrencarem el projecte.
La major part de les activitats d'aquesta fase les podem agrupar en 5 grups :

Organització del projecte :

Inclou totes les activitats necessàries per poder arrencar el projecte :

- Definir els equips de treball
- Contractar i formar el personal
- Desenvolupar els criteris de selecció d'eines
- Visites a empreses que ofereixen serveis de *hosting* de magatzems de dades

- Desenvolupament dels recursos de comunicació
- Desenvolupament dels estàndards de metadades
- Creació de processos de resolució de conflictes
- Establiment de processos de revisió de l'assoliment de les fites

Requeriments d'alt nivell

Aquesta fase inclou l'anàlisi de les activitats de negoci amb els executius i quadres de presa de decisions de GlobalHard i posarà de manifest :

- Entendre les iniciatives estratègiques clau de cada departament
- Identificar els indicadors clau de cada departament o les mètriques d'èxit de cada iniciativa estratègica.
- Determinar els processos clau de cada departament en els que volem causar un impacte.
- Determinar el impacte potencial de l'accés a la informació en les mètriques d'èxit de cada departament.

Establiment de la prioritat de les necessitats del negoci

Davant la potencial multiplicitat de demandes de tots els departaments de GlobalHard, aquesta tasca establirà les prioritats de les necessitats del negoci.

Desenvolupament d'un pla de projecte i del seu abast

Basant-nos en les dades del punt anterior, desenvoluparem l'abast del projecte i el planificarem.

Construcció de la infraestructura tècnica

Aquesta tasca inclou l'adquisició i instal·lació de sistemes base de programari i maquinari necessaris per la implementació de l'arquitectura del magatzem de dades. Inclou :

- Desenvolupament del pla tècnic d'arquitectura
- Redistribució dels servidors existents i instal·lació de la base de dades
- Selecció i instal·lació de l'eina *front-end*
- Adquisició i instal·lació dels servidors web i d'aplicacions

- Selecció i instal·lació de l'eina ETL

La fase d'implementació :

En aquesta fase començarà la implementació concreta de les àrees de vendes, *marketing*, finances i recursos humans. L'enfocament interactiu es durà a terme de la següent manera : per cada àrea, hi ha sis grups d'activitats bàsiques que es descriuen a continuació. El temps d'implementació final de cada àrea variarà en funció de la complexitat definida en la fase de requeriments.

Requeriments dels usuaris

Els usuaris i els seus requeriments impacten qualsevol decisió presa durant tot el procés d'implementació del magatzem de dades. Això inclou totes les activitats necessàries per tal d'assolir un coneixement profund del negoci.

- Selecció de candidats i realització d'entrevistes
- Realització d'auditories de dades i anàlisi dels resultats
- Documentació dels descobriments
- Publicació dels requeriments
- Establiment de prioritats i abast quan sigui necessari

Modelatge

Després de la recopilació dels requeriments cal fer dues activitats : el disseny lògic i l'anàlisi detallada de les fonts de dades i des les regles de transformació.

- Modelatge de les dimensions : traducció dels requeriments del negoci en dissenys *star schema* o *snow flake* i avaluació i disseny de les taules agregades.
- Anàlisi de les fonts de dades : identificació de les fonts d'informació i de les regles de transformació.

Disseny físic

De la mateixa manera que els plànols d'una casa és converteixen en la casa real, aquesta tasca converteix el model lògic en una base de dades física.

Disseny físic de la base de dades : Inclou tasques com :

- definició d'estàndards

- convenció de noms
- estimació de la grandària de la base de dades
- desenvolupament d'un pla d'índexs
- desenvolupament d'un pla de particions
- disseny físic de taules i columnes

Implementació de la base de dades : inclou tasques com :

- l'optimització del SGBD
- la construcció del sistema físic d'emmagatzematge
- dimensió de l'espai per taules i índexs i la seva creació

Disseny de les capes de dades i del seu desenvolupament

Aquesta fase és la clau de la construcció del magatzem de dades. Durant aquesta tasca es dissenyen i desenvolupen els processos que carreguen les dades dels sistemes operatius al magatzem de dades :

Disseny i desenvolupament de les capes :

- Disseny i desenvolupament de l'estratificació per cada taula
- Definició i implementació de les metadades d'estratificació
- Desenvolupament de la càrrega inicial
- Desenvolupament de les càrregues periòdiques
- Disseny i implementació del *data cleansing & profiling* i els processos d'agregació de dades
- Desenvolupament dels processos de garantia de la qualitat de les dades

Càrrega i validació de la base de dades :

- Definició i configuració de l'entorn de producció
- Càrrega de dades inicial
- Càrrega de dades històriques
- Processos de validació i de garantia de la qualitat de les dades

Ajustament de les prestacions del SGBD :

- Definició de consultes de *benchmark*
- Revisió d'índexs i agregacions
- Configuració fina del sistema
- Monitoratge i control de la base de dades

Aplicacions :

En aquesta fase generarem les plantilles pel *front-end*. Si és necessari tornarem a definir informes. El temps necessari per dur-la a terme dependrà força de la definició de requeriments.

- Especificació de les aplicacions finals per l'usuari : categorització i definició de l'estratègia d'accés per part de l'usuari (interfície corporativa o bé eina d'accés directe o bé eina basada en entorn web)
- Desenvolupament de les aplicacions d'usuari : Disseny i desenvolupament de les plantilles per les aplicacions i les interfícies corporatives.
- Formació dels usuaris principals. Comença la formació dels usuaris que més en contacte estaran amb el magatzem de dades.
- Manteniment de les metadades. Documentació i introducció de les metadades.

Posada en marxa

En aquesta fase es posa en marxa el magatzem de dades i es facilita el seu accés a la comunitat d'usuaris.

- Planificació del desplegament :

  desenvolupament d'estratègies de formació i suport

  desenvolupament d'un pla de desplegament

- Dur a terme un test general i complet de tot el sistema : dur a terme un test general de tot el sistema des de l'estratificació de les dades fins a les plantilles i formularis de les aplicacions finals dels usuaris.

- Desplegament : configuració i test de la infraestructura de client, configuració de la seguretat i formació dels usuaris.

La fase de creixement :

En aquesta fase, a diferència dels sistemes tradicionals, els canvis es veuran com un indicador d'èxit assolit i no com un fracàs. Aquesta fase es dedica a continuar adreçant les necessitats d'evolució i creixement derivades de les demandes dels usuaris.

El manteniment i el creixement del magatzem de dades requerirà una estratègia molt diferent de la tradicional. En sistemes operatius tradicionals el desplegament significa que el projecte s'ha acabat i disposem dels recursos per la següent implementació. Normalment, en projectes de magatzems de dades, les necessitats de recursos no s'acostumen a reduir d'una manera significativa en passar de la fase de desenvolupament a la fase de manteniment i creixement i a diferència d'altres projectes de desenvolupament, el magatzem de dades no es considera mai "acabat". Caldrà :

- Dotar els usuaris d'un suport permanent
- Dotar els usuaris de formació continuada
- Mantenir la infraestructura tècnica
- Monitorar el rendiment de les consultes d'usuari
- Monitorar el rendiment de l'estratificació de les dades
- Monitorar l'èxit del sistema
- Identificar els canvis i les noves necessitats
- Identificar les necessitats de noves fonts d'informació (internes i externes)
- Identificar les necessitats de nous departaments

Diagrama de Gannt

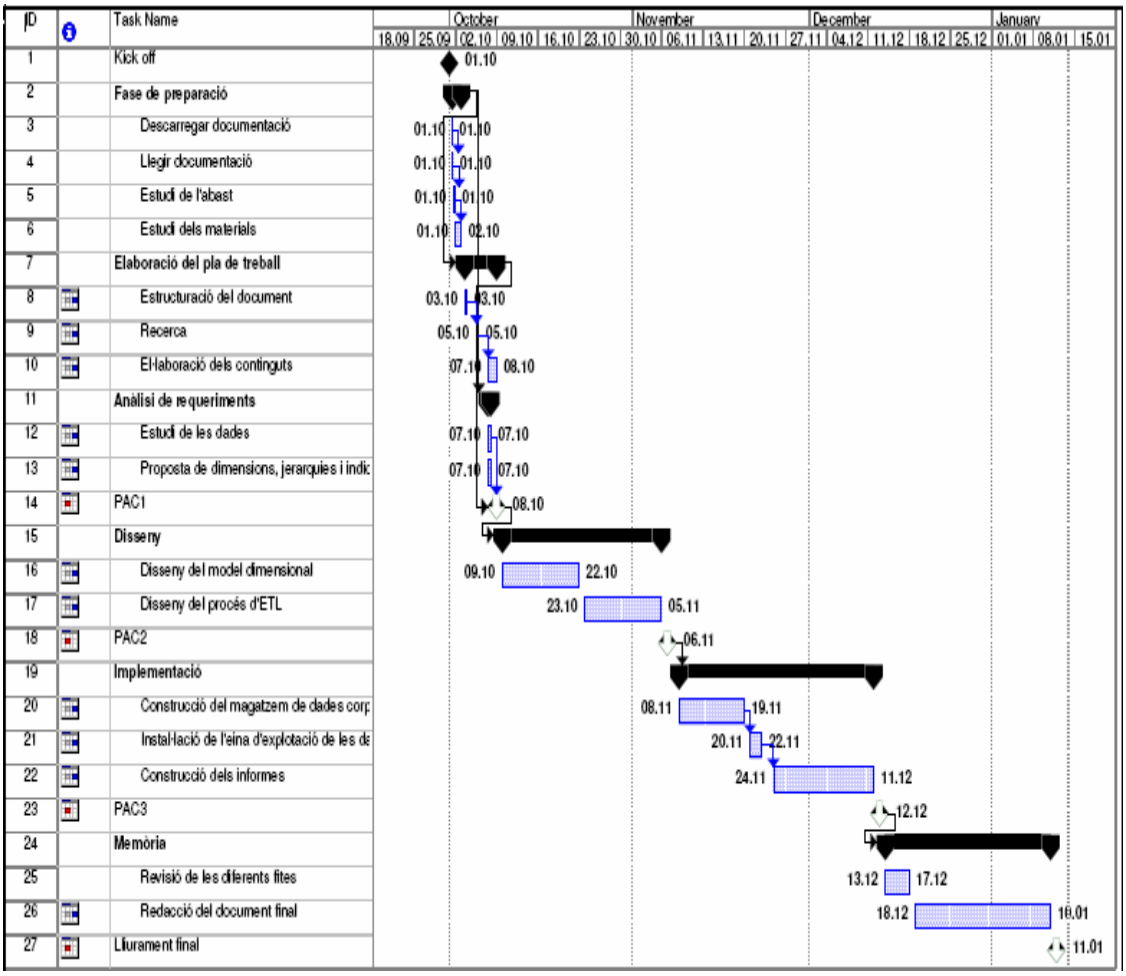


Figura 1. Diagrama de GANTT de la planificació del projecte

Llistat de productes obtinguts :

Magatzem de dades amb arquitectura de 3 capes que recull totes les dades de les diferents bases de dades operacionals de Globahard.
Procés d'importació / depuració / unificació de dades : ETL.

4 *Workbooks* amb informes per cada departament :

Recursos humans :

Distribució empleats
Distribució sous / departament
Distribució sous / categories
Distribució sous / categories / mitjanes

Finances :

Pressupostos anuals

Marketing :

Distribució Clients / Regions – Famílies / Productes
Distribució Canals de venda – Sectors / Clients

Facturació :

Facturació / productes
Facturació / regions
Facturació / canals
Facturació / sectors
Facturació / segments
Facturació / temps

L'estructura actual basada en jerarquies i carpetes que responen a les necessitats actuals és molt flexible i ens permet d'ampliar-la i/o modificar-la en qualsevol moment. De fet aquests informes només són un punt de partida i a mida que els diferents departaments vagin desenvolupant noves necessitats / requeriments anirem ampliant aquesta llista així com anirem adaptant les jerarquies i carpetes a través del *discoverer administrator.*

Anàlisi

En aquesta secció descriurem els resultats de l'anàlisi de les quatre bases de dades dels diferents sistemes operatius inicials i la proposta conceptual que se'n deriva i que després dissenyarem :

Dimensions :

L'ordenació jeràrquica dels diferents atributs i les seves mètriques es basa en el següent concepte:

Tots els atributs que pertanyen a una mateixa dimensió es poden ordenar d'una manera jeràrquica, de manera que d'entrada podem dir que són del mateix tipus, a més, cada escala és una granularitat menor o major dins de la jerarquia.
Aquesta agrupació jeràrquica en dimensions és el què ens permetrà a posteriori realitzar *drill-downs*, entre altres funcionalitats, possibilitant l'anàlisi *ad-hoc* de les dades en els diferents reports.

Construirem les següents dimensions :

Temps :

Categoria temps :

- Any
- Trimestre
- Mes

Categoria alternativa temps :

- Any
- Trimestre
- Mes
- Setmana

Categoria alternativa temps fiscal :

- Any fiscal
- Trimestre fiscal
- Mes fiscal

Categoria relativa temps :

- Any fins ara (mesos)
- Any anterior fins ara (mesos)

Productes :

Categoria productes :

- Família
- Producte

Canals

Categoria canals :

- Canal

Clients

Categoria clients regions :

- Regió
- Client

Categoria clients sectors :

- Sector
- Client

Categoria clients segments :

- Segment
- Client

Empleats

Categoria empleats departaments :

- Departament

- Empleat

Categoria empleats categories :

- Categoria
- Empleat

Categoria derivada empleats antiguitat :

- Antiguitat
- Empleat

Categoria derivada empleats classe de sou :

- Classe de sou
- Empleat

Indicadors :

- Marge de venda per venda
- Marge de venda mitjà per producte
- Marge de venda mitjà per família
- Marge de venda mitjà per canal
- Marge de venda mitjà per client
- Marge de venda mitjà per regió
- Marge de venda mitjà per sector
- Marge de venda mitjà per segment
- Vendes per producte
- Vendes per família
- Vendes per canal
- Vendes per client
- Vendes per regió

- Vendes per sector
- Vendes per segment
- Número d'empleats per departament
- Número d'empleats per categoria
- Número d'empleats per classe de sou
- Número d'empleats per antiguitat
- Sou mitjà per categoria
- Sou mitjà per departament
- Sou mitjà per antiguitat
- Sou mitjà per classe de sou
- Pressupost mitjà per departament
- Diferencial mitjà per departament
- Benefici per departament
- Benefici per empleat i departament
- Benefici en funció del sous i departaments
- Benefici per empleat
- Benefici en funció del sous

* Tots aquests indicadors es podran observar en funció del temps

Diagrames del model conceptual

Bases de dades originals de Globalhard :

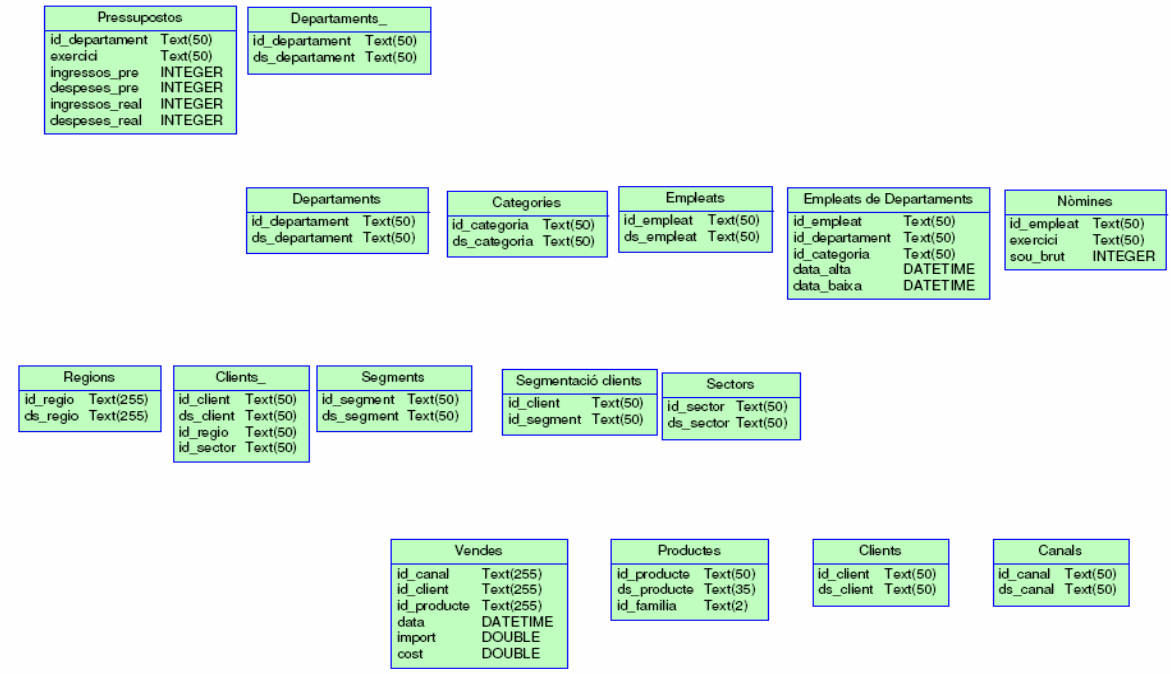


Figura 2. Diagrama de les bases de dades originals de Globalhard

Capa d'importació (IMP) :

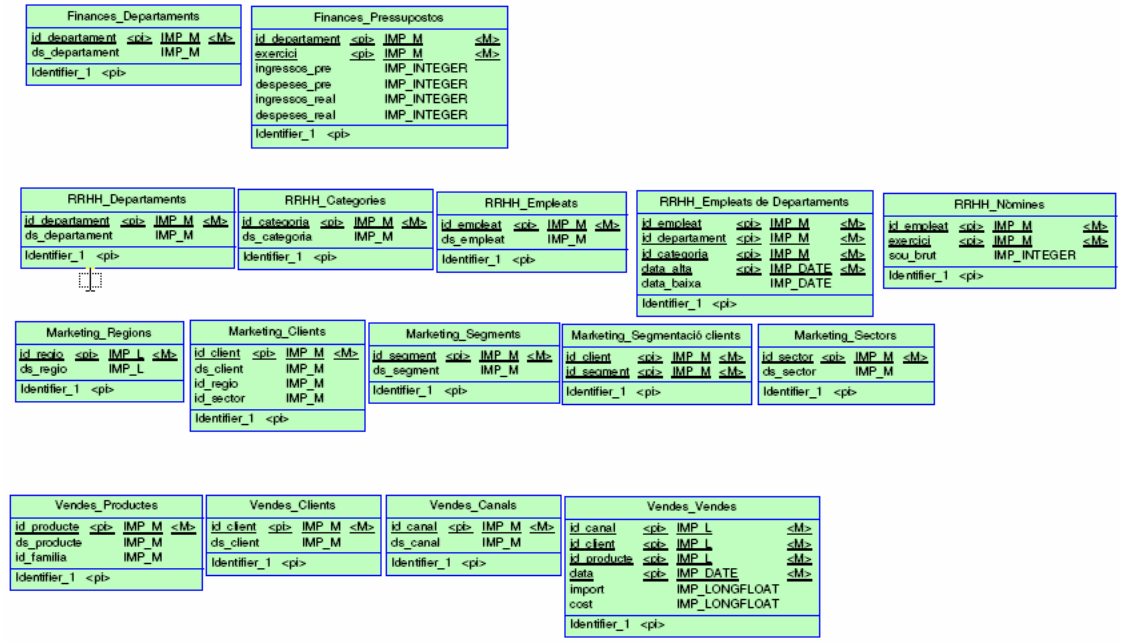


Figura 3. Diagrama de la capa IMP

Capa d'STAGE :

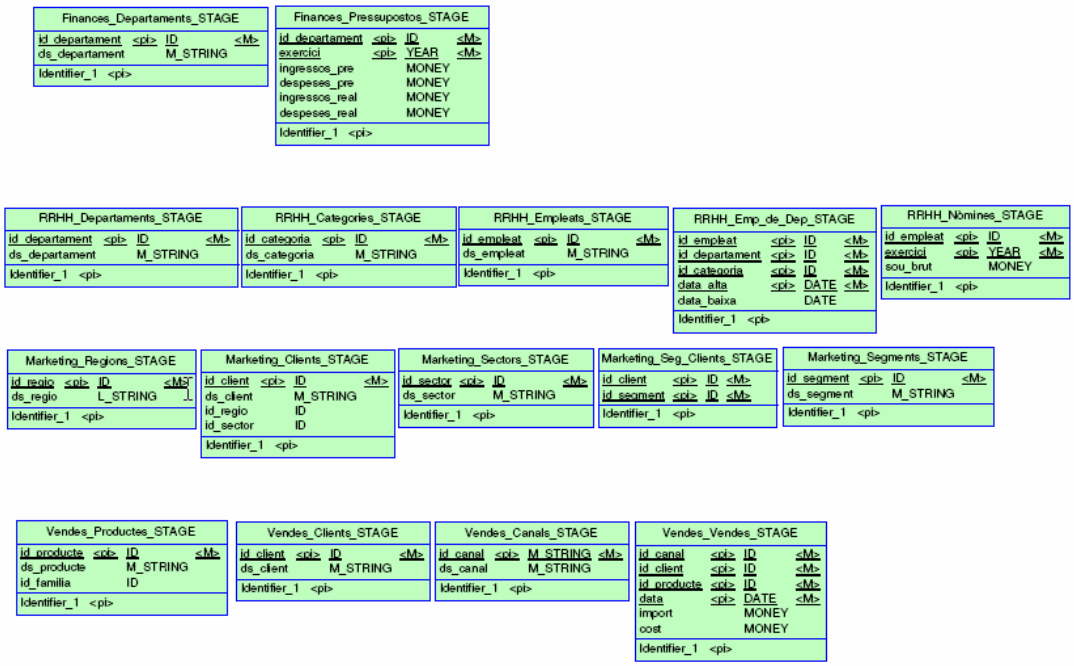


Figura 4. Diagrama de la capa STAGE

Capa del magatzem pròpiament dit (BASE) :

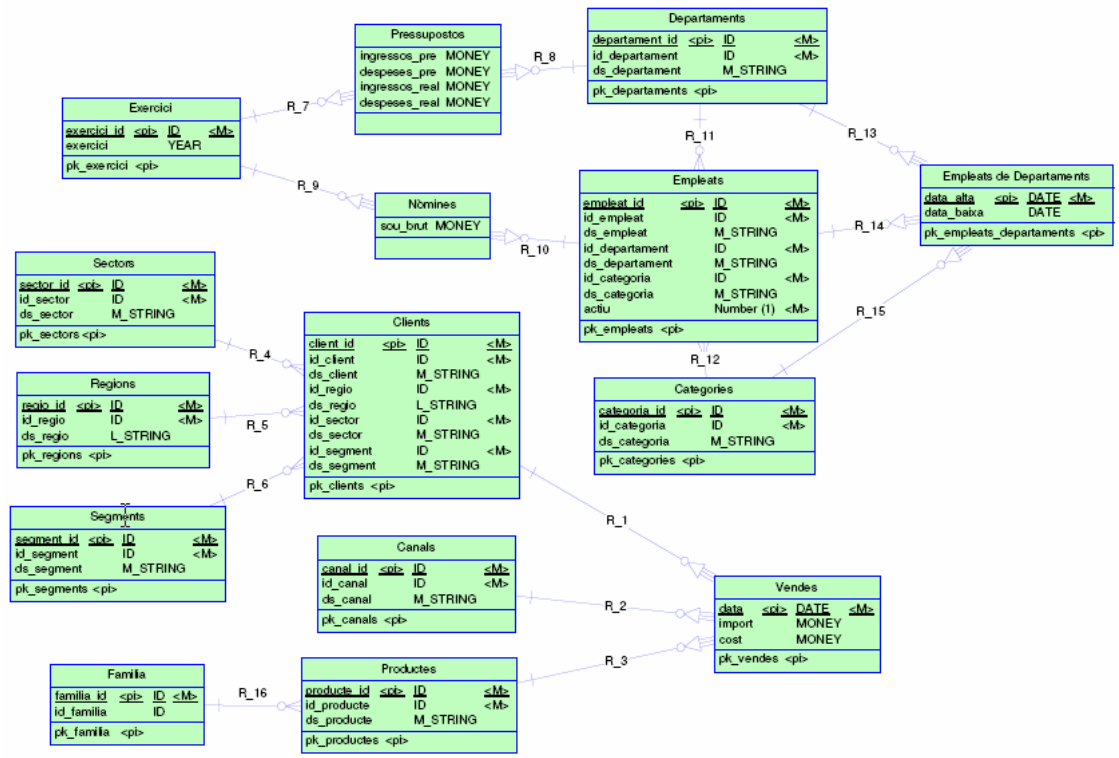


Figura 5 Diagrama de la capa BASE

Diagrama general de la base de dades amb les tres capes :

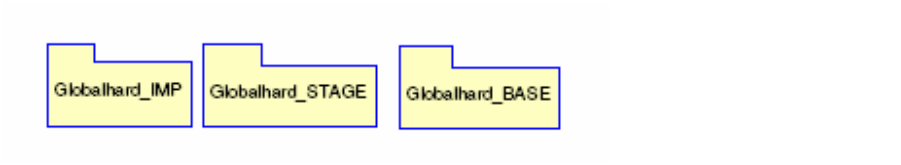


Figura 6. Diagrama de la les tres capes

Disseny

Per confeccionar el disseny del magatzem de dades s'ha optat per un model mixt. Davant les opcions d'un disseny *snow-flake* pur o un disseny *star* pur s'ha optat per un model mixt on les dimensions de cada taula de fets estan en forma star però aquestes taules de l'esquema star a la vegada tenen taules de referència formant un snow flake de dos nivells constant. Això ens proporciona dos avantatges :
Les dimensions estan encapsulades totes en una taula de manera que no cal fer *joins* per accedir als seus continguts.
Les taules de referència són independents i es poden mantenir sense requerir que formin part activa d'una dimensió.
Per exemple pot ser que no tinguem cap client del segmentX tot i així podem tenir les dades del segmentX al magatzem de dades.
Això com és evident ens crea redundància a la base de dades però la desnormalització és un fet característic dels magatzems de dades.

A la transformació del model operatiu al disseny del magatzem de dades podem observar :

Les taules clients i departaments que apareixien en dues bases de dades es fusionen en una de sola.

A totes les taules s'ha introduït un identificador del tipus Number(9) que seguint la convenció de noms s'anomenarà *nomdelataula*_id.

S'han unificat els tipus de dades en 7 dominis :

- Id : Number(9)
- M_string : Varchar2(50)
- L_string : Varchar2(255)
- S_string : Varchar2(10)
- Money : Number(10,2)
- Date : Date
- Year : Char(4)

S'ha introduït l'entitat Exercici que modela un any.

S'ha introduït l'entitat Familia que modela una família de productes.

Tot i que a la base de dades operativa la interrelació entre les entitats client i segment es tracta com si fos M : N, una observació detallada del contingut de la taula de la interrelació sembla indicar el contrari. A falta d'aclarir-ho conceptualment i no a nivell tècnic hem optat per convertir-la en una interrelació 1:N on un segment té N clients però un client pertany a 1 segment i prou. La majoria dels dubtes que tindrem en modelar el nostre magatzem no els hem de resoldre des d'un punt de vista tècnic o de la implementació existent dels sistemes operacionals que pot ser fruït de circumstàncies que en aquests moments en són alienes o simplement errònia. Els dubtes els hem d'aclarir a nivell de negoci i a ser possible directament amb el propietari de les dades o un dels seus operadors habituals que han de

conèixer bé els conceptes que s'han modelat als sistemes operacionals (amb més o menys fortuna).

Podem observar tres taules de fets :

- Vendes
- Nòmines
- Pressupostos

i una taula d'història : Empleats de Departament

La informació actual de quins empleats estan a quin departament es guardarà també a la taula empleats durant el procés d'ETL.

La taula de fets Nòmines té dues dimensions : Empleats (empleat, departament, categoria) i Exercici

La taula de fets Pressupostos té dues dimensions : Departaments i Exercici

La taula de fets Vendes té tres dimensions : Clients (client, regió, sector, segment), Canals i Productes (producte, família).

Aquest disseny és la bastida sobre la qual muntarem el nostre magatzem de dades.

Donat el volum de dades relativament petit i que l'actualització només és necessària a nivell mensual no és aconsellable un procés de càrrega en forma de deltes. En un model d'aquest tipus en el procés d'extracció només es carreguen les dades noves o modificades dels sistemes operatius des de la última actualització. Tot i així en el cas de GlobalHard farem un import complert mensual la matinada del primer dia de cada mes. Donat el volum de dades reduït la finestra de temps necessària per fer-lo és més que suficient. En cas de requerir-ho sempre podem en el futur canviar a un procés basat en deltes tot i que això suposa un nivell de complexitat força major en tot el procés d'extracció i càrrega del magatzem.

Per sota de la capa BASE que és la que estem modelant hi haurà dues capes més. La capa STAGE i la capa IMP que a nivell físic les modelarem amb Schema propi a la base de dades. Les dades del sistemes operatius s'importaran 1:1 a la capa IMP on tenim una imatge exacte de la interfície d'import del magatzem de dades. Per convenció no deixarem entrar cap dada al magatzem sinó és a través d'aquesta capa. En aquesta capa totes les taules modelades tenen els camps del tipus Varchar2 la grandària del qual s'adapta al camp que volem importar i es sobredimensiona de manera que puguem importar els registres fins a aquesta capa gairebé sempre.
En aquesta capa tampoc tenim cap *constraint* implementat ni cap relació entre les taules.

El procés d'ETL extreu les dades de les bases de dades operacionals i les carrega en aquesta capa IMP.

En la següent capa STAGE es llegeixen les dades de la capa IMP i es fan diferents tipus de comprovacions per cadascun dels registres de les entitats de la capa IMP. Així per exemple es comproven les claus primàries de cada entitat de manera que no importarem registres duplicats. També comprovarem l'obligatorietat del camp, els

tipus de dades i la llargada i precisió de manera que la conversió dels valors Varchar2 importats a IMP cap a la capa STAGE sigui possible. Els registres que no compleixin els criteris de qualitat necessaris per estar a la capa STAGE els deixarem a la capa IMP i els marcarem amb un codi d'error elevat. Així en aquesta capa tindrem una vista dels sistemes operacionals sense errors greus. Finalment l'últim pas serà moure les dades de la capa STAGE a la capa BASE i per fer-ho també farem comprovacions contra les taules de referència de manera que un valor que no tingui referència el marcarem seguint la convenció amb diferents valors (-1 si és un número, '-' si és un string, '11.11.1111' si és un dia etc. Amb això permetem que valors desconeguts facin el seu camí fins al final del magatzem. Això és una decisió política que permet que intenta obtenir com més registres millor a la capa BASE. Això ho fem així perquè una venda per exemple que s'hagi importat al magatzem amb un canal desconegut, no ens interessa deixar-la a STAGE ja que llavors no sortirà a les estadístiques i seria com si no s'hagués produït. Això ens provoca una distorsió molt més gran que no pas el fet de tenir una venta amb un canal desconegut. Ara bé, el nostre disseny tenim que les dimensions tenen referències contra taules de referència on els *constraints* no permeten valors desconeguts. És per això que substituirem els valors no coneguts amb els què hem esmentat anteriorment, que sí són presents a les taules de referència com a valors comodí per valors desconeguts o erronis.

Els identificadors artificials que generem pel magatzem de dades també es generen en el salt de les dades de la capa STAGE cap a la capa BASE.
D'aquesta manera també complim el nostre objectiu de tenir a la capa STAGE una vista dels sistemes operacionals ja que les claus primàries de la capa STAGE són les claus primàries d'aquests sistemes. A la capa BASE però les claus primàries són totes del mateix tipus N9 per convenció i generades automàticament pel magatzem. Això ens encapsula més els nostres registres de manera que no depenem en absolut dels sistemes operacionals que d'un dia per l'altre podrien canviar la llargada, el tipus, etc d'una clau primària sense que això ens afectes en gran mesura. O per exemple podríem tenir dos clients que compartissin el mateix magatzem i per dos clients diferents haguessin emprat el mateix identificador.

És també en el salt de STAGE a BASE on es calcularan tots els valors agregats derivables a partir d'altres valors importats.

També és en aquest moment on farem comprovacions de regles de negoci sobre les registres importats com per exemple que la data d'alta d'un empleat en un departament no pot ser posterior a la data de baixa del mateix empleat en el mateix departament.

En funció dels resultats realitzats en les diferents proves informarem el camp de valor d'error de cada registre dotant així a cada registre d'un valor qualitatiu que reflexa la qualitat de la informació que conté. En cas d'haver trobat errors els codificarem seguint la convenció de manera que observant un registre siguem capaços no només de veure la seva qualitat sinó també de desxifrar-ne els errors trobats en les diferents proves successives.

Els processos ETL de moviment de dades funcionaran un darrera l'altra en l'ordre següent :

1. *Import* a IMP
2. Moviment de IMP a STAGE

3. Moviment de STAGE a BASE

Per tal de poder exportar dades cap al món exterior el magatzem també disposarà del Schema EXP i per convenció tot el què s'exporti al món exterior haurà de passar per aquesta capa.

Aquesta estratificació tan estricte ens aportarà molts avantatges en créixer el volum del magatzem de dades de manera que no només serà operatiu sinó que en la seva organització interna seguirà convencions lògiques i de fàcil interpretació per part dels desenvolupadors que es puguin incorporar en el futur al projecte.

Procés d'ETL (Extracció transformació i càrrega de dades al magatzem)

A continuació donarem una breu descripció del procés ETL de càrrega de dades de les bases de dades de producció al magatzem de dades.

S'ha creat una connexió ODBC amb la base de dades del magatzem de dades : una Oracle XE 10g. S'ha anomenat la connexió ODBC XE. Usuari : Globahard_imp, Password: Globalhard_imp.
D'ara en endavant ens referirem a la bd magatzem de dades com XE.

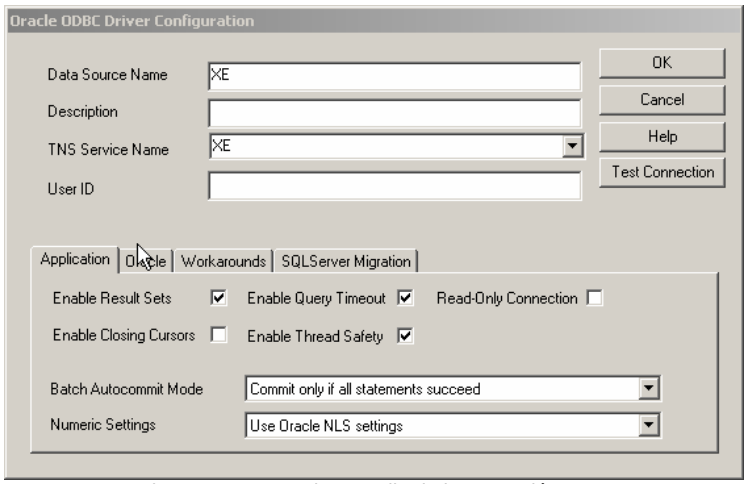


Figura 7. Captura de pantalla de la connexió ODBC

S'han creat *database links* a les bases de dades de producció amb taules 1 : 1 al schema Globalhard_imp de la XE.

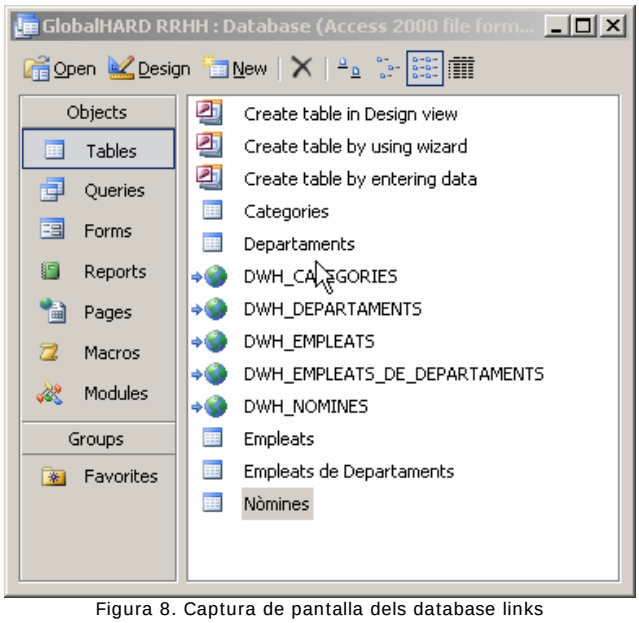


Figura 8. Captura de pantalla dels database links

S'han creat *delete* i *append queries* a les bases de dades de producció que esborren les taules homònimes del schema imp de la XE i hi afegeixen els registres.

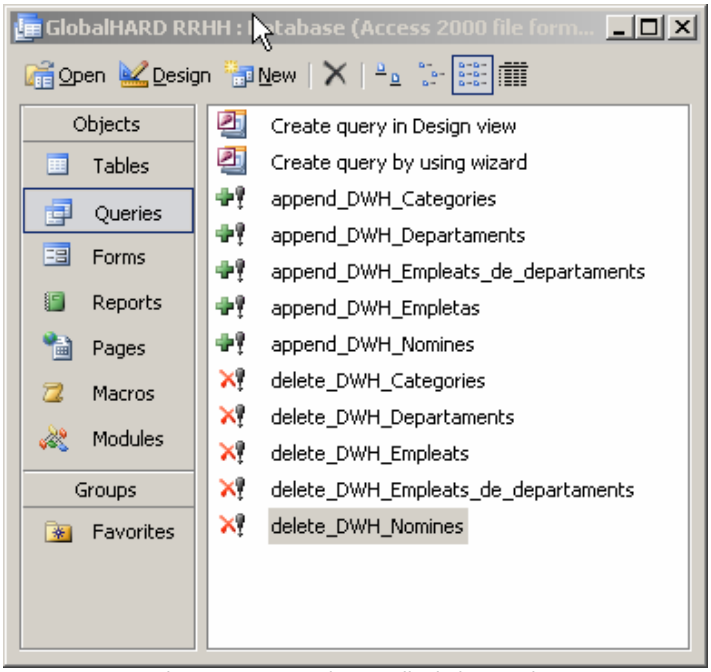


Figura 9. Captura de pantalla de les queries

S'ha creat una macro d'exportació que executa les *queries* de neteja i inserció de registres

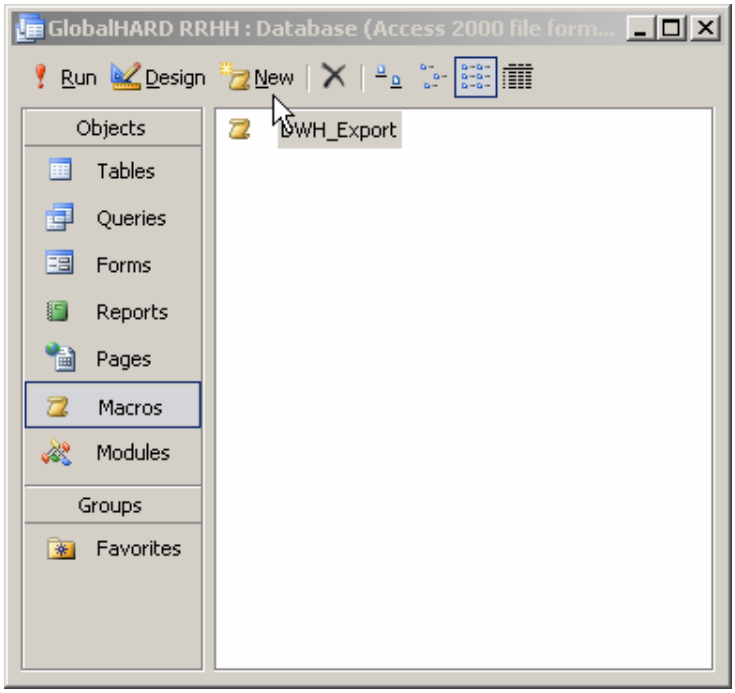


Figura 10. Captura de pantalla de la macro

Així fent clic a aquesta macro a cadascuna de les bases de dades de producció es carregaran les dades a la capa d'importació del magatzem. A partir d'allà ja les tindrem al magatzem i les processarem. El fet de tenir-les aïllades a un schema propi, que d'altra banda no és visible a cap usuari final del magatzem ens permet aïllar el procés i fer que el puguem realitzar tan sovint com vulguem, o que falli si hi ha problemes de xarxa, etc. Sense que això ens afecti les altres capes productives del magatzem. Un cop estiguem segurs que les dades de la capa imp són íntegres continuarem amb el procés de transformació.

A continuació una vista de les taules de la capa imp del magatzem :

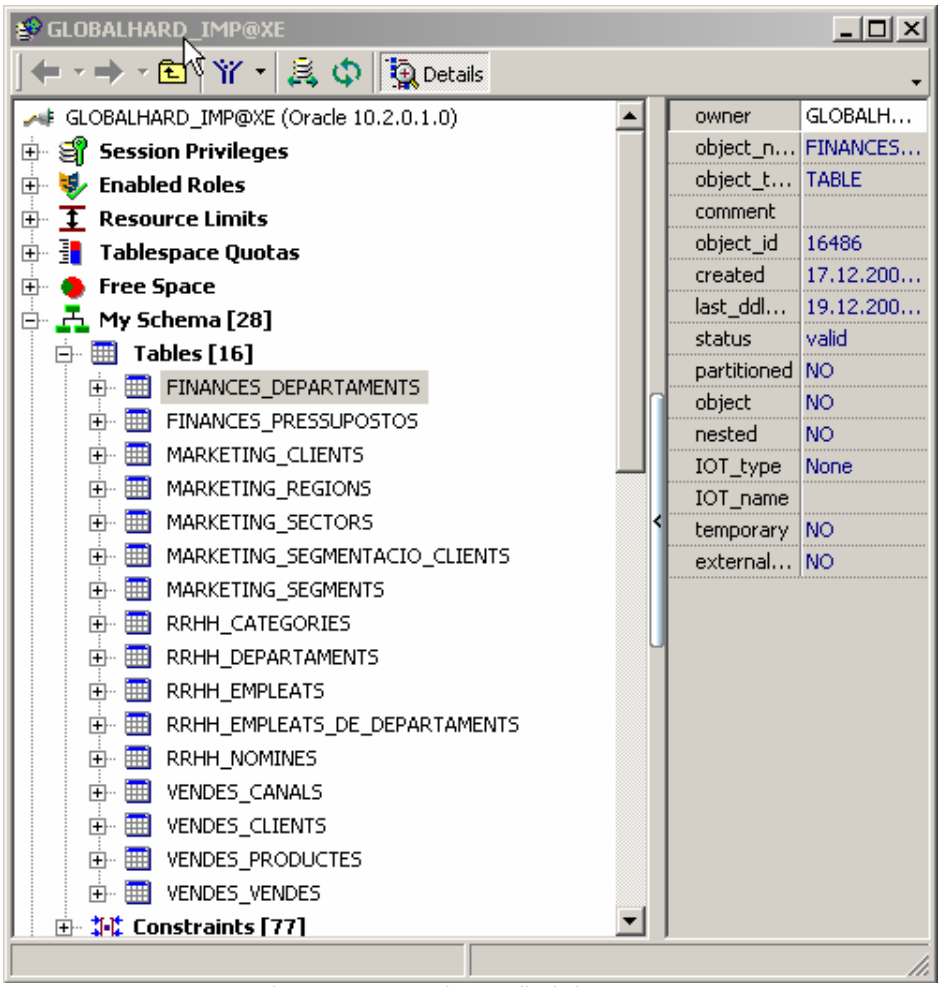


Figura 11. Captura de pantalla de la capa IMP

Fixem-nos que hem prefixat les taules amb la base de dades de procedència ja que ara les tenim totes a un mateix schema i així evitarem duplicitats de nom i possibles confusions.

Un cop tenim les dades a la capa (on tots els camps de totes les taules són del tipus varchar2, cosa que permetrà que pràcticament acceptem qualsevol tipus de registre que ens exportin) caldrà fer el següent pas que és convertir les dades al seu tipus de dada correcte. Així d'un 2 en format varchar2 en farem un 2 en format number(9,0) que és el nostre estàndard pels enters al magatzem.

És aquí on cal fer un parèntesi i explicar els diferents tipus d'usuaris que tenim a la nostra XE:

Cada usuari propietari de les taules té drets per crear taules, modificar-les etc. El què podríem anomenar drets DLL. Aquests usuaris només els farem servir per desenvolupar o modificar el magatzem però mai per moure registres o llegir dades. Aquest usuaris els tenim encapsulats i només poden modificar els seus propis *schemas*. Amb això evitem treballar amb un usuari totpoderós tipus *system* que en cas d'error ens podria fer un desastre d'abast global.

Per llegir les dades del magatzem tenim dos usuaris. L'usuari *readonly* que seria un lector tot poderós ja que li hem donat drets per veure qualsevol taula però no pot modificar res.
L'usuari *globalhard* que només veu les taules de la capa final del magatzem i serà el què farem servir des del *discoverer* per analitzar les dades.

Finalment tenim un usuari que hem anomenat *master*, que si bé no té drets DLL en cap altre schema, és totpoderós pel què fa a veure, inserir, actualitzar i esborrar registres. A més a més a nivell de base de dades. Amb aquest usuari podem llegir al schema *imp* i escriure la schema *stage*, etc.
A més a més aquest usuari posseeix els *packages* i les *procedures* que regiran la transformació de les dades a capes més superiors.
A continuació podem veure els dos *packages* i les dues *procedures* de l'usuari *master* que ens serviran per processar les dades.

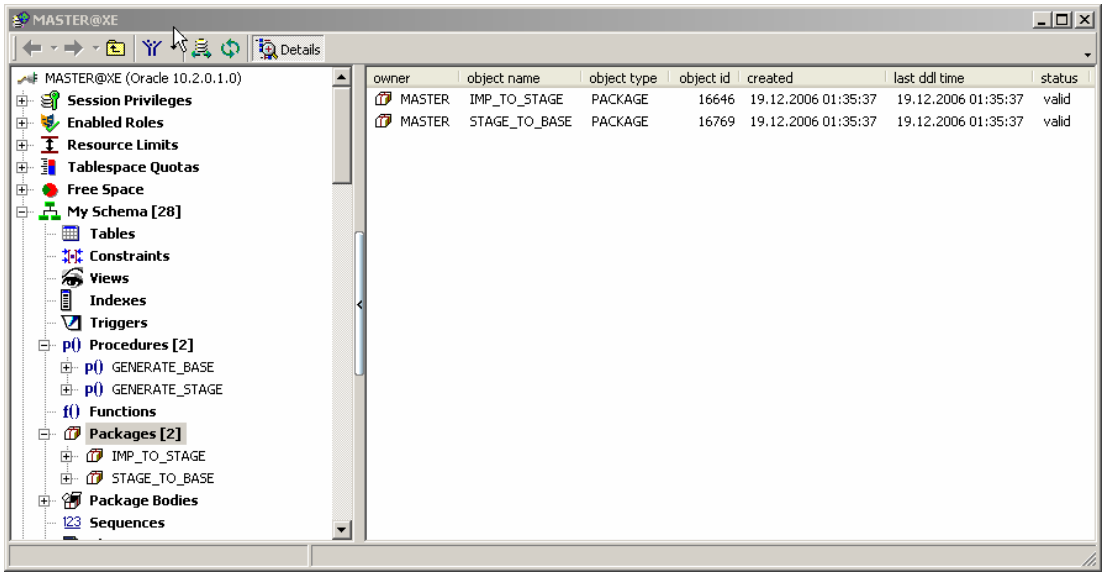


Figura 12. Captura de pantalla dels packages d'ETL

Els *packages* contenen a la vegada les *procedures* de transformació de les dades.

Però per gaudir de comoditat i poder escalar a base de pocs clics, un cop tinguem les dades exportades a la capa *imp*, només ens cal executar la *procedure* imp_to_stage, amb la qual es netejarà la capa *stage* i es s'inseriran les noves dades de la capa *imp*. En aquesta fase es converteixen les dades al seu tipus correcte. A la capa *stage* intentem modelar el què serien les bases de dades de producció amb els ids originals, els tipus de dades correctes i els constraints de negoci pertinents.

Un cop hem executat aquest procés tindrem la capa *stage* poblada.
A continuació un model conceptual d'aquesta capa.

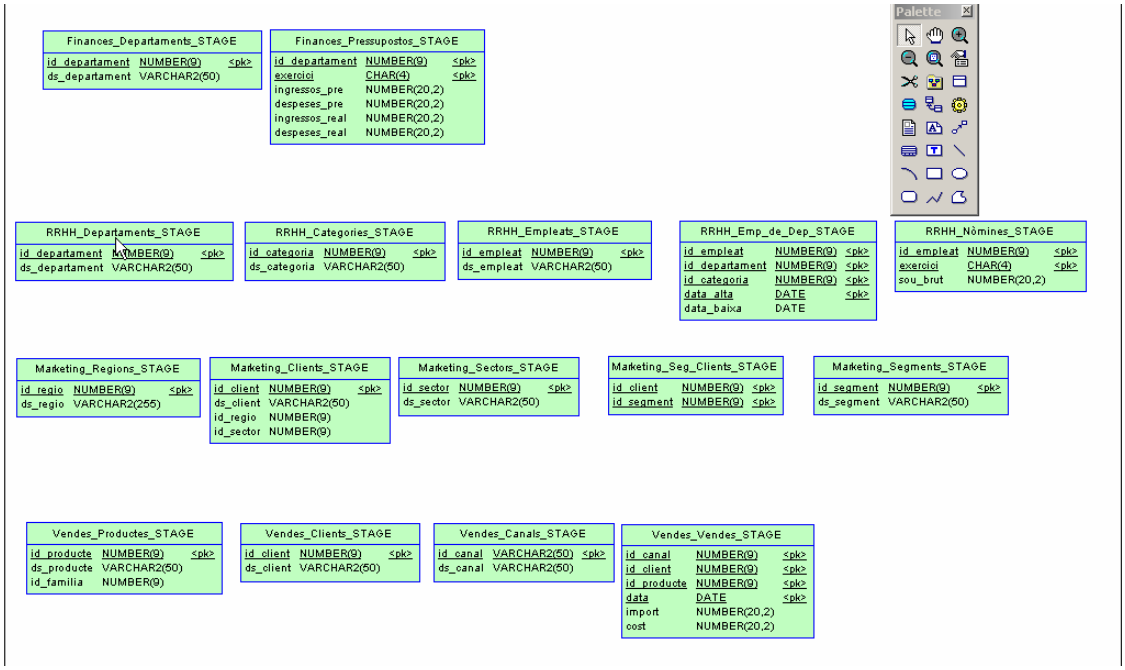


Figura 13. Diagrama de la capa STAGE

Ara ens queda el pas final, on poblarem la capa base que seria el magatzem de dades pròpiament dit on farem les següents transformacions :
Unificació de taules, creació d'ids propis del magatzem, creació de relacions entre les taules. Replicació de dades per tal d'evitar haver de fer massa joins entre les taules de fets i les taules de dimensions, etc.

Per fer-ho hem encapsulat totes les transformacions en un *package* i hem tornat a crear una *procedure* anomenada stage_to_base que quan l'executem s'encarrega de tota la transformació : neteja la capa base i la torna a poblar a partir de la capa *stage*.

La capa base final conté *sequences* i *triggers* que ens ajuden en el procés de creació de nous ids propis del magatzem. Això és una decisió que si bé després de l'anàlisi de les dades originals també ens podríem haver estalviat i treballar amb els ids originals, crec que és una bona política abstreure'ns dels ids de producció i crear-nos els nostres, això fa que tots siguin del mateix tipus N(9,0) i ens independitza de possibles canvis que es puguin donar en el futur en les bases de dades de producció. Tot i així també continuarem incorporant a les taules finals els ids de producció però com un camp més ja que podria ser que els usuaris els necessitessin alhora de fer

alguna cerca, etc. Aquesta és una decisió política que és qüestionable ja que ens dóna molta feina i fa el procés més difícil però que repeteixo l'experiència m'ha demostrat que és encertada. A més a més aconseguim unificar la nomenclatura del magatzem ja que hem adoptat la convenció de noms on l'id de qualsevol entitat és el nom singular de l'entitat més un guió a baix més les lletres ID. Així si tenim una taula anomenada Entitats, l'id serà Entitat_ID. En el cas de Globalhard no ens ha passat que tinguem taules amb claus primàries compostes i rareses d'altra mena però en la pràctica això és dóna sovint i amb aquest procés d'abstracció d'identificadors eliminem totes aquestes impureses i guanyem en velocitat alhora de fer joins etc.

Veiem a continuació un model lògic de la capa base :

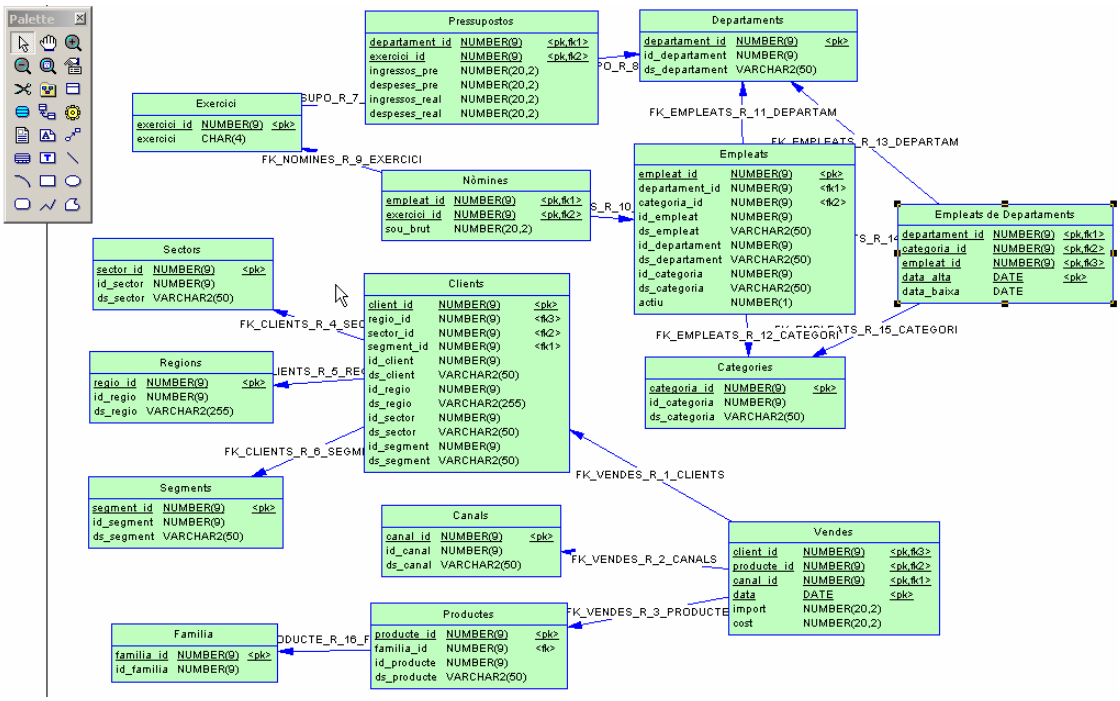


Figura 14. Diagrama de la capa BASE

En aquesta secció no entrarem a descriure un altre cop les dimensions una per una ni els detalls de cadascuna de les taules ja que ens concentrem en el procés ETL. Podem però a tall d'exemple fer una ullada a la taula Categories on veurem que tenim el camp Categoria_ID que és l'id que hem generat, el camp id_categoria que seria l'id de producció original i el camp ds_categoria que seria la descripció original de la base de dades de producció. Fixem-nos que és un model mixt *Star*, *Snowflake*, on tenim les taules de fets, les taules de dimensió, i finalment les taules de control. Així d'una taula de fets qualsevol fent només un *join* podem accedir a qualsevol dada que necessitem.

Arribats a aquest punt ja tenim el magatzem de dades carregat i a punt per treballar-hi des del *discoverer*, o qualsevol altra eina que decidim en el futur.

Diagrama de l'arquitectura programari

- **Windows 2003 Server** : sistema operatiu del servidor/s.
- **Windows XP Professional** : sistema operatiu del client/s.
- **Microsoft Access 2003** : programari client per fer una ullada a les dades subministrades per GlobalHard. No és estrictament necessari ja que podríem explorar les dades amb qualsevol ETL, però com que disposem de l'Office doncs l'emprarem.
- **Microsoft Word 2003** : per escriure la documentació del projecte.
- **Oracle Database 10g Express Edition** : base de dades on implementarem el magatzem de dades. És una versió amb funcionalitat limitada però gratuïta de la base de dades Oracle 10g.
- **SQL Navigator 5** : eina que ens permet explorar la base de dades, així com moltes altres tasques. No és indispensable però és el client amb el que normalment ens connectem a una base de dades Oracle.
- **Sybase Power Designer 11** : és una eina CASE que ens permet dissenyar bases de dades d'una manera gràfica. Així elaborem el model conceptual que llavors podem convertir en model físic sobre qualsevol dels SGBD suportats.
- **Oracle Discoverer Administrator** : és l'eina que emprarem per la definició de les dimensions i la generació dels informes multidimensionals.
- **Oracle Discoverer Desktop** : és el *front-end* client mitjançant el qual els usuaris podran connectar-se al magatzem.

Diagrama de l'arquitectura maquinari

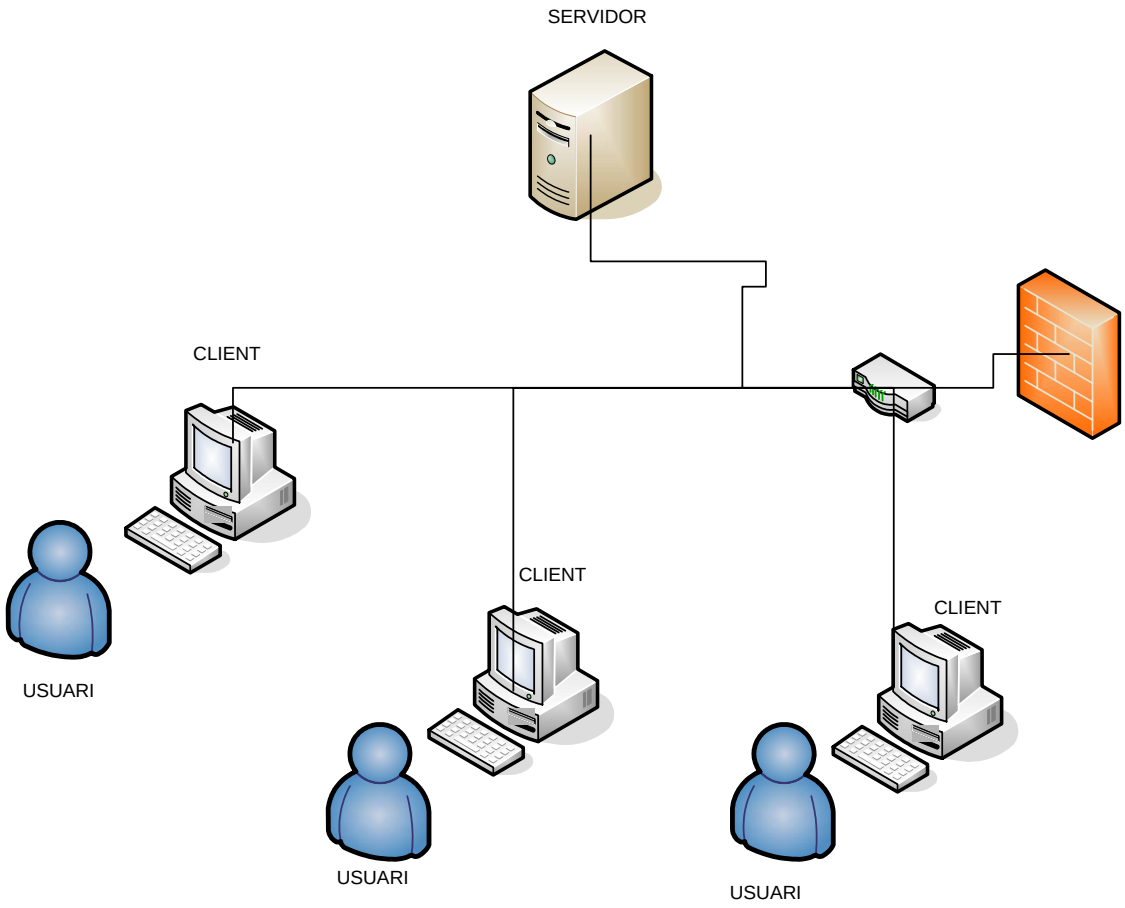


Figura 15. Diagrama de l'arquitectura de maquinari

De moment començarem amb el maquinari indispensable i en cas de necessitar-ne de més potent adaptarem la configuració inicial a una de més potent amb més maquinari, etc.

Així al servidor sobre un sistema operatiu Windows 2003 hi instal·larem :
La base de dades Oracle 10g Express Edition
El Oracle *Discoverer Administrator*

A cada màquina client sobre un sistem operatiu Windows XP Professional hi instal·larem :
Un Oracle *Discoverer Desktop*

A una màquina client que serà des d'on desenvoluparem també hi instal·larem :
Un Sybase Power Designer 11
Un SQL Navigator 5
Un Microsoft Word 2003

Disseny de la base de dades i diagrama del model físic

Capa IMP :

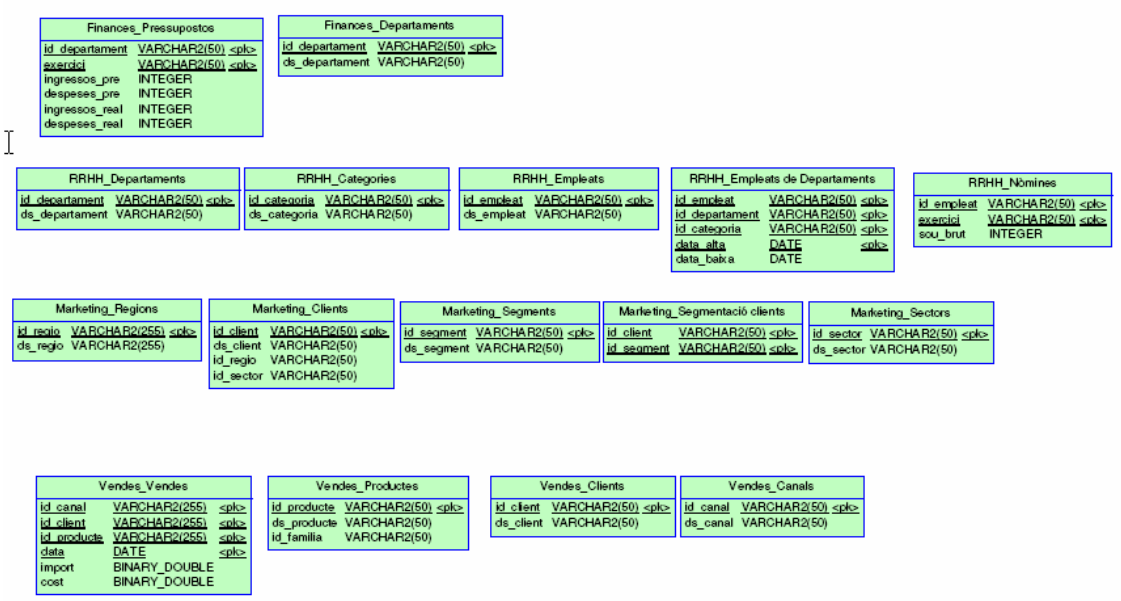


Figura 16. Disseny de la capa IMP

Capa STAGE :

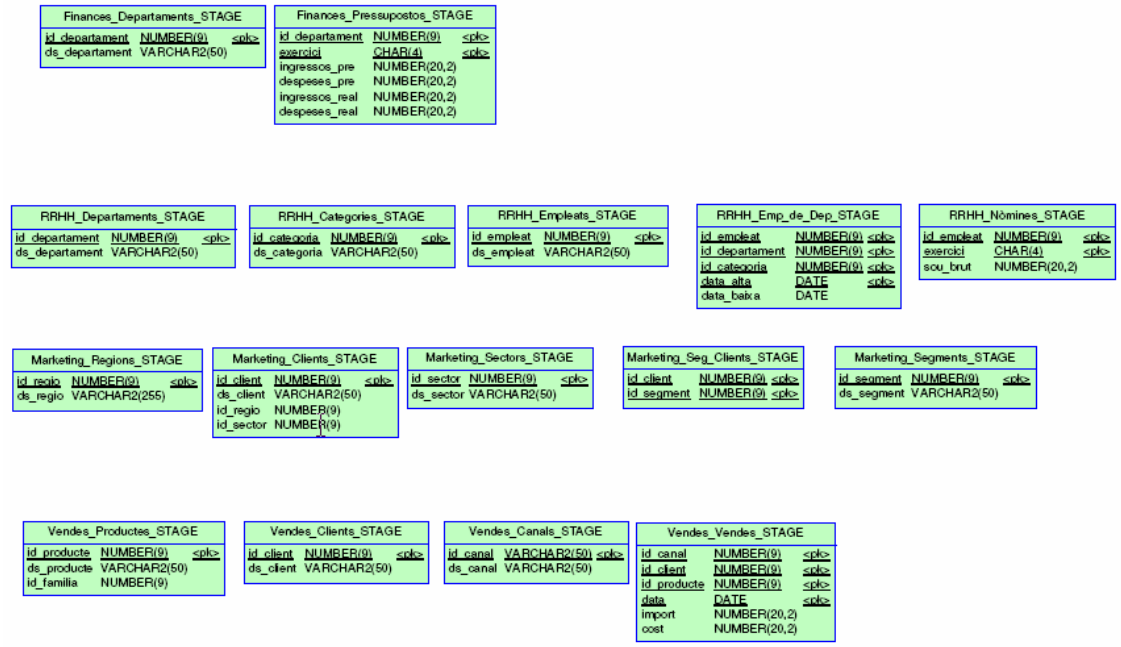


Figura 17. Disseny de la capa STAGE

Capa BASE :

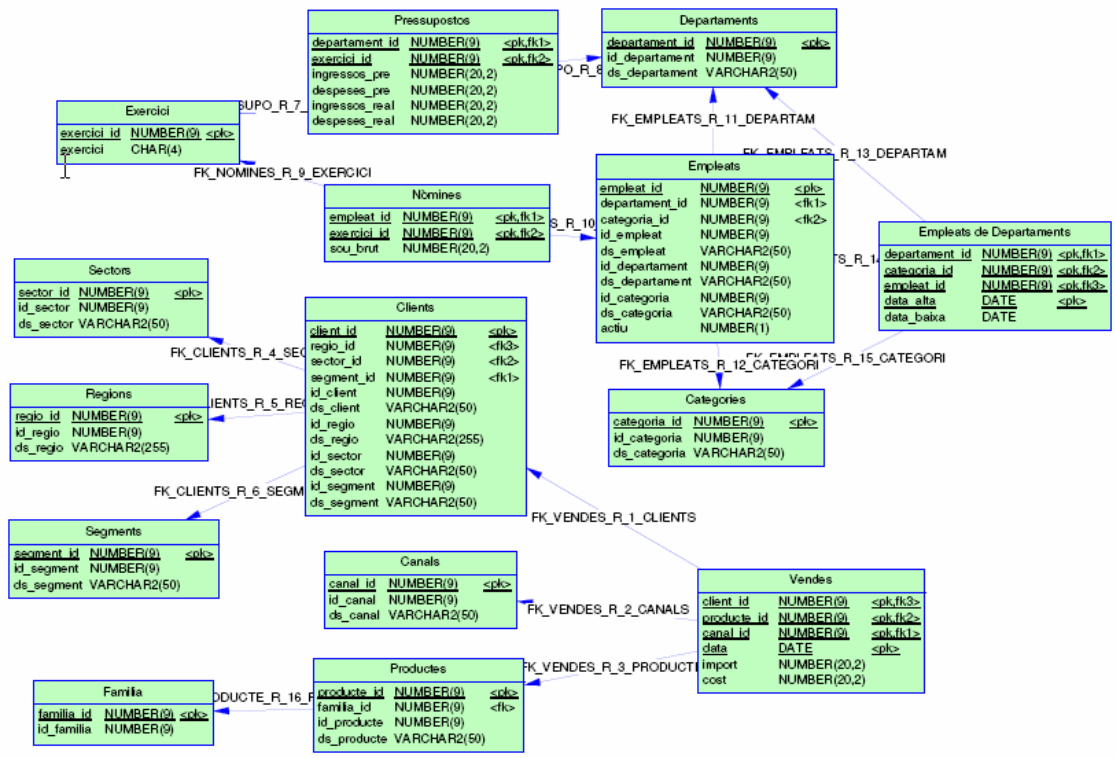


Figura 18. Disseny de la capa BASE

Disseny i descripció del informes creats

Facturació productes : facturació, cost i marge per producte

	Facturació total	Cost total	Marge
Envoy External Keyboard	€65.738.595,63	€63.231.179,88	€2.507.415,75
Envoy External 6X CD-ROM	€61.946.883,71	€59.919.848,13	€2.027.035,58
Deluxe Mouse	€49.021.761,07	€46.665.140,18	€2.356.620,89
Envoy Ambassador	€47.102.433,83	€44.266.567,02	€2.835.866,81
Envoy External 8X CD-ROM	€41.607.344,11	€40.374.543,76	€1.232.800,35
Envoy Executive	€28.253.741,26	€26.730.069,61	€1.523.671,65
Internal 8X CD-ROM	€14.249.988,31	€12.149.487,31	€2.100.501,00
Keyboard Wrist Rest	€10.000.050,39	€8.662.999,80	€1.337.050,59
O/S Documentation Set - French	€9.985.342,64	€8.505.339,11	€1.480.003,53
External 101-key keyboard	€9.855.504,34	€8.547.385,23	€1.308.119,11
Laptop carrying case	€9.158.805,54	€8.290.287,57	€868.517,97
PCMCIA modem/fax 28800 baud	€8.929.337,60	€7.615.684,88	€1.313.652,72
O/S Documentation Set - Spanish	€8.789.746,79	€7.805.731,59	€984.015,20
O/S Documentation Set - Italian	€8.343.887,31	€7.276.786,49	€1.067.100,82
Monitor- 19"" Super VGA	€6.713.739,93	€5.272.758,32	€1.440.981,61
External 6X CD-ROM	€3.548.380,95	€3.173.222,43	€375.158,52
Envoy Standard	€3.381.084,41	€2.650.675,68	€730.408,73
Sentinel Financial	€3.228.835,20	€2.583.729,43	€645.105,77
Internal 6X CD-ROM	€1.969.604,35	€1.592.064,21	€377.540,14
Multimedia speakers- 5" cones	€1.554.220,35	€1.223.816,83	€330.403,52
Unix/Windows 5-user pack	€1.545.819,98	€1.209.221,46	€336.598,52
O/S Documentation Set - English	€1.537.238,15	€1.203.763,50	€333.474,65
Monitor- 15"" Super VGA	€1.377.133,28	€1.077.990,73	€299.142,55
O/S Documentation Set - German	€1.360.010,20	€1.150.089,66	€209.920,54
Mouse Pad	€1.248.174,38	€1.016.836,20	€231.338,18
O/S Documentation Set - Kanji	€833.242,11	€724.981,60	€108.260,51
Sentinel Multimedia	€739.710,22	€605.764,32	€133.945,90
Standard Mouse	€699.302,35	€568.056,25	€131.246,10
Unix/Windows 1-user pack	€474.401,21	€372.077,87	€102.323,34
PCMCIA modem/fax 19200 baud	€460.095,85	€388.070,40	€72.025,45
Sentinel Standard	€452.555,57	€368.687,18	€83.868,39
SIMM- 16MB PCMCIAII card	€212.201,81	€174.013,78	€38.188,03
External 8X CD-ROM	€178.797,76	€145.937,82	€32.859,94
SIMM- 8MB PCMCIAII card	€149.987,64	€122.227,80	€27.759,84
Multimedia speakers- 3" cones	€93.476,95	€73.459,26	€20.017,69

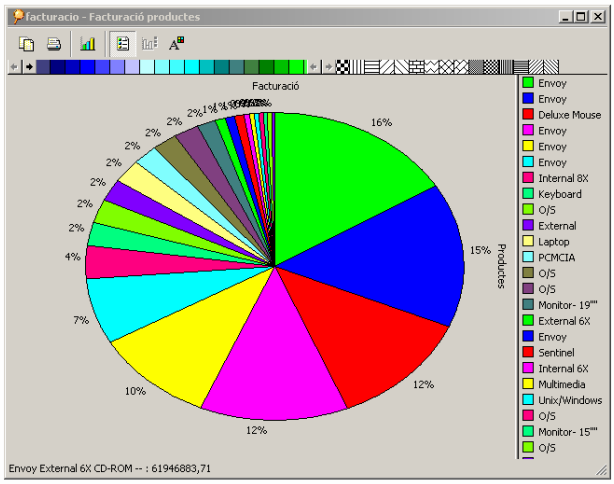


Figura 19. Captura de pantalla de l'informe de facturació i productes

Facturació regions : facturació, cost i marge per regió

	Facturació total	Cost total	Marge
United States	€128.722.085,91	€119.540.515,36	€2.248.672,66
Japan	€31.650.667,38	€29.304.922,84	€792.466,85
United Kingdom	€19.911.679,16	€18.517.128,87	€438.344,87
Singapore	€16.736.818,12	€15.581.267,28	€397.715,21
Canada	€8.166.500,07	€7.591.924,25	€196.436,97
Germany	€7.484.028,26	€6.932.683,79	€168.635,96
France	€4.467.741,05	€4.117.657,55	€111.985,76
Australia	€4.073.010,38	€3.782.007,14	€87.826,63
Italy	€3.801.120,80	€3.507.565,47	€85.342,16
Spain	€1.711.156,88	€1.578.332,59	€43.632,36
Hong Kong	€1.479.131,22	€1.351.127,90	€41.818,64

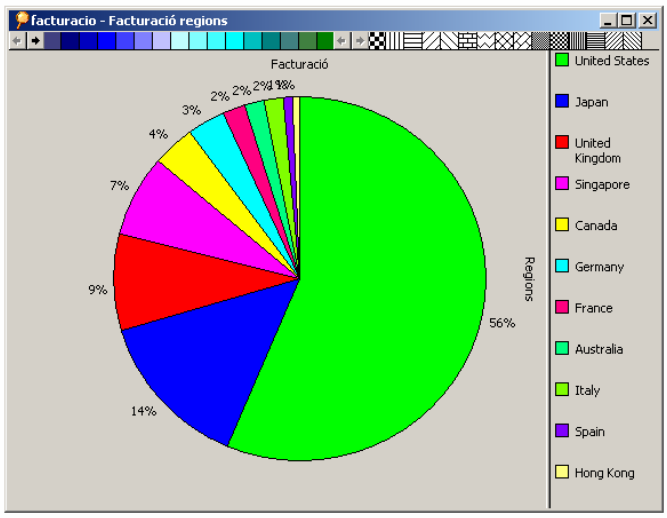


Figura 20. Captura de pantalla de l'informe de facturació i regions

Facturació canals : facturació, cost i marge per canal

	Facturació total	Cost	Marge
Catalog	€336.848.036,19	€313.211.405,85	€6.171.136,06
Internet	€14.188.485,88	€13.075.547,66	€216.190,20
Sales	€25.485.983,28	€23.667.427,05	€269.944,22

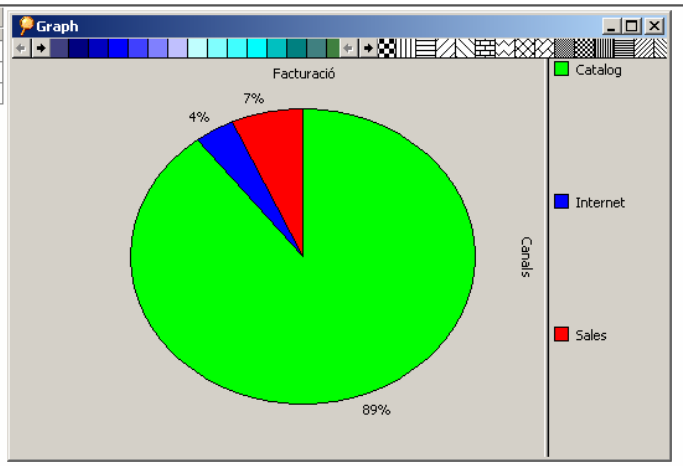


Figura 21. Captura de pantalla de l'informe de facturació i canals

Facturació sectors : facturació, cost i marge per sector

	Facturació total	Cost	Marge
Consultoria	€155.373.111,65	€144.195.759,88	€2.776.636,16
Reseller	€149.345.801,81	€138.566.619,50	€3.175.165,30
Govern	€5.681.055,51	€5.279.918,00	€42.003,06
Manufacturing	€2.689.158,59	€2.490.546,34	€36.303,38
Educació	€307.953,63	€285.619,87	€7.477,97

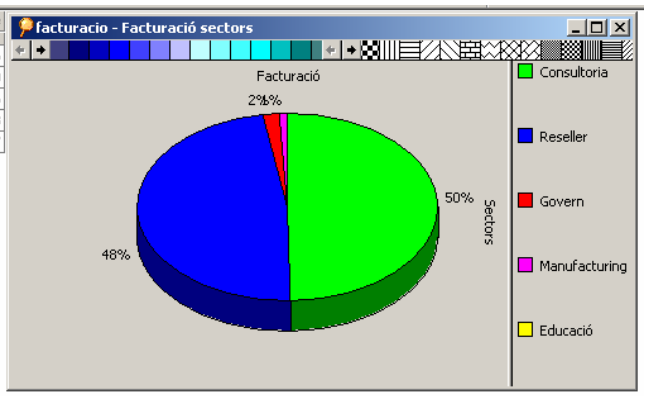


Figura 22. Captura de pantalla de l'informe de facturació i sectors

Facturació segments : facturació, cost i marge per segment

	Facturació total	Cost	Marge
TIPUS B	€398.982.535,19	€370.451.946,87	€6.876.171,71
TIPUS A	€10.538.444,26	€9.780.454,28	€65.493,28
TIPUS C	€1.431.103,04	€1.330.286,56	€29.030,94

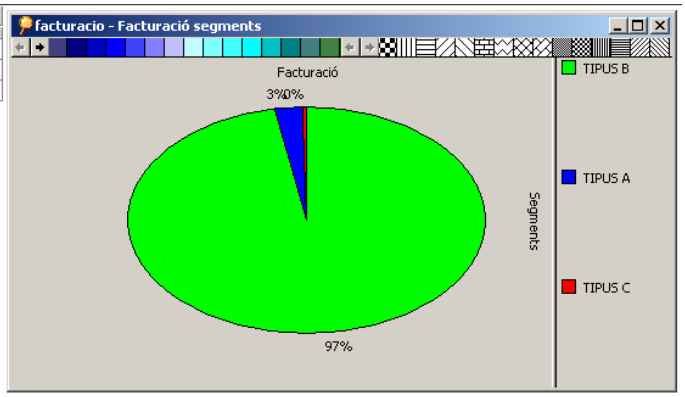


Figura 23. Captura de pantalla de l'informe de facturació i segments

Facturació temps : facturació, cost i marge en funció del temps

	Facturació total	Cost	Marge
1993	€67.881.233,56	€62.814.430,37	€1.388.051,61
1994	€102.091.116,67	€95.213.551,70	€1.825.880,91
1995	€122.212.645,16	€113.685.532,60	€2.042.561,52
1996	€125.544.575,86	€116.183.171,44	€2.014.471,06

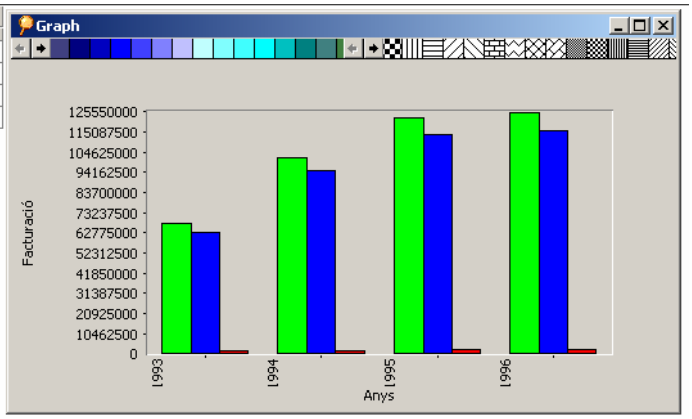


Figura 24. Captura de pantalla de l'informe de facturació i temps

Pressupostos anuals : evolució dels pressupostos (ingressos, despeses, benefici) i comparativa amb la realitat de cada any.

	1993	1994	1995	1996
Ingressos Pressupostats	€320.000.000,00	€350.000.000,00	€390.000.000,00	€375.000.000,00
Despeses Pressupostades	€260.000.000,00	€318.000.000,00	€332.000.000,00	€340.000.000,00
Ingressos Reals	€292.295.218,00	€389.045.322,00	€359.338.634,00	€339.593.636,00
Despeses Reals	€271.265.698,00	€362.390.913,00	€333.605.578,00	€314.472.032,00
Benefici Pressupostat	€60.000.000,00	€32.000.000,00	€58.000.000,00	€35.000.000,00
Benefici Real	€21.029.520,00	€26.654.409,00	€25.733.056,00	€25.121.604,00

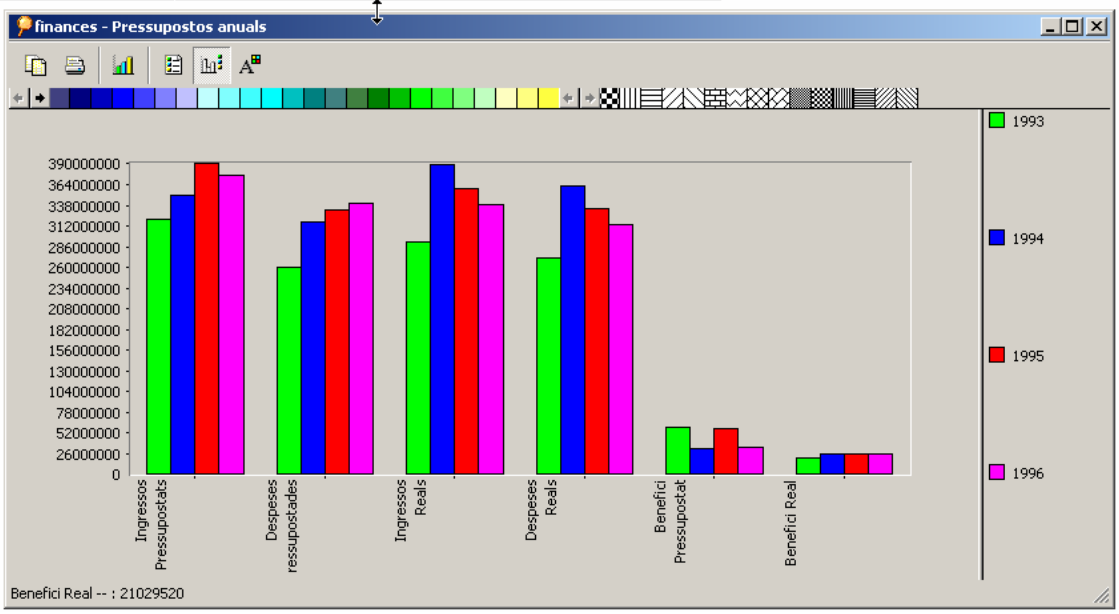


Figura 25. Captura de pantalla de l'informe de pressupostos

Distribució Clients Regions Famílies Productes : Distribució de la facturació dels diferents productes agrupats per famílies segons els diferents clients de les diferents regions.

	Volum facturació								
	4	5	6	7	8	9	10	11	12
Australia	€4.057.245,00	€284.610,02	€89.451,58	€7.159.319,49	€546.797,86	€579.297,31	€20.973,48	€6.001.696,50	€1.662.794,67
Canada	€8.685.651,91	€455.910,22	€166.625,92	€16.506.707,37	€867.341,81	€858.514,39	€36.777,07	€13.119.261,15	€2.901.817,23
France	€4.485.167,29	€327.360,30	€120.541,87	€10.679.525,11	€634.616,48	€673.951,18	€19.975,68	€9.174.151,48	€2.128.404,71
Germany	€9.896.493,34	€558.496,02	€181.939,39	€19.381.858,72	€1.072.935,98	€1.068.168,36	€33.260,49	€15.335.149,54	€3.440.964,26
Hong Kong	€1.848.124,57	€150.802,13	€44.136,58	€3.255.565,47	€276.145,25	€275.568,65	€7.530,78	€2.945.794,27	€783.020,89
Italy	€4.544.824,01	€303.851,74	€105.604,03	€8.909.010,31	€611.233,69	€601.575,31	€19.338,74	€6.677.887,93	€1.900.294,42
Japan	€45.109.645,99	€2.416.129,71	€826.135,91	€83.432.464,51	€4.177.954,40	€4.688.817,90	€198.966,44	€65.134.173,98	€14.846.730,30
Singapore	€18.895.079,83	€1.029.338,26	€337.763,31	€34.620.590,25	€1.791.309,29	€1.855.754,02	€69.222,44	€26.656.946,89	€6.171.205,71
Spain	€1.957.684,54	€133.319,00	€48.036,34	€3.797.426,78	€280.485,09	€281.355,13	€10.504,18	€3.232.725,50	€748.022,73
United Kingdom	€22.856.452,49	€1.295.024,70	€453.891,08	€43.981.419,75	€2.444.825,54	€2.446.505,45	€88.292,82	€35.517.719,84	€8.012.511,68
United States	€148.929.665,06	€8.595.413,15	€3.024.428,96	€287.229.861,80	€16.147.531,04	€16.622.680,68	€624.599,64	€229.888.481,95	€53.191.654,76

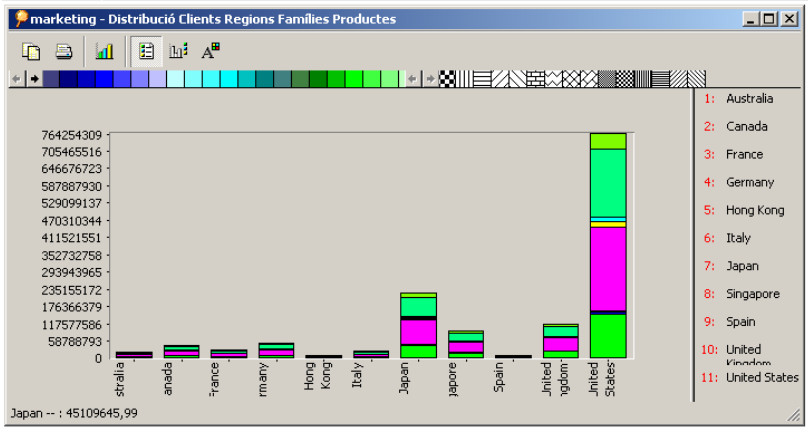


Figura 26. Captura de pantalla de l'informe de distribució de clients regions famílies i productes

Distribució Canals de venda Sectors Clients : Distribució en funció dels canals de venda i els clients agrupats per sectors.

	Volum Facturació		
	Catalog	Internet	Sales
Consultoria	€601.964.453,60	€21.864.652,23	€45.625.624,34
Educació	€1.428.471,88	€19.815,65	€13.403,60
Govern	€22.263.082,03	€305.882,64	€287.211,89
Manufacturing	€12.036.270,04	€257.319,39	€248.343,82
Reseller	€609.620.965,04	€21.101.067,41	€43.236.247,21

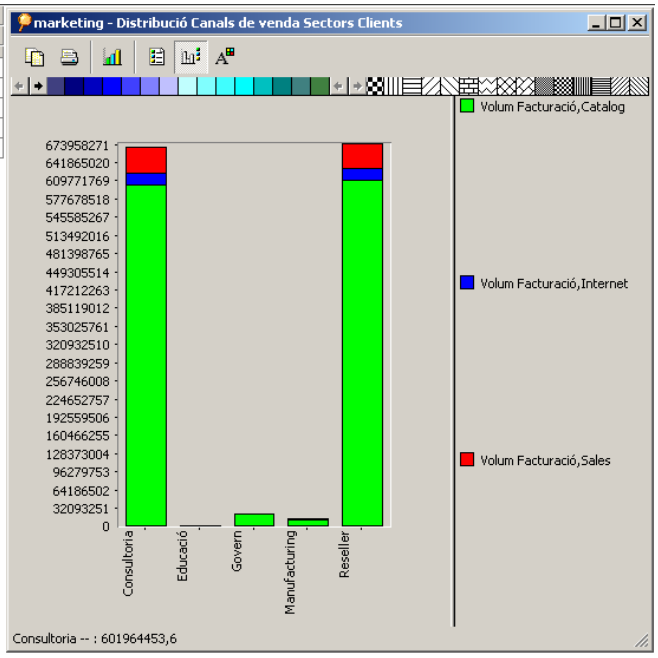
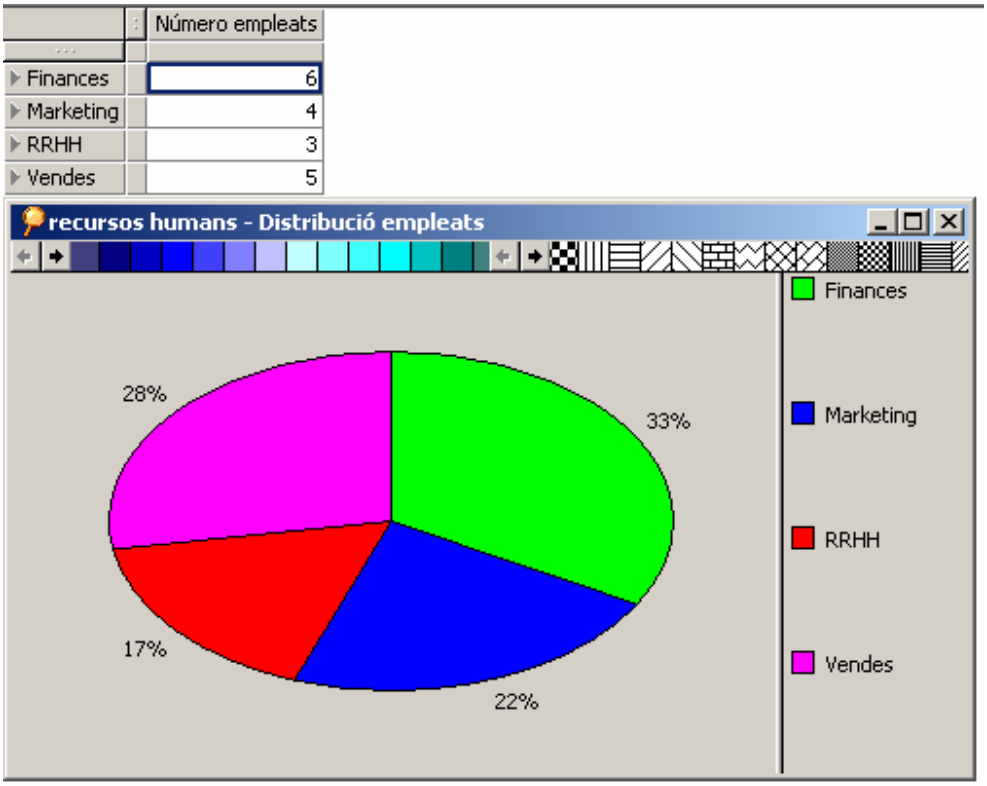


Figura 27. Captura de pantalla de l'informe de distribució de canals de venda sectors i clients

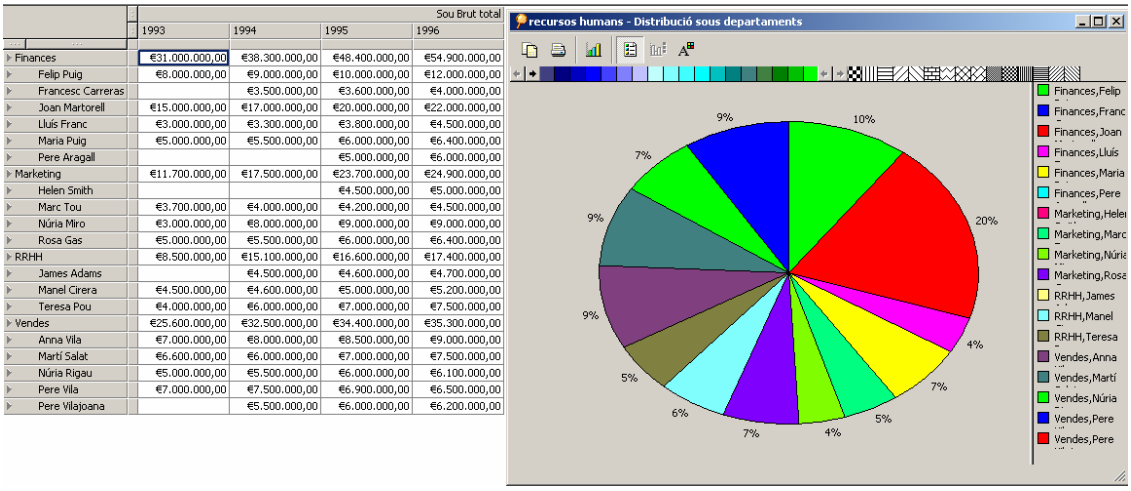
Distribució empleats : distribució dels empleats segons els departaments.



	Número empleats
Finances	6
Marketing	4
RRHH	3
Vendes	5

Figura 28. Captura de pantalla de l'informe de distribució d'empleats

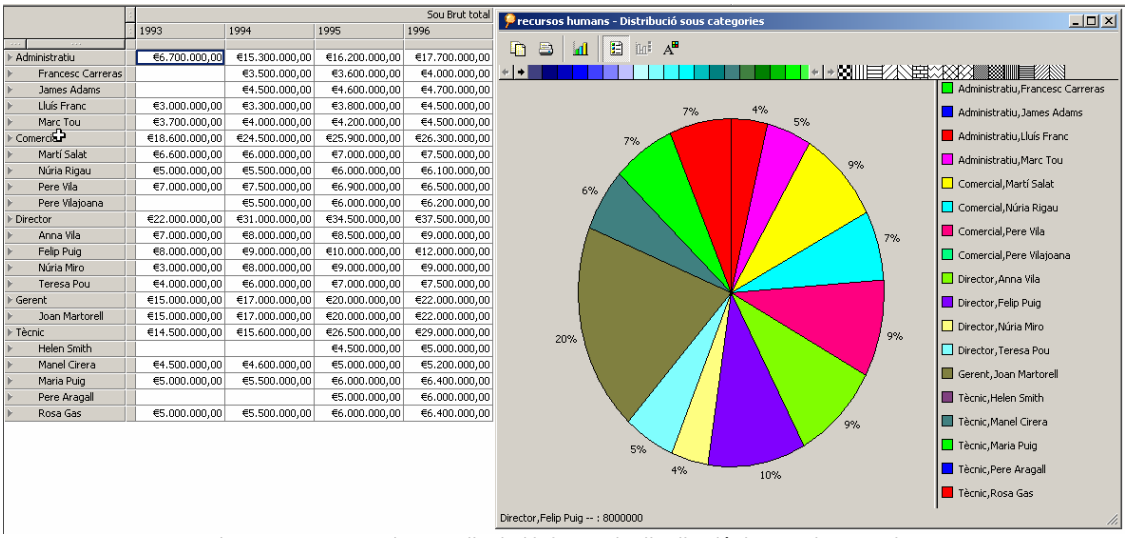
Distribució sous departament : Distribució de sous per departaments / empleats.



	Sou Brut total			
	1993	1994	1995	1996
Finances	€31.000.000,00	€38.300.000,00	€48.400.000,00	€54.900.000,00
Felip Puig	€8.000.000,00	€9.000.000,00	€10.000.000,00	€12.000.000,00
Francesc Carreras		€3.500.000,00	€3.600.000,00	€4.000.000,00
Joan Martorell	€15.000.000,00	€17.000.000,00	€20.000.000,00	€22.000.000,00
Lluís Franc	€3.000.000,00	€3.300.000,00	€3.800.000,00	€4.500.000,00
Maria Puig	€5.000.000,00	€5.500.000,00	€6.000.000,00	€6.400.000,00
Pere Aragall			€5.000.000,00	€6.000.000,00
Marketing	€11.700.000,00	€17.500.000,00	€23.700.000,00	€24.900.000,00
Helen Smith			€4.500.000,00	€5.000.000,00
Marc Tou	€3.700.000,00	€4.000.000,00	€4.200.000,00	€4.500.000,00
Núria Miro	€3.000.000,00	€8.000.000,00	€9.000.000,00	€9.000.000,00
Rosa Gas	€5.000.000,00	€5.500.000,00	€6.000.000,00	€6.400.000,00
RRHH	€8.500.000,00	€15.100.000,00	€16.600.000,00	€17.400.000,00
James Adams		€4.500.000,00	€4.600.000,00	€4.700.000,00
Manel Cirera	€4.500.000,00	€4.600.000,00	€5.000.000,00	€5.200.000,00
Teresa Pou	€4.000.000,00	€6.000.000,00	€7.000.000,00	€7.500.000,00
Vendes	€25.600.000,00	€32.500.000,00	€34.400.000,00	€35.300.000,00
Anna Vila	€7.000.000,00	€8.000.000,00	€8.500.000,00	€9.000.000,00
Martí Salat	€6.600.000,00	€6.000.000,00	€7.000.000,00	€7.500.000,00
Núria Rigau	€5.000.000,00	€5.500.000,00	€6.000.000,00	€6.100.000,00
Pere Vila	€7.000.000,00	€7.500.000,00	€6.900.000,00	€6.500.000,00
Pere Vilajoana		€5.500.000,00	€6.000.000,00	€6.200.000,00

Figura 29. Captura de pantalla de l'informe de distribució de sous

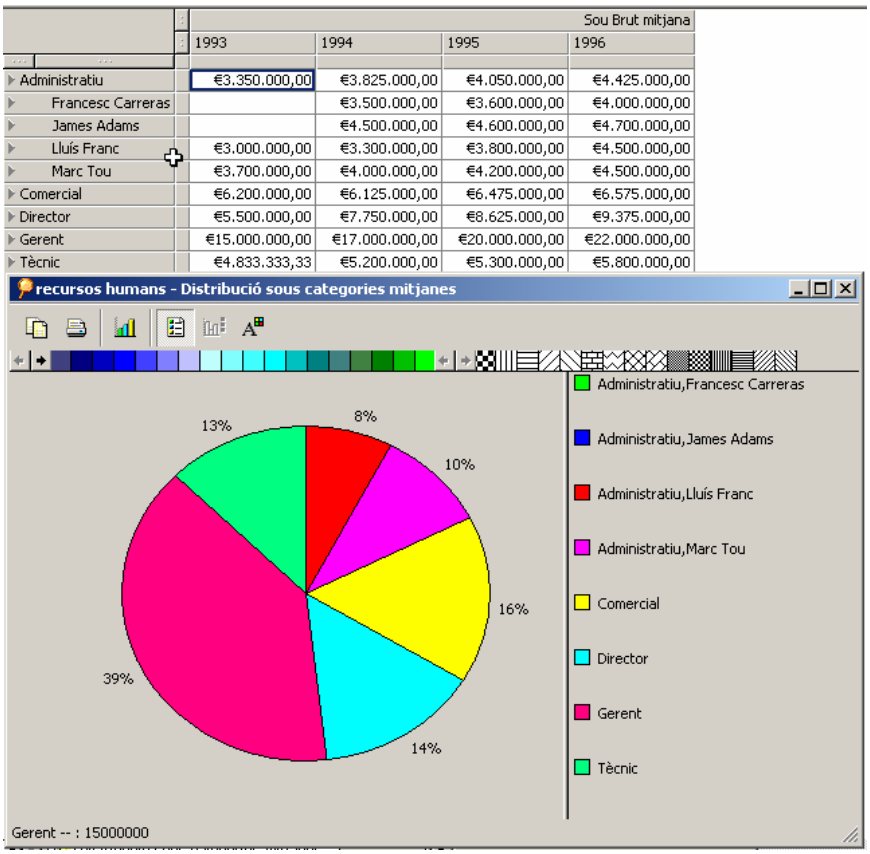
Distribució sous categories : Distribució de sous per categories / empleats.



	Sou Brut total			
	1993	1994	1995	1996
Administratiu	€6.700.000,00	€15.300.000,00	€16.200.000,00	€17.700.000,00
Francesc Carreras		€3.500.000,00	€3.600.000,00	€4.000.000,00
James Adams		€4.500.000,00	€4.600.000,00	€4.700.000,00
Lluís Franc	€3.000.000,00	€3.300.000,00	€3.800.000,00	€4.500.000,00
Marc Tou	€3.700.000,00	€4.000.000,00	€4.200.000,00	€4.500.000,00
Comercial	€18.600.000,00	€24.500.000,00	€25.900.000,00	€26.300.000,00
Martí Salat	€6.600.000,00	€6.000.000,00	€7.000.000,00	€7.500.000,00
Núria Rigau	€5.000.000,00	€5.500.000,00	€6.000.000,00	€6.100.000,00
Pere Vila	€7.000.000,00	€7.500.000,00	€6.900.000,00	€6.500.000,00
Pere Vilajoana		€5.500.000,00	€6.000.000,00	€6.200.000,00
Director	€22.000.000,00	€31.000.000,00	€34.500.000,00	€37.500.000,00
Anna Vila	€7.000.000,00	€8.000.000,00	€8.500.000,00	€9.000.000,00
Felip Puig	€8.000.000,00	€9.000.000,00	€10.000.000,00	€12.000.000,00
Núria Miro	€3.000.000,00	€8.000.000,00	€9.000.000,00	€9.000.000,00
Teresa Pou	€4.000.000,00	€6.000.000,00	€7.000.000,00	€7.500.000,00
Gerent	€15.000.000,00	€17.000.000,00	€20.000.000,00	€22.000.000,00
Joan Martorell	€15.000.000,00	€17.000.000,00	€20.000.000,00	€22.000.000,00
Tècnic	€14.500.000,00	€15.600.000,00	€26.500.000,00	€29.000.000,00
Helen Smith			€4.500.000,00	€5.000.000,00
Manel Cirera	€4.500.000,00	€4.600.000,00	€5.000.000,00	€5.200.000,00
Maria Puig	€5.000.000,00	€5.500.000,00	€6.000.000,00	€6.400.000,00
Pere Aragall			€5.000.000,00	€6.000.000,00
Rosa Gas	€5.000.000,00	€5.500.000,00	€6.000.000,00	€6.400.000,00

Figura 30. Captura de pantalla de l'informe de distribució de sous i categories

Distribució sous categories mitjanes : Distribució de la mitjana de sous per categories.



	Sou Brut mitjana			
	1993	1994	1995	1996
Administratiu	€3.350.000,00	€3.825.000,00	€4.050.000,00	€4.425.000,00
Francesc Carreras		€3.500.000,00	€3.600.000,00	€4.000.000,00
James Adams		€4.500.000,00	€4.600.000,00	€4.700.000,00
Lluís Franc	€3.000.000,00	€3.300.000,00	€3.800.000,00	€4.500.000,00
Marc Tou	€3.700.000,00	€4.000.000,00	€4.200.000,00	€4.500.000,00
Comercial	€6.200.000,00	€6.125.000,00	€6.475.000,00	€6.575.000,00
Director	€5.500.000,00	€7.750.000,00	€8.625.000,00	€9.375.000,00
Gerent	€15.000.000,00	€17.000.000,00	€20.000.000,00	€22.000.000,00
Tècnic	€4.833.333,33	€5.200.000,00	€5.300.000,00	€5.800.000,00

Figura 31. Captura de pantalla de l'informe de distribució de sous mitjans per categories

Captures de pantalla del modelatge de dades i jerarquies

Jerarquies

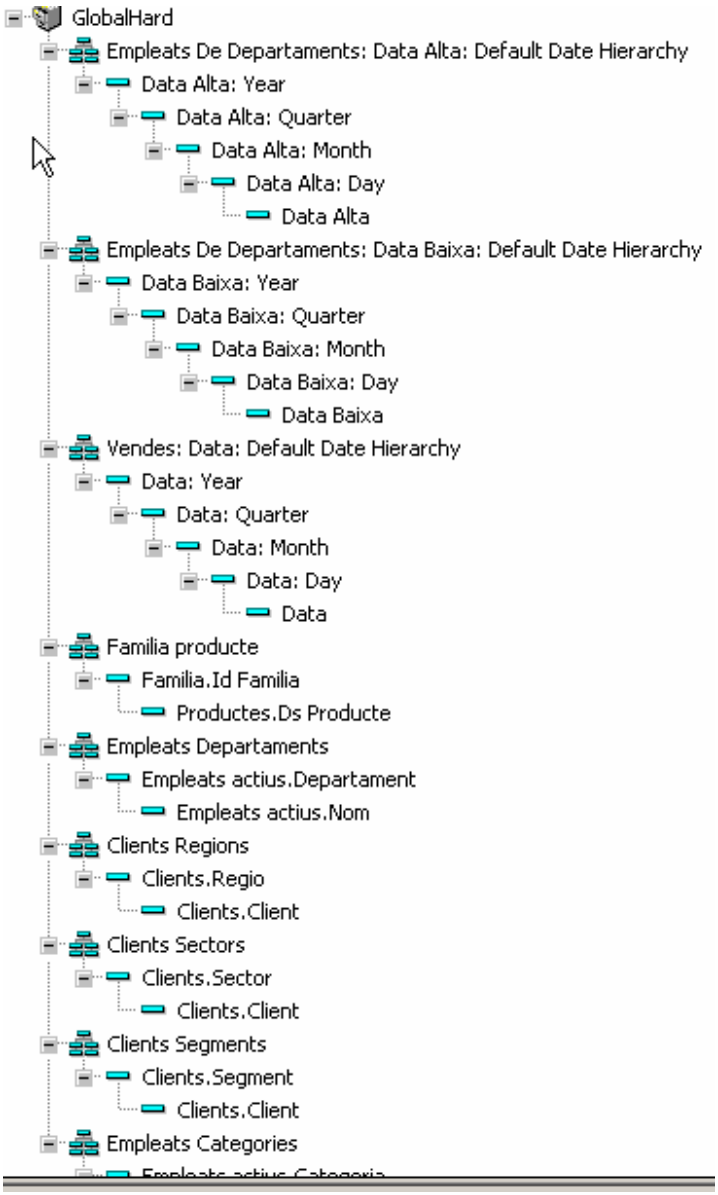


Figura 32. Captura de pantalla de l'editor de jerarquies

Dades

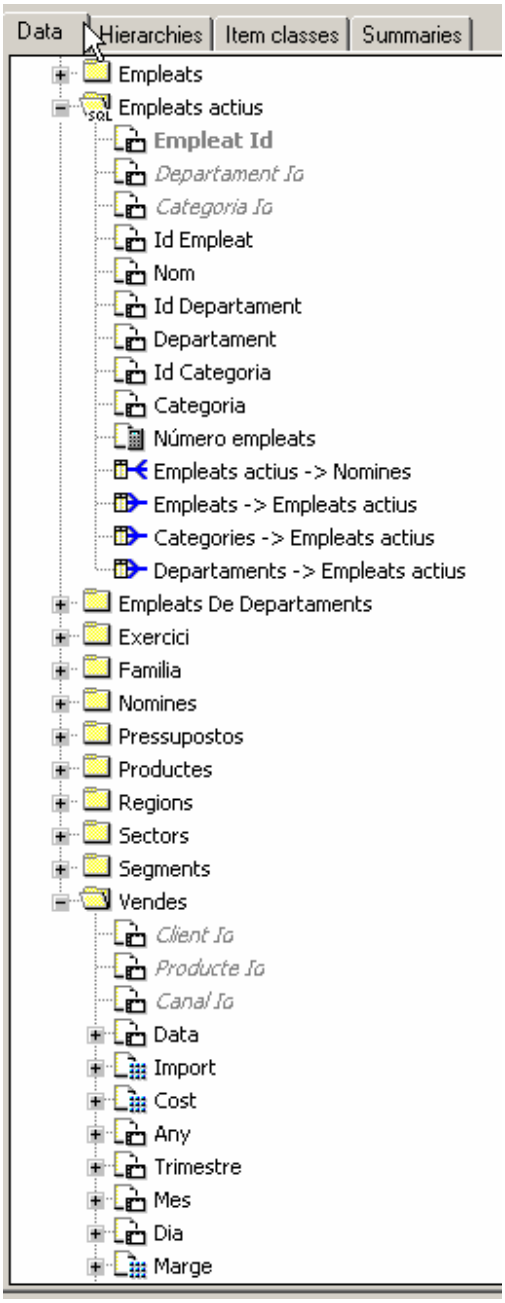


Figura 33. Captura de pantalla de l'editor de carpetes

Conclusions

GlobalHard, com qualsevol altra empresa, recull, emmagatzema i manté grans volums de dades. Tota aquesta informació però, normalment es recull per donar resposta a una necessitat ben concreta, com ara el sistema de nòmines, la facturació, etc. A més a més però és un actiu estratègic per a tota la organització si s'empra per a donar suport a la presa de decisions i com a eina de generació d'informes. El grau d'aprofitament d'aquest actiu estratègic dependrà de la nostra habilitat per desplegar la informació i posar-la a l'abast dels que prenen les decisions.

El veritable canvi esdevé quan els usuaris tenen accés a la informació i no només a les dades i poden dur a terme una anàlisi analítica per executar una presa de decisions efectiva.
El magatzem de dades per si sol facilitarà l'accés a la informació però no resoldrà tota la problemàtica. Caldrà doncs combinar-lo amb un ventall més ampli de serveis de suport a la presa de decisions.

Línies d'evolució futura

En el futur el magatzem de dades haurà de suportar les següents funcionalitats :

Oferir un ampli ventall de serveis de suport a la presa de decisions a totes les unitats de GlobalHard

Facilitar la recopilació, anàlisi i interpretació de les dades institucionals.

Dotar GlobalHard d'informació per a la planificació.

Promoure l'estandardització de les definicions de dades i la seva classificació.

Coordinació amb la comunitat de GlobalHard per establir polítiques, estàndards, procediments i rols pel què fa a la gestió i accés a les dades.

Monitorar i encarregar-se de la integritat de les dades.

Ser un recurs d'informació de manera que sigui senzill identificar la informació existent i com obtenir-la.

Servir de base per a processos de *data-mining*.

Webhousing.

Mantenir el magatzem de dades actual i rellevant a les necessitats canviants de GlobalHard. Re-avaluar els continguts del magatzem monitorant els nivells d'accessos i duent a terme enquestes entre els usuaris. Mantenir una comunicació fluida amb l'administració de GlobalHard, duent a terme anàlisis de necessitats i implementant noves iniciatives d'accés a la informació.

Dotar els usuaris finals del suport adequat i necessari. Amb la formació continuada i el suport en les eines de l'entorn del magatzem de dades. Mantenir una documentació en forma web així com definicions de les metadades que siguin accessibles a tothom.

Glossari

Base de dades multidimensional : base de dades dissenyada per al processament analytic en línia (OLAP), estructurada com un hipercub amb un eix per dimensió.

Business intelligence: habilitat de transformar les dades en informació, i la informació en coneixement, i així optimitzar el procés de presa de decisions dels directius.

CRM : conjunt d'estratègies de negoci, màrqueting, comunicació i infraestructures tecnològiques, dissenyades amb l'objectiu de treballar amb clients de manera individual identificant, comprenent i satisfent les seves necessitats.

data mining : mineria de dades. E Extracció d'informació que es preveu que està amagada en grans bases de dades.

dimensió : en una base de dades relacional, cada camp en un registre representa una dimensió.

EIS : sistema informàtic que permet als executius d'accedir de manera fàcil a la informació interna i externa que influeix en els factors crítics de l'èxit de l'empresa.

ERP : programari que permet d'integrar en un sistema tota la informació de l'empresa.

ETL : Extraction, Transformation & Loading : proces a partir del qual es realitza una extracció de dades d'un sistema original, es transformen i es carreguen a un sistema destí.

Magatzem de dades : base de dades orientada a àrees d'interès de l'empresa, que integren dades de diferents fonts, amb informació històrica i no volàtil que té com a objectiu principal fer de suport en la presa de decisions

OLAP : on line analitic processing. Processament analític en línia : aplicacions que permeten als usuaris de veure, manipular i analitzar bases de dades multidimensionals i navegar-hi.

webhousing :fet de poder consultar el magatzem de dades per web.

Bibliografia

INMON, W. H: *Building the Data Warehouse (3rd Edition)*. New York: Wiley, 1996

CASERTA, Joe; KIMBALL, Ralph: *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. New York: Wiley, 2004

KIMBALL, Ralph; ROSS, Margy: *The Data Warehouse Toolkit. (2nd Edition)*. New York: Wiley, 2002

Enllaços d'interès a Internet

http://datawarehouse.ittoolbox.com
http://www.businessintelligence.com
http://www.oracle.com
http://www.inmoncif.com
http://www.ralphkimball.com
http://www.dwreview.com
http://www.1keydata.com/datawarehousing/datawarehouse.html
http://www.intranetjournal.com/features/datawarehousing.html