



Universitat Oberta
de Catalunya

www.uoc.edu

Memòria

*Construcció i explotació d'un magatzem de dades de
planificació hidrològica*

Joan Carles Vera Hernández
Enginyeria Tècnica d'Informàtica de Sistemes
Consultor: José Ángel Martín Carballo
10 Gener de 2011

Resum

El treball "Construcció i explotació d'un magatzem de dades de planificació hidrològica" consisteix en la construcció d'un magatzem de dades o Data Warehouse, per tal de donar solució a la obtenció d'una sèrie d'informes destinats a La Confederació Hidrogràfica Nord-est (CHNE) amb l'objectiu de poder analitzar i controlar l'estat dels embassaments dins el seu abast geogràfic.

El procés en general consisteix, que a partir d'una o varies fonts de dades es processa aquesta informació mitjançant processos ETL (Extraction, transformation and load) netejant, transformant i carregant les dades generades en un o varis 'data marts' o bades de dades departamentals.

Un cop desada la informa s'explota utilitzant les eines OLAP (Online Analytical Processing), que permeten mostrar la informació en el format desitjat.

El treball es desenvolupa en varies fases o processos:

- Obtenció de les dades aportades per la CHNE
A partir d'arxius Excel, origen de la informació a tractar es transformen en arxius amb format .csv
- Càrrega dels arxius.
El arxius .csv son carregats a la BBDD Oracle en taules temporals per a ser tractats en fases posteriors
- Transformació de les dades
Utilitzant llenguatge PL/SQL les dades de les taules temporals son transformades i copiades a noves taules definitives.
- Explotació de les dades.
Utilitzant eines OLAP com ara Oracle Discoverer i s'ha dissenyat els informes sol·licitats.

El treball s'ha desenvolupat utilitzant les següents eines:

- Excel com a eina de conversió a format .csv
- Sqldr per a carregar els arxius csv.
- PL/SQL per a transformar les dades
- Oracle com a SGBD (Sistema de gestió de BD)
- Oracle Discoverer (Administrator i Desktop) com eina OLAP i de creació d'informes

Taula de contingut

1.	Introducció	5
1.1	Justificació del TFC i context en el qual es desenvolupa: punt de partida i aportació del TFC.	5
1.2	Objectius	6
1.3	Enfocament i mètode seguit.....	7
1.4	Planificació.....	9
1.5	Productes obtinguts	11
1.6	Altres capítols de la memòria	12
2.	Anàlisis	13
2.1	Casos d'ús	13
2.2	El model conceptual.....	14
	Fets.....	16
	Grànul	16
	Dimensions	16
	Mesures	17
2.3	Diagrama	19
3.	Disseny	20
3.1	Diagrama de l'arquitectura software.....	20
3.2	Diagrama de l'arquitectura hardware	23
3.3	Disseny de la base de dades i diagrama del model físic	25
3.4	Processos ETL	28
3.5	Disseny i descripció del informes creats.....	33
4.	Captures de pantalla.....	39
5.	Conclusions	45
6.	Línies d'evolució a futur.....	46
7.	Glossari.....	47
8.	Bibliografia i referències	49

Taula d'il·lustracions

Figura 1. Esquema del cicle de vida.....	7
Figura 2. Fites principals	9
Figura 3. Diagrama de Gann del TFC	10
Figura 4. Cas d'us de l'usuari	13
Figura 5. Cas d'us de l'administrador.....	14
Figura 6. Model Conceptual	19
Figura 7. Diagrama d'arquitectura de programari.....	21
Figura 8. Diagrama d'arquitectura de maquinari	23
Figura 9. Diagrama físic BD DT	25
Figura 10. Taula DT_NET_CAPACITATS.....	26
Figura 11. Taula DT_TMP_EMBASSAMENTS	26
Figura 12. Diagrama BD DW	27
Figura 13. Pantalla de selecció d'unitats de mesura	33
Figura 14. Capacitat total vs capacitat actual per riu i any i mes	39
Figura 15. Capacitat total vs capacitat actual per embassament any i mes	40
Figura 16. Capacitat total vs capacitat actual per comunitat autònoma, any i mes	41
Figura 17. Comparativa anual de l'evolució per comunitat autònoma de la capacitat agregada de tots els seus embassaments.....	41
Figura 18. Comparativa de la capacitat mitjana agregada de tots els embassaments d'una comunitat autònoma entre un any i l'any anterior ...	42
Figura 19. Aportació (Entrada-Sortida) per riu, embassament i any, comparativa amb la mitjana anual d'aportació de l'embassament	42
Figura 20. Aigua diària en reserva disponible a tota la confederació hidrogràfica per any i mes	43
Figura 21. Evolució d'aigua de neu equivalent prevista per any i embassament.....	43
Figura 22. Cassos de desbordament.....	43
Figura 23. Períodes de sequera.....	44

1. Introducció

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

Aquest projecte s'anomena '*Construcció i explotació d'un magatzem de dades de planificació hidrològica*'. Un magatzem de dades consisteix en el que es coneix com una estructura multidimensional de dades, a partir de la qual, la empresa client pot obtenir informació acurada sobre la seva activitat.

Partim de informació que es distribueix en arxius Excel que poden provenir d'un sistema OLTP (OnLine Transaccional Process), que és com es defineixen els sistemes ERP o de gestió típics de les empreses, que es caracteritzen per tenir molta informació en petites unitats, com ara albarans factures, comandes, etc. i molts moviments d'aquest. El podríem anomenar com el sistema que conté '*dades*'.

Cada un d'aquests arxius conté informació mensual relativa a un embassament.

Partint d'aquests arxius no podem obtenir gaire informació del sistema hidrològic, hem d'agrupar, netejar i transformar aquesta informació per a poder obtenir una visió més concreta del nostre sistema. Per això es justifica el desenvolupament d'un magatzem de dades.

Un cop tenim el magatzem o Data Warehouse creat, podem obtenir de ell tota mena d'informació. Podem veure informació relativa a un embassament i un mes evidentment, però també podem veure aquesta mateixa informació de tots els embassaments, o per comunitat autònoma, per província, any, mes, etc.

Podem crear qualsevol informació disponible a la BD, podem agregar informació o podem baixar fins al nivell més baix de detall que permet la granularitat de les dades obtingudes.

El que resulta encara més atractiu es que utilitzant tecnologia OLAP, podrem visualitzar les dades ara transformades en '*informació*' de totes aquestes maneres i això està a l'abast de pràcticament qualsevol usuari sense coneixements informàtics que podrà manipular al seu gust creant gràfics, taules, o quadres de comandament (Dashboards) en pocs minuts.

Gràcies al TFC tenim un procés automatitzat de les dades, substituint els processos manuals actuals que ocupen molts recursos i temps.

També es proveeixen mecanismes de neteja de les dades, i en resum, el magatzem de dades ens permet a partir d'una font de dades desorganitzada, tenir una visió molt concreta i precisa del estat i l'evolució

d'un sistema com el del nostre cas, i a més pot ser explotat i consultat per usuaris sense gaire especialització informàtica.

1.2 Objectius

Els objectius finals d'aquest TFC son:

- Automatitzar la recollida d'informació dels arxius Excel
- Depuració de la informació recollida
- l'anàlisi, disseny i construcció d'un magatzem de dades per a la CHNE
- Explotació de les dades mitjançant eines de Business Intelligence

Resulta obvi que per a aconseguir aquesta fita calen primerament assolir una sèrie de coneixements i tècniques que sense elles no el podríem abordar.

- ✓ La capacitat d'anàlisi que ens permet determinar com hem de manegar la informació des de el seu origen fins a la estructura final a la BD.
- ✓ Coneixements de eines adients per a executar processos de transformació i càrrega de dades (ETL) així com la determinació de quins processos son necessaris per a la obtenció de la màxima qualitat d'informació.
- ✓ Obtenir el coneixement i habilitat per a generar un magatzem de dades multidimensional a través de un sistema de gestió de bases de dades relacionals.
- ✓ Aplicar els conceptes bàsics de la teoria de magatzems de dades per a modelar la informació processada amb fets, dimensions, mesures, jerarquies, etc.
- ✓ L'aprenentatge de l'ús d'eines OLAP com l'Oracle Discoverer, per a, finalment, poder satisfer els requeriments demanats pel que fa als informes sol·licitats.
- ✓ No es un objectiu prioritari, però si un element a considerar, la optimització dels processos i del disseny de forma que el rendiment del magatzem no sigui un element que cridi l'atenció per la seva lentitud.

1.3 Enfocament i mètode seguit.

L'enfocament seguit no difereix gaire de qualsevol altre projecte tradicional de enginyeria de programari.

Per tant s'ha utilitzat en el procés de desenvolupament el mètode del cicle de vida en cascada, que sol ser un dels mètodes mes utilitzats en general per a projectes informàtics.

Aquest mètode destaca per la seva planificació senzilla, el no requerir personal massa especialitzat i el producte obtingut es de gran qualitat.

Es necessari per a poder utilitzar aquest mètode que els requeriments estiguin definits des del inici del projecte.

No obstant hem adoptat una variant d'aquest cicle de vida, anomenada cicle de vida en 'V'.

La diferencia consisteix en que en aquest hi ha dues etapes de retroalimentació entre l'anàlisi i el manteniment i entre el disseny i les proves com es pot apreciar a la Figura 1.

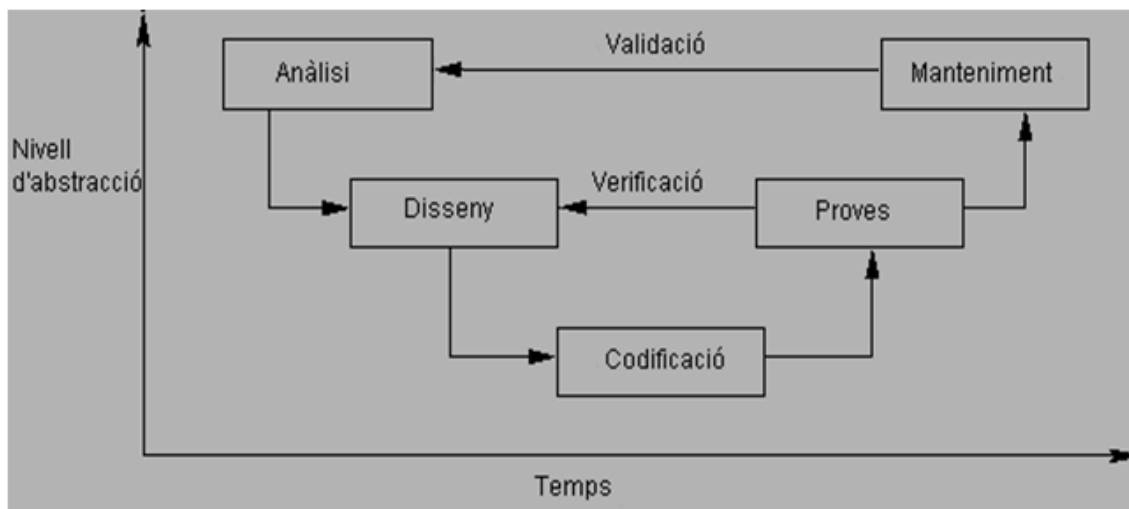


Figura 1. Esquema del cicle de vida

Cada una de las fases del cicle sol anar acompanyada de la documentació adient, que és la base per a poder dur a terme la fase següent. En aquest TFC es disposa la següent documentació:

- Document inicial de requeriments o enunciat del TFC
- Document d'anàlisi

Construcció i explotació d'un magatzem de dades de planificació hidrològica

- Document de disseny
- Processos ETL, que correspon a la implementació o codificació.
- Documentació de producte, relatiu als informes.

Tota la documentació descrita forma part d'aquesta memòria.

No hi ha documentació ni fase de Manteniment degut a les característiques especials del projecte al ser un TFC.

1.4 Planificació

La planificació d'aquest TFC ve donada per les fites ja definides al Pla docent de la assignatura.

Es mostra una taula a continuació que conte les fites definides amb la seva data de inici i final:

Fita	Data Inici	Data Final	Descripció
Inici TFC	21/09/2010		Inici del Treball fi de Carrera
Trobada	30/09/2010	30/09/2010	Trobada presencial amb el consultor
PAC1	23/09/2010	05/10/2010	PAC1 - Elaboració del pla de treball i anàlisi de requeriments
PAC2	06/10/2010	08/11/2010	PAC2 - Anàlisi
PAC3	09/11/2010	20/12/2010	PAC3 - Construcció
Lliurament final	21/12/2010	10/01/2011	Lliurament final de TFC
Debat	24/01/2011	27/01/2011	Període de debat

Figura 2. Fites principals

Aquestes fites s'han de complir obligatòriament i constitueixen en general les dates a on s'ha de lliurar cada una de les parts del projecte.

Cada una d'aquestes fites està desglossada en tasques, el detall de totes les tasques s'ha descrit en el diagrama de Gann que es mostra a continuació en la Figura 3.

Construcció i explotació d'un magatzem de dades de planificació hidrològica

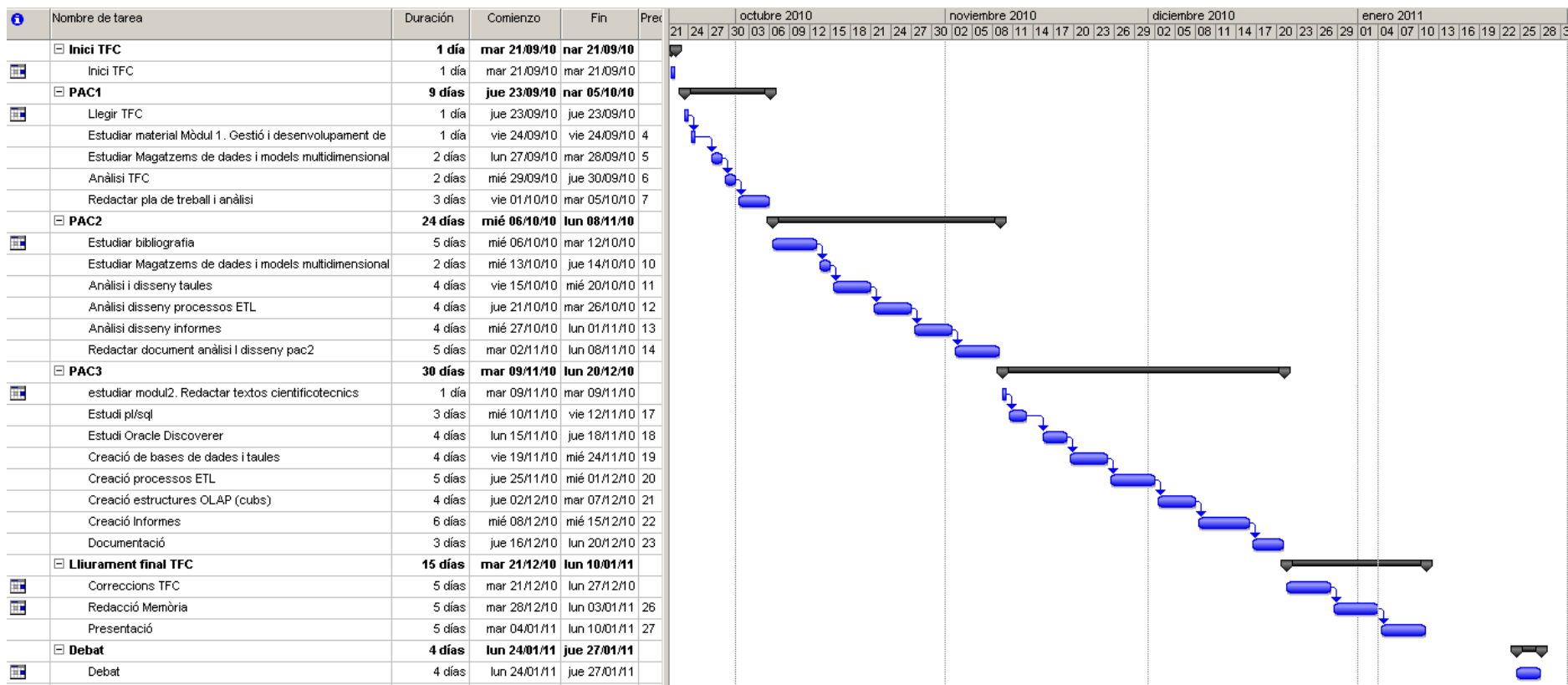


Figura 3. Diagrama de Gann del TFC

1.5 Productes obtinguts

A continuació es detallen els productes obtinguts durant la realització del TFC.

- **Pla de Treball.** S'elabora al començament del projecte per a donar una idea de l'abast del mateix, concretament es defineixen els objectius que es volen aconseguir, es dona una planificació temporal de les taques en que es divideix el projecte, es defineixen els requeriments de tota mena necessaris per a dur-lo a terme, tals com maquinari, programari, requeriments de formació, i d'altres. Finalment es descriuen els riscos que amenacen el bon termini del projecte, intentant proposar una possible solució en cas que el risc es faci realitat.
- **Anàlisi.** Document en que es defineixen les dimensions, jerarquies, taules de fets i les mesures necessàries per a construir la BD multidimensional del nostre model.
També es decideix com es resoldran cada un del informes a realitzar
- **Disseny.** Aquest document va juntament amb l'anterior i es defineix quina serà la estructura i relacions de les taules del model analitzat. Es defineixen també els processos ETL necessaris per a obtenir les dades adients a la BBDD i també es dona una visió general de la arquitectura del sistema plantejat.
- **Accés a la Base de Dades.** Document a on s'explica com accedir tan a la BBDD com a les eines d'administració i reporting, per a visualitzar o crear nous informes.
- **Documentació ETL.** Descripció tècnica detallada del processos ETL. S'explica com es duen a terme la carrega, transformació i extracció de les dades per a adaptar-les al model dissenyat.
- **Documentació Informes.** S'explica el detall d'implementació i us de cada un dels informes demanats.
- **Memòria.** És aquest document.

1.6 Altres capítols de la memòria

En aquest punt es fa una breu descripció de la resta de capítols inclosos en aquest document.

Capítol 2, Anàlisi. En mostra en aquest apartat els elements d'anàlisi com ara diagrames de casos d'ús comentats així com el diagrama conceptual del model que s'ha dissenyat.

Capítol 3, Disseny. En aquest capítol es mostren varis diagrames, entre ells el d'arquitectura de software i hardware, el disseny de la BD i el model físic. També es detallen els informes creats i el processos ETL implementats.

Capítol 4, Captures de pantalla. Imatges dels tots els informes que s'han desenvolupat per aquest TFC.

Capítol 5, Conclusions. Conclusions que s'extreuen un cop finalitzar el TFC.

Capítol 6, Línies de evolució a futur. Es descriu quin es l'escenari a futur d'aquest desenvolupament basat en millores proposades a partir del model actual

2. Anàlisis

2.1 Casos d'us

Tenim dos actors fonamentals en el model

- ❖ Usuari. Utilitza la eina Discoverer Desktop, que li permet accedir als informes.
- ❖ Administrador. Utilitza la eina Discoverer Administrator per a gestionar la estructura del Cub, definir dimensions, jerarquies, etc.

Cas d'us del usuari.

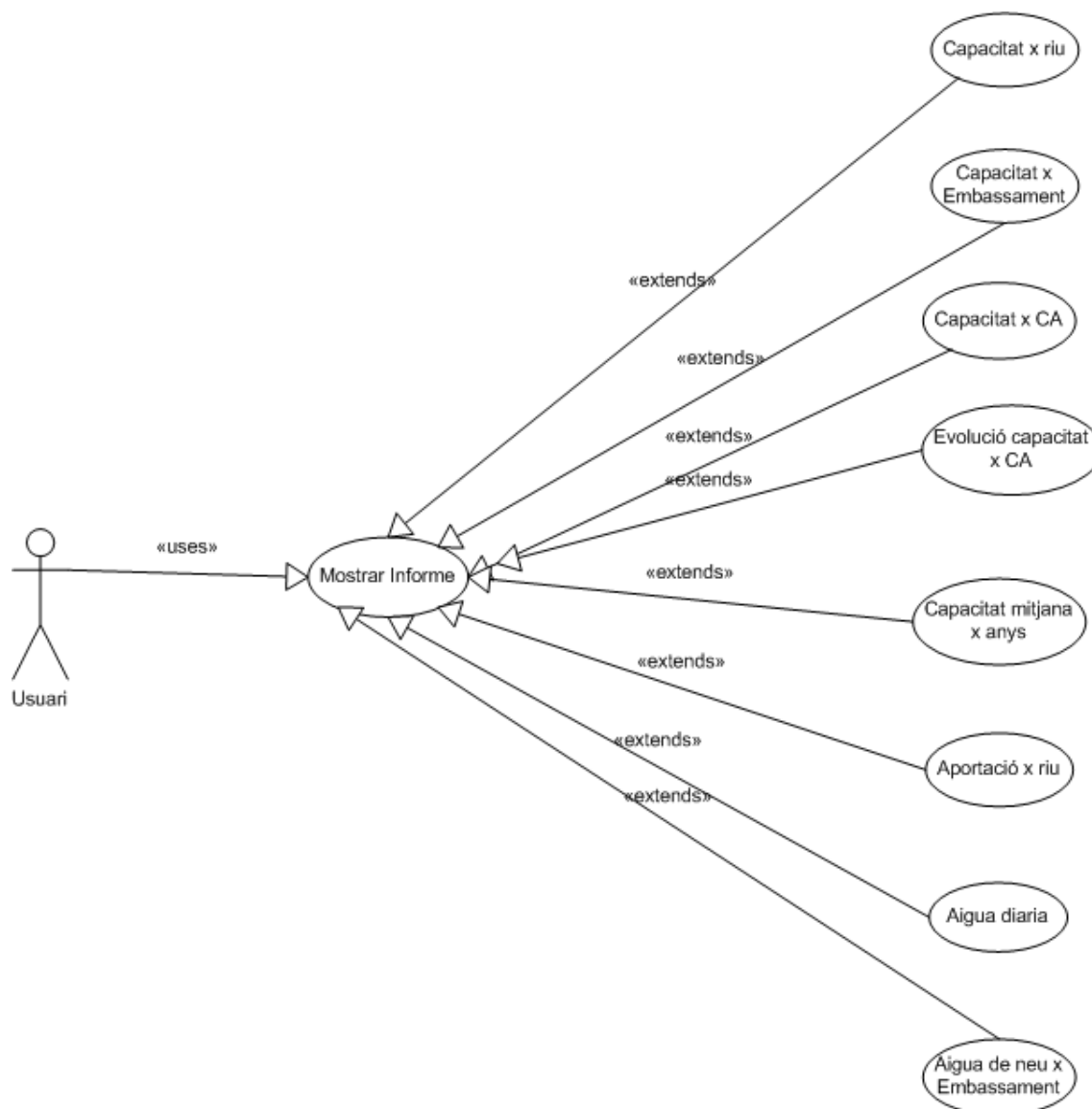


Figura 4. Cas d'us de l'usuari

Cas d'ús del Administrador

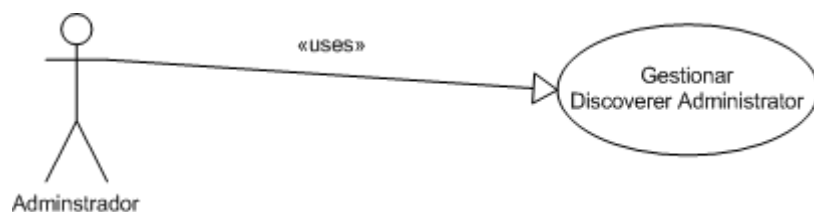


Figura 5. Cas d'us de l'administrador

2.2 El model conceptual

Per a poder especificar amb detall els elements que conformaran el sistema primer hem de veure el que ens demanen:

- **Capacitat total vs capacitat actual en Hm3 per riu i any i mes.** Per a aquest informe cal agrupar els registres per embassament i any sumant capacitats totals i actuals.
- **Capacitat total vs capacitat actual en Hm3 per embassament any.** Per a aquest informe cal agrupar els registres per embassament i any sumant capacitats totals i actuals.
- **Capacitat total vs capacitat actual en Hm3 per comunitat autònoma i any.** Informe similar als anteriors, cal agrupar els registres a un nivell superior, la comunitat autònoma i any sumant capacitats totals i actuals.
- **Comparativa anual de l'evolució per comunitat autònoma de la capacitat agregada de tots els seus embassaments.** Suma de les capacitats actuals dels embassaments agrupant per comunitat i any, es mostra el total de cada any per a mostrar l'evolució
- **Comparativa de la capacitat mitjana agregada de tots els embassaments d'una comunitat autònoma entre un any i l'any anterior (valors i % d'increment any anterior/any).** Informe que ha de mostrar agrupat per comunitat autònoma la mitjana de la seva capacitat, tant de l'any en curs com de l'any anterior. El valor promig no el tenim al sistema, l'hauem de calcular com la (suma de capacitats de tots els registres any /numero de registres)
- **Aportació (Entrada-Sortida) en Hm3 per riu, embassament i any; comparativa amb la mitjana anual d'aportació d'embassament.** Tenim la aportació per riu, embassament, any, no tenim la mitjana anual d'aportació d'embassament. Haurem de

calcular aquesta mitjana com (la suma d'aportacions de tots el registres any /numero de registres)

- **Aigua diària en reserva disponible (Hm³/dia) a tota la confederació hidrogràfica per any.** S'entén com aigua diària la mitjana de la capacitat actual per dia, per a obtenir això podem fer la següent operació (capacitat actual any/numero de dies del mes). Aquest valor l'enregistrarem a la taula de fets donat que el podem utilitzar directament per a aquest informe. Només cal fer l'agregació de suma de la mesura per aquest informe.

- **Evolució dels Hm³ d'aigua de neu equivalent prevista per any i embassament. Mostreu, si n'hi ha, els casos de desbordament de l'embassament.** L'aigua de neu equivalent ens ve donat per la fórmula: (Hm³ d'aigua equivalent per any és igual a la mitjana de la capacitat agregada anual de l'embassament multiplicada per l'estimació de neu de l'any considerat i dividit per 5). L'estimació de neu es dona anualment. En principi es proveeix com a taula de valors, tot i que el plantejament final serà utilitzar un arxiu d'intercanvi en format .csv. Enregistrarem aquest valor calculat a la taula de fets en el primer registre de l'any, es a dir el que pertany al mes de gener. La resta de registres tindran el valor '0' en aquest camp. Així utilitzarem un membre calcular que acumuli aquest valor de forma que el seu valor a qualsevol mes sempre serà el valor del mes de gener (o el primer):

Si d'aquests requeriments en traiem les paraules clau com:

- Riu
- Any, Mes
- Capacitat actual
- Capacitat total
- Comunitat autònoma
- Capacitat mitjana (agregada)
- Aportació (entrada – sortida)
- Embassament
- Aigua diària
- Aigua Neu

I separem els conceptes en dos grups:

- **Mesures** (Capacitat actual, total, aportació, Aigua diària, aigua neu)
- **Dimensions** (Riu, Any,mes, Comunitat, Embassament)

Amb aquesta classificació ja podem determinar el que volem mesurar (Mesures) i per quins criteris volem mesurar (dimensions).

Definim doncs cada un d'aquests elements:

Fets

El primer que cal fer es triar els fets, aquest punt no presenta massa dificultat, donat que ens ve indicat pel arxius que hem de tractar que ens indiquen:

- les capacitats per embassament més els camps que necessitem per d'altres càlculs.

Grànul

Com a següent pas hem de decidir el grànul escaient.

Segons els que es dedueix dels informes demanats tenim que el màxim nivell de detall demanat requereix dades per embassament, any i mes.

Aquest representa l'eix principal de la estructura, ara, hem de definir aquells criteris pels quals volem fer l'anàlisi, es a dir definirem les dimensions.

Dimensions

Dels apartats anterior es desprèn que les dimensions han de ser els valor pels quals volem mesurar i analitzar els fets i son:

- Temps (any, mes)
- Riu i Embassament
- Ubicació (Comunitat autònoma, província, Embassament)

Temps

Aquesta dimensió consta bàsicament dels elements

- Any
- Mes

Que son els que es demanen als informes, no obstant en el processos de transformació afegirem atributs que ens puguin ser d'utilitat en l'anàlisi com:

- Nom del mes (Gener, Febrer, etc).
- Ordre (1,2...,12)
- Estació (Hivern, Estiu, etc.)

Construirem una jerarquia

Any -> Mes

Riu i Embassament

Construirem una jerarquia (RiuEmbass):

Riu-> Embassament

Donat que un riu pot tenir varis embassaments i podrem analitzar per riu o anar al nivell de detall Embassament.

Ubicació

Composa la següent jerarquia:

Comunitat Autònoma -> Província -> Embassament

Ens permetrà veure les dades agrupades per Comunitat baixar fins al nivell Embassament.

Mesures

Finalment definim les mesures que interessin del model

- Capacitat Total (Hm3)
- Aigua Embassada (data actual en Hm3)
- % aigua embassada (calculat)
- Entrada (setmana en Hm3)
- Sortida (setmana en Hm3)
- Increment (calculat)
- Aigua Embassada (Any Anterior Hm3)
- % aigua embassada (Calculat)
- Aigua diària
- Aigua de neu anual

El diagrama del model conceptual ens indica com es relacionen aquest elements,quina es la taula de fets que conte les mesures del nostre model i quines dimensions ens permetran analitzar aquests fets.

També ens explica com s'organitzen les dimensions en jerarquies quan hi ha elements que depenen d'altres.

Així podem veure com es configura una estrella en el model dissenyat, tenim doncs una taula central o taula de fets a on tenim les mesures com ja hem comentat i un conjunt de dimensions que l'envolten com si veiéssim una estrella amb les seves puntes.

Les dimensions son:

- Comunitat Autònoma
- Província
- Riu
- Embassament
- Any
- Mes

I les hem agrupat en jerarquies doncs alguns elements tenen relació amb altres de forma que no poden existir sense l'element de la jerarquia superior.

2.3 Diagrama

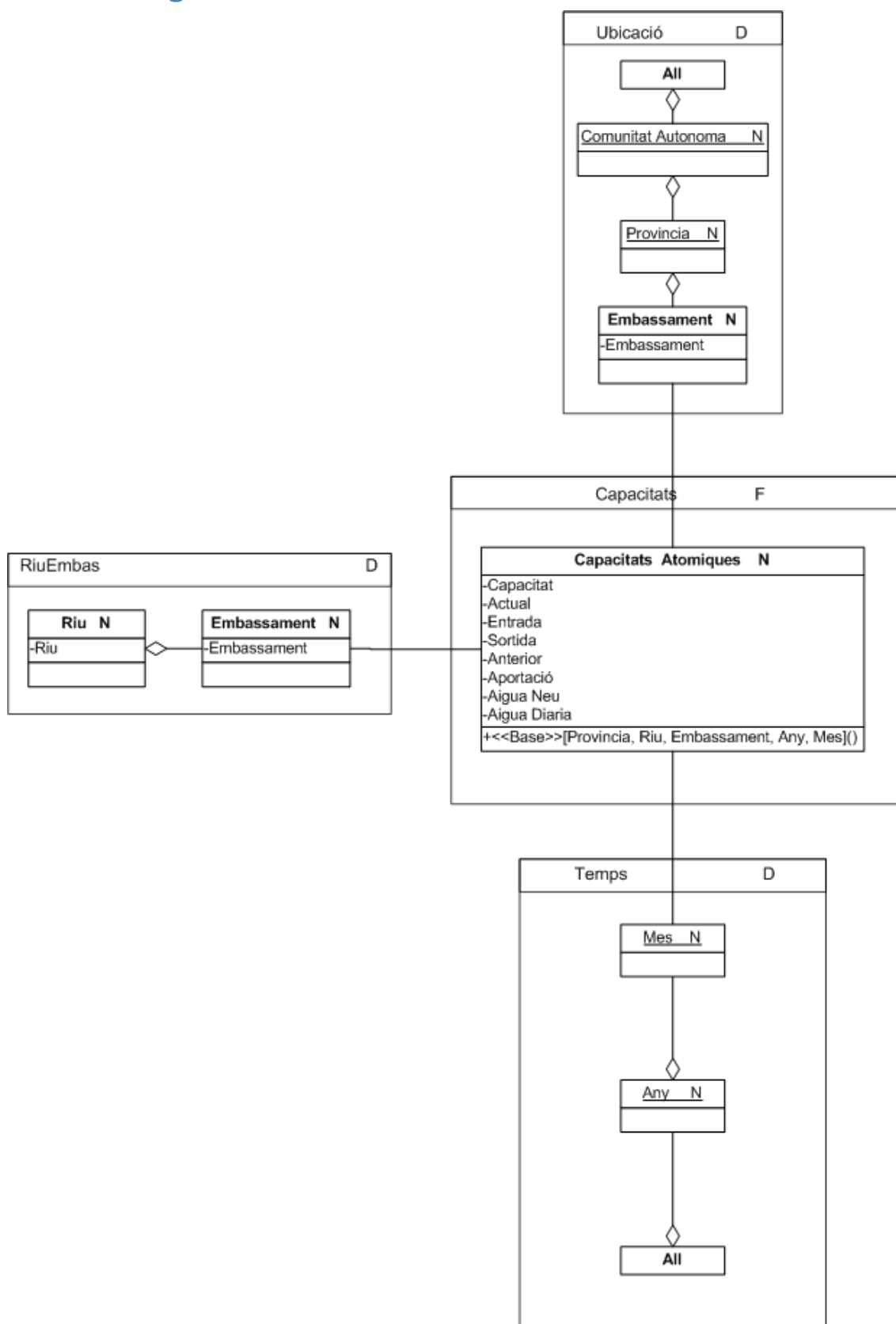


Figura 6. Model Conceptual

3. Disseny

Veiem en aquest capítol els detalls de la implementació física del model proposat.

3.1 Diagrama de l'arquitectura software

En aquest primer punt es detalla els elements de programari que intervenen en el procés quan i com intervenen.

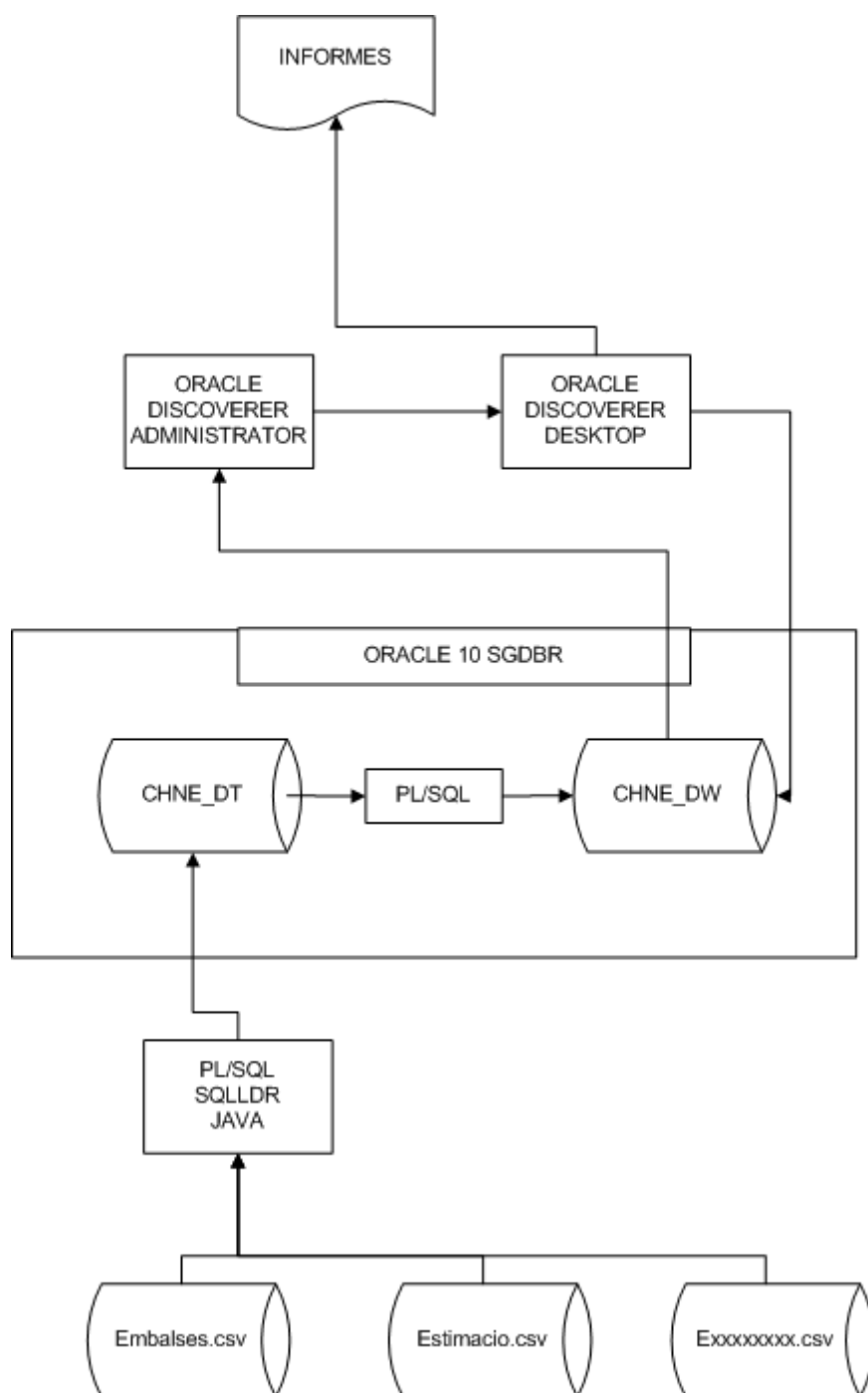


Figura 7. Diagrama d'arquitectura de programari

Inicialment es parteix dels arxius .csv que són els que contenen les dades en estat 'brut'.

Aquests arxius es carreguen mitjançant l'eina Ssqlldr (Sql Loader) que permet inserir dades en format .csv a una BD Oracle.

En el cas dels arxius amb les dades d'embassaments, es dona el cas que són una quantitat d'arxius relativament gran per a ser processats de forma individual, per aquest motiu s'ha desenvolupat un programa escrit en Java

que automatitza la càrrega d'aquests arxius executant par a cada un dels arxius el Sqldr per a la seva càrrega.

En aquesta fase s'omplen taules temporals, un cop tenim les dades carregades procedim a la seva neteja i completeness utilitzant PL/SQL, tot això dins l'entorn de la BD DT (Data)

Finalment en generen les taules amb dades definitives a un altre BD, la DW (Data Warehouse) i el pas d'una BD a un altre es fa amb PL/SQL.

Val a dir que per restriccions del tipus de llicència d'Oracle no s'han pogut crear més BD per tant s'ha simulat el fet de tenir dues BD en només una indicant a cada taula un radical DT o DW per a especificar a quina BD pertany.

Per a descriure el model en forma 'OLAP' o multidimensional utilitzem l'Oracle Discoverer Administrator. Aquesta eina ens permet fer la associació de les taules i camps que volem projectar, la seva relació, jerarquies de dimensions etc.

Finalment amb l'Oracle Discoverer Desktop podem generar els informes. Aquesta eina accedeix al Administrator per a obtenir el model i mitjançant un assistent permet definir el format visual del informe, crear nous camps, condicions, totals, etc.

3.2 Diagrama de l'arquitectura hardware

El diagrama ens explica cada un dels elements del programari comentat al capítol anterior en relació a la seva ubicació física. Es a dir, a quin ordinador o sistema el podem localitzar i la relació entre aquests sistemes.

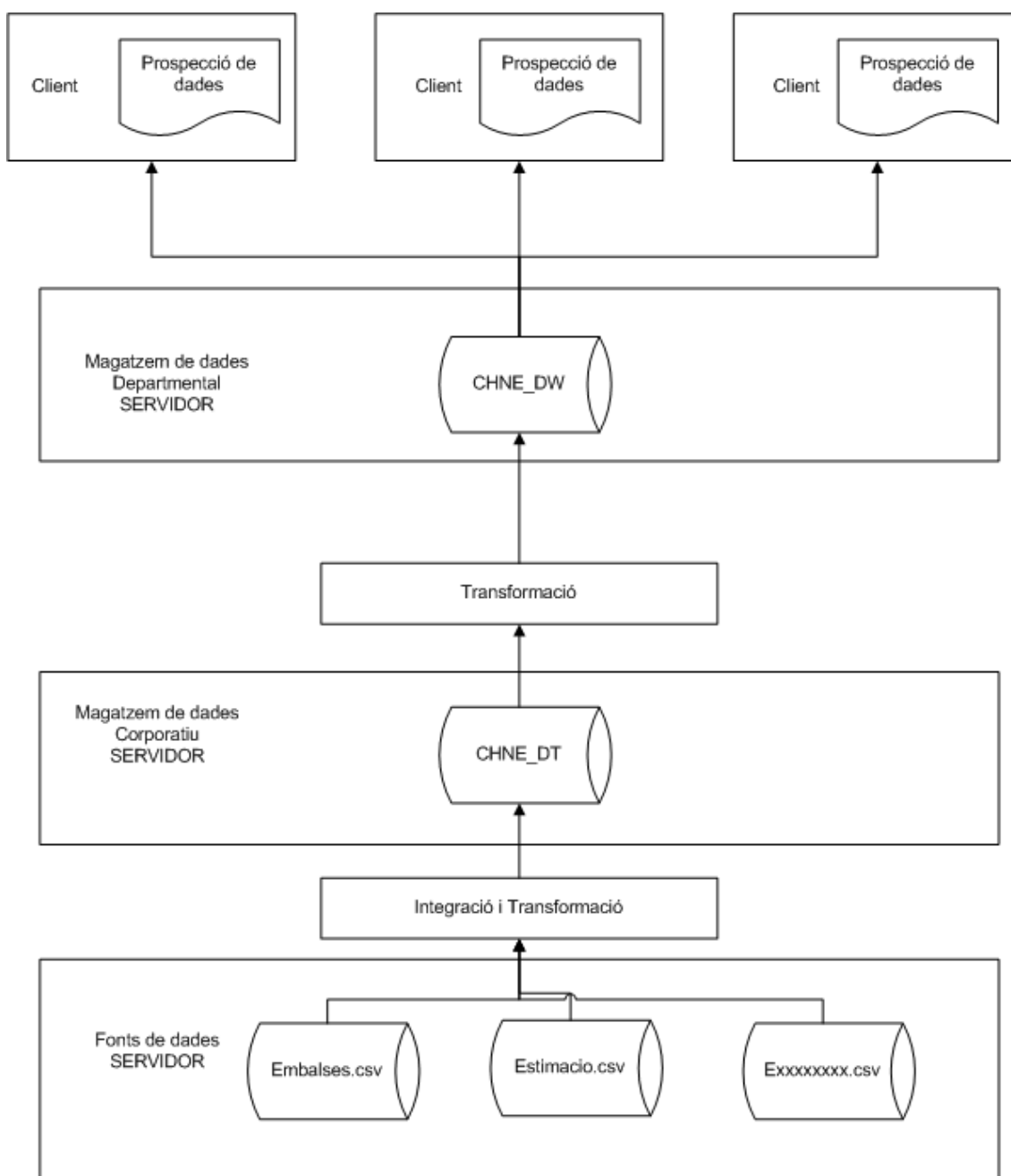


Figura 8. Diagrama d'arquitectura de maquinari

Les dades d'origen es poden ubicar a un sistema servidor, probablement el servidor del ERP corporatiu. En entorns més complexes podria donar-se el cas que les dades origen provinguessin de diferents servidors.

Necessitem doncs un servidor de BD per al magatzem de dades corporatiu, és a dir, a on resideix la BD DT.

Finalment la BD que conte el magatzem de dades departamental pot residir en un altre sistema servidor de BD.

En general els usuaris del sistema han d'utilitzar els seus clients (PC, Mac, etc.) per a consultar o crear informes a partir del data mart amb les eines Oracle Discoverer.

No cal dir que tot el sistema pot residir en el mateix servidor, fins i tot les funcions de client Discoverer, doncs aquest es el muntatge utilitzat en el TFC. No obstant no és la opció més recomanable, per motius de rendiment obvis.

3.3 Disseny de la base de dades i diagrama del model físic

El disseny de la BD ve determinat pel model conceptual ja descrit al capítol 2.

La BD DT (Data) te la següent estructura:

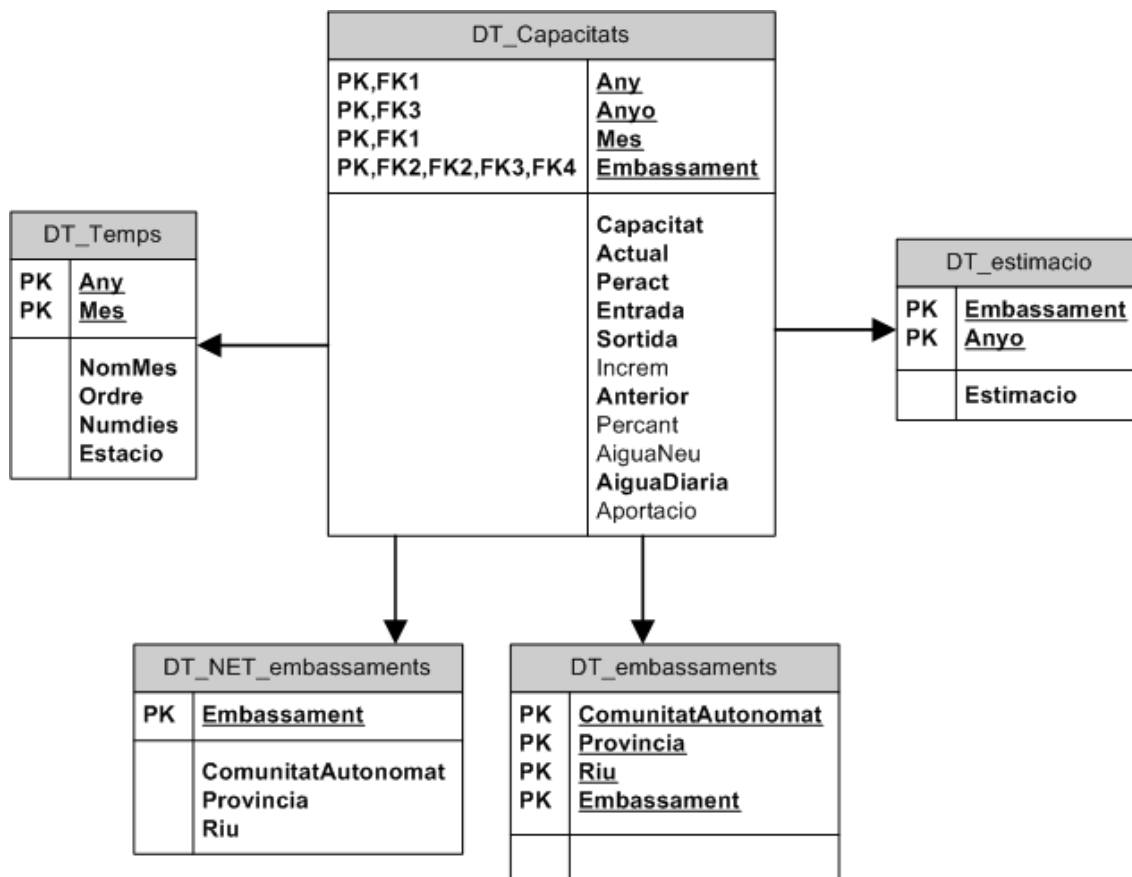


Figura 9. Diagrama físic BD DT

A més s'inclouen a aquesta BD les taules temporals següents:

DT_TMP_CAPACITATS

DT_TMP_Capacitats	
	Embassament
	Riu
	Capacitat
	Actual
	Peract
	Entrada
	Sortida
	Increment
	Anterior
	Percant

Figura 10. Taula DT_NET_CAPACITATS

Aquesta taula s'utilitza com taula per a carregar dades dels arxius .csv.

DT_TMP_EMBASSAMENTS

DT_TMP_embassaments	
	ComunitatAutonomat
	Provincia
	Riu
	Embassament

Figura 11. Taula DT_TMP_EMBASSAMENTS

Aquesta taula conte les dades de l'arxiu embalses.csv

Pel que fa a la BD DW, aquí ja no tenim taules temporals, totes son taules depurades i definitives.

El diagrama és el següent:

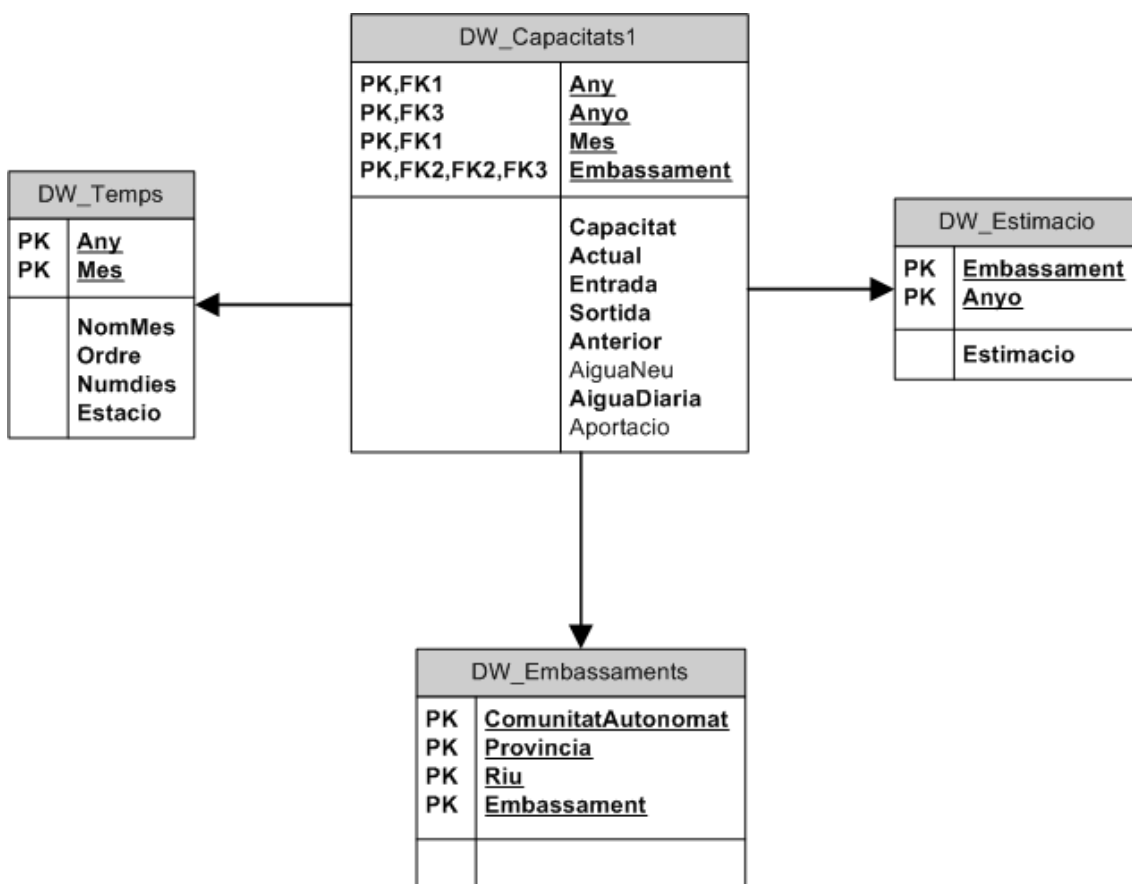


Figura 12. Diagrama BD DW

3.4 Processos ETL

A continuació mostrem la llista dels processos ETL en l'ordre es que aquests s'executen:

- sqlldr system/uoc control=loader0.ctl
- sqlldr system/uoc control=loader1.ctl
- sqlldr system/uoc control=loader2.ctl
- java -jar ExceETL.jar Z:\datos\csv
- sqlplus system/uoc @embassaments.sql
- sqlplus system/uoc @embassaments_clean.sql
- sqlplus system/uoc @capacitats.sql
- sqlplus system/uoc @temps.sql
- sqlplus system/uoc @aiguaneu.sql
- sqlplus system/uoc @complement.sql
- sqlplus system/uoc @anys_anteriors.sql
- sqlplus system/uoc @traspas_capacitats.sql
- sqlplus system/uoc @traspas_embassaments.sql
- sqlplus system/uoc @traspas_temps.sql
- sqlplus system/uoc @traspas_estimacio.sql

sqlldr system/uoc control=loader0.ctl

Aquest procés és de carrega i es descriu a l'arxiu loader0.ctl

Es carrega l'arxiu auxiliar d'embassaments a la taula DT_NET_Embassaments.

Aquest arxiu conte un conjunt de registres d'embassaments i la seva ubicació. Es carrega directament a la taula DT_NET_EMBASSAMENTS. No es l'arxiu proveït pel client i la seva funció es la de afegir noves entrades a mes de les del client així com corregir les entrades que l'arxiu original conté de forma errònia.

sqlldr system/uoc control=loader1.ctl

Aquest procés és de carrega i es descriu a l'arxiu loader1.ctl

Aquest arxiu conte els registres d'embassaments i la seva ubicació originals proveïts pel client. Es carreguen a la taula DT_TMP_EMBASSAMENTS.

Ambdós arxius conformaran la jerarquia de dimensions d'Ubicació (Comunitat autònoma, província, riu i embassament)

sqlldr system/uoc control=loader2.ctl

Aquest procés és de carrega i es descriu a l'arxiu loader2.ctl

Aquest arxiu conte els registres d'estimació d'aigua neu. Es carreguen a la taula DT_ESTIMACIO. Aquesta taula es farà servir per a afegir mesures a la taula de fets en un procés posterior.

java -jar ExcelETL.jar E:\datos\csv

Aquest procés s'ha desenvolupat àdhuc per aquest projecte i consisteix en un procés automàtic de carrega de tots els arxius Excel (ja convertits manualment a format .csv). El requisit d'aquest arxius es que es tingui el format e2*.csv al nom. El directori de on es carreguen els arxius es passa com a paràmetre. El programa bàsicament crida al sqlldr.exe i li passa l'arxiu loader3.ctl i el nom del arxiu a carregar

Aquest arxius contenen les mesures per embassament, any i mes. Es carreguen a la taula DT_TMP_CAPACITATS. És una taula temporal per que aplicarem processos de transformació posteriorment

sqlplus system/uoc @embassaments.sql

Aquest procés es un PL/SQL en el que s'ordena la informació de la taula DT_TMP_EMBASSAMENTS i es passa de forma ordenada a la taula definitiva DT_EMBASSAMENTS.

sqlplus system/uoc @embassaments_clean.sql

Aquest procés es un PL/SQL. El que fa es de la taula que hem omplert abans amb ubicacions d'embassaments DT_NET_EMBASSAMENTS carrega tots el registres a la taula definitiva DT_EMBASSAMENTS.

Si el registre ja existeix l'esborra si no el crea nou. D'aquesta manera es preserva la informació carregada des de l'Excel del client, però si aquesta informació es errònia es pot corregir des de aquesta taula així con afegir noves entrades. Aquest procés es denomina "data cleansing", neteja de dades. Es obvi que la neteja es pot dur a terme des de l'arxiu .csv com directament a la taula.

sqlplus system/uoc @capacitats.sql

Aquest procés es un PL/SQL. L'objectiu d'aquest procés es poder donar a la taula de fets l'any i el mes corresponents a les mesures.

Llegeix la taula temporal DT_TMP_CAPACITATS i genera els camps any i mes a partir de la cadena que figura al Excel, permetent així omplir la PK de la taula e inserir els registres a la taula de fets.

sqlplus system/uoc @temps.sql

Aquest procés es un PL/SQL. Genera la dimensió de temps, materialitzada a la taula DT_TEMPS. Crea un registre per any i mes amb atributs com la estació, el nom del mes i el numero de dies de cada mes.

sqlplus system/uoc @aiguaneu.sql

Aquest procés es un PL/SQL. S'encarrega de actualitzar la taula de fets amb la previsió d'aigua neu segons la formula que ens determina el client. Com que la previsió que es dona és anual, la solució adoptada es enregistrar la mesura calculada al primer registre de cada any i la resta de registres tenen un '0'.

En aquest procés també el calcula la mitjana anual d'aportació del embassament i s'actualitza a cada registre mensual del embassament d'aquell any. Com es una dada anual l'haurem de projectar als informes com AVG(aportacio)

sqlplus system/uoc @complement.sql

Aquest procés es un PL/SQL. La se funció es determinar quins registres de la taula de fets s'han enregistrat sense tenir el seu embassament relacionat a la taula DT_EMBASSAMENTS. Aquest registres es contempen no obstant als informes donat que es crea una entrada 'SENSE DEFINIR' a la taula.

La presencia d'aquests registres amb província, riu o comunitat autònoma 'SENSE DEFINIR' no son desitjables i denoten un valor incorrecte al camp embassament de la taula de fets, que cal corregir, també pot ser un nou embassament. Mentre no es modifiqui la entrada als arxius Excels o les taules de EMBASSAMENTS, el registre es mostrarà con a 'SENSE DEFINIR' a tots els efectes.

sqlplus system/uoc @anys_anteriors.sql

Aquest procés es un PL/SQL. Garanteix que la capacitat actual d'un embassament un any/mes es la mateixa que figura al camp capacitat anterior del mateix embassament però al any següent, mateix mes.

sqlplus system/uoc @traspas_capacitats.sql

Aquest procés es un PL/SQL, que al igual que els processos següents s'encarreguen de traspasar la informació de les taules DT a les taules DW.

Això pretén simular que el magatzem de dades de on farem l'anàlisi no es el mateix de on hem carregat i netejat les dades.

Un raonament que explica aquest fet pot ser el del rendiment o inclús el fet de nomes passar les taules necessàries i/o les files determinades al que representa el repositori departamental.

No es estrictament necessari aquest pas (en aquest treball) donat que la informació de les taules DT definitives ja es correcta.

Aquest procés copia la taula de fets, no passa totes les columnes, donat que algunes son càlculs que es poden obtenir amb la eina Discoverer

sqlplus system/uoc @traspas_embassaments.sql

Aquest procés es un PL/SQL, que copia les dades d'ubicacions d'embassaments

sqlplus system/uoc @traspas_temps.sql

Aquest procés es un PL/SQL, que copia les dades de la dimensió temporal

sqlplus system/uoc @traspas_estimacio.sql

Aquest procés es un PL/SQL, que copia les dades de previsió d'aigua neu.

3.5 Disseny i descripció del informes creats

Es descriu a continuació cada un dels informes elaborats.

Abans de mostrar cada uns dels informes es demana a l'usuari que seleccioni la unitat de mesura en la qual vol que es mostrin les mesures del informe.

Per a aquesta finalitat s'ha definit un paràmetre a cada informe que sol·licita aquesta dada, indicant a l'usuari que ha de seleccionar 'HM3' o 'ML' com a unitats de mesura disponibles. En cas d'indicar qualsevol altre mesura diferent de les possibles l'informe es mostrarà en HM3 que és la unitat bàsica en la que es troben les dades a la BD.

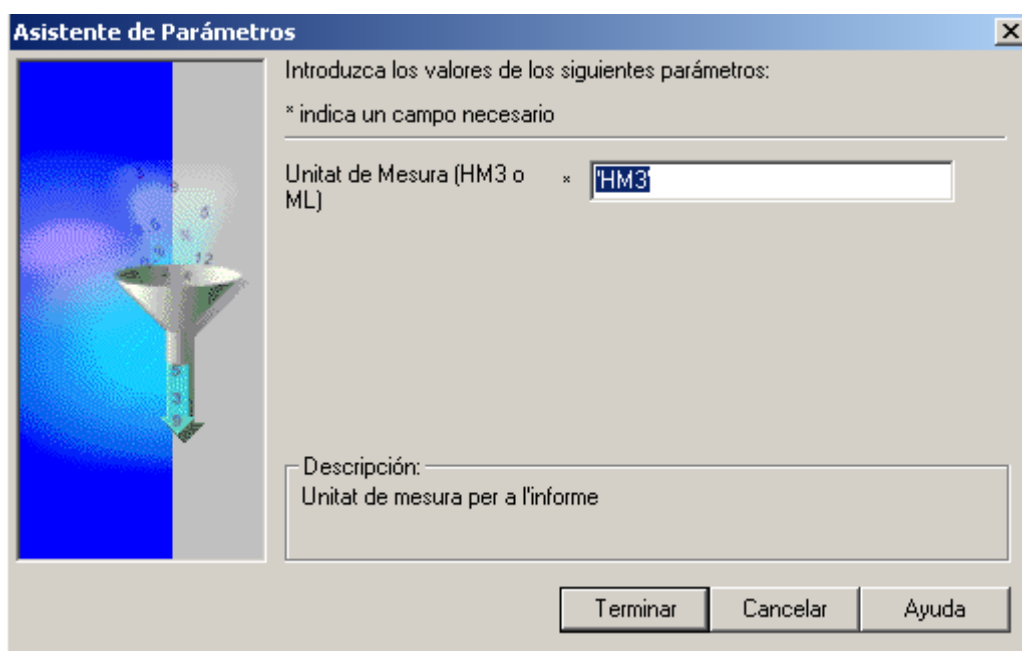


Figura 13. Pantalla de selecció d'unitats de mesura

Aquesta pantalla es mostrarà cada nou informe que es demani o bé cada cop que es refresquin les dades de l'informe.

Capacitat total vs capacitat actual per riu i any

Aquest informe projecte per cada riu la capacitat total i la actual i un camp de percentatge que ens indica el grau de ompliment que tenen els embassaments associats a aquell riu.

Les capacitats per riu totals son la suma de les capacitats totals dels seus embassaments. Igualment en el cas de la capacitat actual.

El camp percentatge es un camp calculat ($\text{Capacitat Actual} / \text{Capacitat total} * 100$)

Podem navegar per l'informe per veure els embassaments individualment, per la jerarquia RiuEmbas.

Es pot filtrar per any/mes.

Capacitat total vs capacitat actual per embassament any

Aquest informe projecte per cada embassament la capacitat total i la actual i un camp de percentatge que ens indica el grau de ompliment que te l'embassament.

Les capacitats totals son la suma de les capacitats totals. Igualment en el cas de la capacitat actual.

El camp percentatge es un camp calculat ($\text{Capacitat Actual} / \text{Capacitat total} * 100$) de cada embassament.

No es pot navegar.

Es pot filtrar per mostrar diferents anys i mesos així como 'tots'.

Capacitat total vs capacitat actual per comunitat autònoma i any

Aquest informe projecte per cada comunitat autònoma la capacitat total i la actual i un camp de percentatge que ens indica el grau de ompliment que te la comunitat.

Les capacitats totals son la suma de les capacitats totals. Igualment en el cas de la capacitat actual.

El camp percentatge es un camp calculat ($\text{Capacitat Actual} / \text{Capacitat total} * 100$) de cada embassament.

Es pot navegar per la jerarquia d'ubicacions per a detallar els valors per Província i finalment per embassament.

Disposem també de filtres que ens mostren només dades corresponents a un any mes.

Comparativa anual de l'evolució per comunitat autònoma de la capacitat agregada de tots els seus embassaments

En aquest informe el que es mostra és la capacitat agregada per Comunitat Autònoma i Any, per a poder visualitzar la evolució es mostren tots els anys disponibles un al costat de l'altre. La qual cosa fa que ràpidament es detecti quina és la tendència al llarg del temps.

La capacitat agregada es calcula com a el promig de la capacitat actual (funció AVG).

No hi ha filtre.

Es permet navegar pels elements de la jerarquia 'Ubicació'.

Es mostra el total per any.

Comparativa de la capacitat mitjana agregada de tots els embassaments d'una comunitat autònoma entre un any i l'any anterior

L'informe ens permet comparar l'estat de la capacitat d'un any respecte a la situació de l'any anterior.

La mesura a comparar és la capacitat actual mitjana agregada, que com a l'informe anterior calculem amb la funció AVG de la Capacitat actual.

La disposició dels elements de l'informe mostra un any, el valor corresponent al període anterior i el % d'increment d'un període respecte l'anterior.

Es permet filtrar per anys.

Es permet navegar per la jerarquia 'Ubicació'.

Aportació (Entrada-Sortida) per riu, embassament i any, comparativa amb la mitjana anual d'aportació de l'embassament

Informe que presenta per any i mes la aportació mensual que fa cada riu. Aquesta aportació es considera com l'aigua que entra a l'embassament menys la que surt. Per tant l'informe mostra la entrada, la sortida i la

diferència. També es demana la comparació amb la mitjana anual d'aportació. Aquesta dada (anual) es calcula en els processos ETL i indica quin és el promig anual d'aportació ($AVG(Entrada-Sortida)$). Al ser un valor anual es projecte a l'informe també com a $AVG(Aportació)$

Es denoten els períodes d'aportació negativa en vermell.

Es pot filtrar per any i mes.

Es permet navegar per la jerarquia '*RiuEmbas*'.

Es totalitzen els camps Entrada, Sortida i Aportació.

Aigua diària en reserva disponible a tota la confederació hidrogràfica per any

Per a aquest informe es necessita l'aigua diària. Aquest valor es calcula als processos ETL i representa el promig diari d'aigua segons la capacitat actual. Consisteix a dividir la capacitat actual entre el numero de dies del mes.

El que es mostra a l'informe es la suma d'aquest valor per Comunitat Autònoma any i mes.

L'informe mostra la evolució per any i mes, tot i que podem veure la informació agregada per any, també podem anar desglossant cada un dels anys en mesos al igual que podem navegar en l'altre eix per províncies i embassaments.

Es permet navegar a l'eix vertical per la jerarquia '*Ubicació*'.

No hi ha filtres, però s'ha implementat navegació a l'eix horitzontal per la jerarquia '*Temps*', podent mostrar d'aquesta manera dades anuals, mensuals o una mescla d'ambdues.

Podem veure el sumatori total de tota la Confederació al peu del informe.

Evolució d'aigua de neu equivalent prevista per any i embassament

Per a mostrar la evolució de l'aigua de neu col·loquem aquest valor anual per a cada un dels anys un al costat del l'altre.

El valor es mostra per embassament i la formula de càlcul es la que ens ha indicat el client (mitjana de la capacitat agregada anual de l'embassament multiplicada per l'estimació de neu de l'any considerat i dividit per 5).

Aquest valor es anual i s'ha enregistrat al primer registre mensual de cada any (la resta de registres tenen un '0'), raó per la qual es projecta com a SUM.

Es filtra per Embassament.

Es demana també mostrar els casos de desbordaments si hi ha. Per a saber si ha desbordament hem de calcular l'aigua total del embassament i entenem que aquesta es l'aigua de neu mes la capacitat actual.

Per aquest motiu s'ha afegit a l'informe l'aigua actual, l'aigua total i la capacitat màxima de l'embassament.

Amb aquestes dades podem calcular el nivell de desbordament i veure si el resultat es negatiu, que indicaria que l'embassament encara no es ple i quina quantitat li falta per al omplir-se.

En canvi si aquest valor es positiu es que l'embassament s'ha desbordat.

L'informe es mostra amb un filtre per embassament.

Si volem saber quins embassaments han tingut l'estat de desbordament ho podem fer mitjançant una condició que eliminarà tots els registres del informe menys els que estan en desbordament.

En qualsevol cas es mostren en color vermell els casos de desbordament.

Períodes de sequera

Es crea un informe per a fer un seguiment exhaustiu de la capacitat actual dels embassaments al llarg del temps.

Si observem els informes d'evolució anual veiem com a tret general que 2001 – 2002 -2003 presenten una tendència creixent, 2004 s'estabilitza i 2005 es decreixent en general.

Així podem veure quina es la evolució general dels estats però no tenim manera de saber en concret quins embassaments estan en un nivell per sota d'un valor determinat. Donat que no seria molt adequat indicar un nivell concret de capacitat ho fem indicant un percentatge.

L'informe doncs te un paràmetre mes a part de la Unitat de Mesura, que és el nivell de sequera que volem veure.

Al mostrar l'informe ens demana el % de capacitat per sota del qual es mostren dades.

Mitjançant el paràmetre i una condició que elimina els registres que no compleixen els valors indicats podem fer un seguiment mensual dels embassaments.

El valor que es mostra per defecte es 10. Això vol dir que veurem aquells embassaments que tenen una capacitat per sota del 10% de la seva

capacitat màxima. Però podem indicar qualsevol valor que sigui coherent per a aconseguir l'objectiu.

Aquest informe pot ser útil per que no només ens mostra aquells embassaments (i no d'altres) que estan a un nivell per sota dels valors indicats sinó que podem veure quina és la seva evolució, és a dir quina és la tendència mensual, cosa que ens serveix de guia per a predir quin pot ser el comportament futur del seu estat.

Els casos més evidents són SAN BARTOLOME, TORT, i especialment SABURO que presenta una capacitat 0 fins a abril de 2004.

A l'informe es confirma la tendència d'increment fins a 2003-2004 i posterior decrement a 2005.

4. Captures de pantalla

S'inclouen les captures de pantalla del informes creats.

Capacitat total vs capacitat actual en Hm3 per riu i any-mes			
Elementos de Página: Año: 2001 ▼ Mes: 07 ▼			
	Capacitat Total	Capacitat Actual	%
▶ AGUAS LIMPIAS	18	16	88,89
▶ AGUAS VIVAS	8	1	12,50
▶ ALBERCOS	33	18	54,55
▶ ALBIÑA	6	4	66,67
▶ ARAGÓN	470	223	47,45
▶ ARANDA	21	14	66,67
▶ ARBA DE LUESIA (DERIVACION)	6	0	0,00
▶ ARGA	21	17	80,95
▶ ASMAT	10	3	30,00
▶ BCO. URDICETO	5	2	40,00
▶ BOLATICA	12	10	83,33
▶ CINCA	829	713	86,01
▶ CIURANA	12	4	33,33
▶ EBRO	2304	1806	78,39
▶ ESCARRA	5	2	40,00
▶ ESERA	92	65	70,65
▶ ESPOT O PEGUERA	6	6	100,00
▶ FLAMISELL	25	6	24,00
▶ FLUMEN	13	2	15,38
▶ GÁLLEGO	96	69	71,88
▶ GUADALOPE	190	137	72,11
▶ GUADALOPE (DERIVACION)	7	6	85,71
▶ GUATIZALEMA	16	12	75,00
▶ HUERVA	7	1	14,29
▶ IBON DE IP	5	5	100,00
▶ IRATI	14	0	0,00
▶ LLADORRE	16	12	75,00
▶ LLAUSET	17	14	82,35
▶ MARTÍN	29	2	6,90
▶ NAJERILLA	68	36	52,94
▶ NAJIMA (DERIVACION)	9	6	66,67
▶ NOGUERA DE TOR	16	13	81,25
▶ NOGUERA PALLARESA	341	235	68,91
▶ NOGUERA RIBAGORZANA	1098	882	80,33

Figura 14. Capacitat total vs capacitat actual per riu i any i mes

Capacitat total vs capacitat actual en Hm3 per embassament any-mes			
Elementos de Página: Anjo: 2001 ▼ Mes: 07 ▼			
	Capacitat Total	Capacitat Actual	%
▶ ALBINA	6	4	66,67
▶ ALCAÑIZ (ESTANCA)	7	6	85,71
▶ ALLOZ	65	37	56,92
▶ BACHIMAÑA (LAGO)	7	6	85,71
▶ BARASONA	92	65	70,65
▶ BASERCA	22	20	90,91
▶ BRAMATUERO ALTO	5	4	80,00
▶ BÚBAL	64	39	60,94
▶ CALANDA	54	36	66,67
▶ CAMARASA	113	64	56,64
▶ CANELLES	688	560	81,40
▶ CASPE	82	63	76,83
▶ CAVALLERS	16	13	81,25
▶ CERTESCÁNS	16	12	75,00
▶ CIURANA	12	4	33,33
▶ CUEVA FORADADA	29	2	6,90
▶ EBRO	540	353	65,37
▶ ESCALES	152	137	90,13
▶ ESCARRA	5	2	40,00
▶ EUGUI	21	17	80,95
▶ GONZALEZ LACASA	33	18	54,55
▶ GRADO (EL)	399	377	94,49
▶ GUIAMETS	10	3	30,00
▶ IP	5	5	100,00
▶ IRABIA	14	0	0,00
▶ LANUZA	17	16	94,12
▶ LLAUSET	17	14	82,35
▶ MAIDEVERA	21	14	66,67
▶ MANSILLA	68	36	52,94
▶ MAR (LAGO)	14	6	42,86

Figura 15. Capacitat total vs capacitat actual per embassament any i mes

Capacitat total vs capacitat actual en Hm3 per comunitat autònoma i any-mes

Elementos de Página: Año: 2001 ▼ Mes: 07 ▼

	Capacitat Total	Aigua Actual	%
▶ ARAGÓN	4273	3307	77,39
▶ CANTABRIA	540	353	65,37
▶ CASTILLA Y LEON	9	6	66,67
▶ CATALUNYA	534	358	67,04
▶ CATALUNYA Y ARAGÓN	232	225	96,98
▶ EUZKADI	225	166	73,78
▶ EUZKADI Y CASTILLA Y LEON	20	20	100,00
▶ NAVARRA	100	54	54,00
▶ NAVARRA Y ARAGÓN	470	223	47,45
▶ RIOJA (LA)	101	54	53,47
	Total Capacitat: 6504	Total Actual: 4766	

Figura 16. Capacitat total vs capacitat actual per comunitat autònoma, any i mes

Comparativa anual de l'evolució per comunitat autònoma de la capacitat agregada de tots els seus

Elementos de Página:

	Volum Anual					
	2001	2002	2003	2004	2005	Total Comunitat
▶ ARAGÓN	94	94	108	106	79	97
▶ CANTABRIA	225	198	281	292	369	278
▶ CASTILLA Y LEON	5	5	4	7	7	6
▶ CATALUNYA	21	29	31	29	26	28
▶ CATALUNYA Y ARAGÓN	112	108	108	108	108	108
▶ EUZKADI	46	47	52	55	55	52
▶ EUZKADI Y CASTILLA Y LEON	18	18	19	19	18	18
▶ NAVARRA	10	14	22	21	22	18
▶ NAVARRA Y ARAGÓN	106	168	314	291	195	227
▶ RIOJA (LA)	14	24	35	36	30	29
Total Anual	69	72	85	84	67	76

Figura 17. Comparativa anual de l'evolució per comunitat autònoma de la capacitat agregada de tots els seus embassaments

Comparativa de la capacitat mitjana agregada de tots els embassaments d'una comunitat autònoma entre un any i l'any anterior (valors i % d'increment any anterior/any)			
Elementos de Página: Año: 2001 ▼			
	Mitjana Anterior	Mitjana Actual	% Increment
▶ ARAGÓN	90,58	94,10	3,88
▶ CANTABRIA	308,50	224,67	-27,17
▶ CASTILLA Y LEON	6,17	5,33	-13,51
▶ CATALUNYA	25,14	21,32	-15,19
▶ CATALUNYA Y ARAGÓN	108,67	111,67	2,76
▶ EUZKADI	53,22	45,83	-13,88
▶ EUZKADI Y CASTILLA Y LEON	18,17	18,17	0,00
▶ NAVARRA	11,94	10,11	-15,35
▶ NAVARRA Y ARAGÓN	201,33	106,17	-47,27
▶ RIOJA (LA)	22,58	14,33	-36,53

Figura 18. Comparativa de la capacitat mitjana agregada de tots els embassaments d'una comunitat autònoma entre un any i l'any anterior

Aportació (Entrada-Sortida) en Hm3 per riu, embassament i any-mes, comparativa amb la mitjana anual d'aportació de l'embassament.				
Elementos de Página: Año: 2001 ▼ Mes: 07 ▼				
	Entrada	Sortida	Aportació	Mitjana Anual Aportacio
▶ AGUAS LIMPIAS	0,00	0,00	0,00	0,0000
▶ AGUAS VIVAS	0,00	0,00	0,00	0,0000
▶ ALBERCOS	0,00	2,20	-2,20	-0,8500
▶ ALBIÑA	0,00	0,00	0,00	0,0000
▶ ARAGÓN	5,50	31,80	-26,30	-6,3167
▶ ARANDA	0,00	1,10	-1,10	-0,3333
▶ ARBA DE LUESIA (DERIVACION)	0,00	0,00	0,00	0,0000
▶ ARGÁ	0,30	0,80	-0,50	-0,3667
▶ ASMAT	0,00	0,00	0,00	0,0167
▶ BCO. URDICETO	0,00	0,00	0,00	0,0000
▶ BOLATICA	0,00	0,00	0,00	0,0000
▶ CINCA	52,00	77,10	-25,10	-1,2750
▶ CIURANA	0,00	0,00	0,00	0,0333
▶ EBRO	115,80	147,50	-31,70	-2,2250
▶ ESCARRA	0,00	0,00	0,00	0,0000
▶ ESERA	9,60	18,70	-9,10	-0,8167
▶ ESPOT O PEGUERA	0,00	0,00	0,00	0,0000
▶ FLAMISELL	0,00	0,00	0,00	0,0000
▶ FLUMEN	0,00	0,00	0,00	0,0000
▶ GÁLLEGO	17,50	21,60	-4,10	-0,0167
▶ GUADALOPE	2,90	6,80	-3,90	-0,4667
▶ GUADALOPE (DERIVACION)	1,00	1,00	0,00	0,0667
▶ GUATIZALEMA	0,30	0,50	-0,20	-0,2167
▶ HUERVA	0,00	0,10	-0,10	-0,0667
▶ IBON DE IP	0,00	0,00	0,00	0,0000
▶ IRATI	0,00	0,00	0,00	0,0000
▶ LLADORRE	0,00	0,00	0,00	0,0000
▶ LLAUSET	0,00	0,00	0,00	0,0000
▶ MARTÍN	0,20	0,20	0,00	0,0833
▶ NAJERILLA	1,60	4,70	-3,10	-1,1167
▶ NAJIMA (DERIVACION)	0,00	0,10	-0,10	-0,0167
▶ NOGUERA DE TOR	1,50	0,40	1,10	0,1000
▶ NOGUERA PALLARESA	22,60	49,30	-26,70	-3,0111
▶ NOGUERA RIBAGORZANA	48,10	52,50	-4,40	-0,7750

Figura 19. Aportació (Entrada-Sortida) per riu, embassament i any, comparativa amb la mitjana anual d'aportació de l'embassament

Aigua diària en reserva disponible (Hm3/dia) a tota la confederació hidrogràfica per any-mes.

Elementos de Página:

	Aigua Diària									
	2001						2002	2003	2004	2005
	07	08	09	10	11	12				
ARAGÓN	106,68	94,52	87,70	90,19	98,30	93,19	1153,07	1328,50	1302,32	965,24
CANTABRIA	11,39	8,32	6,23	5,55	6,17	6,23	78,42	110,95	115,30	145,63
CASTILLA Y LEON	0,19	0,19	0,17	0,16	0,17	0,16	1,91	1,74	2,60	2,73
CATALUNYA	11,55	8,61	7,53	7,03	7,83	7,45	137,30	144,73	136,36	122,62
CATALUNYA Y ARAGÓN	7,26	7,19	7,50	7,26	7,33	7,16	85,03	85,34	85,12	84,90
EUZKADI	5,35	4,90	4,60	3,84	4,27	3,94	55,51	61,69	65,37	65,02
EUZKADI Y CASTILLA Y LEON	0,65	0,61	0,60	0,52	0,63	0,55	7,17	7,34	7,31	7,11
NAVARRA	1,74	1,35	0,93	0,52	0,80	0,58	16,35	25,65	24,62	25,59
NAVARRA Y ARAGÓN	7,19	3,84	1,67	2,45	3,17	2,39	66,28	124,14	115,40	77,11
RIOJA (LA)	1,74	1,23	0,80	0,65	0,60	0,58	18,60	27,29	28,38	23,55
Total Confederació	153,74	130,77	117,73	118,16	129,27	122,23	1619,65	1917,39	1882,77	1519,51

Figura 20. Aigua diària en reserva disponible a tota la confederació hidrogràfica per any i mes

Aigua de neu equivalent prevista per any i embassament

Elementos de Página: Embassament: ALBINA

	2001	2002	2003	2004	2005
Aigua de Neu (Estimació)	0,70	0,80	0,75	0,80	0,72
Aigua Actual	21,00	48,00	45,00	48,00	43,00
Aigua Total (Aigua Neu + Actual)	21,70	48,80	45,75	48,80	43,72
Capacitat Màxima	36,00	72,00	72,00	72,00	72,00
Nivell Desbordament	-14,30	-23,20	-26,25	-23,20	-28,28

Figura 21. Evolució d'aigua de neu equivalent prevista per any i embassament

Aigua de neu equivalent prevista per any i embassament

Elementos de Página: Embassament: NEGRO

	2001
Aigua de Neu (Estimació)	1,20
Aigua Actual	36,00
Aigua Total (Aigua Neu + Actual)	37,20
Capacitat Màxima	36,00
Nivell Desbordament	1,20

Figura 22. Casos de desbordament

Construcció i explotació d'un magatzem de dades de planificació hidrològica

Períodes de sequera																							
Elementos de Página:																							
	2001						2002						2003										
	> 07	> 08	> 09	> 10	> 11	> 12	> 01	> 02	> 03	> 04	> 05	> 06	> 07	> 08	> 09	> 10	> 11	> 12	> 01	> 02	> 03	> 04	>
ALLOZ				4,00											5,00	3,00							
BACHIMAÑA (LAGO)						0,00		0,00	0,00											0,00	0,00	0,00	
BRAMATUERO ALTO								0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00				0,00	0,00	0,00	
BUBAL																							
CAVALLERS								1,00															
CUEVA FORADADA	2,00	2,00	2,00	2,00	2,00					2,00			2,00										
ESCARRA																					0,00		
IRABIA	0,00	0,00	0,00	1,00		1,00																	
MEDIANO																							
MONEVA			0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00													
RESPOMUSO								1,00	0,00	1,00													
SABURO (LAGO)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
SAN BARTOLOMÉ	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
S.MARÍA DE BELSÚE		1,00	0,00	0,00	1,00	0,00	0,00	0,00	1,00	1,00	1,00	1,00	0,00	0,00	0,00	1,00	0,00						
TORT (LAGO)														0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
VADIELLO																							

Figura 23. Períodes de sequera

5. Conclusions

Un cop desenvolupat aquest TFC trobo que ha estat una experiència molt interessant donat que m'ha permès aprendre molts detalls importants en la implementació de tot allò relacionat amb els magatzems de dades:

- La importància de la correcta aplicació pràctica dels conceptes apresos a la teoria, i estic segur que hi ha molt més per aprendre.
- Com hem de gestionar les dades a fi de que es puguin mostrar com l'usuari les necessita.
- Com hem de preveure i minimitzar la falta de correctesa de les dades en el seu origen.
- Els coneixements en l'àrea de les Bases de Dades son fonamentals, en especial l' Sql.

Resulta espectacular veure com quan tot el muntatge està complert i sincronitzat quan fàcil es obtenir 'informació'.

Penso que es el tema dels magatzems de dades en una eina que s'hauria de explotar molt mes en els sistemes informàtics de les empreses actuals, i que malgrat està avui en dia 'de moda' encara hi ha un llarg camí per recórrer.

En definitiva aquest TFC m'ha despertat el interès per aprendre molt mes en aquest àrea i fins i tot en d'altres relacionades com la mineria de dades.

6. Línies d'evolució a futur

Qualsevol projecte sempre es pot millorar, aquest no és una excepció.

Els punts de millora que es proposen, sense fer ampliacions de la base existent son:

- Millorar el sistema de extracció de les dades, de forma més automatitzada, l'ús de arxius d'intercanvi de dades no es molt recomanable. Seria ideal poder tenir les dades a un sistema de BD.
- Un aspecte important a millorar és la codificació de les dades, que és pràcticament inexistent i provoca molts problemes, es proposa codificar els següents elements:
 - Embassaments
 - Rius
 - Comunitats autònomes
 - Províncies
- Una versió definitiva contemplaria afegir dades de forma incremental. Ara es processa tota la informació de cop cada vegada, no hi ha masses dades actualment, però conforme el nivell de dades augmenti podria provocar un problema de rendiment.
- Alguns càlculs al model actual només tenen sentit per a resoldre els requisits actuals. Probablement si ens demanessin fer una comparativa de la capacitat dels últims 3 anys no podríem fer-la sense modificar la estructura del model. Caldria estudiar una nova estructura del model.
- Fer us del llenguatge 'MDX' per a construir informes es una manera d'obtenir més potencia en les consultes, per exemple , per a obtenir comparatives entre períodes sense modificar la estructura de la BD només cal utilitzar la funció 'MDX' *ParallelPeriod*.

7. Glossari

BBDD. Una base de dades o banc de dades (en ocasions abreujada amb la sigles BD o amb l'abreviatura bd) és un conjunt de dades pertanyents a un mateix context i emmagatzemats sistemàticament per al seu posterior ús.

BI. Business Intelligence (BI) es refereix a les tècniques informàtiques utilitzades en la detecció, la cerca i l'anàlisi de les dades empresarials, com ara els ingressos per vendes de productes i / o departaments, o pels costos associats i els ingressos.

CSV. Els arxius CSV (comma separated values) són un tipus de document en format obert senzill per a representar dades en forma de taula, en les que les columnes es separen per comes (o punt i coma a on la coma es el separador decimal) y les files per salts de línia.

Dashboard. Pantalla que fa la funció de quadre de comandament, agrupant diversos informes, generalment en forma de control gràfic, relacionats amb una activitat de negoci, per donar una idea ràpida de la situació del sistema.

Data warehouse. un magatzem de dades és un conjunt de dades orientada a un determinat àmbit (empresa, organització, etc.), integrat, no volàtil i variable en el temps, que ajuda a la presa de decisions en l'entitat en la que s'utilitza.

Data mart. Un Data mart és una versió especial de magatzem de dades (data warehouse). Són subconjunts de dades amb el propòsit d'ajudar a que una àrea específica dins del negoci pugui prendre millors decisions.

ERP. Enterprise Resource Planning, programari d'informació centralitzada orientat a registrar i integrar la majoria dels processos de negocis.

ETL. Extract, transform and load (extracció, transformació i càrrega), en administració de sistemes d'informació i magatzems de dades.

MDX. Les expressions multidimensionals (MDX és l'acrònim de Multidimensional expressions) és un llenguatge de consulta per a bases de dades multidimensionals sobre cubs OLAP, s'utilitza en Business Intelligence per a generar informes per a la presa de decisions basats en dades històriques, amb la possibilitat de canviar la estructura, o permetent rotar el cub.

OLAP. OLAP és l'acrònim en anglès de processament analític en línia (On-line Analytical Processing). És una solució utilitzada en el camp de l'anomenada Intel·ligència empresarial (o Business Intelligence) l'objectiu és agilitzar la consulta de grans quantitats de dades. Per això utilitza

estructures multidimensionals (o Cubs OLAP) que contenen dades resumides de grans bases de dades o Sistemes Transaccionals (OLTP). S'usa en informes de negocis de vendes, màrqueting, informes de direcció, mineria de dades i àrees similars.

OLTP. Son les sigles en anglès de Processament de Transaccions En Línia (Online Transaction Processing) és un tipus de sistemes que faciliten i administren aplicacions transaccionals, usualment per entrada de dades i recuperació i processament de transaccions.

PL/SQL. Llenguatge de programació incrustat en Oracle i PostgreSQL. PL / SQL suporta totes les consultes, ja que la manipulació de dades que es fa servir és la mateixa que en SQL.

SGBD. Els sistemes de gestió de bases de dades (en anglès database management system, abreujat DBMS) són un tipus de programari molt específic, dedicat a servir d'interfície entre la base de dades, l'usuari i les aplicacions que la utilitzen.

SQL. El llenguatge de consulta estructurat o SQL (per les seves sigles en anglès, structured query language) és un llenguatge declaratiu d'accés a bases de dades relacionals que permet especificar diversos tipus d'operacions en aquestes.

SQL*DR. SQL * Loader és una utilitat de carrega usat per moure dades d'arxius externs en la base de dades Oracle.

8. Bibliografia i referències

Ralph Kimball and Margy Ross. *The Data Warehouse Toolkit*. (2nd Edition). Wiley Computer Publishing 2002.

W. H. Inmon. *Building the Data Warehouse* (4th Edition). Wiley Publishing, Inc. 2005

Joe Caserta and Ralph Kimball. *The Data Warehouse ETL Toolkit*. Wiley Publishing, Inc. 2004

Oracle Corporation. *Oracle® Discoverer Administrator Administration Guide*. 2003.

Oracle Corporation. *Oracle9iAS Discoverer Plus Tutorial Version 9.0.2*. January 2002

Oracle Corporation. *Oracle® Business Intelligence Discoverer Desktop User's Guide 10g Release 2 (10.1.2.1) for Windows*. July 2005

Oracle Corporation. *SQL*Plus® Quick Reference Release 10.1*. December 2003

Oracle Corporation. *SQL*Plus® User's Guide and Reference Release 10.1*. December 2003

Inmon Consulting Services. Corporate Information Factory
[<http://www.inmoncif.com>]

Kimball Group
[<http://www.rkimball.com>]

Data Warehousing Review
[<http://www.dwreview.com>]

Data Warehousing site
[<http://www.1keydata.com/datawarehousing/datawarehouse.html>]

Intranet Journal
[<http://www.intranetjournal.com/features/datawarehousing.html>]

Confederación Hidrográfica del Ebro
[<http://www.chebro.es/>]

XPDIAN.COM. Modelling the Data warehouse using the UML
[<http://www.xpdian.com/ModellingtheDataWarehouseusingtheUML.html>]