



MEMÒRIA DEL TFC: MAGATZEMS DE DADES

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

Óscar Sánchez Guiu

Estudis: Enginyeria Tècnica d'Informàtica de Sistemes

Consultor: José Ángel Martín Carballo

10/01/2011

Resum

La Confederació Hidrogràfica Nord-Est, que a partir d'aquest moment anomenarem CHNE, s'encarrega de realitzar el seguiment de l'estat dels diferents embassaments de la seva responsabilitat, actualment aquesta feina es realitza a partir d'un tractament molt manual de les dades, i això suposa una considerable despesa de recursos. Per aquest motiu la CHNE ha demanat automatitzar la recollida i explotació d'aquestes dades.

L'objectiu d'aquest projecte és la creació d'un magatzem de dades que automatitzi la recollida de dades de l'estat dels embassaments proporcionades per la CHNE, i que permeti explotar aquestes dades per tal de poder realitzar les consultes demanades pel client mitjançant eines de Business Intelligence.

Per assolir aquest objectiu s'ha creat una base de dades amb Oracle 10g Express amb l'estructura d'un magatzem de dades, i s'han carregat les dades proporcionades mitjançant els processos d'extracció, transformació i càrrega dissenyats amb PL/SQL. Una vegada carregades les dades, s'han creat mitjançant Oracle Discoverer els diferents informes i gràfics per poder realitzar les consultes demanades pel client.

Com a resultat s'ha obtingut el magatzem de dades i els processos dissenyats per poder carregar i processar automàticament les dades proporcionades, així com els diferents informes per poder explotar aquestes dades.

Els productes obtinguts permetran a la CHNE tenir unificades les dades dels embassaments i poder afegir automàticament les noves dades obtingudes, havent de destinar força menys recursos a aquestes tasques.

D'altra banda també permetrà realitzar l'anàlisi d'aquestes dades mitjançant els informes creats amb l'eina de Business Intelligence, podent obtenir informació per riu, embassament i comunitat autònoma, comparatives de diferents períodes de temps, i fins i tot predir períodes de sequera analitzant la tendència dels resultats en un període de temps determinat.

Paraules clau

Magatzem de dades, Data Warehouse, model multidimensional, planificació hidrològica, Oracle Express, Oracle Discoverer, ETL, PL/SQL.

Àrea del TFC

Magatzems de dades

Índex

	Pag
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 del TFC.....	5
1.3 Enfocament o mètode seguit.....	6
1.4 Planificació del projecte	6
1.4.1 Tasques i Fites	6
1.4.2 Planificació	9
1.4.3 Incidències i riscos.....	12
1.4.4 Material necessari	13
1.5 Productes obtinguts	13
1.6 Altres capítols de la memòria	14
 2. Anàlisi	15
2.1 Model conceptual	15
2.2 Triar el fet.....	15
2.3 Trobar el grànul escaient	16
2.4 Escollir les dimensions que s'utilitzaran en l'anàlisi.....	16
2.5 Trobar els atributs de cada dimensió	17
2.6 Distingir entre descriptors i jerarquies d'agregació.....	18
2.7 Decidir quines són les mesures que interessin.....	18
2.8 Definir Cel·les	18
2.9 Explicitar les restriccions d'integritat.....	19
2.10 Estudiar la viabilitat.....	19
2.11 Diagrama del model conceptual	20
2.12 Diagrama de casos d'us	21
 3. Disseny	23
3.1 Disseny Lògic	23
3.2 Esquema relacional.....	24
3.3 Disseny Físic	25
3.4 Arquitectura del Sistema	27
3.4.1 Arquitectura de Software	27
3.4.2 Arquitectura de Hardware.....	28
3.5 Disseny dels Procediments ETL.....	28
3.5.1 Extracció	28
3.5.2 Transformació i càrrega.....	29
3.6 Disseny i descripció del informes creats	32
3.6.1 Creació de l'àrea de negoci	32
3.6.2 Informes creats.....	34
3.7 Accés al magatzem de dades	45
3.7.1 Accés a la base de dades	45
3.7.2 Accés a Oracle Discoverer.....	45
3.8 Processos de càrrega.....	46

4. Conclusions	47
5. Línies de evolució futures	48
6. Glossari.....	49
7. Bibliografia.....	50
7.1 Publicacions.....	50
7.2 Webs.....	50

Índex de figures

Figura 1: Diagrama de Gantt part 1.....	11
Figura 2: Diagrama de Gantt part 2.....	11
Figura 3: Diagrama de Gantt part 4.....	12
Figura 4: Diagrama de Gantt part 4.....	12
Figura 5: Diagrama del model conceptual	20
Figura 6: Cas d'us 1	21
Figura 7: Cas d'us 2	22
Figura 8: Esquema relacional.....	24
Figura 9: Diagrama del model físic.....	25
Figura 10: Diagrama de l'arquitectura de software.....	27
Figura 11: Diagrama de l'arquitectura de hardware.....	28
Figura 12: Disseny de les taules temporals	30
Figura 13: Estructura àrea de negoci.....	33
Figura 14: Informe 1	35
Figura 15: Informe 2 Part 1	36
Figura 16: Informe 2 Part 2	37
Figura 17: Informe 3	37
Figura 18: Informe 4	38
Figura 19: Informe 5	39
Figura 20: Informe 6	40
Figura 21: Informe 7	41
Figura 22: Informe 8 en Hm3.....	42
Figura 23: Informe 8 en MI.....	43
Figura 24: Informe 9 en MI.....	44

1. Introducció

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

En la constant evolució de les bases de dades estan prenent gran importància els magatzems de dades. Aquests Magatzems i les eines que permeten explotar-los, fan possible l'explotació en temps real del gran volum de dades emmagatzemades, això facilita l'anàlisi de la informació i la seva utilització en la presa de decisions. La funcionalitat que permet aquests tipus de bases de dades fan que cada vegada més organitzacions decideixin implantar aquests sistemes per tal de disposar d'una eina que faciliti la presa de decisions.

La Confederació Hidrogràfica Nord-Est s'encarrega de realitzar el seguiment de l'estat dels diferents embassaments de la seva responsabilitat, actualment aquesta feina es realitza a partir d'un tractament molt manual de les dades, i això suposa una considerable despesa de recursos.

El projecte consisteix en la creació i explotació d'un magatzem de dades de planificació hidrològica. Aquest magatzem de dades permetrà realitzar un seguiment de l'estat dels diferents recursos hidrològics de forma més eficient i facilitar la presa de decisions. Això s'aconseguirà automatitzant la recollida de la informació i explotant aquesta informació amb eines de Business Intelligence adients.

1.2 Objectius del TFC

L'objectiu principal del TFC es donar una solució al problema que planteja l'enunciat proporcionat. Aquesta solució passa per crear un magatzem de dades que automatitzi la recollida de la informació provinent dels fulls Excel proporcionats pel client, i que permeti explotar aquesta informació per tal de poder realitzar les consultes demanades.

Per poder assolir aquest objectiu principal cal primer assolir els següents objectius, que permetran donar solució al problema plantejat:

- Familiaritzar-se amb l'estructura general d'un magatzem de dades i conèixer els aspectes teòrics del seu disseny.
- Adquirir experiència en la construcció i explotació d'un magatzem de dades a partir de la informació disponible en una base de dades transaccional.
- Aprofundir en les tècniques de tractament de dades i la seva integració en el disseny del magatzem de dades.
- Conèixer el llenguatge PL/SQL d'Oracle per poder tractar les dades del magatzem de dades.

- Dominar la càrrega de dades a una base de dades mitjançant SQLLOADER.
- Conèixer l'eina de Business Intelligence Oracle Discoverer, per poder crear els informes sol·licitats i presentar les dades de manera adient.

1.3 Enfocament o mètode seguit

Per poder assolir els diferents objectius dels que es compon el projecte, s'ha dividit en diferents fases de desenvolupament.

En primer lloc es realitzarà una anàlisi de requeriments on es fa una anàlisi exhaustiva de l'enunciat del TFC així com de les fonts de dades proporcionades, i s'enumeren i es descriuen els diferents elements identificats. Amb aquestes dades es crea el model conceptual i el diagrama de casos d'us.

Posteriorment es fa el disseny tècnic del magatzem de dades, aquest inclou el disseny lògic i físic, així com la descripció de l'arquitectura del sistema i el disseny dels processos ETL (extracció transformació i càrrega).

Un cop dissenyat el sistema es realitza la implementació de la base de dades. En aquesta fase es creen tots els elements de la base de dades, com són les taules, els procediments emmagatzemats, disparadors... També en aquesta fase es creen els processos d'automatització de la càrrega i transformació de les dades.

Finalment, quan les dades ja estan carregades a la base de dades es configura l'eina d'exploració de dades i es creen els diferents informes i gràfics per satisfer les consultes demanades pel client.

1.4 Planificació del projecte

La planificació del projecte recull les diferents tasques de les que es compon i la planificació estimada per a cada una d'aquestes tasques.

També es descriuen els recursos necessaris, els riscos, les possibles incidències i els plans de contingències.

1.4.1 Tasques i Fites

En aquest apartat es defineixen les diferents tasques de les que es compon el projecte.

En base als objectius marcats a l'enunciat per cada entrega i a les dates concretes d'aquestes entregues s'han definit les següents tasques i fites:

1. Definició del TFC

1.1 Lectura i anàlisi de l'enunciat del TFC. Es fa una lectura acurada de l'enunciat i s'analitza el problema que planteja.

1.2 Recopilació i lectura de la documentació. Es recopila els apunts de l'assignatura així com tota la documentació addicional associada al TFC, i es fa una primera lectura amb l'objectiu de endinsar-se en les seves característiques.

1.3 Definició dels objectius i les tasques del projecte. Un cop llegida la documentació i en base al que demana l'enunciat, es defineixen els objectius i s'estimen les tasques del projecte.

2. Instal·lació del Software necessari

2.1 Descarrega del software necessari. Es descarrega el software Virtualbox i els discs virtual proporcionats per l'assignatura, així com tota la documentació associada.

2.2 Instal·lació i prova del software. S'instal·la el software descarregat seguint les instruccions proporcionades i es prova el bon funcionament de l'entorn.

3. Trobada presencial del TFC. S'assisteix a la trobada presencial del TFC amb l'objectiu de aclarir possibles dubtes amb el consultor.

4. Realització de Pla de treball (PAC1)

4.1 Recerca de informació addicional i bibliografia recomanada. Com a complement a la documentació entregada, es fa una recerca d'informació i de la bibliografia recomanada mitjançant Internet, això permetrà ampliar els coneixements i tenir més recursos de consulta a l'hora de realitzar el TFC.

4.2 Redacció de l'esborrany del pla de treball. Amb la informació analitzada referent al treball, es redacta l'esborrany del pla de treball, definint les diferents tasques del TFC i fent una planificació estimada d'aquestes.

4.3 Redacció de l'esborrany del document d'anàlisi preliminar. Després d'haver fet l'anàlisi exhaustiva de l'enunciat del TFC així com les fonts de dades proporcionades, es fa una anàlisi preliminar amb l'enumeració i breu descripció dels elements identificats, amb aquesta anàlisi es redacta l'esborrany del document d'anàlisi preliminar.

4.4 Entrega de l'esborrany del pla de treball i del document d'anàlisi preliminar. S'entrega l'esborrany del pla de treball i del document d'anàlisi preliminar al consultor uns dies abans de l'entrega definitiva perquè el pugui revisar i proposar possibles correccions.

4.5 Correcció de l'esborrany del pla de treball i del document d'anàlisi preliminar i entrega final. Amb les possibles correccions indicades pel

consultor, es rectifica el pla de treball i del document d'anàlisi preliminar i es presenta la versió definitiva a la bústia de l'aula.

5. Anàlisi de requeriments i disseny tècnic (PAC2)

5.1 Lectura exhaustiva de la documentació. Es fa una lectura exhaustiva de la documentació recopilada del projecte referent als requeriments i el disseny tècnic.

5.2 Redacció de l'esborrany de l'anàlisi de requeriments. En base a la lectura de la documentació i de l'anàlisi preliminar es fa la redacció de l'esborrany de l'anàlisi de requeriments.

5.3 Redacció de l'esborrany del disseny tècnic. En base a la lectura de la documentació es fa la redacció de l'esborrany del disseny tècnic.

5.4 Entrega de l'esborrany de l'anàlisi de requeriments i del disseny tècnic. S'entrega l'esborrany de l'anàlisi de requeriments i del disseny tècnic al consultor uns dies abans de l'entrega definitiva perquè el pugui revisar i proposar possibles correccions.

5.5 Correcció de l'esborrany de l'anàlisi de requeriments i del disseny tècnic i entrega final. Amb les possibles correccions indicades pel consultor, es rectifica l'anàlisi de requeriments i el disseny tècnic i es presenta la versió definitiva a la bústia de l'aula.

6. Implementació.

6.1 Construcció de la base de dades. En base a l'anàlisi de requeriments i del disseny tècnic es crea la base de dades del projecte.

6.2 Carrega de dades. Es carreguen les dades proporcionades amb l'enunciat a la base de dades del projecte que s'ha creat, aquesta tasca inclou tot el tractament necessari perquè les dades siguin coherents.

6.3 Instal·lació de l'eina d'explotació de dades. S'instal·la l'eina d'explotació de dades segons les instruccions indicades i s'estudia la documentació per fer-la servir.

6.4 Construcció dels informes i anàlisi de la informació. Mitjançant l'eina d'explotació de dades es realitzen els informes i s'analitzen els resultats obtinguts, si es necessari es modifiquen els informes.

6.5 Redacció de l'esborrany de la PAC3 d'implementació. En base als resultats obtinguts es redacta l'esborrany del document amb les diferents fases d'implementació realitzades.

6.6 Entrega de l'esborrany de la PAC3 d'implementació. S'entrega l'esborrany de la PAC3 d'implementació al consultor uns dies abans de l'entrega definitiva perquè el pugui revisar i proposar possibles correccions.

6.7 Correcció de l'esborrany de la PAC3 d'implementació i entrega final.

Amb les possibles correccions indicades pel consultor, es rectifica la implementació, i es presenta la versió definitiva a la bústia de l'aula.

7. Redacció de la memòria

7.1 Integració de tot el projecte a la memòria. S'integra tota la informació del projecte a la memòria i es redacten els capítols i les conclusions.

7.2 Revisió dels aspectes formals i ortogràfics de la memòria.

8. Realització de la presentació

8.1 Síntesis de la memòria. Es fa la tria de les parts més representatives de la memòria del projecte que s'inclouran a la presentació.

8.2 Creació de la presentació. Amb la informació obtinguda de la síntesi es crea la presentació intentant complir els requisits establerts a l'enunciat.

8.3 Revisió dels aspectes formals i ortogràfics de la presentació.

9. Entrega de la memòria i la presentació. Dintre dels terminis establerts es fa l'entrega de la versió definitiva de la memòria i la presentació a la bústia de l'aula.

10. Tribunal d'avaluació. Es contesten en el temps establert les possibles preguntes del tribunal d'avaluació intentant defensar la solució proposada a la memòria del treball.

Les fites concretes del TFC són les dates en les que s'ha de fer una entrega o be realitzar una activitat, són les següents:

Fita	Data
Trobada presencial	30/09/10
PAC1	5/10/10
PAC2	8/11/10
PAC3	20/12/10
Lliurament final	10/01/2011
Inici debat virtual	24/01/2011
Fi debat virtual	27/01/2011

1.4.2. Planificació

En aquest apartat es descriu la planificació temporal estimada per a les tasques del TFC. En base al temps disponible, aquest s'ha assignat a les diferents tasques en funció de la càrrega de treball que s'espera que tinguin.

S'estima que es disposarà per dedicar al TFC 2 hores els dies laborables i 3 hores els dies de cap de setmana i festius. Amb aquesta estimació i la càrrega estimada de cada tasca s'ha fet la següent taula amb les tasques i la seva atribució d'hores corresponent.

Tasca	Període	Hores
1. Definició del TFC		
1.1 Lectura i anàlisi de l'enunciat del TFC	23/9	1
1.2 Recopilació i lectura de la documentació	Del 23/9 al 25/9	6
1.3 Definició dels objectius i les tasques del projecte	26/9	3
2. Instal·lació del Software necessari		
2.1 Descarrega del software necessari	27/9	2
2.2 Instal·lació i prova del software	28/9	1
3. Trobada presencial del TFC	30/9	2
4. Realització de Pla de treball (PAC1)		
4.1 Recerca de informació addicional i bibliografia recomanada	28/9	1
4.2 Redacció de l'esborrany del pla de treball	Del 29/9 al 2/10	4
4.3 Redacció de l'esborrany del document d'anàlisi preliminar	Del 29/9 al 2/10	3
4.4 Entrega de l'esborrany del pla de treball i del document d'anàlisi preliminar	2/10	-
4.5 Correcció de l'esborrany del pla de treball i del document d'anàlisi preliminar i entrega final	Del 3/10 al 5/10	3
5. Anàlisi de requeriments i disseny tècnic (PAC2)		
5.1 Lectura exhaustiva de la documentació	Del 3/10 al 9/10	16
5.2 Redacció de l'esborrany de l'anàlisi de requeriments	Del 10/10 al 18/10	21
5.3 Redacció de l'esborrany del disseny tècnic	Del 19/10 al 1/11	32
5.4 Entrega de l'esborrany de l'anàlisi de requeriments i del disseny tècnic	1/11	-
5.5 Correcció de l'esborrany de l'anàlisi de requeriments i del disseny tècnic i entrega final	Del 2/11 al 8/11	16
6. Implementació		
6.1 Construcció de la base de dades	Del 9/11 al 16/11	18
6.2 Carrega de dades	Del 17/11 al 28/11	28
6.3 Instal·lació de l'eina d'explotació de dades	Del 29/11 al 1/12	6
6.4 Construcció dels informes i anàlisi de la informació	Del 2/12 al 9/12	20
6.5 Redacció de l'esborrany de la PAC3 d'implementació	Del 10/12 al 13/12	10

6.6 Entrega de l'esborrany de la PAC3 d'implementació	13/12	-
6.7 Correcció de l'esborrany de la PAC3 d'implementació i entrega final	Del 14/12 al 20/12	16
7. Redacció de la memòria		
7.1 Integració de tot el projecte a la memòria	Del 21/12 al 26/12	14
7.2 Revisió dels aspectes formals i ortogràfics de la memòria	Del 27/12 al 28/12	4
8. Realització de la presentació		
8.1 Síntesis de la memòria	Del 29/12 al 31/12	6
8.2 Creació de la presentació	Del 1/1 al 8/1	22
8.3 Revisió dels aspectes formals i ortogràfics de la presentació	Del 9/1 al 10/1	5
9. Entrega de la memòria i la presentació		
9. Entrega de la memòria i la presentació	10/1	-
10. Tribunal d'avaluació		
10. Tribunal d'avaluació	Del 24/1 al 27/1	8

Amb aquestes tasques i les dates s'ha creat un diagrama de Gantt per poder veure més gràficament la planificació del TFC. S'ha dividit el diagrama per fites perquè es pugui veure les diferents tasques amb claredat:

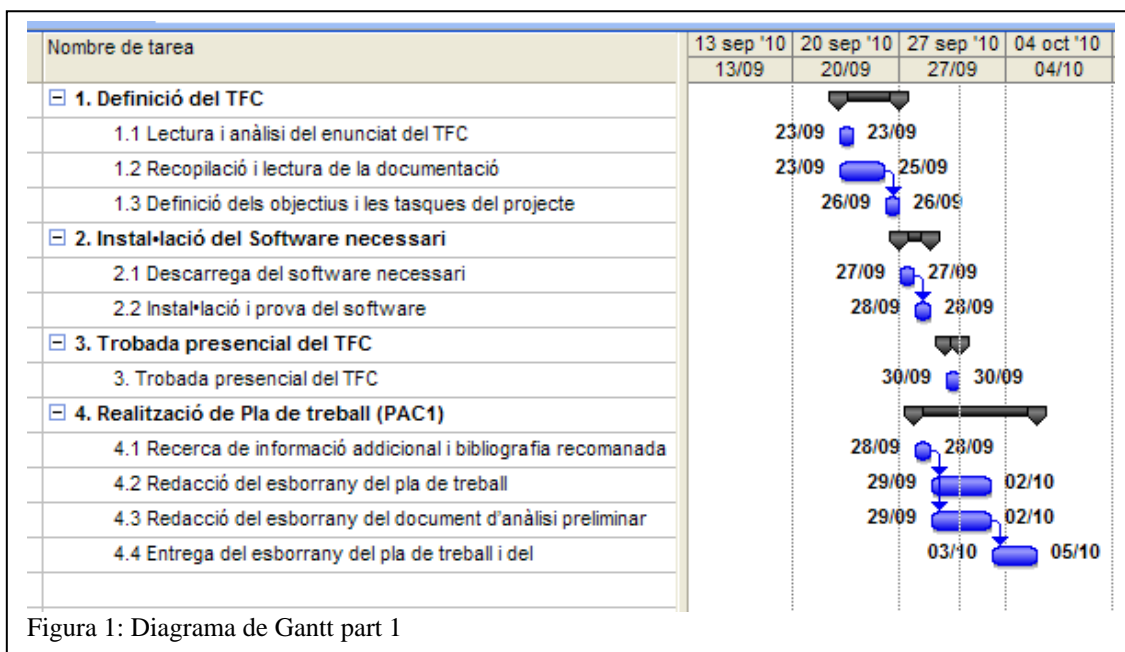


Figura 1: Diagrama de Gantt part 1

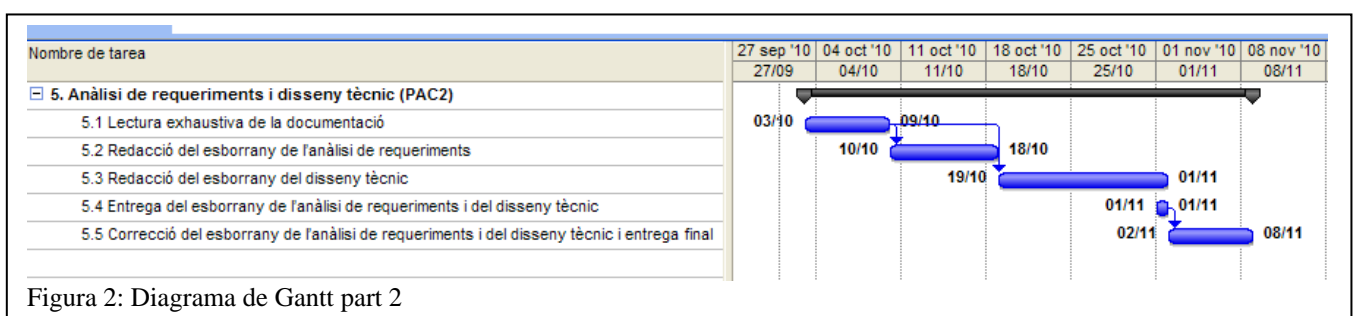
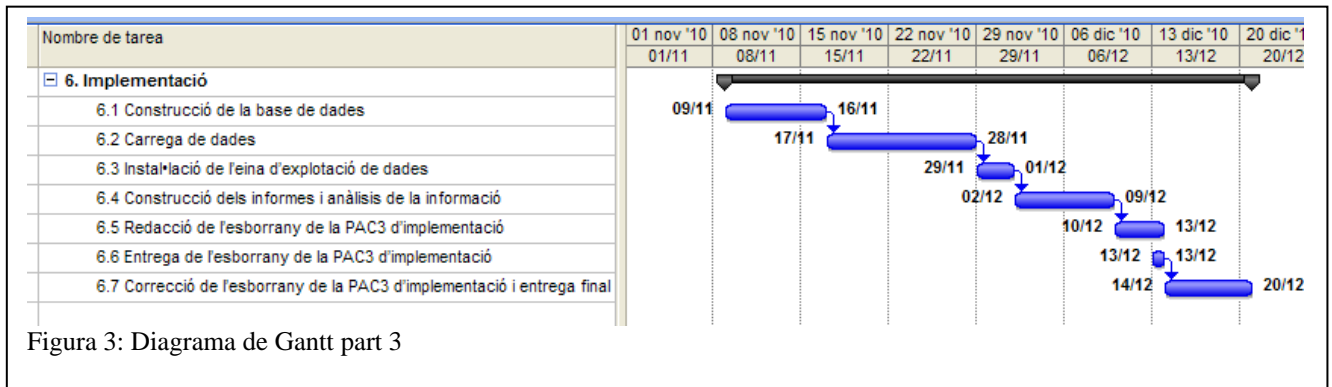


Figura 2: Diagrama de Gantt part 2



1.4.3. Incidències i riscos

En aquest apartat s'analitzen les possibles incidències que puguin afectar al desenvolupament del TFC. Per a cada possible incidència es proposa un pla de contingències per evitar que afecti de manera important al desenvolupament del TFC.

Falta de temps en dies laborables

Degut a les característiques de la meua feina, poden haver setmanes en les que no pugui disposar els dies laborables del temps previst a la planificació. En aquest cas es proposa traslladar aquestes hores perdudes al cap de setmana, la planificació te assignades 6 hores al cap de setmana, però si fos necessari es podria arribar fins a les 16 hores.

Fallada de l'equip informàtic o pèrdua de dades

Una fallada de l'equip informàtic que es farà servir per realitzar el TFC podria provocar un endarreriment de la feina prevista i fins i tot una pèrdua de dades. Per tal de evitar aquesta situació s'instal·larà l'entorn de programari en dos equips diferents, per poder fer servir un o altre en cas d'incidència. Per evitar una pèrdua de dades es farà servir un disc dur extern per guardar una còpia de seguretat dels diferents arxius del TFC.

La planificació establerta no es pot complir

Si durant el desenvolupament del TFC es veu que les hores previstes a la planificació no són suficients per realitzar les tasques serà necessari detectar aquesta situació el més aviat possible i replantejar la planificació, afegint hores al cap de setmana o fins i tot hi ha la possibilitat de demanar dies lliures a la feina. En tot cas es comunicarà al consultor el canvi en la planificació inicial.

1.4.4. Material necessari

En aquest apartat es descriuen les necessitats de maquinari i programari per poder realitzar el TFC, així com els coneixements previs.

Coneixements previs

Per realitzar aquest TFC es necessari tenir uns coneixements generals de bases de dades, així com coneixement de programació amb llenguatge PL/SQL d'Oracle.

Requeriments de maquinari

Punt de treball estàndard de la UOC.

Requeriments de programari

Sistema operatiu que pugui executar VirtualBox.

VirtualBox.

Base de dades Oracle 10g Express.

Eina Business Intelligence Oracle Discoverer.

Eina Case tipus Oracle Designer.

Processador de Textos.

Programa per realitzar presentacions tipus PowerPoint.

Programa de gravació d'escriptori.

1.5 Productes obtinguts

Com a resultat de la realització del projecte, s'han generat una sèrie de documents i fitxers que es descriuen a continuació:

- Un disc virtual de VirtualBox que conté el magatzem de dades creat i els diferents informes.
- Scripts de càrrega i transformació de dades ubicats a la màquina virtual.
- Arxius de control de la carrega de dades amb SQLLOADER ubicats a la màquina virtual.

- Arxius de les fonts de dades en format CSV, que són els arxius d'estadístiques dels embassaments, l'arxiu d'embassaments i el de l'estimació d'aigua de neu, ubicats a la màquina virtual.
- Un document explicatiu sobre com accedir a la base de dades del magatzem de dades creada a Oracle i a l'eina d'anàlisi/reporting Discoverer.
- Un document explicatiu del funcionament dels processos de càrrega.
- Memòria final del TFC.
- Presentació virtual en format vídeo.

1.6 Altres capítols de la memòria

En els següents capítols de la memòria s'aprofundeix en les diferents fases de la realització del projecte. En primer lloc es tracta l'anàlisi de requeriments realitzat, a continuació es presenta el disseny tècnic triat per la realització del magatzem de dades i dels informes.

Com a resultat del projecte es presenten les captures de pantalla obtingudes dels diferents informes creats.

Per acabar s'expliquen les conclusions a les que s'ha arribat amb la realització del projecte i les possibles línies d'evolució futures.

2. Anàlisi

2.1 Model conceptual

L'enunciat del projecte demana crear un magatzem de dades que automatitzi la recollida de la informació provinent dels fulls Excel proporcionats, i que permeti explotar aquesta informació per tal de poder realitzar les consultes indicades.

Les dades proporcionades es poden dividir d'una banda en dades fixes referents als embassaments, tals com nom de l'embassament, riu que l'abasteix, comunitat autònoma a la que pertany, etc. D'altra banda hi ha les dades variables que formen les estadístiques de cada embassament i que varien en funció del període de temps, tals com capacitat actual de l'embassament, entrada i sortida d'aigua o previsió d'aigua de neu.

Estudiant aquestes dades s'ha decidit que l'esquema més adient pel disseny del magatzem de dades és l'esquema en estrella. Aquesta elecció és deguda a que aquest esquema es caracteritza per tenir una taula de fets que conté les dades per l'anàlisi rodejada de taules de dimensions amb informació complementària dels registres de la taula de fets. En aquest cas hi haurà dues taules de fets que compartiran algunes taules de dimensions, la primera taula de fets estarà formada per les dades variables que formen les estadístiques dels embassaments i la segona taula de fets per les dades de l'estimació d'aigua de neu. Les taules de dimensions estaran formades per totes les dades fixes complementàries.

La simplicitat d'aquest esquema permet que el rendiment i velocitat de la base de dades sigui millor que amb altres tipus d'esquemes, degut a que permet obtenir la informació amb consultes no molt complicades.

2.2 Triar el fet

El primer pas en el disseny del model conceptual és la tria del fet objecte de l'anàlisi, el fet és un conjunt d'esdeveniments amb dades numèriques associades. En el cas del projecte aquestes dades numèriques són les estadístiques de cada embassament contingudes en el fitxers proporcionats.

De les dades proporcionades i les consultes que demana l'enunciat s'extreu que podem diferenciar dos fets diferenciats, d'una banda un fet és l'estat d'un embassament determinat en un mes i any determinat. D'altra banda tenim el fet de l'estimació d'aigua de neu prevista per un embassament i any determinat

Aquests dos fets formaran dues taules de fets en el disseny en estrella, una contindrà les dades estadístiques de cada embassament a nivell de mes i l'altra les dades de la previsió d'aigua de neu per embassament a nivell d'any.

2.3 Trobar el grànul escaient

Per tal de trobar el grànul més escaient pel disseny de la base de dades, s'ha d'intentar trobar un equilibri entre el nivell de detall i el rendiment de la base de dades, per assolir aquest objectiu s'ha d'escollir un grànul amb el màxim nivell de detall possible que ens garanteixi una mínima pèrdua d'informació, però que d'altra banda no tingui un impacte excessiu en el rendiment de la base de dades.

En el cas concret del projecte i tenint en compte la informació que demana obtenir el client, es definiran dos grànuls diferents, el grànul més escaient per la taula de fets de les estadístiques dels embassaments és les dades d'un embassament determinat que pertany a una comunitat autònoma determinada en un any i mes concret.

Per la taula de fets de l'estimació d'aigua de neu en canvi el grànul més escaient és l'estimació d'aigua de neu d'un embassament per un any concret.

Tenim per tant que per la taula de fets de les estadístiques les dades estan calculades a nivell de mes i any. I per la taula de fets de l'estimació d'aigua de neu les dades estan calculades a nivell d'any

2.4 Escollir les dimensions que s'utilitzaran en l'anàlisi

Un cop trobat els grànuls escaients, s'han de trobar les dimensions que complementen les dades de les taules de fets.

De l'anàlisi de les dades proporcionades als fitxers i les necessitats del client, s'han identificat les següents dimensions de la taula de fets de les estadístiques:

Comunitat autònoma: contindrà les dades de les diferents comunitats autònomes de les que es tenen dades.

Riu: contindrà les dades dels diferents rius dels que es tenen dades.

Embassament: contindrà les dades dels diferents embassaments dels que es tenen dades.

Mesos: contindrà les dades dels diferents mesos que es poden consultar.

Anys: contindrà les dades dels diferents anys que es poden consultar.

Província: Contindrà les dades de les diferents províncies.

Encara que per realitzar els informes demanats no sigui necessari, s'ha diferenciat entre comunitat autònoma i província degut a que pot ser interessant poder distingir les dades per províncies, en previsió que en un futur es demanin informes amb aquest nivell de detall.

La taula de fets de l'estimació d'aigua de neu compartirà la dimensió Embassament i la dimensió Anys amb l'altra taula de fets per poder determinar a quin embassament i any pertany l'estimació.

2.5 Trobar els atributs de cada dimensió

Per fer la tria dels atributs de cada dimensió s'ha tingut en compte que siguin descriptius i entenedors i que siguin d'utilitat a l'hora de realitzar els informes demanats pel client.

A continuació es fa una breu descripció dels diferents atributs escollits per a cada dimensió:

Dimensió Comunitat autònoma

IdComunitat: Codi numèric que identifica de forma única la comunitat autònoma.

Nom: Nom de la comunitat autònoma.

Dimensió Riu

IdRiu: Codi numèric que identifica de forma única el riu.

Nom: Nom del riu.

Dimensió Embassament

IdEmbassament: Codi numèric que identifica de forma única l'embassament.

Nom: Nom de l'embassament.

Dimensió Mesos

IdMes: Codi numèric que identifica de forma única el mes.

Mes: Nom del mes.

Dimensió Província

IdProvíncia: Codi numèric que identifica de forma única la província.

Nom: Nom de la província.

Dimensió Anys

IdAny: Codi numèric que identifica de forma única l'any.

AnyE: Enter que representa un any concret.

2.6 Distingir entre descriptors i jerarquies d'agregació

Dels atributs de cada dimensió es poden diferenciar els que són descriptors i els que poden formar jerarquies d'agregació. Els atributs descriptors serveixen simplement per seleccionar aquest atribut, els que permeten definir jerarquies d'agregació a més serveixen per agrupar atributs.

En el cas d'aquest projecte no s'ha establert cap jerarquia d'agregació.

2.7 Decidir quines són les mesures que interessin

Les mesures que seran més útils per l'elaboració del magatzem de dades són les següents estadístiques de cada embassament:

Capacitat total: Quantitat d'aigua total que pot contenir un determinat embassament.

Capacitat actual: Quantitat d'aigua que conté l'embassament actualment.

Capacitat any anterior: Quantitat d'aigua que contenia l'embassament l'any anterior en aquest mateix mes i any.

Entrada mensual: Quantitat d'aigua que ha rebut l'embassament en el més.

Sortida mensual: Quantitat d'aigua que ha sortit de l'embassament en el mes.

Aigua neu: Aigua de neu prevista per un any i un embassament.

Aquestes mesures són les variables que ens permetran resoldre les diferents consultes que demana l'enunciat. Les mesures Capacitat total, Capacitat actual, Capacitat any anterior, Entrada mensual i Sortida mensual pertanyen a la taula de fets d'estadístiques. La mesura Aigua de neu pertany a la taula de fets de l'estimació d'aigua de neu.

2.8 Definir Cel·les

En aquest apartat es defineixen les cel·les que contenen mesures interessants pel posterior anàlisi de les dades.

En el cas de projecte hi han dues cel·les que contenen mesures a analitzar, d'una banda tenim la cel·la que correspon a l'estat d'un embassament en un mes concret. D'altra banda tenim la cel·la que correspon a l'estimació d'aigua de neu per un embassament en un any concret.

2.9 Explicitar les restriccions d'integritat

En aquest apartat es defineixen les restriccions d'integritat, per fer-ho s'estableixen les bases. Aquestes bases indiquen quines dimensions identifiquen les cel·les.

En el cas del projecte, per a definir de forma única la cel·la de les estadístiques d'un embassament només són necessàries tres de les sis dimensions establertes. Amb l'embassament, el mes i l'any es pot determinar de forma única les estadístiques d'un embassament en un mes concret, això es degut a que un embassament només pot pertànyer a una comunitat autònoma i estar abastit per un riu concret. El cas de la dimensió província és igual que en el cas de la comunitat autònoma, un embassament només pot pertànyer a una província.

Per definir de forma única la cel·la de l'estimació d'aigua de neu són necessàries les dimensions Embassament i Anys, ja que cada embassament té definida una estimació d'aigua de neu per cada any.

2.10 Estudiar la viabilitat

Per estudiar la viabilitat del model conceptual plantejat s'ha de fer una estimació de l'espai que ocuparan les dades, per així poder valorar si el magatzem de dades podrà tindre la funcionalitat desitjada.

Per calcular aquest espai s'han d'estudiar la quantitat d'informació que contenen les fonts de dades. Els fitxers proporcionats contenen dades d'uns 50 embassaments diferents durant 54 mesos, llavors si calculem per la base {embassament, mes, any} obtenim un espai de 50×54 (2700 possibles cel·les).

Si estimem que els identificadors numèrics de les 6 dimensions ocuparan uns 20 bytes i que les mesures ocuparan 25 bytes més al ser dades numèriques també, tenim que les dades contingudes al sistema ocuparan 45×2.700 bytes = 121.500 bytes / 1024 = 118,65 Kbytes.

Per la base {embassament, any} hem de considerar que hi han dades de l'estimació d'aigua de neu de 5 anys i de 8 embassaments, encara que la resta d'embassament es considera que l'estimació és 1. Si fem el càlcul obtenim un espai de 50×5 (250 possibles cel·les). Si estimem que els identificadors numèrics de les 2 dimensions ocuparan uns 8 bytes i que la mesura ocuparà uns 4 bytes més al ser també una dada numèrica, tenim que les dades contingudes al sistema ocuparan 12×250 bytes = 3.000 bytes / 1024 = 2,92 Kbytes.

També faltaria calcular l'espai que ocuparien les dimensions, però com és molt inferior ja es pot valorar la viabilitat amb les dades obtingudes. El resultat demostra que el volum de dades pot ser processat sense problemes per un equip destinat a aquesta tasca. Tampoc hi hauria problema a mig termini en continuar introduint dades cada mes, perquè el creixement del volum de dades no seria molt significatiu.

2.11 Diagrama del model conceptual

Un cop establerts els passos anteriors en el model conceptual del magatzem de dades, el següent esquema representa el resultat obtingut:

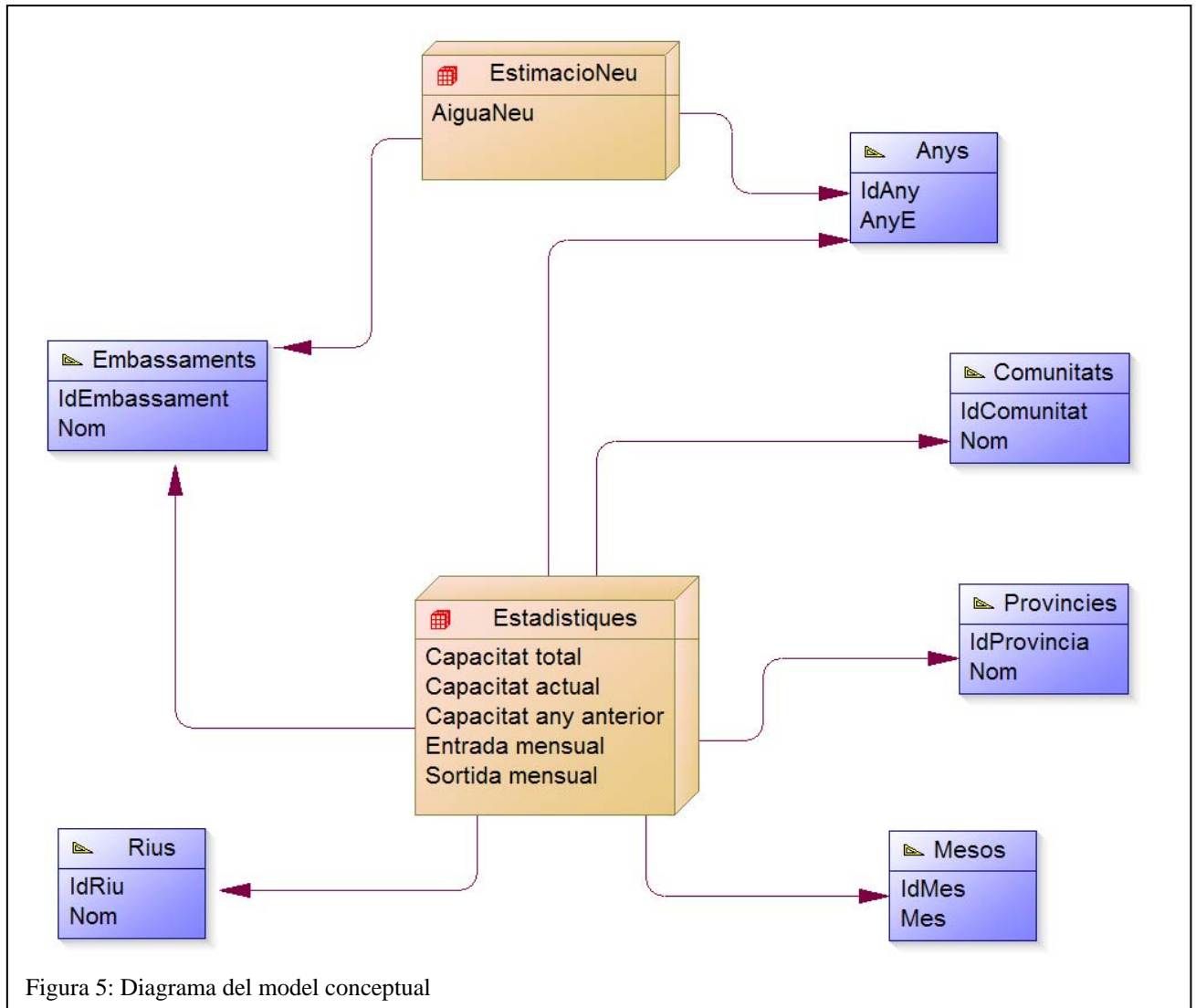


Figura 5: Diagrama del model conceptual

2.12 Diagrama de casos d'us

El següent diagrama de casos d'us representa els diferents usuaris del magatzem de dades i les funcions que poden realitzar:

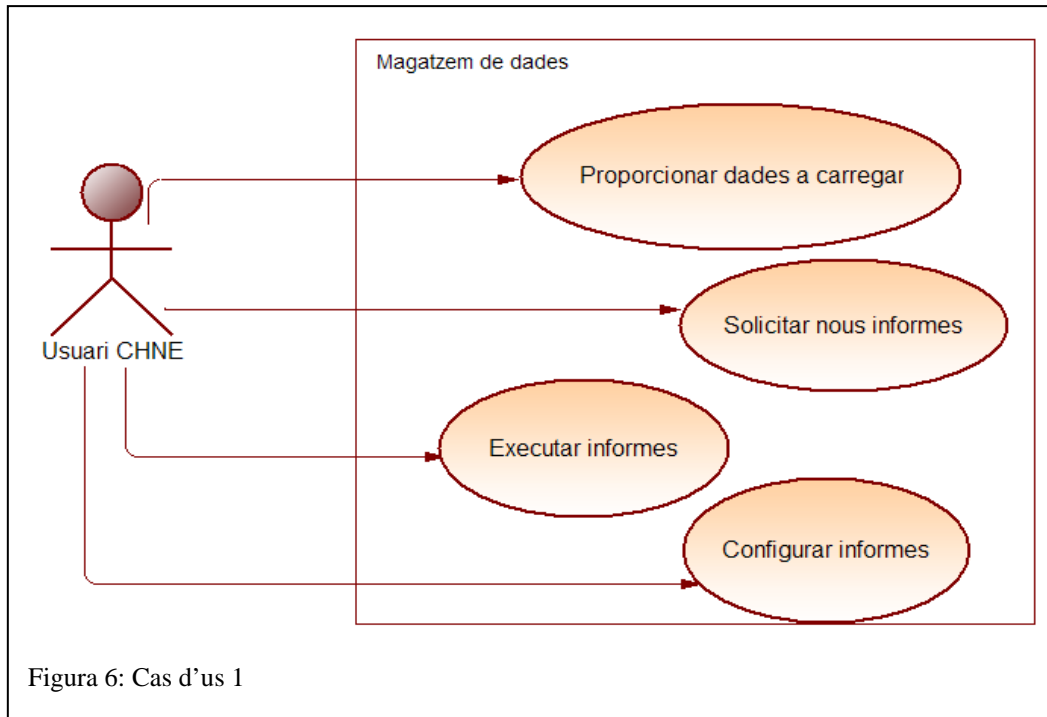


Figura 6: Cas d'us 1

Els usuaris de la CHNE són els encarregats de proporcionar les dades en format Excel que es carregaran al magatzem de dades. A més són qui demanen els diferents informes sobre les dades, configuren els paràmetres i executen els existents.

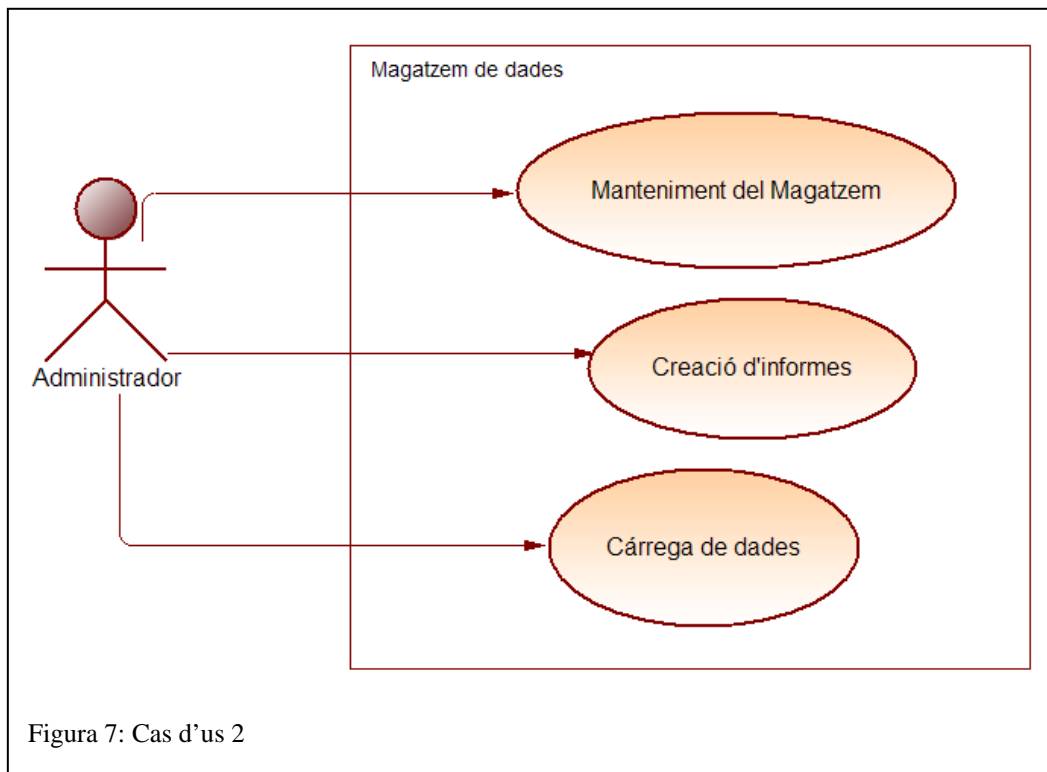


Figura 7: Cas d'us 2

L'administrador és l'encarregat de carregar les dades proporcionades per la CHNE i de crear els diferents informes que es sol·liciten. També és l'encarregat del manteniment de del magatzem de dades, aquesta tasca inclou la gestió d'usuaris, la realització de còpies de seguretat i el seguiment del rendiment i espai disponible del magatzem de dades.

3. Disseny

3.1 Disseny Lògic

Una vegada fet el model conceptual multidimensional s'ha de passar aquest disseny al model lògic. Per dissenyar aquest model es fa servir una altre cop una estrella, per implementar-la s'ha de crear una taula per a cada fet on cada fila és una cel·la de l'espai multidimensional, i també s'ha de crear una taula per a cada dimensió.

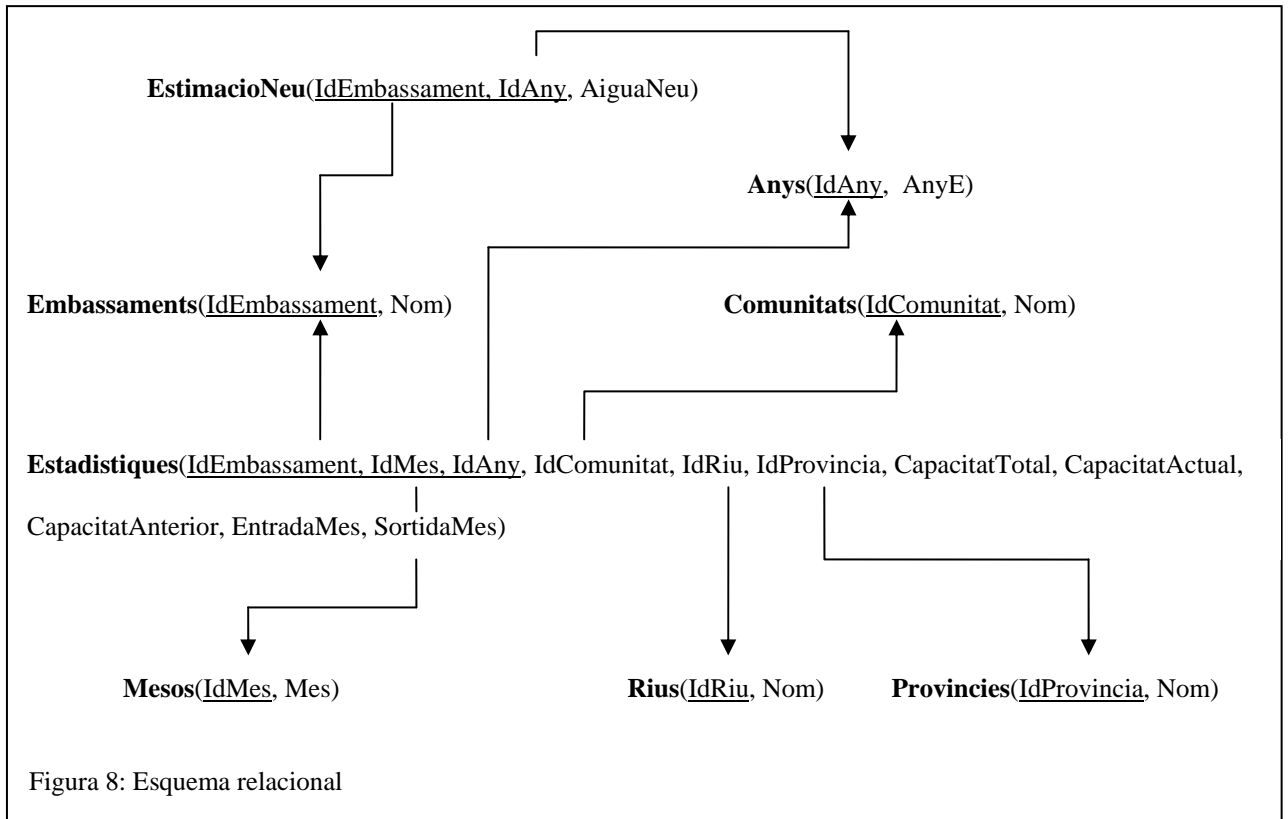
Cada taula del fet estarà lligada amb les taules de dimensions corresponents mitjançant claus foranes i contindrà a més d'aquestes claus foranes les diferents mesures establertes en el model conceptual. Les taules de dimensions contindran els diferents descriptors definits.

La clau primària de cada taula del fet estarà formada per les claus foranes que formen part de la base de la cel·la atòmica, en el cas del fet de les estadístiques seran els atributs de mes, any i embassament i en el cas del fet de l'estimació de neu seran els atributs d'any i embassament.

Com la major part del volum de dades es concentra a les taules del fet, es poden definir substituïts de les claus primàries de les dimensions, això permet reduir la grandària de les taules de fets ja que la seva clau primària està formada per les claus foranes de les taules de dimensions. Fer servir substituïts també permet que possibles canvis en els atributs identificadors de les dimensions no afectin a totes les dades ja introduïdes. Per a aquest motiu s'han definit atributs numèrics com a clau primària per a cada taula de dimensió.

3.2 Esquema relacional

Amb les dades de l'apartat anterior s'ha realitzat l'esquema relacional següent:



3.3. Disseny Físic

A partir del disseny lògic obtingut es crearan les taules que es mostren en el següent diagrama del model físic:

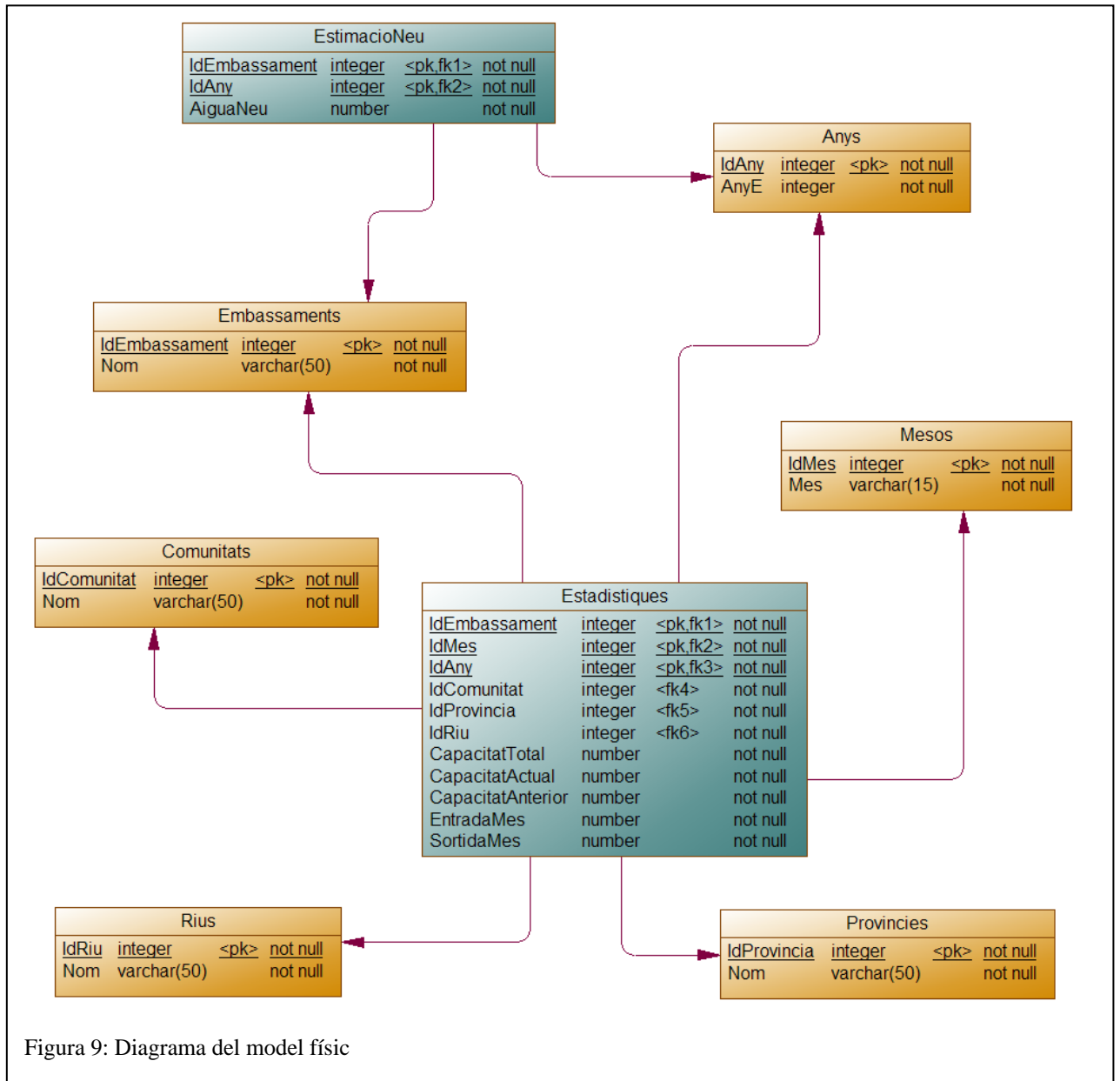


Figura 9: Diagrama del model físic

La taula Estadístiques representa una de les taules del fet i és el centre de l'estrella, com a atributs té les claus foranes de les taules de dimensions i les mesures. De la mateixa manera la taula EstimacioNeu representa l'altra taula del fet, com a atributs té les claus foranes de les taules de dimensions i les mesures corresponents.

Els atributs que corresponen a les mesures tenen com a tipus de dades number, que permet introduir valors numèrics amb decimals. A les fonts de dades algunes mesures com les capacitats i l'estimació d'aigua de neu tenen valors enters, però s'ha escollit aquest tipus de dades per evitar possibles problemes si en futures mesures es decideix introduir valors amb decimals.

La resta de taules representen les diferents dimensions, i com a atributs tenen un identificador substitut i les dades pròpies de la dimensió.

Per determinar la nul·litat dels diferents atributs s'ha tingut en compte que en les fonts de dades la majoria dels embassaments tenen totes les dades necessàries, i en els casos que no estiguin complertes, s'ompliran aquests atributs amb uns valors determinats per no perdre les mesures d'aquest embassament al descartar-lo. Per aquest motiu s'ha decidit posar la restricció de no nul·litat a tots els atributs de les taules. Com es resoldran aquests casos que s'hagi de completar alguna dada es tractarà en el capítol corresponent a l'extracció, transformació i càrrega de les dades.

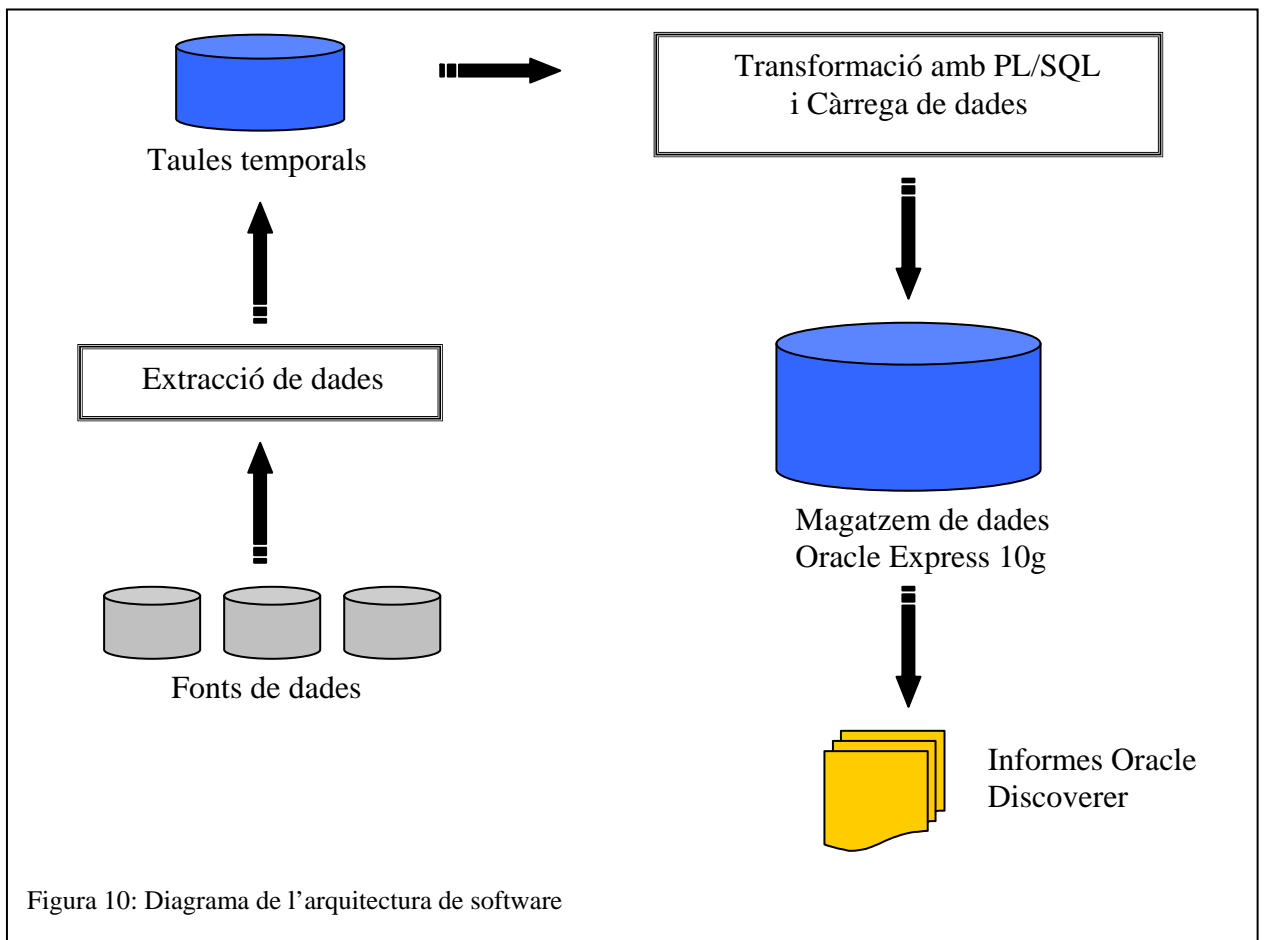
3.4. Arquitectura del Sistema

En aquest apartat es descriu el funcionament general del magatzem de dades i es mostren els diagrames de l'arquitectura de software i hardware del sistema.

El funcionament del magatzem de dades comença per un procés d'extracció, transformació i càrrega de les dades que contindrà el magatzem, aquestes dades s'extrauen de les fonts de dades i es carreguen en taules temporals, per posteriorment ser transformades mitjançant processos PL/SQL i carregades en les taules definitives. Un cop carregades, les dades ja poden ser explotades pels usuaris mitjançant informes creats amb l'eina d'explotació Oracle Discoverer.

3.4.1 Arquitectura de Software

El següent diagrama mostra els diferents components de software i processos que intervenen en el funcionament del magatzem de dades:



3.4.2 Arquitectura de Hardware

El següent diagrama mostra els diferents components de hardware que intervenen en el funcionament del magatzem de dades:

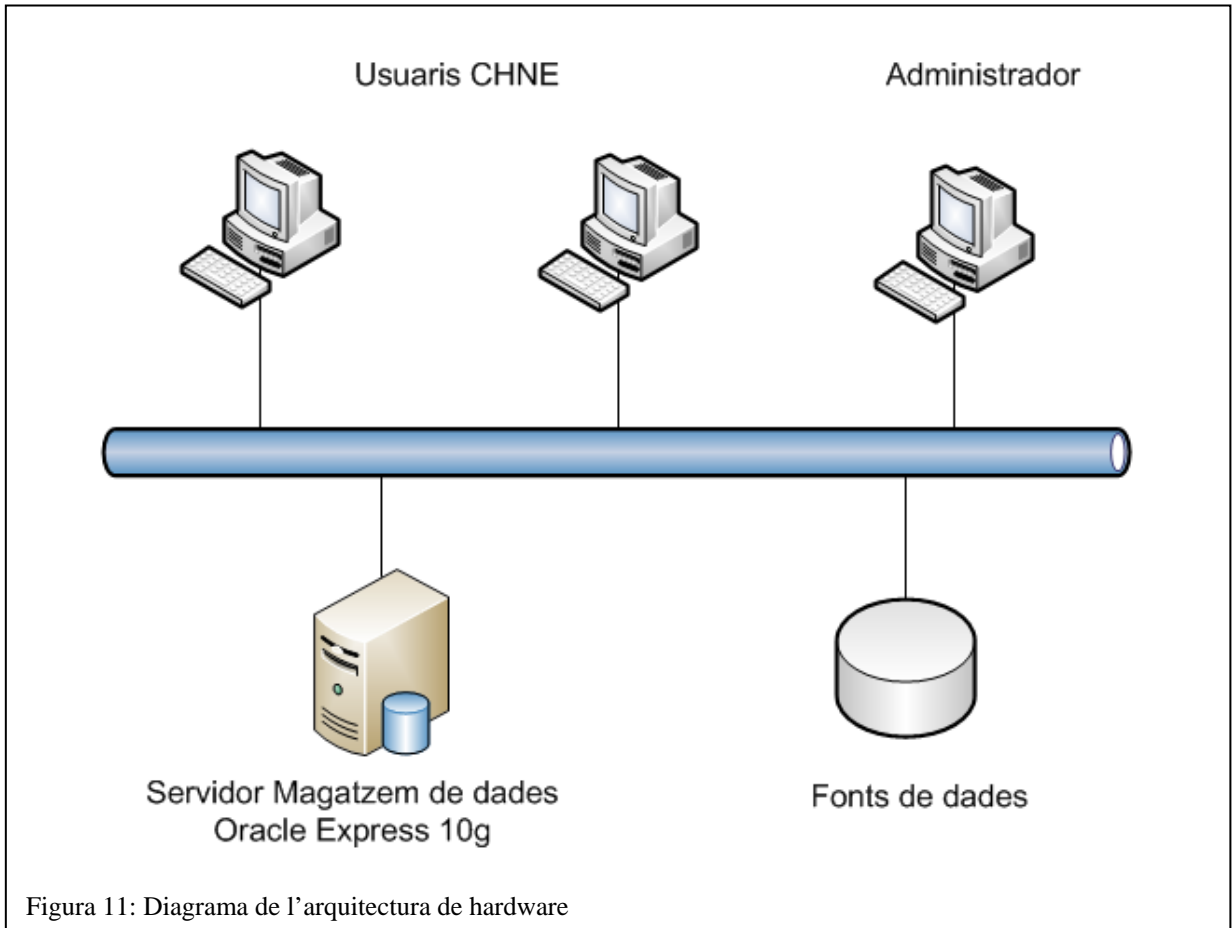


Figura 11: Diagrama de l'arquitectura de hardware

3.5 Disseny dels Procediments ETL

Aquest apartat descriu els procediments utilitzats per la incorporació de les dades proporcionades pel client al magatzem de dades. Aquest procediment rep el nom d'ETL, que es correspon amb les sigles amb anglès d'extracció, transformació i càrrega, i són els passos necessaris per la incorporació de les dades que es volen explotar amb el magatzem de dades.

3.5.1 Extracció

El procés d'extracció consisteix en la recopilació de les diferents dades dels sistemes d'origen que es volen incorporar al magatzem de dades, aquest procés també inclou

l'adaptació d'aquets orígens de dades a un format preparat per la transformació i la càrrega de les dades.

En el cas del projecte no cal recopilar les dades de diferents sistemes d'origen ja que ens les proporciona el client en una sèrie de fitxers d'Excel en format .xls.

Sí és necessari adaptar aquests fitxers a un format preparat per la transformació i la càrrega de les dades, en aquest cas consisteix a guardar els diferents fitxers en format .csv, que és text delimitat per comes fàcilment interpretable pel Sistema de Gestió de Bases de Dades que en el nostre cas és Oracle 10g Express.

Les úniques dades no proporcionades en aquests fitxers són l'estimació d'aigua de neu per any. Aquestes dades es proporcionen en format de text i es necessari guardar-les en format .csv com la resta de dades, per així homogeneïtzar de les fonts de dades.

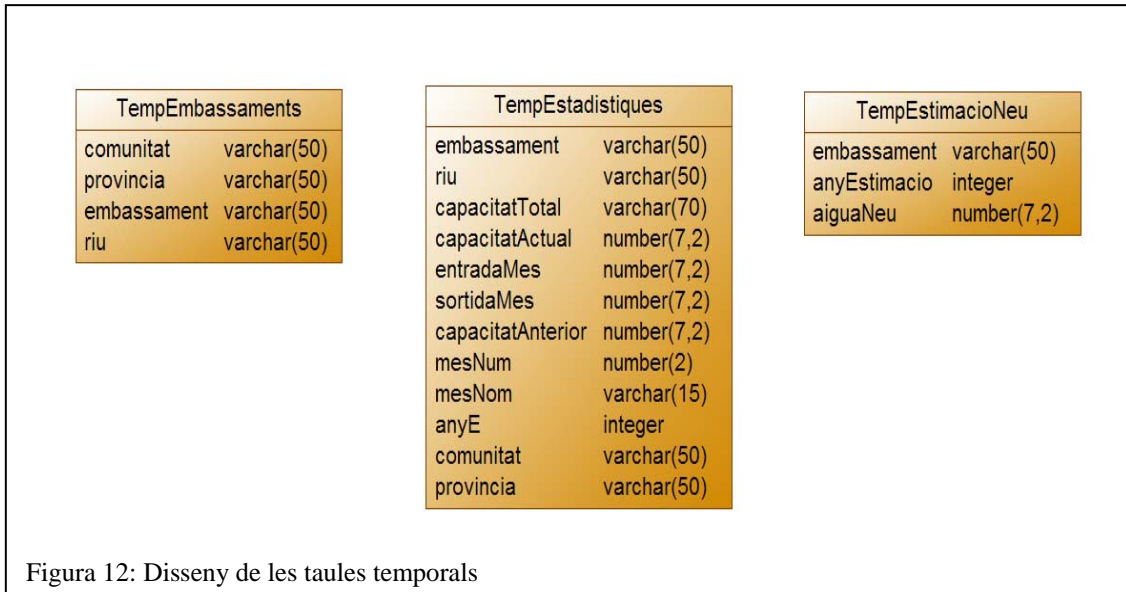
3.5.2 Transformació i càrrega

Com a part del procés ETL, la transformació s'encarrega de fer els canvis necessaris en les dades perquè puguin ser carregades, aquests canvis poden ser comprovar i corregir possibles errors en el format o contingut de les dades, completar camps que estiguin buits o discriminar quines dades es carreguen i quines no.

En el cas del projecte s'ha decidit que primer es carregaran les dades en tres taules temporals mitjançant l'eina SQL*LOADER i posteriorment es farà la transformació i càrrega final a les taules del magatzem de dades.

En aquestes taules temporals es carregaran les dades en brut directament de les fonts de dades i posteriorment es realitzarà la transformació de les dades de forma automàtica mitjançant l'execució de procediments PL/SQL. Un cop finalitzats aquests procediments i amb les dades amb el format i contingut correcte, es procedirà a la càrrega d'aquestes a les taules definitives

El disseny de les taules temporals en les que es carregaran les dades en brut és el següent:



A la taula TempEmbassaments es carregaran les dades del fitxer embassaments.csv, a la taula TempEstadistiques es carregaran les dades dels diferents fitxers que contenen les estadístiques dels embassaments. Per últim a la taula TempEstimacioNeu es carregaran les dades del fitxer EstimacioNeu.csv

Un cop carregades les dades en les taules temporals, s'ha de fer la transformació per comprovar i corregir possibles errors en el format o contingut de les dades i completar camps que estiguin buits. A continuació es descriuen els diferents canvis i correccions realitzades en les dades en el procés de transformació:

- Amb el procediment emmagatzemat insert_TempEstadistiques que s'executa en el procés de càrrega dels fitxers d'estadístiques dels embassaments a la taula TempEstadistiques, s'extreu de la capçalera de cada arxiu el mes i l'any al que pertanyen les estadístiques i s'insereix als registres corresponents. En aquest procés també es substitueix les mesures d'entrada i sortida que estiguin buides per un 0.
- El procediment emmagatzemat transform_TempEmbassaments s'encarrega de diferenciar els valors que corresponen a una comunitat autònoma dels que corresponen a una província, per així poder completar les dades de cada embassament amb la comunitat autònoma i província a la que pertanyen. També s'encarrega de posar en majúscula les dades per homogeneïtzar-les amb la resta de les fonts de dades.
- El procediment emmagatzemat transform_TempEstimacioNeu s'encarrega d'incloure a la taula TempEstimacioNeu tots els embassaments que apareixen a la taula TempEstadistiques i que no tenen estimació d'aigua de neu, en aquestos casos es considera que l'estimació és 1, com s'indica a l'enunciat. En aquest procés també es posa en majúscules les dades i es corregeixen alguns errors concrets als noms dels embassaments perquè coincideixin amb la resta de dades.

- El procediment emmagatzemat `transform_TempEstadistiques` s'encarrega d'afegir a les estadístiques de cada embassament les dades de la comunitat autònoma i província a la que pertanyen des de la taula de `tempEmbassaments`, en cas que l'embassament no aparegui a la taula s'ha decidit incorporar les dades de l'embassament igualment, d'aquesta manera no es perden aquestes dades. En aquestos casos la comunitat autònoma de l'embassament i la província s'ompliran amb la paraula `BUIDA`, per indicar que no es tenen dades de a que comunitat i província pertany l'embassament. També es tracten errors específics en alguns noms d'embassament per tal que es puguin associar a la comunitat autònoma i província corresponents. A més, en aquest procés es transforma el valor numèric del mes en el nom del mes corresponent.

En el procés de transformació s'ha hagut de prendre decisions per donar solució als diferents problemes trobats amb les fonts de dades, és important explicar les mesures preses ja que el fet de decidir tractar aquests problemes d'una manera o d'un altra pot influir en els resultats generats pels informes.

S'han detectat errors de coherència a les fonts de dades entre l'arxiu d'embassaments i els arxius d'estadístiques. Alguns noms d'embassaments i rius no apareixen escrits igual a tots dos arxius, això provoca que a l'assignar la comunitat autònoma i província a la que pertany un embassament, en aquestos casos s'assigni `BUIDA` al no coincidir les dades. Com aquest fet afecta al resultat dels informes que obtenen informació per comunitat autònoma s'ha decidit corregir aquests errors.

Tenint en compte que modificar les fonts de dades no és recomanable, s'ha decidit corregir els errors en la fase de transformació, un cop les dades ja estiguin carregades a les taules temporals. Com les diferències als noms no tenen una pauta comuna, ha calgut tractar cada cas de forma específica.

A les dades proporcionades de l'estimació de l'aigua de neu també hi ha un cas que el nom de l'embassament no és coherent amb les altres fonts de dades. Per no perdre aquesta estimació també s'ha decidit tractar de manera específica aquest cas en la fase de transformació, per tal que el nom de l'embassament coincideixi amb la resta de dades.

Aquesta no és la solució ideal, ja que no s'haurien de tractar els errors de forma individual al codi, però s'ha preferit aquesta solució a perdre dades de cara a l'anàlisi. El més adient seria demanar una correcció de les dades per part del client, però en aquest cas això no és possible.

La resta de procediments emmagatzemats s'encarreguen de carregar les dades a les taules definitives un cop han estat transformades i corregides.

Tots aquests procediments de transformació i càrrega, així com la neteja de les taules temporals, s'executen automàticament mitjançant un arxiu executable, les instruccions per la seva configuració i execució es detallen en el capítol processos de càrrega.

3.6 Disseny i descripció del informes creats

Als següents apartats es descriu el procés de configuració d'Oracle Discoverer i la creació del diferents informes creats.

3.6.1 Creació de l'àrea de negoci

Amb Oracle Discoverer Administrator s'ha creat un nou EUL (nivell d'usuari final) per a l'usuari CHNE de la base de dades.

Un cop creat el nou EUL s'ha creat una àrea de negoci per l'usuari CHNE on s'han afegit les diferents taules que contenen les dades a explotar. El resultat d'aquest procés es la següent estructura:

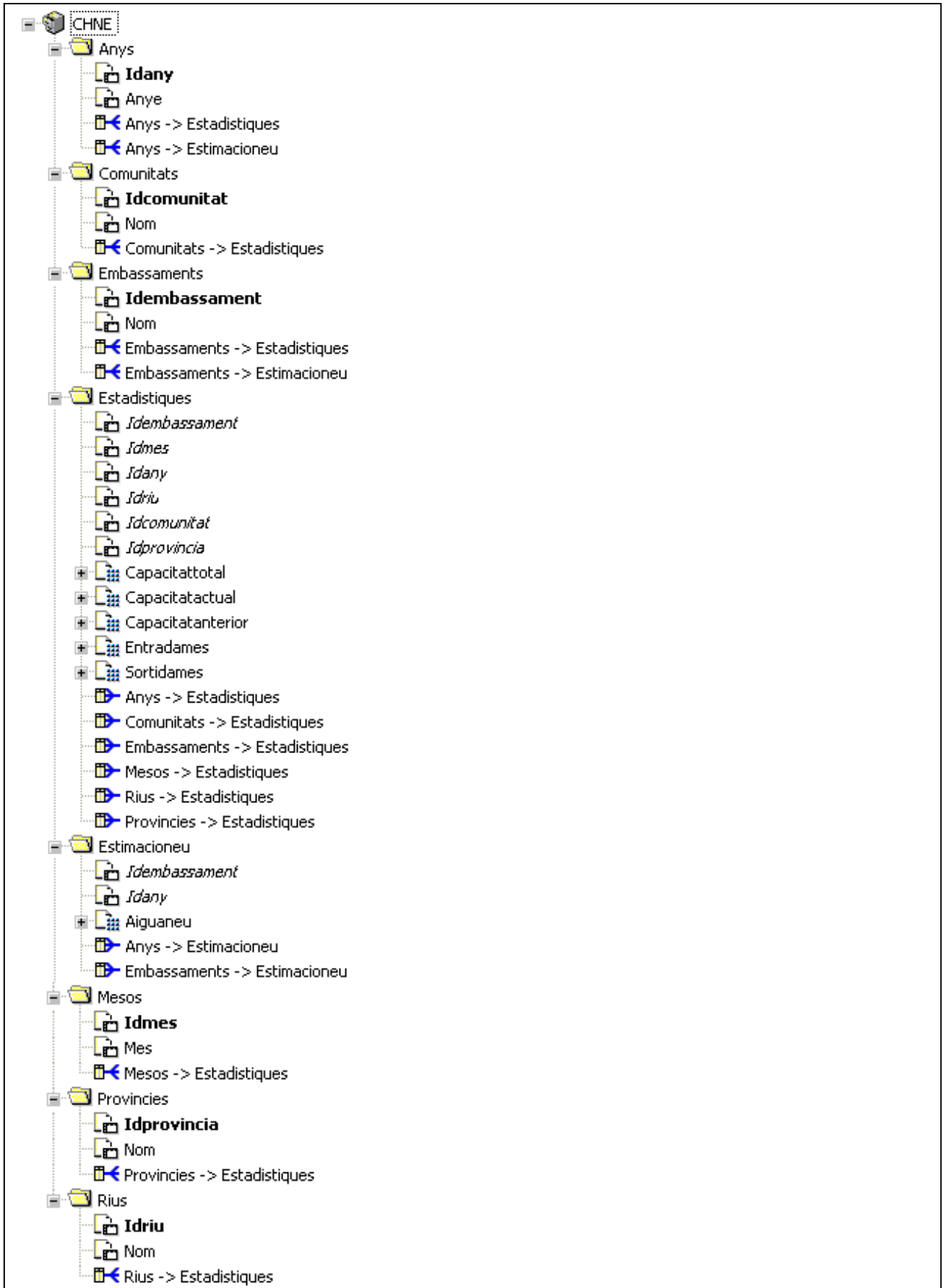


Figura 13: Estructura àrea de negoci

3.6.2 Informes creats

Un cop preparada l'àrea de negoci s'han creat els diferents informes mitjançant Oracle Discoverer Desktop.

Per realitzar els diferents informes s'ha creat un llibre de treball on cada fulla del llibre conté un informe.

Per complir amb el que demana l'enunciat, es mostren els volums d'aigua dels resultats tant en Hm3 (hectòmetres cúbics) com en Ml (milions de litres). Com les mesures contingudes al magatzem de dades estan en Hm3, per tal de poder mostrar als informes els volums d'aigua en Ml s'han creat càlculs en Oracle Discoverer per cada cas concret. S'ha decidit fer-ho així per evitar tenir les mesures duplicades al magatzem de dades i estalviar així aquest espai.

A continuació es mostren les captures de pantalla dels diferents informes creats per mostrar els diferents indicadors demanats, així com una explicació de com han estat dissenyats:

Informe 1: Capacitat total vs capacitat actual en Hm3 i en Ml per riu i any-mes.

Aquest informe mostra la comparativa entre la suma de la capacitat actual que tenen els embassaments als que abasteix un mateix riu i la suma de la capacitat total que poden tenir aquests embassaments. L'informe s'ha dissenyat en forma de matriu, a la part esquerra es poden veure els diferents rius, en la part superior es pot escollir el mes i l'any del que es vol generar l'informe. Els resultats apareixen en la part central tant en Hm3 com en Ml. Per fer aquest informe només s'han fet servir els càlculs per mostrar els resultats en Ml.

Capacitat total vs capacitat actual en Hm3 i en Ml per riu i any-mes				
Elementos de Página: Mes: 07 Julio ▼ Any: 2001 ▼				
Riu	Capacitat Actual Hm3	Capacitat Actual Ml	Capacitat Total Hm3	Capacitat Total Ml
AGUAS LIMPIAS	16	16.000	18	18.000
AGUAS VIVAS	1	1.000	8	8.000
ALBERCOS	18	18.000	33	33.000
ALBINA	4	4.000	6	6.000
ARAGÓN	223	223.000	470	470.000
ARANDA	14	14.000	21	21.000
ARBA	0	0	6	6.000
ARGA	17	17.000	21	21.000
ASMAT	3	3.000	10	10.000
BARRANCO DE IP	5	5.000	5	5.000
CALDARÉS	4	4.000	5	5.000
CALDERÉS	6	6.000	7	7.000
CERTESCÁNS	12	12.000	16	16.000
CINCA	713	713.000	829	829.000
CIURANA	4	4.000	12	12.000
EBRO	1.806	1.806.000	2.304	2.304.000
ESCARRA	2	2.000	5	5.000
ÉSERA	65	65.000	92	92.000
FLAMISEL	0	0	11	11.000
FLAMISELL	11	11.000	21	21.000
FLUMEN	2	2.000	13	13.000
GÁLLEGO	69	69.000	96	96.000
GUADALOPE	143	143.000	197	197.000
GUATIZALEMA	12	12.000	16	16.000
HUERVA	1	1.000	7	7.000
IRATI	0	0	14	14.000
LLAUSET	14	14.000	17	17.000
MARTÍN	2	2.000	29	29.000
NAJERILLA	36	36.000	68	68.000
NÁJIMA	6	6.000	9	9.000
NOG. PALLARESA	235	235.000	341	341.000
NOG. RIBAGORZANA	882	882.000	1.098	1.098.000
NOGUERA DE TORT	13	13.000	16	16.000
PEGUERA	6	6.000	6	6.000
PENA	15	15.000	19	19.000
PIEDRA	30	30.000	79	79.000
PENA	15	15.000	19	19.000
PIEDRA	30	30.000	79	79.000
SALADO	37	37.000	65	65.000
SANTA ENGRACIA	52	52.000	72	72.000
SEGRE	74	74.000	101	101.000
SOTÓN	101	101.000	189	189.000
URDICETO	2	2.000	5	5.000
ZADORRA	110	110.000	147	147.000

Figura 14: Informe 1

Informe 2: Capacitat total vs capacitat actual en Hm3 i en Ml per embassament any-mes.

Aquest informe mostra la comparativa entre la capacitat actual que té un embassament i la capacitat total que pot tenir l'embassament. L'informe s'ha dissenyat en forma de matriu, a la part esquerra es poden veure els diferents embassaments, en la part superior es pot escollir el mes i l'any del que es vol generar l'informe. Els resultats apareixen en la part central tant en Hm3 com en Ml. Per fer aquest informe només s'han fet servir els càlculs per mostrar els resultats en Ml.

Capacitat total vs capacitat actual en Hm3 i en Ml per embassament any-mes					
Elementos de Página: Mes: 07 Julio Any: 2001					
Embassament	Capacitat Actual Hm3	Capacitat Actual Ml	Capacitat Total Hm3	Capacitat Total Ml	
ALBINA	4	4.000	6	6.000	
ALCAÑIZ (ESTANCA)	6	6.000	7	7.000	
ALLOZ	37	37.000	65	65.000	
BACHIMAÑA (LAGO)	6	6.000	7	7.000	
BARASONA	65	65.000	92	92.000	
BASERCA	20	20.000	22	22.000	
BRAMATUERO ALTO	4	4.000	5	5.000	
BÚBAL	39	39.000	64	64.000	
CALANDA	36	36.000	54	54.000	
CAMARASA	64	64.000	113	113.000	
CANELLES	560	560.000	688	688.000	
CASPE	63	63.000	82	82.000	
CAVALLERS	13	13.000	16	16.000	
CERTESCÀNS	12	12.000	16	16.000	
CIURANA	4	4.000	12	12.000	
CUEVA FORADADA	2	2.000	29	29.000	
EBRO	353	353.000	540	540.000	
ESCALES	137	137.000	152	152.000	
ESCARRA	2	2.000	5	5.000	
EUGUI	17	17.000	21	21.000	
GONZALEZ LACASA	18	18.000	33	33.000	
GRADO (EL)	377	377.000	399	399.000	
GUIAMETS	3	3.000	10	10.000	
IP	5	5.000	5	5.000	
IRABIA	0	0	14	14.000	
LANUZA	16	16.000	17	17.000	
LLAuset	14	14.000	17	17.000	
MAIDEVERA	14	14.000	21	21.000	
MANSILLA	36	36.000	68	68.000	
MAR (LAGO)	6	6.000	14	14.000	
MEDIANO	336	336.000	430	430.000	
MEQUINENZA	1.228	1.228.000	1.534	1.534.000	
MONEVA	1	1.000	8	8.000	
MONTEAGUDO	6	6.000	9	9.000	

Figura 15: Informe 2 Part 1

NEGRO	6	6.000	6	6.000
OLIANA	74	74.000	101	101.000
PENA	15	15.000	19	19.000
PEÑA (LA)	14	14.000	15	15.000
RESPOMUSO	16	16.000	18	18.000
RIBARROJA	205	205.000	210	210.000
SABURO (LAGO)	0	0	11	11.000
SAN BARTOLOMÉ	0	0	6	6.000
SANTA ANA	165	165.000	236	236.000
SANTOLEA	38	38.000	54	54.000
S.MARÍA DE BELSUÉ	2	2.000	13	13.000
SOBRÓN	20	20.000	20	20.000
SOTONERA	101	101.000	189	189.000
TERRADETS	19	19.000	23	23.000
TORCAS (LAS)	1	1.000	7	7.000
TORT (LAGO)	5	5.000	7	7.000
TRANQUERA (LA)	30	30.000	79	79.000
TREMP O TALÁRN	152	152.000	205	205.000
ULLIVARRI	110	110.000	147	147.000
URDICETO	2	2.000	5	5.000
URRUNAGA	52	52.000	72	72.000
VADIELLO	12	12.000	16	16.000
YESA	223	223.000	470	470.000

Figura 16: Informe 2 Part 2

Informe 3: Capacitat total vs capacitat actual en Hm3 i en Ml per comunitat autònoma i any-mes.

Aquest informe mostra la comparativa entre la suma de la capacitat actual que tenen els embassaments d'una mateixa comunitat autònoma i la suma de la capacitat total que poden tenir aquests embassaments. L'informe s'ha dissenyat en forma de matriu, a la part esquerra es poden veure les diferents comunitats autònomes, en la part superior es pot escollir el mes i l'any del que es vol generar l'informe. Els resultats apareixen en la part central tant en Hm3 com en Ml. Per fer aquest informe només s'han fet servir els càlculs per mostrar els resultats en Ml.

Capacitat total vs capacitat actual en Hm3 i en Ml per comunitat autònoma i any-mes				
Elementos de Página: Mes: 07 Julio Any: 2001				
Comunitat	Capacitat Actual Hm3	Capacitat Actual Ml	Capacitat Total Hm3	Capacitat Total Ml
ARAGÓN	1.319	1.319.000	1.816	1.816.000
BUIDA	2.680	2.678.000	3.371	3.366.000
CANTABRIA	353	353.000	540	540.000
CASTILLA LEON	6	6.000	9	9.000
CATALUÑA	77	77.000	111	111.000
LA RIOJA	54	54.000	101	101.000
NAVARRA	277	277.000	556	556.000

Figura 17: Informe 3

Informe 4: Comparativa anual de l'evolució per comunitat autònoma de la capacitat agregada de tots els seus embassaments.

Aquest informe mostra l'evolució de la suma de la capacitat actual que tenen els embassaments d'una mateixa comunitat autònoma en els diferents anys dels que es tenen mesures. L'informe s'ha dissenyat en forma de matriu, a la part esquerra es poden veure les diferents comunitats autònomes. Els resultats apareixen en la part central per any, tant en Hm3 com en Ml. Per fer aquest informe només s'han fet servir el càlcul per mostrar els resultats en Ml.

Comparativa anual de l'evolució per comunitat autònoma de la capacitat agregada de tots els seus embassaments											
Elementos de Página:											
Comunitat	Any	2001		2002		2003		2004		2005	
		Capacitat Hm3	Capacitat Ml	Capacitat Hm3	Capacitat Ml	Capacitat Hm3	Capacitat Ml	Capacitat Hm3	Capacitat Ml	Capacitat Hm3	Capacitat Ml
ARAGÓN		6.606	5.807.000	13.223	11.934.000	16.594	13.764.000	14.476	12.751.000	10.361	9.069.000
BUIDA		14.416	13.541.000	29.712	26.250.000	32.055	29.569.000	28.825	26.798.000	26.783	23.491.000
CANTABRIA		1.348	1.348.000	2.381	2.381.000	3.367	3.203.000	3.221	3.221.000	4.425	4.425.000
CASTILLA LEON		32	11.000	58	15.000	53	9.000	74	18.000	83	21.000
CATALUÑA		297	254.000	872	755.000	902	783.000	832	676.000	760	627.000
LA RIOJA		172	159.000	567	511.000	828	732.000	775	616.000	716	660.000
NAVARRA		811	793.000	2.410	2.310.000	4.442	4.170.000	3.662	3.559.000	3.003	2.919.000

Figura 18: Informe 4

Informe 5: 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).

Aquest informe mostra la comparativa de la suma de la capacitat actual mitjana que tenen els embassaments d'una mateixa comunitat autònoma en un any respecte a la que tenien l'any anterior. També mostra quin increment hi ha hagut d'un any a l'altre i el percentatge d'aquest increment, en cas que l'increment tingui un valor negatiu indica que ha disminuït la capacitat agregada respecte a l'any anterior. L'informe s'ha dissenyat en forma de matriu, a la part esquerra es poden veure les diferents comunitats autònomes i a la part superior es pot escollir l'any del que es vol generar l'informe. Els resultats apareixen en la part central, tant en Hm3 com en Ml. Per fer aquest informe s'han fet servir els càlculs per mostrar els resultats en Ml, també s'ha fet servir un càlcul per mostrar l'increment i un altre càlcul per mostrar el percentatge d'increment. A més s'ha establert una condició que fa que no es tinguin en compte els resultats que tinguin la capacitat agregada anterior igual a 0. S'ha decidit establir aquesta condició degut a que aquests resultats provocaven un error de divisió per 0, i que no són rellevants per la comparativa, ja que el fet que la mitjana agregada sigui 0 indica que no hi han dades d'aquesta comunitat en l'any anterior i per tant no te sentit la comparativa.

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

Elementos de Página: Any: 2001 ▼

Comunitat	Cap. Mitjana Agregada Ant. Hm3	Cap. Mitjana Agregada Ant. Ml	Cap. Mitjana Agregada Act. Hm3	Cap. Mitjana Agregada Act. Ml	Increment Hm3	Increment Ml	% Increment
ARAGÓN	49,69	49.692,98	58	57947,37	8,25	8.254,39	16,61%
BUIDA	89,29	89.285,71	86	85785,71	-3,50	-3.500,00	-3,92%
CANTABRIA	308,50	308.500,00	225	224666,67	-83,83	-83.833,33	-27,17%
CASTILLA LEON	6,17	6.166,67	5	5333,33	-0,83	-833,33	-13,51%
CATALUÑA	33,58	33.583,33	25	24750,00	-8,83	-8.833,33	-26,30%
LA RIOJA	22,58	22.583,33	14	14333,33	-8,25	-8.250,00	-36,53%
NAVARRA	81,41	81.411,76	47	47294,12	-34,12	-34.117,65	-41,91%

Figura 19: Informe 5

Informe 6: Aportació (Entrada-Sortida) en Hm3 i en Ml per riu, embassament i anys; comparativa amb la mitjana anual d'aportació de l'embassament.

Aquest informe mostra una comparativa entre l'aportació d'aigua per any i mes i la mitjana anual d'aportació de l'embassament, es a dir la diferencia entre l'entrada i sortida de cada mes i la diferencia mitjana entre la suma de les entrades de l'any i les sortides. L'informe s'ha dissenyat en forma de matriu, a la part esquerra es poden veure els diferents rius i els seus corresponents embassaments, i a la part superior es pot escollir l'any del que es vol generar l'informe. Els resultats apareixen en la part central, tant en Hm3 com en Ml, es pot veure l'aportació dels diferents mesos de l'any i la mitjana d'aportació anual. Per fer aquest informe s'ha fet servir un càlcul per mostrar l'aportació en Hm3 i un altre càlcul per l'aportació en Ml. Per mostrar l'aportació mitjana s'han creat dos totals que calculen la mitjana, un per mostrar el resultat en Hm3 i l'altre per Ml.

S'ha decidit fer servir una condició per tal que no apareguin a l'informe els embassaments que tinguin una mitjana anual d'aportació igual a 0, ja que aquestos casos són deguts a la falta de dades d'aquests embassaments i no aporten informació a aquest informe.

La captura de pantalla és una mostra de les dades, ja que l'extensió de l'informe fa difícil mostrar els resultats complerts. En la captura de pantalla es mostren només alguns mesos de l'any i la mitjana anual per una mostra de rius i embassaments.

Aportació (Entrada-Sortida) en Hm3 i en Ml per riu, embassament i any-mes; comparativa amb la mitjana anual d'aportació de l'embassament								
Elementos de Página:		Any: 2001 ▼						
Riu	Embassament	Mes	11 Noviembre		12 Diciembre		Mitjana Anual d'Aportació Hm3	Mitjana Anual d'Aportació Ml
			Aportació Hm3	Aportació Ml	Aportació Hm3	Aportació Ml		
ALBERCOS			-0,10	-100,00	-0,40	-400,00	-0,85	-850,00
GONZALEZ LACASA			-0,10	-100,00	-0,40	-400,00	-0,85	-850,00
ARAGÓN			2,10	2100,00	-6,80	-6800,00	-6,32	-6316,67
YESA			2,10	2100,00	-6,80	-6800,00	-6,32	-6316,67
ARANDA			0,10	100,00	0,10	100,00	-0,33	-333,33
MAIDEVERA			0,10	100,00	0,10	100,00	-0,33	-333,33
ARGA					-0,20	-200,00	-0,44	-440,00
EUGUI					-0,20	-200,00	-0,44	-440,00
ASMAT			0,10	100,00			0,10	100,00
GUIAMETS			0,10	100,00			0,10	100,00
CINCA			-0,70	-700,00	-6,00	-6000,00	-2,55	-2550,00
GRADO (EL)			-0,10	-100,00	-4,10	-4100,00	-1,45	-1450,00
MEDIANO			-0,60	-600,00	-1,90	-1900,00	-1,10	-1100,00
CIURANA			0,20	200,00			0,20	200,00
CIURANA			0,20	200,00			0,20	200,00
EBRO			42,70	42700,00	0,00	0,00	-8,90	-8900,00
EBRO			5,10	5100,00	-0,40	-400,00	-9,78	-9783,33
MEQUINENZA			41,80	41800,00	-3,50	-3500,00	-1,63	-1633,33
RIBARROJA			-4,20	-4200,00	3,90	3900,00	2,52	2516,67
ÉSERA			0,50	500,00	-0,80	-800,00	-0,82	-816,67
BARASONA			0,50	500,00	-0,80	-800,00	-0,82	-816,67
GÁLLEGO			-1,70	-1700,00	-2,40	-2400,00	-0,05	-50,00
BÚBAL			-1,10	-1100,00	0,30	300,00	-0,60	-600,00
LANUZA					-1,20	-1200,00	0,00	0,00
PEÑA (LA)			-0,60	-600,00	-1,50	-1500,00	0,55	550,00
GUADALOPE			0,20	200,00	0,30	300,00	-1,33	-1333,33
ALCAÑIZ (ESTANCA)							0,13	133,33
CALANDA					0,10	100,00	-1,15	-1150,00
CASPE			-0,20	-200,00	-0,10	-100,00	-0,47	-466,67
SANTOLEA			0,40	400,00	0,30	300,00	-0,17	-166,67
GUATIZALEMA			-0,10	-100,00	-0,10	-100,00	-0,26	-260,00
VADIELLO			-0,10	-100,00	-0,10	-100,00	-0,26	-260,00

Figura 20: Informe 6

Informe 7: Aigua diària en reserva disponible (Hm3/dia i Ml/dia) a tota la confederació hidrogràfica per any-mes.

Aquest informe mostra per dia la suma de la capacitat actual de tots els embassaments de la confederació hidrogràfica en un mes i any concret.

L'informe s'ha dissenyat en forma de matriu, a la part esquerra es poden veure els diferents anys, i a la part superior es pot escollir el mes del que es vol generar l'informe. Els resultats apareixen en la part central, tant en Hm3 com en Ml. Per fer aquest informe s'ha fet servir un càlcul per mostrar l'aigua diària disponible en Hm3 i un altre càlcul per mostrar-la en Ml.

Aigua diària en reserva disponible (Hm ³ /dia i Ml/dia) a tota la confederació hidrogràfica per any-mes			
Elementos de Página:		Mes: 07 Julio ▼	
Any		Aigua Diària Disponible Hm ³	Aigua Diària Disponible Ml
2001		158,87	158.866,67
2002		131,47	131.466,67
2003		151,40	151.400,00
2004		167,70	167.700,00
2005		123,60	123.600,00

Figura 21: Informe 7

Informe 8: Mostrar un gràfic de l'evolució dels Hm³ i Ml d'aigua de neu equivalent prevista per any i embassament. Mostreu, si n'hi ha, els casos de desbordament de l'embassament.

Aquest informe mostra l'evolució de la previsió d'aigua de neu equivalent per any i embassament, amb les dades obtingudes es genera un gràfic.

L'informe s'ha dissenyat en forma de matriu, a la part esquerra es poden veure els diferents anys, i a la part superior es pot escollir l'embassament del que es vol generar l'informe. Els resultats apareixen en la part central.

Per mostrar la previsió d'aigua de neu equivalent s'ha creat un càlcul amb la fórmula indicada a l'enunciat.

Per mostrar els casos de desbordament d'un embassament, s'han fet servir un càlcul que mostra un valor negatiu en cas que no hi hagi desbordament i un valor positiu ressaltat en vermell en cas que hi hagi desbordament per aquest embassament en un any concret. Aquest càlcul s'ha fet sumant la capacitat mitjana de l'embassament en aquest any més l'estimació d'aigua de neu per aquest any, al valor obtingut s'ha restat la capacitat total de l'embassament, per així poder determinar si s'ha pogut superar la capacitat total.

Per facilitar la claredat de les dades representades al gràfic, s'ha dividit l'informe en dos fulls, un pel càlcul en Hm³ i un altre pel càlcul en Ml.

Evolució dels Hm3 d'aigua de neu equivalent prevista per any i embassament (Desbordaments ressaltats en vermell)

Elementos de Pàgina: Embassament: MAIDEVERA ▾

Any	Aigua Neu Equivalent Hm3	Desbordament Hm3
2001	10,00	-6,00
2002	6,93	-8,33
2003	13,73	0,17
2004	17,64	1,64
2005	8,25	-4,25

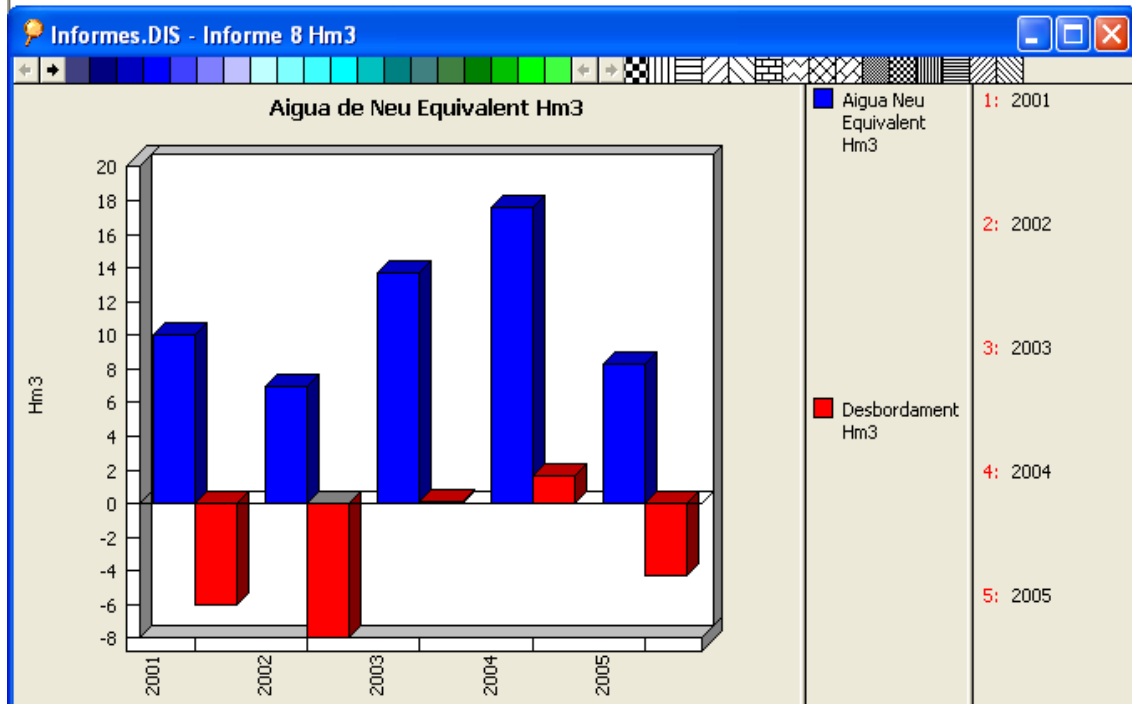


Figura 22: Informe 8 en Hm3

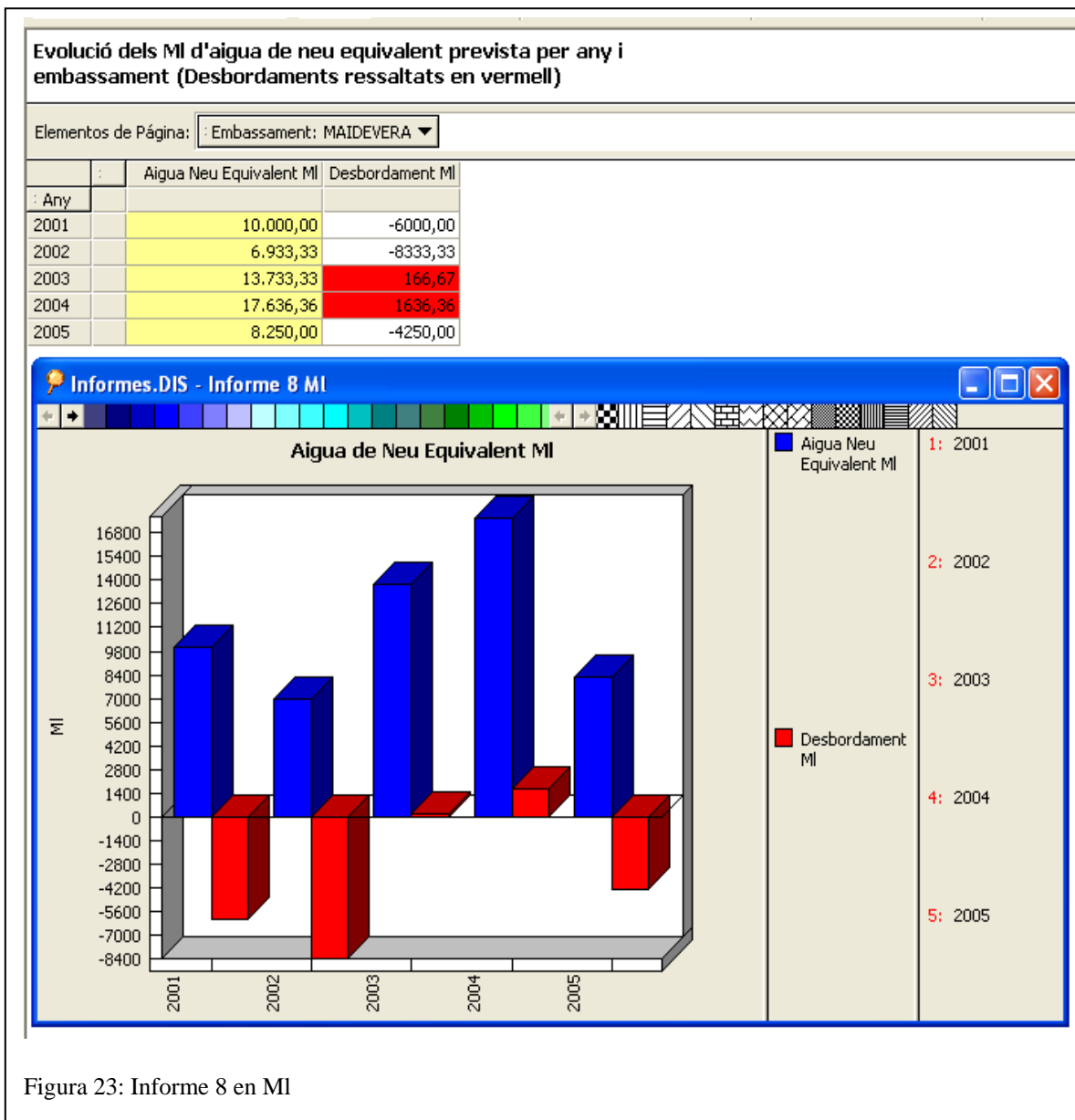
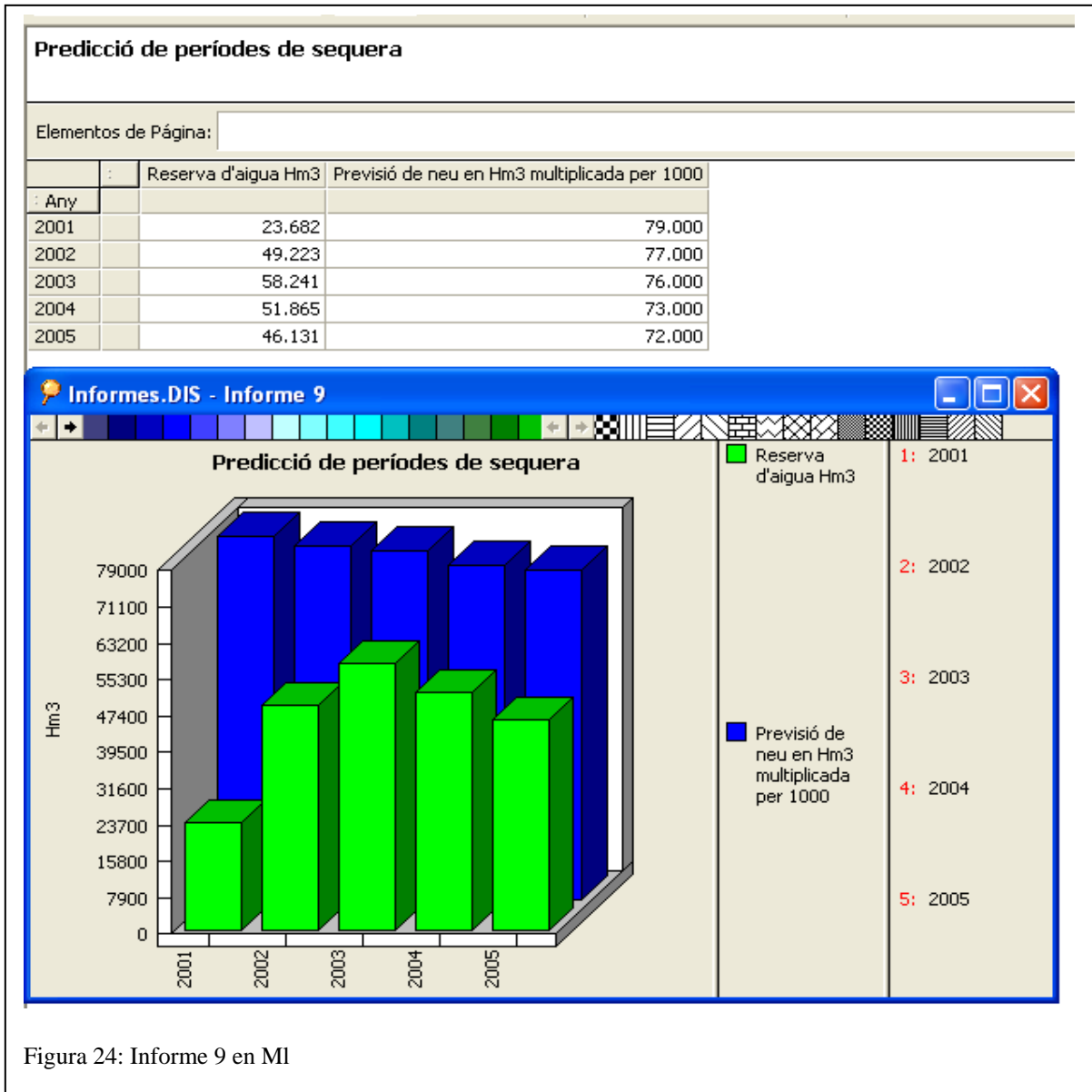


Figura 23: Informe 8 en MI

Informe 9: Mostrar com podem predir, a partir del data mart desenvolupat, períodes de sequera.

Estudiant les dades contingudes al magatzem de dades creat és possible predir períodes de sequera. Amb l'ajuda d'informes creats amb Oracle Discoverer és possible estudiar la tendència de l'estat dels diferents embassaments comparant diferents anys. Per exemple, amb les dades proporcionades dels anys entre el 2001 i 2005, s'ha creat un informe que compara en els diferents anys l'estimació d'aigua de neu i la reserva d'aigua de tota la confederació hidrogràfica. Amb aquestes dades es pot veure quina és la tendència en un període, si tant la reserva d'aigua com l'estimació de neu tendeixen a disminuir es pot predir un possible període de sequera. Per il·lustra-ho també s'ha creat un gràfic:



A les dades recollides en l'informe i el gràfic creat es pot observar que l'estimació d'aigua de neu dels diferents anys és força constant, però té una tendència a disminuir. El que sí que es veu més clar és que entre l'any 2001 i 2002 es surt d'un període de sequera ja que la reserva d'aigua de la confederació hidrogràfica augmenta considerablement a l'any 2002 en relació a l'any anterior. A partir de l'any 2003 la tendència en canvi és a disminuir la reserva d'aigua cada any, això amb combinació amb la tendència a la disminució de l'aigua de neu, fa que es pugui predir que es produirà un període de sequera.

Les dades de l'estimació d'aigua de neu s'han multiplicat per 1000 per poder representar-les en el mateix gràfic que les dades de la reserva d'aigua.

3.7. Accés al magatzem de dades

En aquest apartat es descriuen les diferents instruccions necessàries per accedir al magatzem de dades.

3.7.1 Accés a la base de dades

Per la creació de la base de dades del magatzem de dades s'ha creat un usuari nou anomenat **CHNE** que té com a password **uoc**. Amb aquest usuari s'han creat les diferents taules i els procediments necessaris.

Per accedir a la base de dades es pot fer mitjançant l'accés directe a la pàgina inicial de la base de dades d'Oracle que hi ha a l'escriptori de la màquina virtual i introduint l'usuari i el password que anteriorment s'han indicat.

També es possible accedir a la base de dades mitjançant el programa SQLDeveloper, aquest programa es troba a E:\TFC\sqldeveloper. En aquest cas la connexió a la base de dades ja està configurada i només cal desplegar la connexió CHNE per accedir a la base de dades.

3.7.2 Accés a Oracle Discoverer

Tant per accedir a l'Oracle Discoverer Administrator com a l'Oracle Discoverer Desktop, s'ha d'introduir les següents dades de connexió:

Usuario: **CHNE**

Contraseña: **uoc**

Conectar: **//localhost:1521/XE**

Els informes estan guardats a la base de dades. Per obrir els informes s'ha de seleccionar l'opció "Abrir un libro de trabajo existente", seleccionar després l'opció "Base de Datos" i seleccionar el llibre que apareix anomenat Informes.DIS.

3.8 Processos de càrrega

Per realitzar la càrrega de les dades proporcionades al magatzem de dades, s'han creat diferents processos ETL que carreguen les dades a taules temporals de la base de dades, les transformen i finalment les carreguen a les taules definitives.

Per automatitzar aquest procés s'ha creat un directori anomenat CHNE a E:\TFC, aquest directori conté d'una banda tots els arxius de dades proporcionats en format CSV i d'altra banda diferents fitxers encarregats de la seva càrrega i transformació.

A continuació s'enumeren els diferents fitxers que hi ha al directori i la seva funció:

Carregar_Dades.bat: És l'arxiu executable que conte totes les sentències i crides encarregades de carregar i tractar les dades contingudes als fitxers CSV.

loader1.ctf, loader2.ctf i loader3.ctf: Són els fitxers de control que fa servir la crida de SQLLOADER amb les instruccions adients en cada cas.

esborra_tablesTemp.sql: Conté la crida al procediment encarregat d'esborrar les taules temporals abans d'afegir noves dades.

insert_tempestadistiques.sql: Conté la crida al procediment encarregat de tractar cada fitxer de dades estadístiques dels embassaments carregats.

transformacio_carrega.sql: Conté la crida als procediments que un cop carregades les dades a les taules temporals s'encarreguen de transformar les dades i carregar-les a les taules definitives.

Aquest directori també conté arxius que crea SQLLOADER al carregar les dades, són els fitxers .log que recullen el resum del resultat de la càrrega i els fitxers .bad que recullen els registres que no s'han carregat.

Per iniciar el procés de càrrega i transformació de les dades, primer s'ha de configurar el fitxer executable **Carregar_Dades.bat**, aquest fitxer ha de contenir una instrucció com "sqlldr CHNE/uoc control=loader2.ctf data=eXXXXXXXXX.csv" per cada fitxer de dades d'estadístiques, on només s'ha de modificar el nom del fitxer de dades.

Per als fitxers dels embassaments i l'estimació d'aigua de neu no cal configurar els fitxers sempre que es mantingui el seu nom i estructura.

Un cop configurat només cal executar el fitxer **Carregar_Dades.bat** per posar en marxa el procés, quan acaben tots els processos es tanca automàticament la finestra de la línia de comandes.

4. Conclusions

Una vegada acabat tot el procés de creació i explotació del magatzem de dades, s'ha arribat a les següents conclusions:

- S'ha aconseguit conèixer els aspectes teòrics del disseny d'un magatzem de dades, aquests coneixements s'han posat en pràctica en l'anàlisi de requeriments i el disseny del magatzem de dades realitzat, donant lloc a una base de dades orientada a l'explotació de les dades mitjançant una eina de Business Intelligence.
- S'ha aconseguit conèixer l'ús del motor de bases de dades Oracle 10g Express per crear l'estructura del magatzem de dades, i també dominar les bases del seu llenguatge de programació PL/SQL, mitjançant el qual s'han pogut crear els diferents procediments emmagatzemats necessaris pel correcte funcionament del sistema proposat.
- S'ha dissenyat una sèrie de processos ETL per tal d'automatitzar l'extracció, transformació i càrrega de les dades proporcionades. Aquests processos permeten tractar les dades proporcionades pel client i carregar-les al magatzem de dades per la seva posterior explotació. Tots aquests passos es realitzen automàticament i de forma transparent a l'usuari encarregat de proporcionar les dades a carregar, això permet que aquesta tasca la pugui fer un usuari sense un nivell avançat de coneixements en el magatzem de dades, ja que es possible fer la càrrega seguint uns senzills passos.
- S'ha començat a conèixer l'ús de l'eina de Business Intelligence Oracle Discoverer, mitjançant la qual s'han creat una sèrie d'informes dissenyats per mostrar la informació demanada a l'enunciat, a partir de les dades contingudes al magatzem de dades. Aquests informes permetran als usuaris realitzar les diferents consultes demandes i obtenir la informació de forma senzilla i en un entorn gràfic que presenta les dades de forma adient.
- La planificació i estudi previ del projecte establert s'ha pogut complir sense pràcticament cap desviació. Això reflecteix una correcta planificació de les tasques a realitzar i de la càrrega de treball de cada una d'elles.

Com a conclusió final es pot dir que s'han complert els diferents objectius proposats inicialment, destacant l'objectiu principal que era donar solució al problema plantejat a l'enunciat. Com a producte final s'entrega un sistema de magatzem de dades que compleix amb els requeriments demanats pel client i que permetrà realitzar un seguiment de l'estat dels diferents recursos hidrològics de forma més eficient i facilitarà la presa de decisions.

A nivell personal, la realització del projecte m'ha permès començar a conèixer aquest tipus de bases de dades orientades a l'explotació de dades i poder veure la diferència amb les bases de dades transaccionals amb les que he pogut treballar en altres assignatures. D'altra banda el projecte m'ha permès posar en pràctica molts dels coneixements adquirits durant els estudis realitzats, això hem provoca gran satisfacció al comprovar com són d'útils a la pràctica aquests coneixements, que tants sacrificis han suposat adquirir-los durant aquests anys.

A més, crec que aquests nous coneixements adquirits hem poden ser de gran utilitat en un futur professional, degut a la forta aposta que estan fent les organitzacions per aquest tipus de sistemes.

5. Línies d'evolució futures

En aquest apartat es tractaran les possibles millores que es podrien realitzar en el sistema i també el diferents problemes trobats i que s'haurien de solucionar en un futur.

Per tal d'evitar els problemes trobats a les fonts de dades, detallats en altres capítols de la memòria, s'hauria d'homogeneïtzar i completar les diferents fonts de dades, per així facilitar la càrrega d'aquestes dades i tenir una informació més detallada recollida al magatzem de dades. Per aquest motiu es recomana al client que es revisin, corregeixin i actualitzin les fonts de dades.

Com a resum dels problemes trobats amb les fonts de dades, i amb l'objectiu de servir com a guia per la correcció i actualització de les dades, es destaca el següent:

- No tots els embassaments que apareixen als fitxers d'estadístiques es troben al fitxer d'embassaments, per tant no es pot assignar a tots la seva comunitat autònoma i província.
- Alguns noms d'embassaments o rius difereixen entre els arxius d'estadístiques, el d'embassaments i les dades de l'estimació d'aigua de neu.
- Les dades de l'estimació d'aigua de neu no es proporcionen en el mateix format que la resta de fonts de dades.

Un aspecte que es podria millorar en un futur en el magatzem de dades és el control d'errors en el procés de transformació i càrrega de les dades. En el disseny d'aquest procés s'ha tingut en compte el control de possibles errors continguts a les fonts de dades proporcionades i s'ha comprovat el correcte funcionament dels diferents processos amb aquestes dades. Encara que s'ha tingut en compte una sèrie de possibles errors, al incorporar més dades al sistema podria fer aparèixer nous casos no contemplats en el tractament d'errors actual. Per aquest motiu es recomana fer un seguiment d'aquests possibles nous errors i també que, dintre del possible, s'intenti no fer canvis en l'estructura i el format de les fonts de dades.

En un futur també es podria plantejar fer canvis al magatzem de dades per millorar el rendiment d'algunes consultes. En aquest moment no representa un problema ja que amb el volum de dades que hi ha actualment, els informes s'executen en un temps raonable. Però a mig termini, si el volum de dades creix molt, s'hauria de plantejar optimitzar el magatzem de dades prenent mesures com poden ser ajustar una sèrie de paràmetres del sistema, o bé definir índex per a millorar els accessos a les taules.

6. Glossari

Business Intelligence: Conjunt de metodologies, aplicacions i tecnologies que permeten reunir, depurar i transformar dades dels sistemes transaccionals e informació des estructurada en informació estructurada, per la seva explotació directa o pel seu anàlisi i conversió en coneixement, ajudant així a la presa de decisions.

CHNE: La Confederació Hidrogràfica Nord-Est, es una entitat encarregada de realitzar el seguiment de l'estat dels diferents embassaments de la seva responsabilitat.

Data Warehouse: Denominació en anglès de Magatzem de dades.

Dimensió: Component del model multidimensional que representa un punt de vista que s'utilitzarà en l'anàlisi de les dades.

ETL: Sigles amb anglès d'extracció, transformació i càrrega. És el procés que permet obtenir dades de diferents fonts i sistemes, tractar-les i carregar-les a la base de dades.

Magatzem de dades: És una col·lecció de dades orientades a un determinat àmbit, que estan emmagatzemades en una base de dades dissenyada per afavorir la seva anàlisi. L'objectiu d'aquests sistemes és ajudar en la presa de decisions a les organitzacions.

Model Multidimensional: Tipus de disseny de base de dades que consisteix a concebre les dades que volem analitzar en termes de fets i dimensions d'anàlisi, de manera que les podem situar en un espai n-dimensional.

Oracle 10g Express: Versió gratuïta del gestor de base de dades relacional desenvolupat per Oracle Corporation.

PL/SQL: Llenguatge de programació propi d'ORACLE, que suporta totes les consultes i manipulació de dades que es fan servir en SQL, i a més inclou característiques com maneig de variables, estructures modulars, control de flux , control d'excepcions, etc..

Oracle Discoverer: Conjunt d'eines orientades a l'anàlisi i l'explotació de les dades.

7. Bibliografia

En aquest apartat s'enumeren les fonts d'informació consultades per la realització del projecte.

7.1 Publicacions

- Àngels Rius Gavídia, Montse Serra Vizern. *Magatzems de dades i models multidimensionals*. Barcelona, Fundació per a la Universitat Oberta de Catalunya, Primera edició: setembre 2003.
- Javier Jofre González-Granda. *PL/SQL Básico*. INNOVA Desarrollos Informáticos, Fecha Edición: Junio de 2004.

7.2 Webs

- Atica, Universidad de Murcia. *SQL - Formatos de Fecha (Obtener nombre del mes)*. [<http://www.um.es/atica/sql---formatos-de-fecha-1>, 16/11/2010].
- Department of Statistics and Computer Information Systems Baruch College, City University of New York. *ORACLE SQL*Plus: An Introduction and Tutorial*. [<http://cisnet.baruch.cuny.edu/holowczak/oracle/sqlplus/>, 19/11/2010].
- Lampharin Solutions. *Instalar los Guest Additions en una maquina virtual con Ubuntu*. [<http://lampharin.wordpress.com/2008/04/22/instalar-los-guest-additions-en-una-maquina-virtual-con-ubuntu/>, 10/11/2010].
- ORACLE. *External Tables Concepts*. [http://download.oracle.com/docs/cd/B10500_01/server.920/a96652/ch11.htm, 18/11/2010].
- ORACLE. *SQL*Plus Documentation*. [http://www.oracle.com/technology/docs/tech/sql_plus/index.html, 19/11/2010].
- ORACLE. *PL/SQL User's Guide and Reference, 10g Release 2 (10.2)*. [http://download.oracle.com/docs/cd/B19306_01/appdev.102/b14261.pdf, 15/11/2010].
- Oracle Faq's. *External Tables*. [<http://www.oraFAQ.com/node/848>, 18/11/2010].
- Oracle Faq's. *SQL*Loader FAQ*. [http://www.oraFAQ.com/wiki/SQL*Loader_FAQ, 18/11/2010].

- Oracleutilities. *SQL Loader*. [<http://www.oracleutilities.com/OSUtil/sqlldr.html>, 20/11/2010].
- Sinnexus. *Business Intelligence*. [http://www.sinnexus.com/business_intelligence/, 4/01/2011].
- Wikipedia. *Extract, transform and load*. [http://es.wikipedia.org/wiki/Extract,_transform_and_load, 20/11/2010].
- Wikipedia. *Hectómetro cúbico*. [http://es.wikipedia.org/wiki/Hect%C3%B3metro_c%C3%ABico, 12/11/2010].