



Memòria del projecte – DW

Creació i explotació d'un magatzem de dades per l'anàlisi de l'estat dels diferents embassaments del la CHNE (Confederació Hidrogràfica del Nord-est).

Alumne: Lluís Gómez de Argila

Estudis: Enginyeria Tècnica de Informàtica de Sistemes

Tutor: Pere Juanola Juanola

Data: 10 de gener de 2011

RESUM

El projecte que es presenta, vol ser una solució a les necessitats de la CHNE de falta de manejabilitat en la informació que actualment disposa i de l'obtenció de informació automàtica.

Per poder homogeneïtzar aquesta informació s'ha cregut adient crear un magatzem de dades (Data Warehouse) dins la plataforma Oracle 10g. La càrrega i l'explotació d'aquestes dades també es faran amb productes d'Oracle (Sqlloader i Oracle Discovery respectivament). Per finalitzar la part referent a dades, la transformació d'aquestes es fa en una altre plataforma (concretament MVS) per motius de agilitat en la programació d'algorismes de transformació de dades.

Respecte la informació a obtenir, aquesta es ceneix a uns models proporcionats en l'enunciat. Hi ha dos tipus de informes que van acompanyats amb uns gràfics per una millor comprensió i visibilitat. La flexibilitat de l'eina, però, fa que no sigui difícil obtenir nous informes a partir de la base de dades creada i de les seves relacions.

Així mateix, aquesta informació es pot obtenir en Hm3 o en milions de litres. Un cop l'usuari es connecta al producte d'explotació de dades (Oracle Discovery) pot escollir en quina magnitud vol fer les consultes.

INDEX DE CONTIGUTS

1.- INTRODUCCIÓ.....	7
1.1.- Definició de Data Warehouse.....	7
1.2.- Justificació del DW com a TFC.....	8
1.3.- Justificació del TFC i context en el qual es desenvolupa: punt de partida i aportació del TFC.....	8
1.4.-Objectius del TFC.....	9
1.5.-Enfocament i mètode seguit.....	9
1.6.-Planificació.....	10
1.6.-Productes obtinguts.....	11
1.7.-Breu descripció dels altres capítols de la memòria.....	12
2.- ANÀLISI.....	13
2.1.- Casos d'us.....	13
2.2.- Fonts de dades.....	15
2.2.1.- <i>Fulls Excel mensuals</i>	15
2.2.2.- <i>Altres fitxers</i>	16
2.2.3.- <i>Estimació de neu</i>	17
2.3.- Disseny conceptual. Model Entitat/Relació.....	18
3.- DISSENY.....	19
3.1.- Diagrama d'arquitectura de software.....	19
3.2.- Disseny de taules.....	20
3.2.1.- <i>Disseny físic</i>	20
3.2.1.1.- <i>Taules a crear</i>	21
3.2.1.2.- <i>Càlcul d'espai</i>	23
3.2.2.- <i>Creació de taules</i>	24
3.3.- Processos ETL(extracció, transformació i càrrega de dades).....	30
3.3.1.- <i>Anàlisi previ de les dades de partida</i>	30
3.3.2.- <i>Càrrega de la taula de rius, comunitats i dates</i>	32
3.3.3.- <i>Creació del fitxer de moviments i càlcul de totals i mitjanes</i>	32
3.3.4.- <i>Càrrega d'embassaments i totals per comunitat</i>	34
3.3.5.- <i>Càrrega de la taula de moviments mensuals</i>	34
3.3.6.- <i>Procés de càrrega de les taules</i>	34
3.3.7.- <i>Altres consideracions</i>	36

3.4.- Obtenció de informes.....	38
3.4.1.- Capacitat total vs capacitat actual per riu i any - mes.....	38
3.4.2.- Capacitat total vs capacitat actual per embassament i any - mes.....	39
3.4.3.- Capacitat total vs capacitat actual per comunitat i any - mes.	40
3.4.4.- Comparativa anual de l'evolució per CA de la capacitat agregada de tots els seus embassaments.....	41
3.4.5.- Comparativa de la capacitat mitjana d'una comunitat entre un any i l'anterior.	42
3.4.6.- Aportació per riu, embassament i any - mes. Comparativa amb la mitjana anual d'aportació de l'embassament.	43
3.4.7.- Aigua diària en reserva disponible a tota la confederació.....	44
3.4.8.- Gràfic de l'evolució d'aigua de neu equivalent prevista per any i embassament.....	45
3.4.9.- Predicció de períodes de sequera.	46
4.- CAPTURES DE PANTALLES.....	47
4.1.-CÀRREGUES DE DADES.....	47
4.1.-ARRANJAMENT DE DADES I CREACIÓ DE TAULES.	48
5.- CONCLUSIONS I LINIES D'EVOLUCIÓ.....	49
6.- GLOSSARI.....	50
7.- BIBLIOGRAFIA.....	51
7.1.- PUBLICACIONS.	51
7.2.- WEBS.	51
8.- ANNEXOS.	52

INDEX DE FIGURES I ILUSTRACIONS

Figura 1: Diagrama de Gantt de planificació.....	10
Figura 2: Taula de planificació per tasques/dates.....	10
Figura 3: Diagrama de casos d'ús per l'usuari final.....	13
Figura 4: Diagrama de casos d'ús per l'usuari administrador final.....	14
Figura 5: Diagrama de casos d'ús per l'usuari analista.....	14
Figura 6: Mostra de fulls Excel amb dades a carregar.....	15
Figura 7: Full Excel d'algunes relacions comunitat/embassament/riu.....	16
Figura 8: Constants pel càlcul d'estimació de neu.....	17
Figura 9: Fórmula pel càlcul d'estimació de neu.....	17
Figura 10: Diagrama conceptual. Model Entitat/Relació.....	18
Figura 11: Diagrama d'arquitectura de software.....	19
Figura 12: Diagrama del disseny físic de la base de dades.....	20
Figura 13: Definició de la taula ANNO.....	21
Figura 14: Definició de la taula FECHA.....	21
Figura 15: Definició de la taula AUTONOMIA.....	21
Figura 16: Definició de la taula RIUS.....	21
Figura 17: Definició de la taula EMBASSAMEN.....	22
Figura 18: Definició de la taula MOVIMENS.....	22
Figura 19: Definició de la taula CAPAANU.....	23
Figura 20: Càlcul de l'espai necessari per la creació de la base de dades.....	23
Figura 21: Script per la creació de taules.....	25
Figura 22: Creació de la taula ANNO.....	25
Figura 23: Creació de la taula FECHA.....	26
Figura 24: Creació de la taula AUTONOMIA.....	26
Figura 25: Creació de la taula RIUS.....	27
Figura 26: Creació de la taula EMBASSAMEN.....	27
Figura 27: Creació de la taula MOVIMENS.....	28
Figura 28: Creació de la taula CAPAANU.....	29
Figura 29: Actuacions a fer en el ETL.....	31
Figura 30: Algorisme pel càlcul de totals per Comunitat.....	33
Figura 31: Algorisme pel càlcul de la mitjana d'aportació per embassament.....	33
Figura 32: Script exemple de carrega de dades a partir d'un fitxer txt.....	35
Figura 33: Script de modificació de la dada 'estimació de neu' a la taula MOVIMENS.....	37

Figura 34: Disseny de l'informe capacitat total vs capacitat actual per riu i any - mes.....	38
Figura 35: Disseny de l'informe capacitat total vs capacitat actual per embassament i Any – mes.....	39
Figura 36: Disseny de l'informe capacitat total vs capacitat actual per Comunitat Autònoma i any – mes.....	40
Figura 37: Disseny de l'informe comparativa anual de l'evolució per CA de la capacitat Agregada de tots els seus embassaments.....	41
Figura 38: Disseny de l'informe comparativa de la capacitat mitjana d'una comunitat entre un any i l'anterior.....	42
Figura 39: Disseny de l'informe aportació per riu, embassament i any - mes. Comparativa amb la mitjana anual d'aportació de l'embassament.....	43
Figura 40: Disseny de l'informe aigua en reserva disponible a tota la confederació.....	44
Figura 41: Disseny de l'informe amb gràfic de l'evolució d'aigua de neu equivalent prevista per any i embassament.....	45
Figura 42: Disseny de l'informe amb gràfic de prediccions de sequera.....	46
Figura 43: Captura de pantalla per la carrega de les taules EMBASSAMEN I MOVIMENS.....	47
Figura 44: Captura de pantalla per la modificació de la taula MOVIMENS.....	48

1.- INTRODUCCIÓ.

El present treball, correspon al lliurament final del treball de fi de carrera (en endavant TFC) de la Enginyeria Tècnica d'Informàtica de Sistemes (Pla 2003) cursat a la UOC.

En aquesta memòria, donarem una visió general de lo que es Data Warehouse (en endavant DW) , i en crearem un a partir d'uns requeriments proporcionats i a partir d'unes dades de càrrega.

1.1.- Definició de Data Warehouse.

Podem definir a un DW com un magatzem de dades amb les següents característiques:

- Orientats a temes: Les dades estan organitzades de tal manera que tots les dades relatives al mateix tema queden unides entre sí.
- Variant en el temps: Els canvis produïts al llarg del temps, han de quedar registrats en el DW perquè els informes que es generin reflecteixin aquests canvis.
- No volàtil: La informació ni es modifica ni s'esborra. Només es consulta i es manté per futures consultes. Periòdicament es van fent carregues a partir de les bases de dades transaccionals que es mantenen diàriament en l'operativa empresarial.
- Integrat: El DW conté les dades de tots els sistemes operacionals de l'organització i aquestes dades han de ser consistents.

A més, els mitjans per obtenir i analitzar aquestes dades, per extreure-les, transformar-les i carregar-les, així com les diferents formes per realitzar la gestió de dades, també són components essencials de un DW.

Per finalitzar aquesta petita introducció i definició d'un DW, apuntarem que la funció principal es donar resposta a les necessitat d'usuaris experts, fent servir Sistemes de Suport a Decisions (DSS), Sistemes de informació executiva (EIS) o eines per fer consultes o informes. Els usuaris finals poden fer fàcilment consultes sobre els magatzems de dades sense modificar o afectar la operativa del sistema.

1.2.- Justificació del DW com a TFC.

Tal com en el seu dia es va raonar, personalment, vaig escollir aquesta àrea pel TFC per quatre motius principals:

- Experiència professional: Encara que no estigui involucrat en cap DW de la empresa, sí que les aplicacions que porto, envien dades a DW's de la corporació.
- Experiència acadèmica: Havent cursat assignatures com Minería de Dades, Intel·ligència Artificial (I i II), vaig creure que lo més relacionat amb aquests temes era el DW.
- Possible sortida professional: He comprovat que són bastant demandats els experts en Business Intelligencie (altre manera d'anomenar el DW).
- Conèixer una nova eina: El poder desenvolupar un producte mitjançant Oracle, també va influir en la meua decisió per escollir el DW com a TFC.

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

Actualment, la “Confederación Hidrogràfica Nord-est” (CHNE) es una entitat que s'encarrega de la gestió de l'aigua i infraestructures hidràuliques en el nord-est de la península. Aquesta àrea, engloba las CC.AA. De La Rioja, Navarra, Aragó i Catalunya. La seva activitat principal es realitzar un seguiment de l'estat dels diferents embassaments, comprovar tendències i predir períodes de sequera.

Actualment, es pot dir que la tasca a fer pels tècnics no està mecanitzada i les dades que venen en fulls de càlcul Excel són tractades mitjançant eines ofimàtiques. Aquesta feina es molt feixuga i crea una despesa de temps empleat excessiva.

L'aportació principal del projecte es la creació d'un magatzem de dades que permeti automatitzar la recollida de la informació provinent dels fulls de càlcul i explotar aquesta informació. La principal explotació d'aquestes dades serà obtenir informes comparatius.

La solució proposada es la creació d'una base de dades que servirà com a Data Warehouse. Aquesta base de dades tindrà una carrega inicial a partir de tots els fulls Excel que tenim.

1.4.-Objectius del TFC.

L'objectiu principal es la creació d'un sistema de emmagatzematge de dades (DW) a partir d'uns requeriments o informes que s'han d'extreure i a partir d'un model de fitxers que ens proporciona un usuari d'una entitat en concret (en aquest cas la CHNE).

Com altres objectius –secundaris- es pot considerar fer una simulació de la creació d'un projecte professional amb una planificació, unes dates d'entrega, unes entregues parcials (PAC's) amb el seu vist-i-plau i/o amb les seves correccions i finalment la posta en marxa i l'acceptació per part de l'usuari del producte final.

1.5.-Enfocament i mètode seguit.

L'enfocament i mètode seguit per a la realització d'aquest projecte es correspon amb el cicle de vida clàssic d'un projecte on el projecte es divideix en etapes. Aquestes etapes, que en un cercle estrictament professional serien aprovades pels usuaris, rectificant en el cas de que haguessin errors, seran aprovades pel tutor i/o fent les indicacions necessàries per la millora del TFC.

Aquestes etapes son entregues de documents (PAC's). Concretament, hi ha les següents entregues:

- PAC1: Planificació del projecte.
- PAC2: Disseny conceptual i físic de la base de la dades.
- PAC3: Creació de la base de dades, processos de extracció, transformació i càrrega i producte final.
- Memòria i presentació virtual: Document final (una síntesi del tres documents anteriors) i un vídeo fent una mica de manual d'usuari.

1.6.-Planificació.

La planificació del treball, s’ha basat en les entregues parcials (PAC’s) a més dels productes a entregar de la memòria, el vídeo i la participació en el debat.

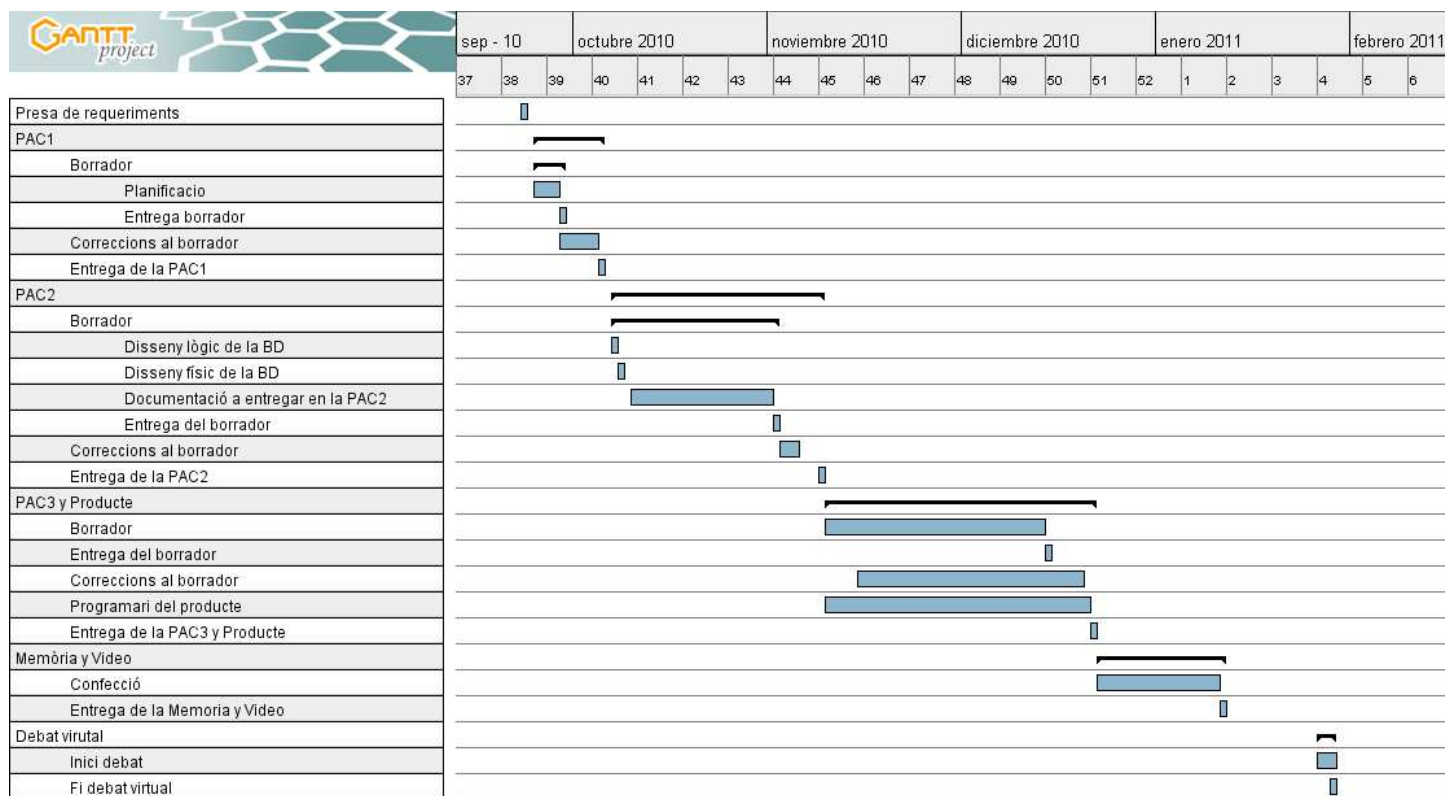


Figura 1: Diagrama de Gantt de planificació.

En un altre format, la planificació seria com segueix :

TITOL	INICI	LLIURAMENT
PAC1	23-09-2010	05-10-2010
PAC2	06-10-2010	08-11-2010
PAC3 I PRODUCTE	09-11-2010	20-12-2010
MEMORIA I VIDEO	21-12-2010	10-01-2011
DEBAT VIRTUAL	24-01-2011	27-01-2011

Figura 2: Taula de planificació per tasques/dates.

1.6.-Productes obtinguts.

- Fitxer per crear les taules en l'entorn Oracle (Annexat a aquest document).
- Un fitxer amb les dades a carregar cada taula creada (en total set fitxers –annexats a aquest document-).
- Uns fitxers de control per utilitzar la utilitat SQLLOADER i carregar físicament les dades dels fitxers del punt anterior a les taules creades.
- Uns fitxers on es veuen els resultats de les carregues fetes.
- Un fitxer per calcular i carregar l'estimació de neu per embassament -any (Annexat a aquest document).
- Tres entregues per fer un seguiment del projecte (PAC's).
- Memòria final (el present document).
- Presentació virtual amb vídeo.
- La relació de informes que es desitja obtenir, que es la següent:
 - Capacitat total vs capacitat actual en Hm3 per riu i any - mes.
 - Capacitat total vs capacitat actual en Hm3 per embassament any - mes.
 - Capacitat total vs capacitat actual en Hm3 per CA any - mes.
 - 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).
 - Aportació (Entrada -Sortida) en Hm3 per riu, embassament i any - mes; 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.
 - Gràfica de l'evolució dels Hm3 d'aigua de neu equivalent prevista per any i embassament. Mostrar, si n'hi ha, els casos de desbordament de l'embassament.
 - Mostrar prediccions de períodes de sequera.

1.7.-Breu descripció dels altres capítols de la memòria.

A continuació, es fa un petit resum dels propers capítols:

- **CAPITOL 2 – ANÀLISI:** Analitza els casos d'us, crea el model conceptual i analitza les dades de les que s'alimentarà la futura base de dades.
- **CAPITOL 3 – DISSENY:** Fa referència al software i el hardware utilitzats, la base de dades dissenyada, tot el que fa referència a l'extracció, transformació i càrrega de dades (processos ETL) i un disseny i descripció dels informes creats.
- **CAPITOL 4 – CAPTURES DE PANTALLES:** Mostra les pantalles més significatives del procés de la creació física de la base de dades i la seva carrega per donar una idea més palpable i no tant teòrica de tot lo explicat en el capítol anterior.
- **CAPITOL 5 – CONCLUSIONS I LINIES D'EVOLUCIÓ:** Com tot projecte, un cop acabat ja es susceptible de fer-li millores o bé de integrar-lo amb altres aplicacions. En aquest capítol s'intentarà de donar una visió de tot lo remarcable i millorable.
- **CAPITOL 6 – GLOSSARI.**
- **CAPITOL 7 – BIBLIOGRAFIA.**
- **CAPITOL 8 – ANNEXOS.**

2.- ANÀLISI.

En aquest punt es farà un anàlisi dels conceptes més importants en que ha de girar el magatzem de dades. Així mateix s'analitzaran les relacions entre les persones i aquests magatzem mitjançant els casos d'us.

Finalment, s'analitzaran les dades que disposem per obtenir –com a mínim- els informes que es sol·liciten.

2.1.- Casos d'us.

Un Diagrama de casos d'us, mostra la relació entre persones físiques i els casos d'ús del sistema. Representa la funcionalitat que ofereix el sistema i la interrelació amb les persones.

Tenint en compte les característiques del producte final del TFC, s'exposen tres diagrames de casos d'us diferents, associats al següents usuaris/actors que interactuaran amb el producte final d'aquest TFC:

- Usuari final.
- Usuari Administrador.
- Usuari Analista.

L'usuari final, serà qui exploti les dades del magatzem i com a molt sol·licitar nous informes

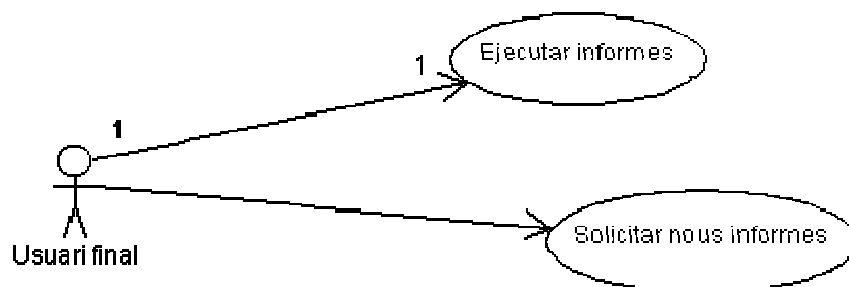


Figura 3: Diagrama de casos d'ús per l'usuari final.

Seguidament, existeix la figura de l'administrador que serà l'encarregat de tot lo intrínsec del sistema (còpies de seguretat, control d'espai dels tablespaces, alta -baixa d'usuaris, etc.).

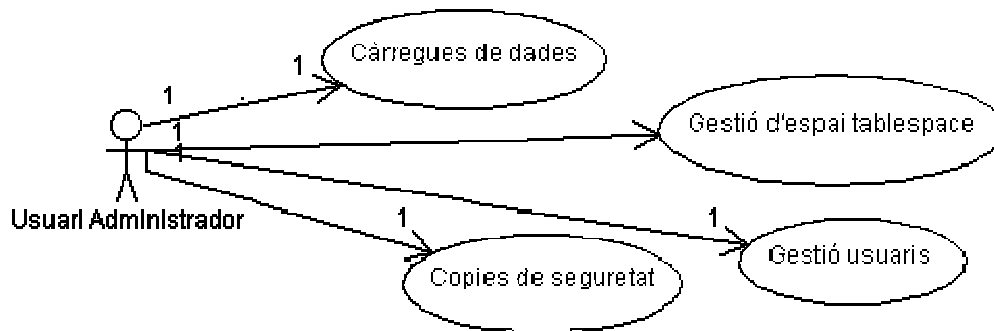


Figura 4: Diagrama de casos d'ús per l'usuari administrador.

Finalment, tenim l'actor de l'analista (que podria ser l'usuari final segons el volum d'usuaris) que faria de pont entre l'administrador del sistema i l'usuari final. La part més important que seria la gestió de nous informes si es poden extreure a partir de la base de dades ja existent.

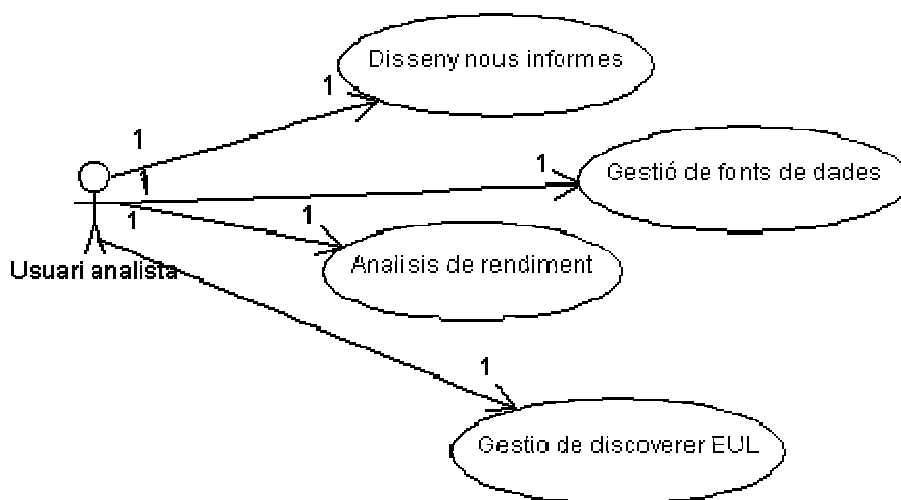


Figura 5: Diagrama de casos d'ús per l'usuari analista.

2.2.- Fonts de dades.

2.2.1.- Fulls Excel mensuals.

SITUACIÓN DE LOS EMBALSES EL 30 DE JULIO DE 2001											
AGUA EMBALSADA						VARIACIÓN SEMANAL			HACE 1 AÑO		
EMBALSES	RÍOS	Capacidad		En esta fecha		Entrada	Salida	Incremento		HACE 1 AÑO	
		Total	En esta fecha	%							%
		Hm³	Hm³		Hm³	Hm³		Hm³	Hm³		
10	ALBINA	ALBINA	6	4	67	---		0		4	67
11	ALCANIZ (ES)	GUADALOPE	7	6	86	1	1	0		4	57
12	ALLOZ	SALADO	65	37	57	0	2,4	-3		30	46
13	BACHIMAÑA	CALDERÉS	7	6	86	---		0		5	71
14	BARASONA	ÉSERA	92	65	71	9,6	18,7	-9		53	58
15	BASERCA	NOG. RIBAG	22	20	91	3,8	3,1	1		19	86
16	BRAMATUER	CALDARÉS	5	4	80	---		0		4	80
17	BÚBAL	GÁLLEGO	64	39	61	4,2	7,3	-3		35	55
18	CALANDA	GUADALOPE	54	36	67	1,8	3,6	-2		13	24
19	CAMARASA	NOG. PALLA	113	64	57	6,9	20	-13		67	59
20	CANELLES	NOG. RIBAG	688	560	81	24	24	0		494	72
21	CASPE	GUADALOPE	82	63	77	0,7	1,9	-2		10	12
22	CAVALLERS	NOGUERA DI	16	13	81	1,5	0,4	1		14	88
23	CERTESCÁN	CERTESCÁN	16	12	75	---		-1		12	75
24	CIURANA	CIURANA	12	4	33	0	0	0		3	25
25	CUEVA FORA	MARTÍN	29	2	7	0,2	0,2	0		4	14
26	EBRO	EBRO	540	353	65	0,9	24,8	-24		387	72
27	ESCALES	NOG. RIBAG	152	137	90	2,7	8,3	-6		128	84
28	ESCARRA	ESCARRA	5	2	40	---		-1		3	60
29	EUGUI	ARGA	21	17	81	0,3	0,8	0		19	90
30	GONZALEZ L	ALBERCOS	33	18	55	0	2,2	-3		20	61
31	GRADO (EL)	CINCA	399	377	94	34,3	42,8	-9		392	98
32	GUIAMETS	ASMAT	10	3	30	0	0	0		1	10
33	IP	BARRANCO I	5	5	100	---		0		3	60

Figura 6: Mostra de fulls excel amb dades a carregar.

Aquests fulls, representen la principal font de informació de la que disposarem per dissenyar i carregar el futur DW. Existeix un full mensual des de juliol de 2001 fins a desembre de 2005 (ambdós inclosos). En un principi també podríem obtenir les dades de juliol del 2000 a juny del 2001 ja que a cada informe tenim la columna de l'any anterior.

A partir d'aquest informe ja podem veure que disposem de la informació d'embassament, riu, capacitat de l'embassament, capacitat en una data en concret i el % que representa sobre la capacitat total, flux (entrada -sortida) d'aigua que hi hagué en el mes i com hem dit en el paràgraf anterior, les dades de l'any anterior de capacitat en una data en concret i el % que representa sobre la capacitat total.

2.2.2.-Altres fitxers..


	A	B	C	D	E	F
1	C. A.	Embalse	Río			
2	ARAGÓN					
3		Pina	Ebro			
4		Ardisa	Gállego			
5		Almochuel	Aguas Vivas			
6		Moneva	Aguas Vivas			
7		Las Torcas	Huerva			
8		Caspe	Guadalope			
9		Tranquera	Piedra			
10	Zaragoza	Maidevera	Aranda			
11		Cienfuens	Flumen			
12		Santa Mª Belsué	Flumen			
13		Joaquín Costa (Barasona)	Esera			
14		Mediano	Cinca			
15		Santa Ana	Noguera-Ribagorzan a			
16		La Sotonera	Sotón			
17		El Grado I	Cinca			
18		Búbai	Gállego			
19		Vadiello	Guatizalem a			
20		Lanuzza	Gállego			
21		Torrollón	Grande			
22		Valdabra	Valdabra			
23	Huesca	Rialb	Calcón			
24		Cueva Foradada	Martín			
25		Gallipuéñ	Guadalopillo			
26		Pena	Pena			
27		Santolea	Guadalope			
28		Estanca de Alcañiz	Guadalope (derivación)			
29	Teruel	Calanda	Guadalope			
30	NAVARRA					

Figura 7: Full excel d'algunes relacions comunitat/embassament/riu.

Aquest segon fitxer o full excel, es un full únic, on la informació principal es la relació Comunitat - Embassament - Riu. Com es veurà mes endavant, entre aquest full i els anteriors sorgeixen una sèrie de incongruències. Aquestes incongruències de informació es quelcom que es pot considerar normal dins la creació d'un DW ja que si tenim diverses fonts de dades sense estar normalitzades, aquesta poden ser incongruents. Es part del DW (en el processos d'extracció, transformació i càrrega, nomenats en el punt 1.1) normalitzar aquestes dades.

2.2.3.- Estimació de neu.

Embassament	Estimació neu 2001	Estimació neu 2002	Estimació neu 2003	Estimació neu 2004	Estimació neu 2005
Maidereva	5	4	4	5	3
Santa Ana	4	4	4	3	3
Búbal	4	3	3	2	3
Valdabra	4	5	5	2	2
Calanda	3	3	3	3	3
Mansilla	3	2	1	2	2
Vicarías	4	5	5	5	5
Oliana	2	1	1	1	1
RESTA	1	1	1	1	1

Figura 8: Constants pel càlcul d'estimació de neu.

Finalment, tenim aquest quadre d'estimació de neu que amb la fórmula proporcionada fem una transformació de neu a aigua embassada.

La fórmula a la que hem fet referència:

$$H_{aigua_neu_equivaleu_anual_embassament} = estimació_neu_embassament \frac{capacitat_anual_embassament_neu}{5}$$

Figura 9: Fórmula pel càlcul d'estimació de neu.

Així doncs, amb els fulls excels mensuals, el full excel de comunitat - embassament i la taula d'estimació de neu, tenim totes les dades per crear un DW que pugui obtenir els informes sol·licitats en el punt 1.6.

2.3.- Disseny conceptual Model Entitat/Relació.

El model conceptual no es mes que un esquema de les dades i de les associacions o interrelacions que s'han obtingut a rel del coneixement dels requeriments i especificacions de lo que es vol obtenir al final.

Tot DW ha de ser multidimensional, que consisteix en concebre les dades que analitzarem en termes de fets i dimensions d'anàlisi, de manera que les podem situar en un espai n -dimensional. El model multidimensional en que ens mourem, consta de dos tipus de dades: els FETS que volem analitzar i les DIMENSIONS que utilitzarem per analitzar-los. Un fet, no res mes que un objecte d'anàlisi i una dimensió representa un punt de vista que utilitzarem en el anàlisi de dades. Dins de una dimensió podem distingir grups de instàncies segons tamany (granularitat).

En el nostre particular DW, ens trobem amb les dimensions de Comunitats, Embassaments, Rius, Temps i com a fets totes les dades relatives a les variacions de les capacitats i moviments mensuals del embassaments. Tal com s'explica, aquestes variacions es poden analitzar des del 'prisma' de rius, comunitat, temps, etc. (o sigui des de cada dimensió).

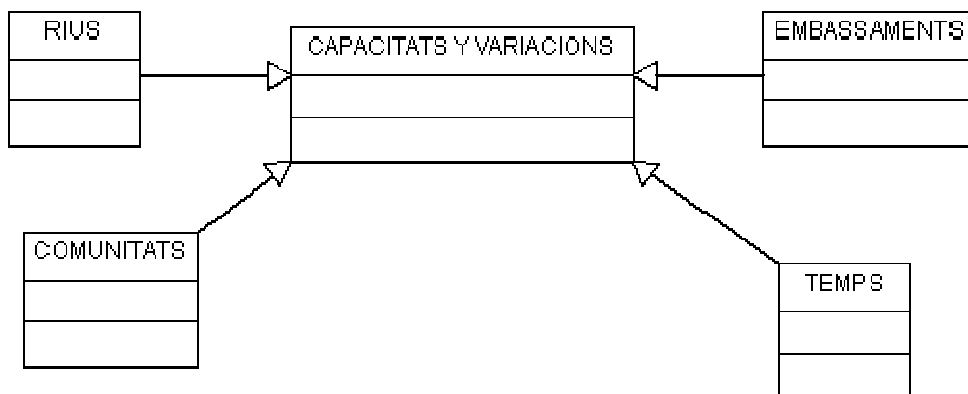


Figura 10: Diagrama conceptual. Model Entitat/Relació.

3.- DISSENY.

En aquest capítol, ja donarem forma tant a la base de dades a partir del model conceptual com als informes que es desitgen obtenir.

3.1.- Diagrama d'arquitectura de software.

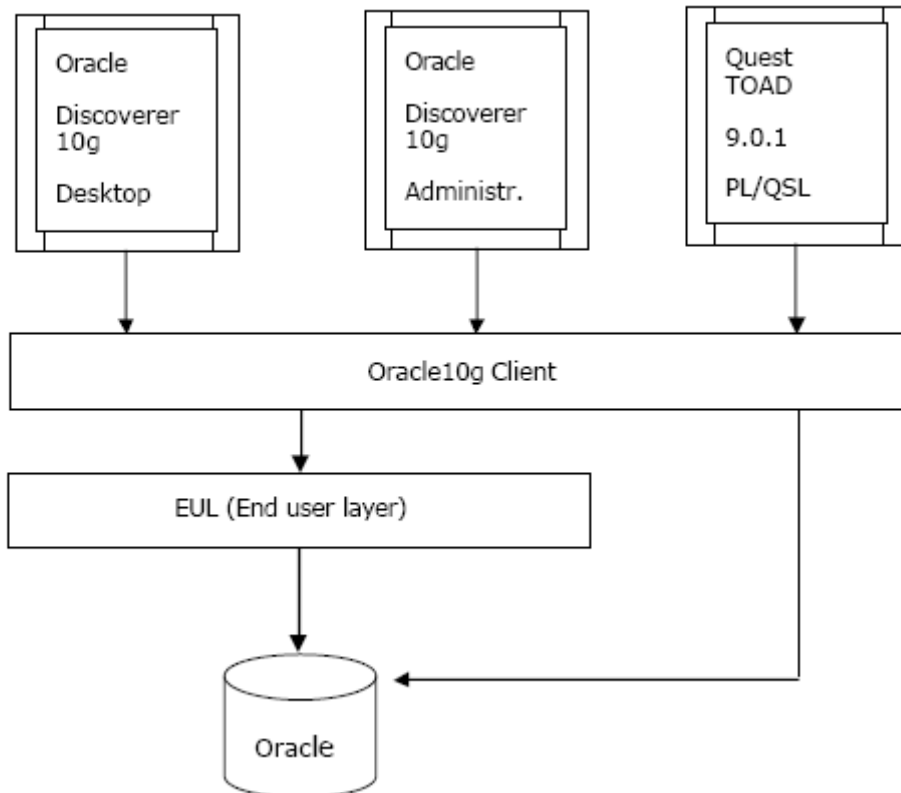


Figura 11: Diagrama d'arquitectura de software.

Aquest diagrama ens mostra l'arquitectura de la plataforma ORACLE amb les seves utilitats associades com son el Discoverer Desktop (per explotar dades i crear nous informes a partir de les relacions creades) i el Discoverer Administrador (per a crear noves relacions) i lògicament el PL/SQL per carregar dades, fer consultes, etc.

3.2.- Disseny de taules.

Un cop concebut el disseny conceptual i coneixent les possibles limitacions de la plataforma en la que es treballa, es passarà a fer un disseny de les taules on residiran les dades a explotar.

3.2.1.- Disseny físic.

En el nostre particular, el disseny físic no varia massa del conceptual. Tant sols hem afegit una taula de totals per Comunitat Autònoma i hem dividit la taula de temps en una d'any i una altre de mes. No es un disseny particularment complexa en que variï massa el disseny conceptual del lògic i del físic. Per tant, el disseny físic, queda com segueix:

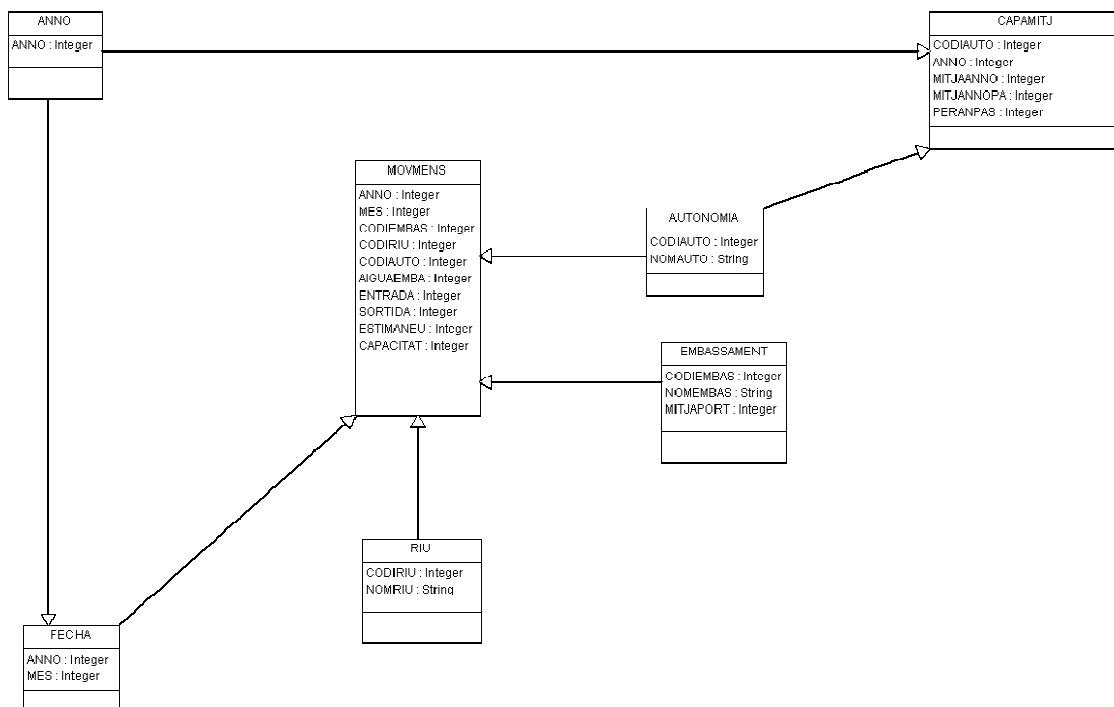


Figura 12: Diagrama del disseny físic de la base de dades.

3.2.1.1.- Taules a crear.

TAULA ANNO: Definició dels anys

<i>NOM ATRIBUT</i>	<i>TIPUS</i>	<i>DESCRIPCIO</i>	<i>ORIGEN</i>
ANNO	NUM(4)	Any	Fulls Excel

Figura 13: Definició de la taula ANNO.

TAULA FECHA: Definició dels mesos dins els anys

<i>NOM ATRIBUT</i>	<i>TIPUS</i>	<i>DESCRIPCIO</i>	<i>ORIGEN</i>
ANNO	NUM(4)	Any	Fulls Excel
MES	NUM(2)	Mes	Fulls Excel

Figura 14: Definició de la taula FECHA.

TAULA AUTONOMIA: Definició de les diferents autonomies

<i>NOM ATRIBUT</i>	<i>TIPUS</i>	<i>DESCRIPCIO</i>	<i>ORIGEN</i>
CODIAUTO	NUM(2)	Codi id. De auto	Seqüencial
DESCAUTO	VARCHAR(50)	Nom de autonomia	Full 'embalses'

Figura 15: Definició de la taula AUTONOMIA.

TAULA RIUS: Definició dels diferents rius

<i>NOM ATRIBUT</i>	<i>TIPUS</i>	<i>DESCRIPCIO</i>	<i>ORIGEN</i>
CODIRIU	NUM(2)	Codi id. De riu	Seqüencial
DESCRIU	VARCHAR(50)	Nom del riu	Fulls Excel

Figura 16: Definició de la taula RIUS

TAULA EMBASSAMEN: Definició dels diferents embassaments.

<i>NOM ATRIBUT</i>	<i>TIPUS</i>	<i>DESCRIPCIO</i>	<i>ORIGEN</i>
CODEMBA	NUM(2)	Codi id. D'embassament	Seqüencial
DESCEMBA	VARCHAR(50)	Nom de l'embassament	Fulls Excel
MITJAAPO	NUM(7,2)	Mitjana d'aportació anyal.	Calculat

Figura 17: Definició de la taula EMBASSAMEN.

TAULA MOVIMENS: Moviments mensuals dins els embasaments.

<i>NOM ATRIBUT</i>	<i>TIPUS</i>	<i>DESCRIPCIO</i>	<i>ORIGEN</i>
ANNO	NUM(4)	Any del moviment	Fulls Excel
MES	NUM(2)	Mes del moviment	Fulls Excel
CODEMBAS	NUM(2)	Codi embassament	Transformat a partir de fulls excel
CODIRIU	NUM(2)	Codi riu	Transformat a partir de fulls excel
CODIAUTO	NUM(2)	Codi autonomia	Calculat a partir del full 'embalses'.
CAPACITAT	NUM(7,2)	Capacitat de l'embassament	Fulls Excel
AIGUAEMBAS	NUM(7,2)	Aigua embassada en el any/mes	Fulls Excel
ENTRADA	NUM(7,2)	Entrada d'aigua en el any/mes	Fulls Excel
SORTIDA	NUM(7,2)	Sortida d'aigua en el any/mes	Fulls Excel
ESTIMANEU	NUM(7,2)	Estimació de increment d'aigua per neu	Calculat a partir de la taula d'estimació de neu.

Figura 18: Definició de la taula MOVIMENS.

TAULA CAPAANU: Totals anyals per Comunitat Autònoma.

<i>NOM ATRIBUT</i>	<i>TIPUS</i>	<i>DESCRIPCIO</i>	<i>ORIGEN</i>
ANNO	INT(4)	Any dels totals	Fulls Excel
CODIAUTO	INT(2)	Codi de l'autonomia.	Calculat a partir del full 'embalses'
MITJANU	INT(7,2)	Mitjana d'aigua	Calculat
MITAANANT	INT(7,2)	Mitjana d'aigua l'any passat	Calculat
PERDIF	INT(5,2)	Percentatge inc/dec	Calculat

Figura 19: Definició de la taula CAPAANU.

3.2.1.2.- Càlcul d'espai.

TAULA	LONGITUD TUPLA (APROX.)	NUMERO DE TUPLAS (APROX.)	ESPAI NECESARI
ANNO	4	5	20
FECHA	6	60	360
AUTONOMIA	52	7	364
RIUS	52	50	2.600
EMBASSAMENTS	60	60	3.600
MOVIMENS	50	3420	171.000
CAPAANU	25	35	875

Figura 20: Càlcul de l'espai necessari per la creació de la base de dades.

Segons els nostres càlculs, es necessari un espai de 180K per contenir les dades en la nostra base de dades. A aquest espai, s'ha d'afegir tot aquell espai que el gestor de base de dades requereix per la creació d'índexs primaris, claus estrangeres, etc.

Però en el pitjor dels casos, aquest espai no superaria mai les 200K. Aquest espai es totalment assumible per qualsevol equip mitjanament modern.

3.2.2.-Creació de taules.

En aquest punt, es procedirà a comentar les taules creades en la base de dades a carregar per obtenir els documents sol·licitats en l'enunciat del TFC.

Per a la creació de les taules, executarem el següent script:

```
CREATE TABLE "ANNO"  
("ANNO" NUMBER(4,0) NOT NULL ENABLE,  
 CONSTRAINT "ANNO_PK" PRIMARY KEY("ANNO") ENABLE)  
/  
  
CREATE TABLE "FECHA"  
("ANNO" NUMBER(4,0) NOT NULL ENABLE,  
 "MES" NUMBER(2,0) NOT NULL ENABLE,  
 CONSTRAINT "FECHA_FK" FOREIGN KEY ("ANNO") REFERENCES "ANNO"("ANNO")ENABLE,  
 CONSTRAINT "FECHA_PK" PRIMARY KEY ("ANNO","MES") ENABLE)  
/  
  
CREATE TABLE "AUTONOMIA"  
("CODIAUTO" NUMBER(2,0) NOT NULL ENABLE,  
 "DESCAUTO" VARCHAR2(50) NOT NULL ENABLE,  
 CONSTRAINT "AUTONOMIA_PK" PRIMARY KEY ("CODIAUTO") ENABLE)  
/  
  
CREATE TABLE "RIUS"  
("CODIRIU" NUMBER(2,0) NOT NULL ENABLE,  
 "DESCRIU" VARCHAR2(50) NOT NULL ENABLE,  
 CONSTRAINT "RIUS_PK" PRIMARY KEY ("CODIRIU") ENABLE)  
/  
  
CREATE TABLE "EMBASSAMEN"  
("CODIEMBA" NUMBER(2,0) NOT NULL ENABLE,  
 "DESCEMBA" VARCHAR2(50) NOT NULL ENABLE,  
 "MITJAAPO" NUMBER(7,2) NOT NULL ENABLE,  
 CONSTRAINT "EMBASSAMEN_PK" PRIMARY KEY ("CODIEMBA") ENABLE)  
/  
  
CREATE TABLE "MOVIMENS"  
("ANNO" NUMBER(4,0) NOT NULL ENABLE,  
 "MES" NUMBER(2,0) NOT NULL ENABLE,  
 "CODIEMBAS" NUMBER(2,0) NOT NULL ENABLE,  
 "CODIRIU" NUMBER(2,0) NOT NULL ENABLE,  
 "CODIAUTO" NUMBER(2,0) NOT NULL ENABLE,  
 "CAPACITAT" NUMBER(7,2),  
 "AIGUAEMBAS" NUMBER(7,2),  
 "ENTRADA" NUMBER(7,2),  
 "SORTIDA" NUMBER(7,2),  
 "ESTIMANEU" NUMBER(7,2),  
 CONSTRAINT "MOVIMENS_FK1" FOREIGN KEY ("ANNO","MES") REFERENCES "FECHA"("ANNO","MES")ENABLE,  
 CONSTRAINT "MOVIMENS_FK2" FOREIGN KEY ("CODIEMBAS") REFERENCES "EMBASSAMEN"("CODIEMBA")ENABLE,  
 CONSTRAINT "MOVIMENS_FK3" FOREIGN KEY ("CODIRIU") REFERENCES "RIUS"("CODIRIU")ENABLE,  
 CONSTRAINT "MOVIMENS_FK4" FOREIGN KEY ("CODIAUTO") REFERENCES "AUTONOMIA"("CODIAUTO")ENABLE,  
 CONSTRAINT "MOVIMENS_PK" PRIMARY KEY ("ANNO","MES","CODIEMBAS","CODIRIU","CODIAUTO") ENABLE)  
/
```



```
CREATE TABLE "CAPAANU"  
("ANNO" NUMBER(4,0) NOT NULL ENABLE,  
"CODIAUTO" NUMBER(2,0) NOT NULL ENABLE,  
"MITJANU" NUMBER(7,2),  
"MITAANANT" NUMBER(7,2),  
"PERDIF" NUMBER(5,2),  
CONSTRAINT "CAPAANU_FK1" FOREIGN KEY ("ANNO") REFERENCES "ANNO"("ANNO")ENABLE,  
CONSTRAINT "CAPAANU_FK2" FOREIGN KEY ("CODIAUTO") REFERENCES "AUTONOMIA"("CODIAUTO")ENABLE,  
CONSTRAINT "CAPAANU_PK" PRIMARY KEY ("ANNO","CODIAUTO") ENABLE)  
/
```

Figura 21: Script per la creació de taules.

A la vegada, per poder executar aquest script, es tindrà que anar a la consola del DOS, connectar-se com el usuari amb la comanda SQLPLUS usuari, seguidament demana la paraula de pas i ja es pot introduir la comanda START amb el nom del arxiu on està l' script mostrat mes amunt.

Aquest fitxer està inclòs en l'annex d'aquest document amb el nom de CREATAU.txt

Un cop executat l' script ja es poden consultar les taules i tenen el següent format

TAULA ANNO



Nombre De Columna	Tipo De Dato	Nulo	Valor Por Defecto	Clave
ANNO	NUMBER(4,0)	No	-	1 -

Figura 22: Creació de la taula ANNO.

Taula de dimensió on estaran informats els anys en que tenim informació a la nostre base de dades.

TAULA FECHA

Nombre De Columna	Tipo De Dato	Null	Valor Por Defecto	Clave
ANNO	NUMBER(4,0)	No	-	
MES	NUMBER(2,0)	No	-	

Figura 23: Creació de la taula FECHA

Taula de dimensió on estaran informats els mesos dels anys en que tenim informació a la nostre base de dades. Conté com a clau forana la dada 'anno' que es relaciona amb la taula de dimensió ANNO.

TAULA AUTONOMIA

Nombre De Columna	Tipo De Dato	Null	Valor Por Defecto	Clave
CODIAUTO	NUMBER(2,0)	No	-	
DESCAUTO	VARCHAR2(50)	No	-	

Figura 24: Creació de la taula AUTONOMIA

Taula de dimensió on estaran informades les comunitats autònomes de que tenim informació. La taula, té com informació un codi d'autonomia que calcularem en el procés de càrrega i el nom de la comunitat.

TAULA RIUS

Nombre De Columna	Tipo De Dato	Hulo	Valor Por Defecto	Clave
CODIRIU	NUMBER(2,0)	No	-	
DESCRIU	VARCHAR2(50)	No	-	

Figura 25: Creació de la taula RIUS

Taula de dimensió on estaran informats els rius del que tenim informació. La taula, té com informació un codi de riu que calcularem en el procés de càrrega i el nom del riu.

TAULA EMBASSAMENTS

Nombre De Columna	Tipo De Dato	Hulo	Valor Por Defecto	Clave
CODIEMBA	NUMBER(2,0)	No	-	
DESCEMBA	VARCHAR2(50)	No	-	
MITJA APO	NUMBER(7,2)	No	-	

Figura 26: Creació de la taula EMBASSAMENTS

Taula de dimensió on estaran informats els embassaments dels que tenim informació. La taula, té com informació un codi d'embassament que calcularem en el procés de càrrega i el nom de l'embassament i la mitjana anual d'aportació. Aquesta dada també serà calculada en el procés de càrrega.

TAULA MOVIMENTS

Tabla	Datos	Índices	Modelo	Restricciones	Permisos	Estadísticas	Valor
Agregar Columna		Modificar Columna		Cambiar Nombre de Columna		Borrar Col	
Nombre De Columna	Tipo De Dato	Null	Valor Por Defecto	Clave			
ANNO	NUMBER(4,0)	No	-				
MES	NUMBER(2,0)	No	-				
CODIEMBAS	NUMBER(2,0)	No	-				
CODIRIU	NUMBER(2,0)	No	-				
CODIAUTO	NUMBER(2,0)	No	-				
CAPACITAT	NUMBER(7,2)	Yes	-				
AIGUAEMBAS	NUMBER(7,2)	Yes	-				
ENTRADA	NUMBER(7,2)	Yes	-				
SORTIDA	NUMBER(7,2)	Yes	-				
ESTIMANEU	NUMBER(7,2)	Yes	-				

Figura 27: Creació de la taula MOVIMENTS.

Aquesta taula, es una **taula de fets** on gira tot el disseny de la base de dades i del DW a dissenyar. Conté totes les claus foranes amb les taules de dimensions relacionades. I a més te la dada repetitives de capacitat d'embassament (per alleugerar els càlculs en la obtenció de informes) i la dada calculada de estimació de neu (que també es repetitiva per tots els mesos d'un embassament/any). En concret, té les següents dades:

- ANNO: Any de la informació. Clau forana de la taula FECHA.
- MES: Mes de la informació. Clau forana de la taula FECHA.
- CODIEMBAS: Codi d'embassament. Clau forana de la taula EMBASSAMENT.
- CODIRIU: Codi de riu. Clau forana de la taula RIUS.
- CODIAUTO: Codi de comunitat. Clau forana de la taula AUTONOMIA.
- CAPACITAT: Capacitat total de l'embassament.
- AIGUAMEBAS: Capacitat puntual (en un any/mes) d'un embassament.
- ENTRADA: Aportació positiva en Hm3 per un any/mes.
- SORTIDA: Aportació negativa en Hm3 per un any/mes.
- ESTIMANEU: Estimació de neu per embassament

TAULA CAPAANU

Tabla Datos Índices Modelo Restricciones Permisos Estadísticas Valor					
Agregar Columna		Modificar Columna		Cambiar Nombre de Columna	Borrar Columna
Nombre De Columna	Tipo De Dato	Null	Valor Por Defecto	Clave	
ANNO	NUMBER(4,0)	No	-		
CODIAUTO	NUMBER(2,0)	No	-		
MITJANU	NUMBER(7,2)	Yes	-		
MITAANANT	NUMBER(7,2)	Yes	-		
PERDIF	NUMBER(5,2)	Yes	-		
				1 -	

Figura 28: Creació de la taula CAPAANU.

Taula agregada de totals que es calcularan a l'hora de fer els processos de càrrega perquè l'extracció de dades quan es demanin certs informes no sigui tant costosa. En aquesta taula calcularem la mitja agregada de tots els embassaments d'una comunitat autònoma en un any concret. A més, tindrem la informació de la mitja de l'any anterior i el percentatge de increment.

3.3.- Processos ETL(extracció, transformació i càrrega de dades).

En aquest punt, ens centrarem en les carregues de taules a partir dels fitxer d'entrada (en format Excel) que disposem. Prèviament a aquesta càrrega, ha d'haver-hi un procés d'extracció i transformació. I prèviament a la transformació s'ha de fer un anàlisi per veure que i com s'ha de transformar.

3.3.1.- Anàlisi previ de les dades de partida.

Tal com hem indicat –una mica per sobre- en el punt 2.1 i 2.2 de incongruències en les dades a carregar, fent un anàlisi exhaustiu d'aquestes dades, hem creat aquest quadre que recull totes aquestes anormalitats trobades amb la solució a adoptar.

CONFLICTE	SOLUCIÓ
Hi ha embassaments en el fitxer C.A./embassament que no s'utilitzen en els fulls mensuals.	No es carregaran els embassaments del fitxer d'embassaments que no estiguin en el full mensual ja que tindria les dades a zero i afectaria a mitjanes. Es consideren dades 'escombraria' i han d'estar a fora del DW.
Hi ha un embassament a la previsió de neu (Vicarias) que es correspon a l'embassament Monteagudo ja que el seu nom correcte es "Monteagudo de las Vicarias" .	Es tractarà com el pantà Monteagudo.
Els embassaments que estan en els fulls mensuals i que no estan en el fitxer d'embassaments no sabem a quina província/comunitat pertanyen.	Es farà una recerca de tots aquests embassaments que estan sense comunitat i se'ls informarà.
En el fitxer d'embassaments en hi ha un que pertany a dues Comunitats Autòniques (Rialb).	Eliminarem el registre que indica que pertany a Osca.
Hi ha dos rius que fan referència al mateix: Flamisel i Flamisell.	Eliminarem el registre del riu Flamisel ja que el correcte es el Flamisell.
En els fitxers mensuals se'ns referència una 'variación semanal'.	Es tractarà com a variació mensual.

CONFLICTE	SOLUCIÓ
A nivell dada, tant l'atribut Entrada Hm com Salida Hm quan ve sense informar, a vegades ve com zero, a vegades a blancs i a vegades a '—'.	Es carregaran tots aquests valors com a únic valor 'NULL'.
Venen quantitats d'entrada com a negatives i les utilitza com un increment de la sortida.	A les quantitats negatives d'entrada deixarem la entrada com a zero i sumarem aquesta quantitat a la sortida mensual.
A l'apartat de variació setmanal ens ve la suma algebraica de l'entrada i la sortida sense un ajustament de decimals.	Obviarem aquesta dada que ens ve i la diferència la calcularem quan tinguem que fer-la servir. I, calculant-la tindrem la precisió que requereix, cosa que en aquests moments no tenim.
Els caràcters accentuats i les 'Ñ' poden quedar mal transcrits un cop es facin les càrregues automàtiques.	Les 'Ñ' es canviaran per 'NY' i els caràcters accentuats se'ls hi traurà l'accent.
A priori, l'any 2001 no el tenim sencer. Nomes tenim dades a partir de juliol.	Podem recuperar les dades de gener a juliol de 2001 a partir de les dades de gener a juliol de 2002 en les columnes que corresponen a les dades de l'any anterior.
Com en el punt anterior es podria recuperar mig any del 2000 a partir de les dades del any anterior que tenim en els fulls excel de 2001	No es creu convenient ja que tindríem dades de nomes mig any i poden afectar als càlculs de totals i mitjanes.
No quadren les capacitats del any anterior amb el que correspon. Per exemple per l'embassament Uribarri el febrer de 2002 hi ha una capacitat de 89. En canvi al febrer de 2003, a la columna de capacitat any anterior hi ha una capacitat de 85.	S'ignoraran les dades de 'capacitat any anterior' i 'percentatge any anterior'. Es calcularan en un procés especial.
Al quadre d'estimació de neu, apareix un embassament (Valdabra) que no està ni en el fitxer 'embalses' ni al fulls Excel.	S'ignorarà aquest embassament.

Figura 29: Actuacions a fer en el ETL.

Òbviament, en un entorn de treball professional, cada punt s'analitzaria amb els usuaris finals i es determinaria quina solució a adoptar en cada un d'ells.

3.3.2.- Càrrega de la taula de rius, comunitats i dates.

Per fer la càrrega de la taula de comunitats, es crearà un fitxer a partir del rebut amb el nom 'embalses'. Dins aquests fitxer estan les comunitats amb les que tenim que treballar. Aquestes comunitats les codificarem segons el seu ordre alfabètic.

Per la càrrega de rius es tindran en compte els fitxers 'embalses' i un fitxer qualsevol de moviments mensuals. Ja em dit en el punt anterior que hi ha rius diferents en els dos fitxers i el sistema els tindrà que recollir tots. Un cop tinguem la llista de tots els rius (eliminant el Flamisel), els codificarem de igual manera que les comunitats. Els ordenarem alfabèticament i els hi assignarem un codi numèric seqüencial.

Per carregar els fitxers de dates (anno i mes) es crearan fitxers manualment.

3.3.3.- Creació del fitxer de moviments i càlcul de totals i mitjanes.

Per crear el fitxer de moviments, refundrem els fitxers excel rebuts en un sol i en cada línia de informació, afegirem el codi de riu i el codi de comunitat a qui pertany l'embassament. Eliminarem d'aquest full de totals el percentatge de capacitat i tal com hem dit en la taula de 'problemes - solucions', també eliminarem el capacitat de l'any anterior i el percentatge de l'any anterior (nomes deixarem les de gener a juny del 2002 per carregar el gener a juny de 2001). Un cop estigui tot això en un procés carregarem les dades que acabem de mencionar.

Una vegada tenim el fitxer de moviments, tindrem que calcular els totals per crear la taula agregada de totals CAPAANU. L'algorisme per crear aquesta taula es el següent:

```
READ FILE(ENTRADA) INTO(RENTRADA);
DO WHILE (FIN_RENTRADA = 0);
  SELECT (RENTRADA.ANNO);
  WHEN (2001) I=1; WHEN(2002) I=2; WHEN(2003) I=3; WHEN(2004) I=4; WHEN(2005) I=5;
  END;
  J = RENTRADA.CODIAUTO;
  CONTADOR(I,J)=CONTADOR(I,J) + 1;
  CAPAANO(I,J) =CAPAANO(I,J) + RENTRADA.CAPAMES;
  READ FILE(ENTRADA) INTO(RENTRADA);
END;
DO I = 2 TO 5;
  DO J = 1 TO 7;
    CAPAANOAN(I,J) =CAPAANO(I-1,J);
  END;
END;
DO I = 1 TO 5;
  DO J = 1 TO 7;
    RSORTIDA.ANNO = I + 2000;
    RSORTIDA.CODIAUTO = J;
    RSORTIDA.MITJANO = CAPAANO(I,J)/CONTADOR(I,J);
    IF I = 1 THEN DO;
```


Treball fi de carrera (TFC) – DATA WAREHOUSE

```
RSORTIDA.MITJANOPAS = CAPAANOAN(I,J)/CONTADOR(I-1,J);
RSORTIDA.PERCENDIF = (RSORTIDA.MITJANO-RSORTIDA.MITJANOPAS)/
    RSORTIDA.MITJANOPAS;
RSORTIDA.PERCENDIF = RSORTIDA.PERCENDIF * 100;
END;
ELSE DO;
    RSORTIDA.MITJANOPAS = 0;
    RSORTIDA.PERCENDIF = 0;
END;
WRITE FILE (SORTIDA) FROM(RSORTIDA);
END;
END;
```

Figura 30: Algorisme pel càlcul de totals per Comunitat.

Es fa una lectura de tots els fitxers rebuts. En un primer pas, es van acumulant en una matriu de dues dimensions (any i comunitat) el nombre de registres d'aquella autonomia i la seva capacitat. Aquest càlcul també serveix per saber la capacitat de l'any anterior.

Seguidament es tracta aquesta taula per treure la mitja anual capacitat que la calcularem dividint la capacitat anual acumulada i el nombre de registres de informació per aquella comunitat. Anàlogament, la mitjana de capacitat de l'any passat es calcularà igual i amb aquestes dues dades es calcularà el percentatge diferencial amb la fórmula $(MITJA_ANY - MITJA_ANY_PASSAT) / MITJA_ANY_PASSAT$.

També tindrem que calcular mitjana anual d'aportació de l'embassament. Per fer aquest càlcul, utilitzarem el següent algorisme:

```
READ FILE(ENTRADA) INTO(RENTRADA);
DO WHILE (FIN_RENTRADA = 0);
SELECT (RENTRADA.ANNO);
    WHEN (2001) I=1; WHEN(2002) I=2; WHEN(2003) I=3; WHEN(2004) I=4; WHEN(2005) I=5;
END;
J = RENTRADA.CODIPANTA;
CONTADOR2(I,J)=CONTADOR2(I,J) + 1;
ENTRADA2(I,J)=ENTRADA2(I,J) + ENTRADA0;
SORTIDA2(I,J)=SORTIDA2(I,J) + SORTIDA0;
READ FILE(ENTRADA) INTO(RENTRADA);
END;
DO I = 1 TO 5;
    DO J = 1 TO 60;
        RSORTID2.ANNO = I + 2000;
        RSORTID2.CODIEMBAS = J;
        RSORTID2.ENTRAD4 = ENTRADA2(I,J);
        RSORTID2.SORTID4 = SORTIDA2(I,J);
        WRITE FILE (SORTID2) FROM(RSORTID2);
    END;
END;
/* TREU ELS PROMITJOS DELS EMBASSAMENT */
DO J=1 TO 60;
AUX = 0;
    DO I = 1 TO 5;
        AUX = AUX + ENTRADA2(I,J) - SORTIDA2(I,J);
    END;
AUX = AUX/5;
RSORTID2.ANNO = 9999;
RSORTID2.CODIEMBAS = J;
RSORTID2.ENTRAD4 = AUX;
RSORTID2.SORTID4 = 0;
WRITE FILE (SORTID2) FROM(RSORTID2);
END;
```

Figura 31: Algorisme pel càlcul de la mitjana d'aportació per embassament.

Es similar a l'algorisme anterior. En aquest, en un primer pas, anem acumulant tots els registres per any/embassament. En un segon pas, fem la diferència entre l'entrada i la sortida i ho dividim pels cinc anys en que tenim informació. Ja tenim la mitjana d'aportació anual d'un embassament.

3.3.4.- Càrrega d'embassaments i totals per comunitat.

Per carregar els embassaments tindrem en compte un fitxer qualsevol de moviments mensuals. Els codificarem igual que els rius i/o CA's (ordenats alfabèticament i assignant-li un codi seqüencial) i per cada un d'ells li informarem la mitjana anual d'aportació de l'embassament calculat en el segon algorisme.

Els totals de Comunitat es carregaran a partir d'un fitxer que tindrà el mateix format que la taula a carregar. Aquest fitxer el deixarà el primer algorisme que hem descrit anteriorment.

3.3.5.- Càrrega de la taula de moviments mensuals.

Per finalitzar ja podem carregar els moviments mensuals. Hi ha que indicar que en els moviments carreguem per cada un d'ells la capacitat de l'embassament. Pot ser que sigui una dada repetitiva i redundant però això estalviarà de fer unes consultes molt més dures, en especial per obtenir els tres primers informes.

3.3.6.-Procés de càrrega de les taules.

En aquest punt, ens centrarem a com carregar les dades seguint la utilitat '**SQL LOADER**'.

En els fitxers annexat a aquest document, hi ha un joc de fitxer (un per cada taula) que serviran per carregar les taules abans creades.

El funcionament, es el següent: Per cada taula, executarem la comanda sqlldr amb els següents paràmetres:

Sqlldr userid=user/password@connexió control=fitxercntl.ctl log=fitxerlog.txt

Per cada taula, tenim un fitxer de control que te el següent format (poso d'exemple la carrega de la taula moviments).

```
LOAD DATA
BADFILE "E:\SCRIPTS\Cargataules\movimens1.txt"
DISCARDFILE "E:\SCRIPTS\Cargataules\movimens2.txt"
INFILE "E:\Cargas\DADES A CARREGAR\Variacions.prn"
INTO TABLE MOVIMENS
FIELDS TERMINATED BY ";"
TRAILING NULLCOLS
(ANNO DECIMAL EXTERNAL,
MES DECIMAL EXTERNAL,
CODIEMBAS DECIMAL EXTERNAL,
CODIRIU DECIMAL EXTERNAL,
CODIAUTO DECIMAL EXTERNAL,
CAPACITAT DECIMAL EXTERNAL,
AIGUAEMBAS DECIMAL EXTERNAL,
ENTRADA DECIMAL EXTERNAL,
SORTIDA DECIMAL EXTERNAL)
```

Figura 32: Script exemple de carrega de dades a partir d'un fitxer txt.

Aquest fitxer indica al SQLLOADER on està el fitxer de les dades, on volem que deixi el fitxer erroni i les dades que no ha carregat (si s'escau). Així mateix, també li indiquem la adreça d'un fitxer de log per veure les estadístiques del procés. Per últim i per acabar de fer el procés hi tindrà que haver-hi un fitxer .dat amb el mateix nom que el .ctl, encara que aquest primer estigui buit.

En la carpeta on hi ha els fitxers de control (Cargataules) també s'han inclòs els fitxers .dat i els log que es van obtenir en les càrregues.

3.3.7.-Altres consideracions.

Sempre pot haver-hi oblots d'última hora i petits detalls que facin que tota la feina que hem fet de carregues hagi sigut en va. Segon la gravetat d'aquest oblit o d'aquesta errada conceptual, no faria falta tirar enrere tot lo carregat, modificar els processos de carrega i tornar a carregar-ho tot. En el nostre cas en particular, se'ns va descuidar el càlcul de l'estimació de neu al crear el fitxer de carrega de moviments. La solució va ser modificar la taula amb unes instruccions especial mitjançant la comanda START (com a la creació de les taules) prèviament d'haver-nos connectat a la base de dades amb el sqlplus.

```
Update movimens
set estimaneu = aiguaembas
where (codiembas = 28 and
      (anno = 2001 or anno = 2004)) or
      (codiembas = 34 and
      anno <> 2001)
/
update movimens
set estimaneu = (aiguaembas*4)/5
where (codiembas = 28 and
      (anno = 2002 or anno = 2003)) or
      (codiembas = 43 and
      (anno = 2001 or anno = 2002 or anno = 2003)) or
      (codiembas = 8 and
      anno = 2001) or
      (codiembas = 34 and
      anno = 2001)
/
update movimens
set estimaneu = (aiguaembas*3)/5
where (codiembas = 28 and
      anno = 2005) or
      (codiembas = 43 and
      (anno = 2004 or anno = 2005)) or
      (codiembas = 8 and
      (anno = 2002 or anno=2003 or anno=2005)) or
      codiembas = 9 or
      (codiembas = 29 and
      ANNO=2001)
/
update movimens
set estimaneu = (aiguaembas*2)/5
where (codiembas = 8 and
      anno = 2004) or
      (codiembas = 29 and
      (anno = 2002 or anno=2004 or anno=2005)) or
      (codiembas = 36 and
      anno = 2001)
/
```

```
update movimens  
set estimaneu = (aiguaembas)/5  
where estimaneu is null  
/
```

Figura 33: Script de modificació de la dada 'estimació de neu' a la taula MOVIMENS.

Aquest fitxer es troba es el que s'inclou en el annex amb el nom de MODITAU.txt

3.4.- Obtenció de informes.

3.4.1.- *Capacitat total vs capacitat actual per riu i any - mes.*

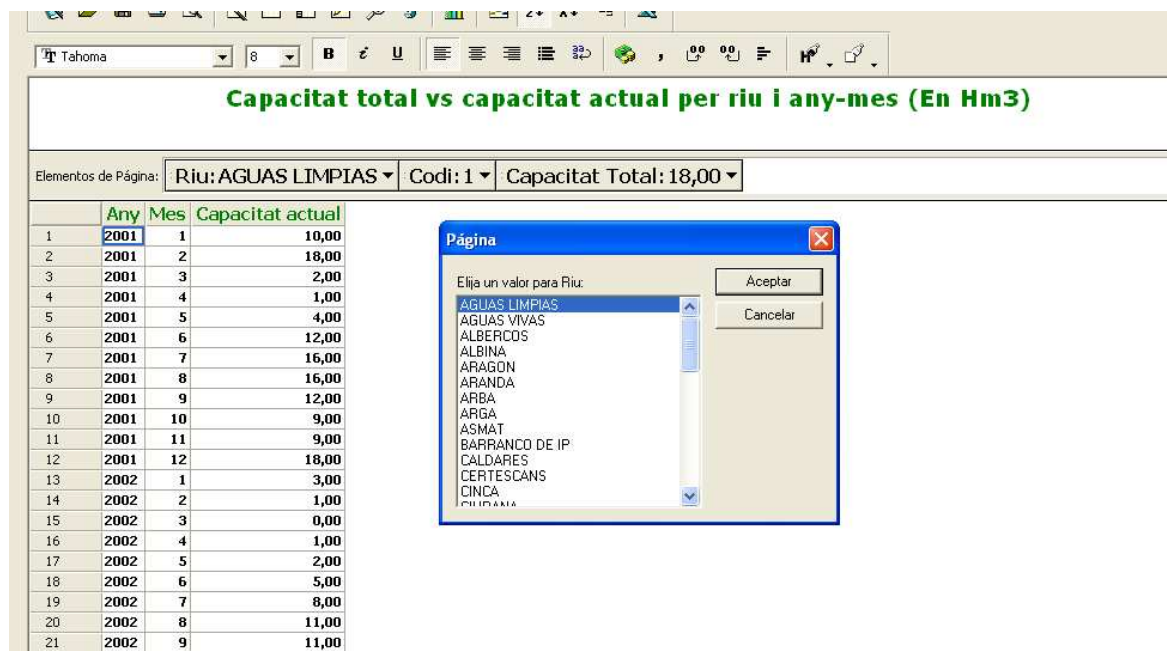


Figura 34: Disseny de l'informe capacitat total vs capacitat actual per riu i any - mes.

En aquest primer informe, es fa una comparativa entre la capacitat total de la dimensió RIUS contra l'aigua embassada en un any - mes. La capacitat, fa referència a tots aquells embassaments que alimenta el riu ja que poden ser mes d'un.

A la capçalera del informe, surt el nom del riu, el seu codi i la capacitat total ja que es una dada invariant.

Si es vol canviar de riu, nomes cal clicar a la capçalera a la dada de riu, i surt un paginable de tots els rius que s'han carregat. Clickant el desitjat es refrescaran les dades i es mostrarà la informació del riu col·leccionat.

3.4.2.- Capacitat total vs capacitat actual per embassament i any - mes.

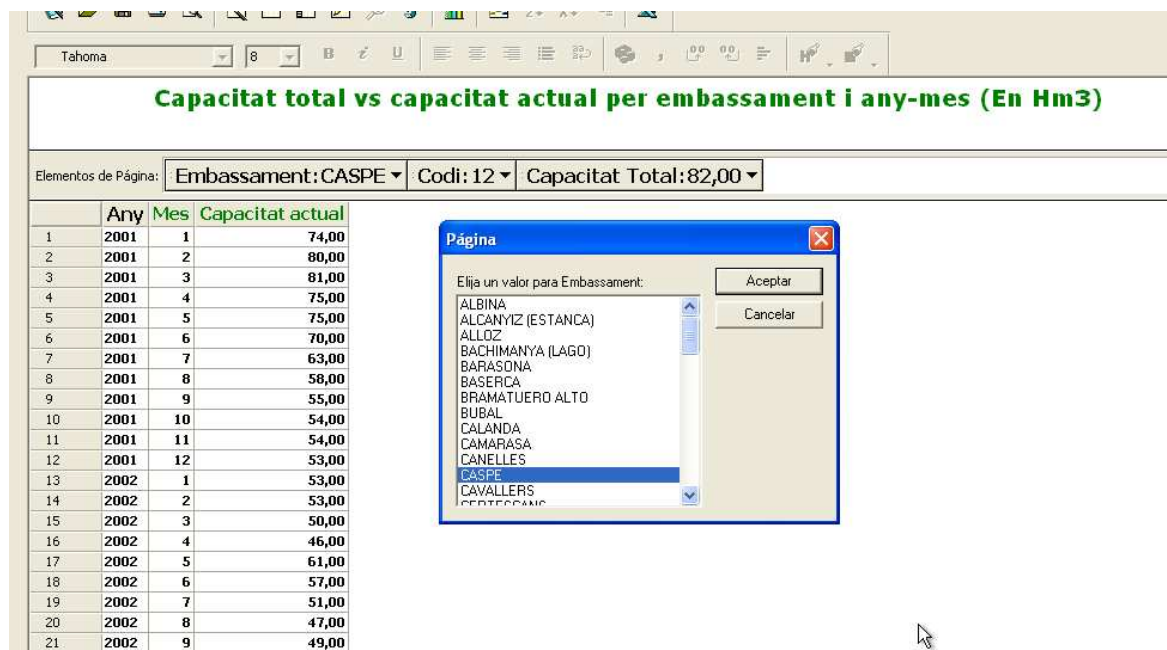


Figura 35: Disseny de l'informe capacitat total vs capacitat actual per embassament i any - mes.

En aquest segon informe (molt similar a l'anterior), es fa una comparativa entre la capacitat total de la dimensió EMBASSAMENTS contra l'aigua embassada en un any - mes. La capacitat, fa referència a la capacitat d'un embassament col·leccionat.

A la capçalera del informe, surt el nom de l'embassament, el seu codi i la capacitat total ja que es una dada invariant.

Si es vol canviar d'embassament, nomes cal clicar a la capçalera a la dada d'embassament, i surt un paginable de tots els embassaments que s'han carregat. Clickant el desitjat es refrescaran les dades i es mostrarà la informació del seleccionat.

3.4.3.- Capacitat total vs capacitat actual per comunitat i any - mes.

	Any	Mes	Capacitat Actual
1	2001	1	3.545,00
2	2001	2	3.586,00
3	2001	3	3.741,00
4	2001	4	3.625,00
5	2001	5	3.760,00
6	2001	6	3.635,00
7	2001	7	3.321,00
8	2001	8	2.943,00
9	2001	9	2.645,00
10	2001	10	2.815,00

Figura 36: Disseny de l’informe capacitat total vs capacitat actual per Comunitat Autònoma i any - mes.

En aquest informe (de la mateixa família que els dos anteriors), es fa una comparativa entre la capacitat total de la dimensió AUTONOMIA contra l’aigua embassada en un any - mes. La capacitat, fa referència a tots aquells embassaments d’aquella autonomia ja que poden ser mes d’un.

A la capçalera del informe, surt el nom de l’autonomia, el seu codi i la capacitat total ja que es una dada invariant.

Si es vol canviar de riu, nomes cal clicar a la capçalera a la dada de comunitat, i surt un paginable de totes les comunitats que s’han carregat. Clickant el desitjat es refrescaran les dades i es mostrarà la informació de l’autonomia col·leccionada.

3.4.4.- Comparativa anual de l'evolució per CA de la capacitat agregada de tots els seus embassaments.

	Any	Capacitat agregada dels embassaments
1	2001	3290,50
2	2002	2930,58
3	2003	3375,75
4	2004	3306,75
5	2005	2452,00

Figura 37: Disseny de l'informe comparativa anual de l'evolució per CA de la capacitat agregada de tots els seus embassaments.

En aquest informe, es fa una comparativa entre la capacitat agregada de la dimensió AUTONOMIA entre els diferents anys. Con que dins la base de dades, la informació està a nivell any - mes, aquesta capacitat agregada es la suma de les capacitats de tots els embassaments d'una comunitat en un any i dividit per dotze.

A la capçalera del informe, surt el nom de l'autonomia i el seu codi.

Si es vol canviar de comunitat, nomes cal clicar a la capçalera a la dada de comunitat, i surt un paginable de totes les comunitats que s'han carregat. Clickant el desitjat es refrescaran les dades i es mostrarà la informació de l'autonomia seleccionada.

3.4.5.- Comparativa de la capacitat mitjana d'una comunitat entre un any i l'anterior.

Comparativa de capacitat mitjana agregada de tots els embassaments d'una CA entre un any i l'anterior (En Ml. de lit)

Elementos de Página: Comunitat: ARAGON Codi: 1

	Any	Mitjana agregada	Mitjana any anterior	Percentatge diferència
1	2001	106.140	0	0,00%
2	2002	94.530	106.140	-10,93%
3	2003	108.890	94.530	15,19%
4	2004	106.660	108.890	-2,04%
5	2005	79.090	106.660	-25,84%

Figura 38: Disseny de l'informe comparativa de la capacitat mitjana d'una comunitat entre un any i l'anterior.

En aquest informe, es fa una comparativa entre la mitjana de la capacitat agregada de la dimensió AUTONOMIA entre els diferents anys. Es un informe molt similar a l'anterior però a més dividim la dada pel nombre d'embassaments que te una comunitat.

A la capçalera del informe, surt el nom de l'autonomia i el seu codi.

Si es vol canviar de comunitat, nomes cal clicar a la capçalera a la dada de comunitat, i surt un paginable de totes les comunitats que s'han carregat. Clickant el desitjat es refrescaran les dades i es mostrarà la informació de l'autonomia seleccionada.

3.4.6.- Aportació per riu, embassament i any - mes. Comparativa amb la mitjana anual d'aportació de l'embassament.

Aportació en Hm3 per riu, embassament i any-mes. Comparativa amb la mitjana anual d'aportació de l'embassament.

Elementos de Página: Anno: 2004 ▼ Mes: 2 ▼

	Codi riu	Riu	Codi emb.	Embassament	Entrada	Sortida	Aportació	Mitja Mensual aportació Emb.
1	1	AGUAS LIMPIAS	39	RESPOMUSO	,00	,00	0,00	0,00
2	2	AGUAS VIVAS	33	MONEVA	,10	,00	0,10	0,00
3	3	ALBERCOS	21	GONZALEZ LACASA	,60	,40	0,20	-0,03
4	4	ALBINA	1	ALBINA	,00	,00	0,00	0,00
5	6	ARAGON	57	YESA	17,60	11,50	6,10	-1,30
6	7	ARANDA	28	MAIDEYERA	,60	,60	0,00	-0,07
7	8	ARBA	42	SAN BARTOLOME	,00	,00	0,00	0,00
8	9	ARGA	20	EUGUI	1,60	1,50	0,10	0,17
9	10	ASMAT	23	GUIAMETS	,00	,00	0,00	0,00
10	11	BARRANCO DE IP	24	IP	,00	,00	0,00	0,00
11	12	CALDARES	4	BACHIMANYA (LAGO)	,00	,00	0,00	0,00
12	12	CALDARES	7	BRAMATUERO ALTO	,00	,00	0,00	0,00
13	13	CERTESCANS	14	CERTESCANS	,00	,00	0,00	0,00
14	14	CINCA	22	GRADO (EL)	9,30	10,00	-0,70	-0,82
15	14	CINCA	31	MEDIANO	12,90	9,30	3,60	-0,78
16	15	CIURANA	15	CIURANA	,00	,00	0,00	-0,05
17	16	EBRO	17	EBRO	4,00	,60	3,40	-1,02
18	16	EBRO	32	MEQUINENZA	104,70	135,60	-30,90	-5,20
19	16	EBRO	40	RIBARROJA	184,80	185,50	-0,70	0,85
20	16	EBRO	46	SOBRON	,00	,00	0,00	0,00

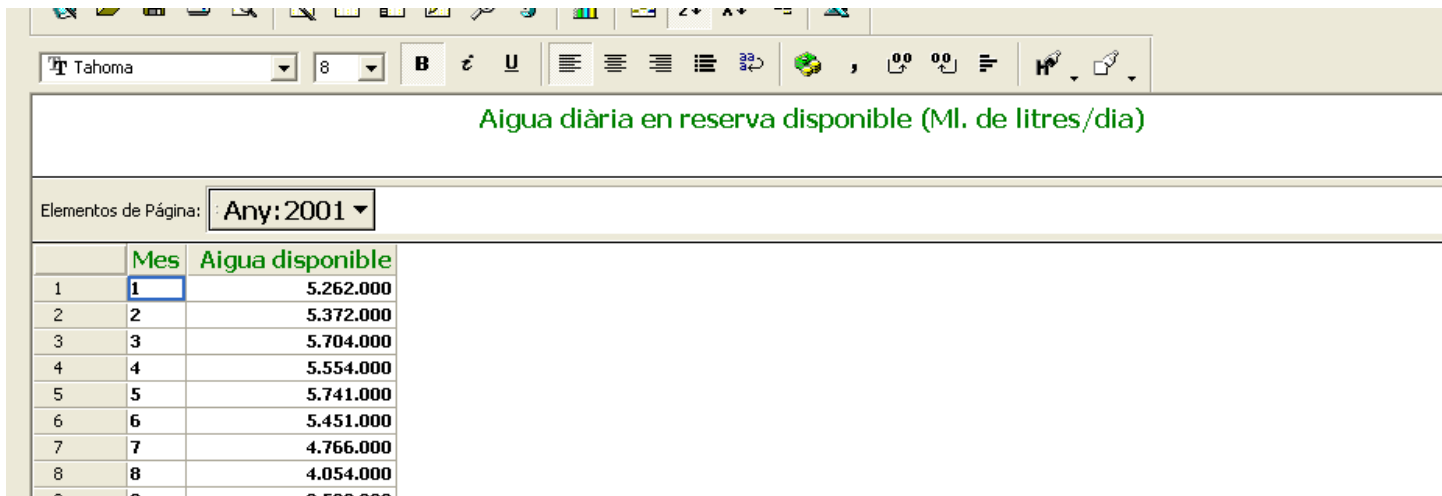
Figura 39: Disseny de l'informe aportació per riu, embassament i any - mes. Comparativa amb la mitjana anual d'aportació de l'embassament.

Aquest es el informe mes complexa doncs barreja informació de rius - embassament i embassament pròpiament dit. Concretament, mostra per un any i un mes seleccionat l'aportació d'un riu a un embassament. I com a dada consultiva ens informa de la mitja mensual de flux d'aportació per l'embassament.

Aquest informe, no deixa de ser els fulls excel d'entrada que han servit per carregar les taules de la base de dades, si mes no, aquells informaven embassament - riu, aquest ho fa riu - embassament i a més mostra la dada que hem calculat prèviament de mitja mensual d'aportació de l'embassament.

A la capçalera del informe, podem seleccionar el mes/any de la informació a extreure.

3.4.7.- Aigua diària en reserva disponible a tota la confederació.



Elements de Pàgina: Any: 2001

	Mes	Aigua disponible
1	1	5.262.000
2	2	5.372.000
3	3	5.704.000
4	4	5.554.000
5	5	5.741.000
6	6	5.451.000
7	7	4.766.000
8	8	4.054.000

Figura 40: Disseny de l'informe aigua en reserva disponible a tota la confederació.

Aquest informe ens mostra el total d'aigua disponible per dia. Con que la informació la tenim a nivell mes, no podem granularitzar mes aquesta dimensió i hem de mostrar la reserva disponible a nivell any - mes.

A la capçalera del informe, podem seleccionar l'any de la informació a extreure.

3.4.8.- Gràfic de l'evolució d'aigua de neu equivalent prevista per any i embassament.

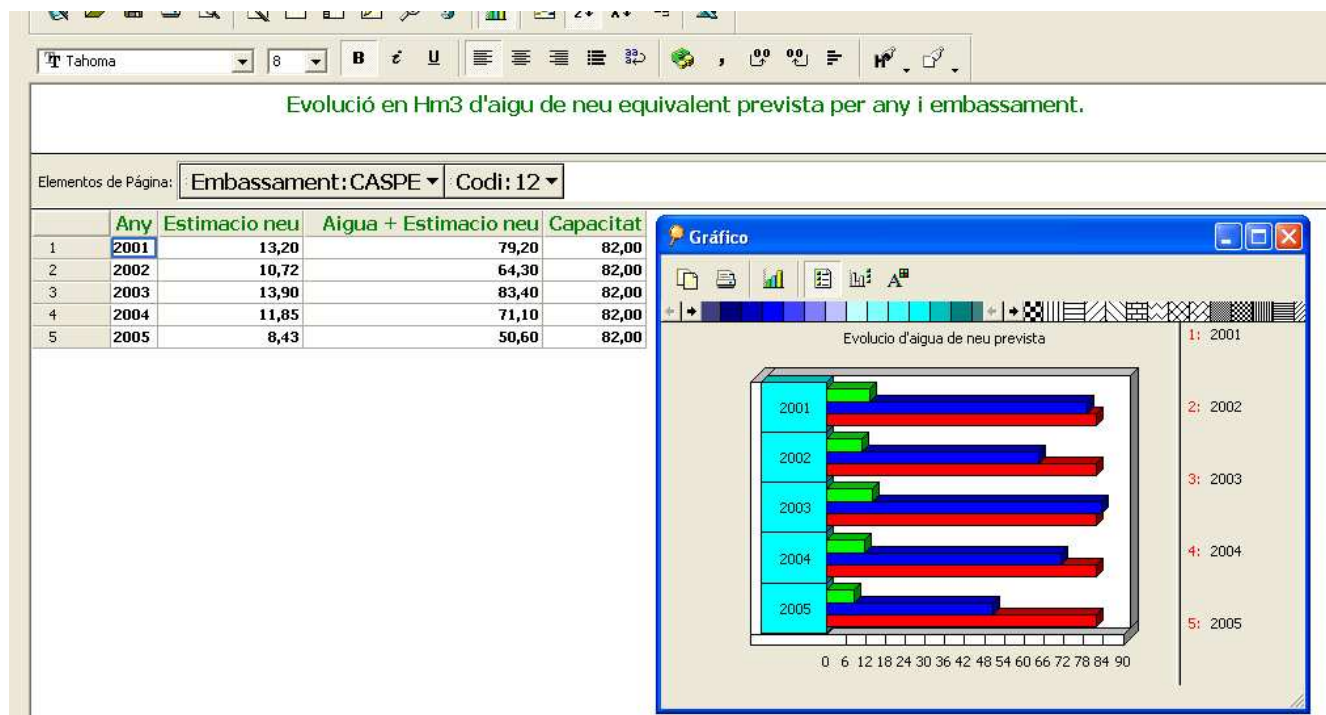


Figura 41: Disseny de l'informe amb gràfic de l'evolució d'aigua de neu equivalent prevista per any i embassament.

Informe amb gràfica associat on per un embassament prèviament seleccionat ens informa per cada any que tenim informació de l'estimació de neu, de l'aigua embassada i de la capacitat total per aquell embassament. Amb aquestes dades es crea una gràfica on es reflexa el risc de desbordament d'aquell embassament.

En l'exemple que mostrem, la franja verda mostra l'estimació de neu, la blava d'aigua + estimació de neu i la vermella de la capacitat. Si la blava rabassa a la vermella (com en l'exemple en l'any 2003) hi ha risc de desbordament. Tampoc es fixa i absolut ja que la dada estimació de neu, es precisament això, una estimació.

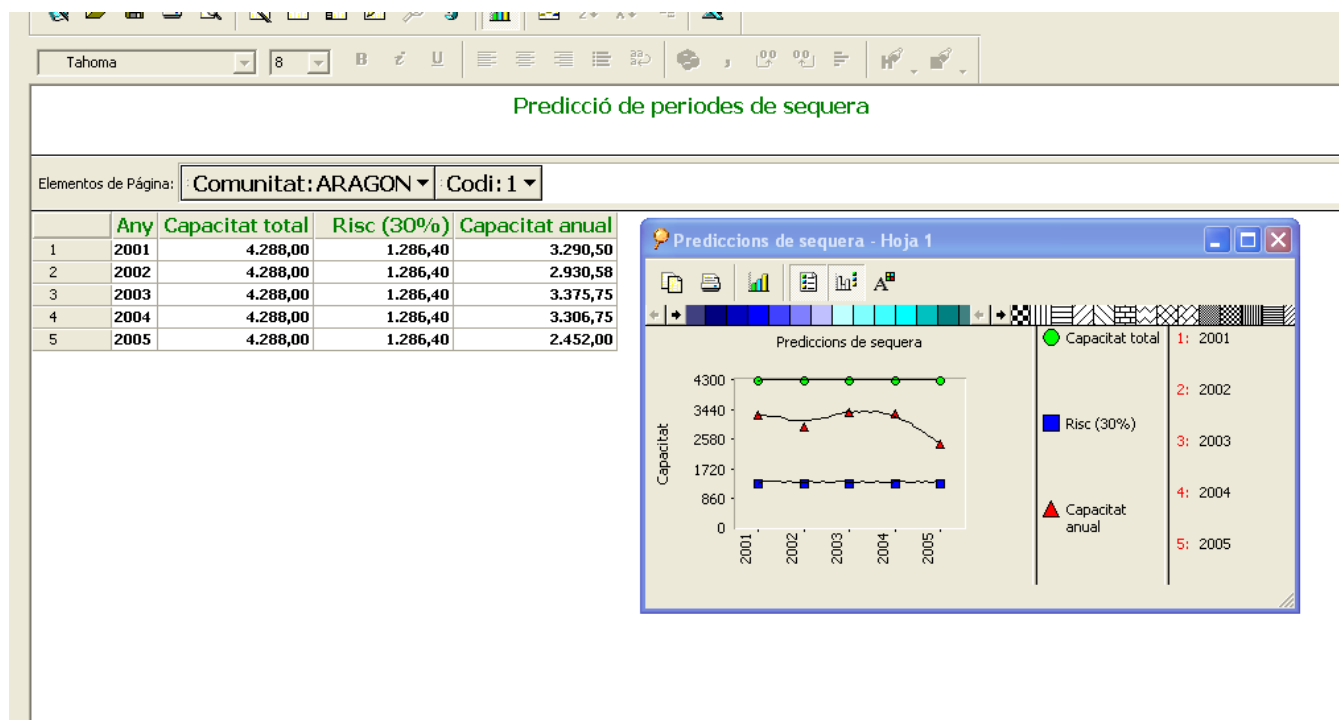
3.4.9.- Predicció de períodes de sequera.

Figura 42: Disseny de l'informe amb gràfic de prediccions de sequera.

Per predir períodes de sequera, s'han suposat dues premisses: la primera que la sequera la definirem a partir de Comunitats. Pot ser que a una comunitat hi hagi sequera però a la del costat no. I la segona premissa es que considerem sequera si la capacitat dels embassaments d'una CA no supera el 30% de la capacitat total. Aquestes premisses, ens donaran la informació de si puntualment una CA té sequera, però ens interessa poder predir i per això s'ha de veure la tendència de les capacitats i si aquesta es va apropant al llindar del 30%.

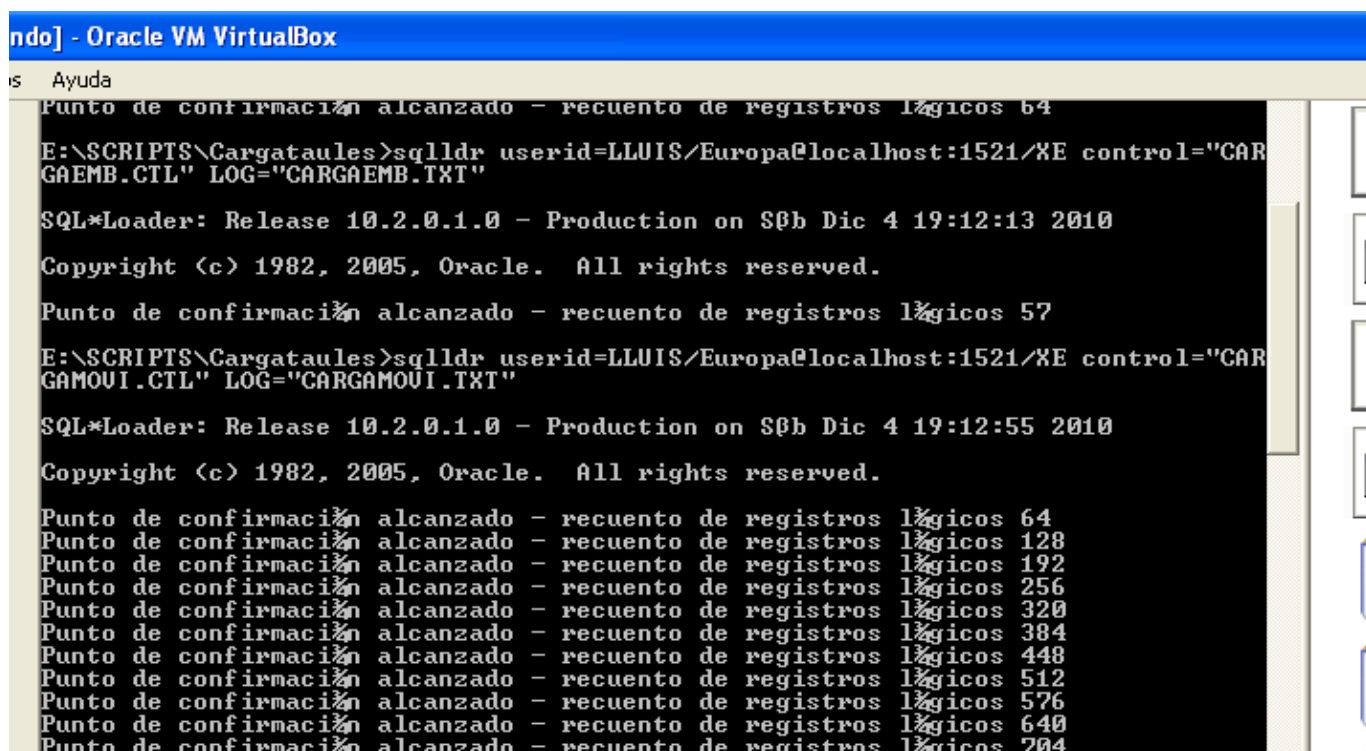
En el cas de l'exemple que mostrem, es veu clarament que per l'Aragó si es repeteixen pel període 2006-2010 les dades del 2001-2005, entraria en sequera.

4.- CAPTURES DE PANTALLES.

En aquest capítol, mostrarem diferents captures de pantalles per que la part explicada de carrega de dades no sigui tant teòrica i es pugui fer una idea mes palpable.

4.1.-CÀRREGUES DE DADES.

Captura de pantalla on es fan dues carregues (la taula de movimens i la taula d'embassaments).



```
ndo] - Oracle VM VirtualBox
Ayuda
Punto de confirmación alcanzado - recuento de registros lógicos 64
E:\SCRIPTS\Cargataules>sqlldr userid=LLUIS/Europa@localhost:1521/XE control="CAR
GAEMB.CTL" LOG="CARGAEMB.TXT"

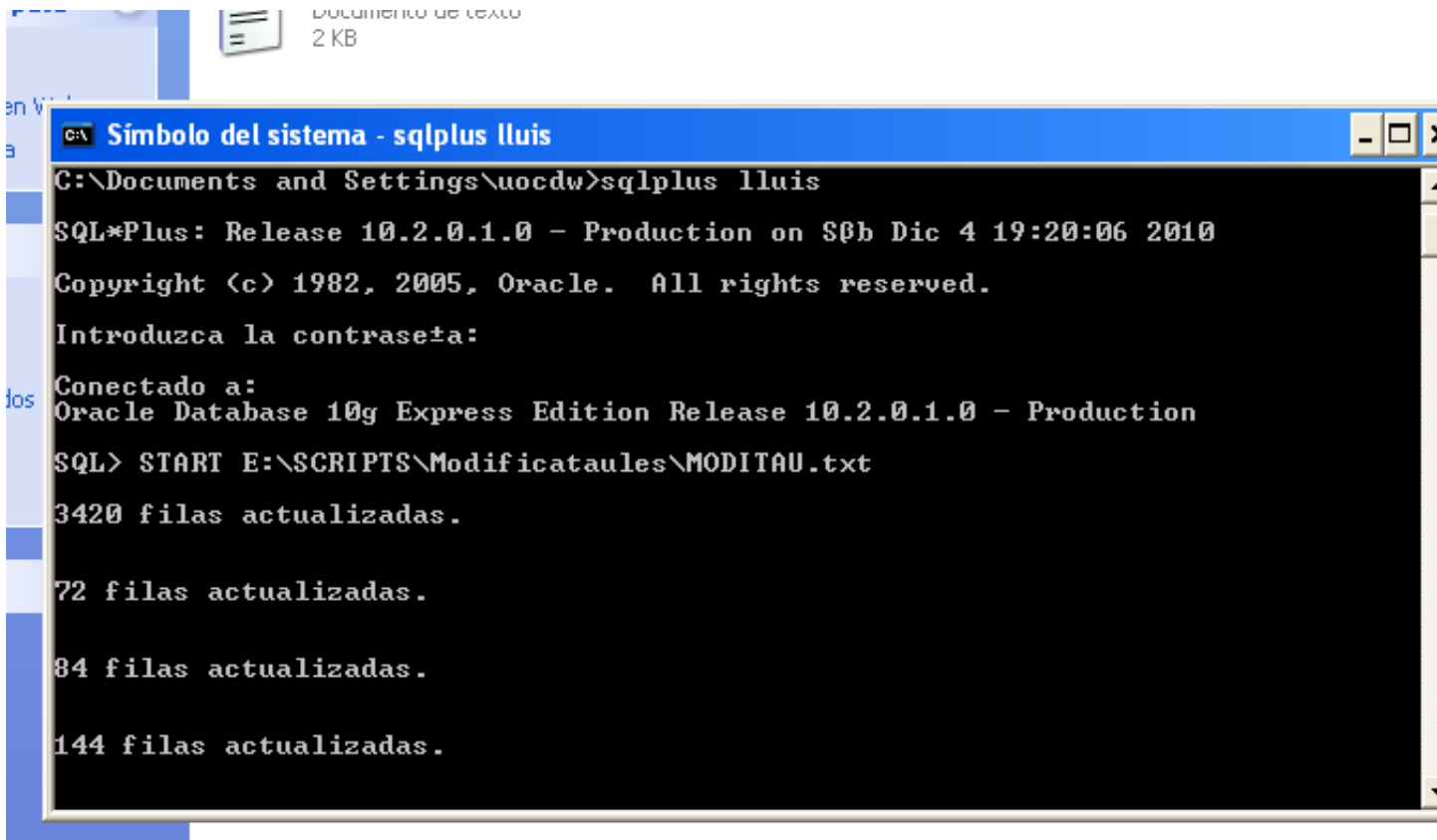
SQL*Loader: Release 10.2.0.1.0 - Production on Sòh Dic 4 19:12:13 2010
Copyright (c) 1982, 2005, Oracle. All rights reserved.
Punto de confirmación alcanzado - recuento de registros lógicos 57
E:\SCRIPTS\Cargataules>sqlldr userid=LLUIS/Europa@localhost:1521/XE control="CAR
GAMOUI.CTL" LOG="CARGAMOUI.TXT"

SQL*Loader: Release 10.2.0.1.0 - Production on Sòh Dic 4 19:12:55 2010
Copyright (c) 1982, 2005, Oracle. All rights reserved.
Punto de confirmación alcanzado - recuento de registros lógicos 64
Punto de confirmación alcanzado - recuento de registros lógicos 128
Punto de confirmación alcanzado - recuento de registros lógicos 192
Punto de confirmación alcanzado - recuento de registros lógicos 256
Punto de confirmación alcanzado - recuento de registros lógicos 320
Punto de confirmación alcanzado - recuento de registros lógicos 384
Punto de confirmación alcanzado - recuento de registros lógicos 448
Punto de confirmación alcanzado - recuento de registros lógicos 512
Punto de confirmación alcanzado - recuento de registros lógicos 576
Punto de confirmación alcanzado - recuento de registros lógicos 640
Punto de confirmación alcanzado - recuento de registros lógicos 704
```

Figura 43: Captura de pantalla per la carrega de les taules EMBASSAMEN I MOVIMENS.

4.1.-ARRANJAMENT DE DADES I CREACIÓ DE TAULES.

Captura de pantalla on es fa aquest arranament de dades (també útil per la creació de taules explicat anteriorment).



```
C:\Documents and Settings\uocdw>sqlplus lluis
SQL*Plus: Release 10.2.0.1.0 - Production on 08 Dic 4 19:20:06 2010
Copyright (c) 1982, 2005, Oracle. All rights reserved.
Introduzca la contraseña:
Conectado a:
Oracle Database 10g Express Edition Release 10.2.0.1.0 - Production
SQL> START E:\SCRIPTS\Modificataules\MODITAU.txt
3420 filas actualizadas.
72 filas actualizadas.
84 filas actualizadas.
144 filas actualizadas.
```

Figura 44: Captura de pantalla per la modificació de la taula MOVIMENS.

5.- CONCLUSIONS I LINIES D'EVOLUCIÓ.

En aquest TFC s'ha pogut demostrar la potencialitat d'un DW en el aspecte de crear i dissenyar informes. Degut al relatiu poc volum de dades no s'ha pogut veure la velocitat real d'un producte de DW en recórrer tot una base de dades de milions de tuples i amb altres taules relacionades. Aquest, no ha sigut el cas però ens hem fet una idea de la potencia en el tractament de dades.

També s'ha pogut comprovar els altres productes associats de crear les relacions per obtenir informes en base a la base de dades creades.

A nivell personal, ha sigut una experiència enriquidora el poder treballar amb un producte amb tant de renom com Oracle.

Tal com s'ha apuntat en el punt 1.7, un cop finalitzada qualsevol aplicació, aquesta, ja pot ser millorada. En el nostre cas, si es tingues que fer una ampliació, millora, etc. Incidiríem en el aspecte de la captura de dades que alimenten el DW. Fins i tot, encara que els processos ETL fossin totalment automatitzats, cada vegada que ens proporcionessin un full excel mensual, el tindríem que revisar per veure que no hi ha cap incongruència que pugui afegir-la a la taula de problemes/solucions.

Per tant, com a principal millora seria el de crear una aplicació transaccional (que fins i tot podria donar-nos les dades a nivell diari o setmanal millorant la granularitat del DW en la dimensió temps) i que aquesta alimentes el DW ja automàticament sense tenir que fer un estudi de les dades de partida doncs totes les comprovacions i validacions lògiques ja es tindrien que haver fet en la nova aplicació.

Poden haver-hi altres millores en el aspecte d'obtenir altres informes, altres gràfics, etc. Però ja seria me feina del usuaris finals i de qui explota les dades que no pas de nosaltres.

6.- GLOSSARI.

Agregació: Activitat de combinar dades des de varies taules per formar una altre unitat de informació mes complexa. Es fa servir freqüentment per respondre a consultes d'un DW de forma mes ràpida i fàcil.

Datawarehouse (DW): Base de dades que emmagatzema una gran quantitat de dades integrats per ser utilitzats per anàlisis per usuaris especialitzats (qui prenen decisions en les empreses).

Dimensió: Entitat independent dins del model multidimensional d'una organització, que serveix com una clau de recerca o com un mecanisme de selecció de dades.

ETL(Extracció, Transformació i Càrrega –load- de dades): Etapes per les que travessen les dades per anar des d'un sistema OLTP (o font de dades utilitzada) al DW dimensional. Per extracció, s'entén al mecanisme pel que les dades son llegides des de la font original. Transformació o neteja, es l'etapa per estandarditzar dades de diferents fonts, normalitzar-les, estructurar-les i netejar de possibles incongruències. L'última etapa, el transport, consisteix en portar les dades llegides i transformades al DW.

Fets: Entitats que es volen analitzar. Aquest anàlisis en un DW sempre s'intenta fer des de varis punts de vista (dimensions) diferents.

Granularitat: Nivell de detall d'una taula. Pot ser tant una taula de fets com una taula de dimensions.

OLTP (on-line Transaction Processing): Sistema transaccional diari (o en detall) que manté les dades operacionals de negoci.

ORACLE: Producte comercial per desenvolupar i explotar un DW.

Taula: Agrupació de dades amb característiques comuns.

Taula de dimensions: Agrupacions de dades per analitzar uns fets en concrets.

Taula de fets: Agrupacions de dades a ser analitzades des de uns punts de vista diferents.

7.- BIBLIOGRAFIA.

7.1.- PUBLICACIONES.

- Material UOC per desenvolupar el TFC.
- Material UOC de l'assignatura Minera de Dades.
- Material UOC de l'assignatura Base de Dades I.
- Material UOC de l'assignatura Comp. comunicativa per a professionals de la informàtica.

7.2.- WEBS.

http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos

<http://enfoquepractico.com/category/datawarehouse/>

<http://www-oei.eui.upm.es/Asignaturas/BD/DYOBD/DataWarehouse.pdf>

<http://informationmanagement.wordpress.com/2007/11/19/disenio-de-un-data-warehouse-estrella-y-copo-de-nieve/>

<http://www.inf.udec.cl/~revista/ediciones/edicion5/datawh.PDF>

8.- ANNEXOS.

S'ha cregut convenient annexar els fitxers que s'han utilitzat tant per crear les taules, com per carregar-les i finalment per modificar-les (en el cas de l'estimació de neu).

- Fitxer per crear les taules en l'entorn Oracle (Annexat a aquest document). Aquest fitxer es troba a la carpeta "CREATAULES"
- Un fitxer amb les dades a carregar cada taula creada (en total set fitxers –annexats a aquest document-). Aquests fitxers es troben a la carpeta "DADES A CARREGAR".
- Uns fitxers per utilitzar la utility SQLLOADER i poder carregar dins les taules els fitxers del punt anterior. En els fitxers .ctl indica la ruta on està cada fitxer de dades a carregar. Aquests fitxers es troben a la carpeta "CARGATAULES".
- Uns fitxers on es veu el resultat de les carregues anteriors. Aquests fitxers es troben a la carpeta "RESULTATS".
- Un fitxer per calcular i carregar l'estimació de neu per embassament - any (Annexat a aquest document). Aquest fitxer es troba a la carpeta "MODIFICATAULES".