

Construcció i explotació d'un magatzem de dades d'informació cinematogràfica

Diego Guerrero Martínez
ETIG

José Ángel Martín Carballo
13 de juny de 2011

Dedicatòria

Aquest treball està dedicat a la meva dona Maribel i els nostres fills Dídac, 6 anys, i Gisela, 3 anys, per donar-me força per a seguir endavant, pel seu suport en els moments dolents, la seva alegria en els bons i la comprensió durant aquest llarg cinc anys de sacrificis i caps de setmana a casa. Sense ells no hagués tingut l'empenta suficient per arribar al final del camí.

Resum

Aquest document és la memòria del treball de de final de carrera que consisteix en la creació d'un magatzem de dades amb informació de les nominacions i premis de diferents festivals de cinema.

Posteriorment a una lectura detinguda de les característiques i requeriments del projecte, es va preparar la planificació i elaborat l'anàlisi amb el que es van desenvolupar aquest treball.

Partint de la informació continguda a uns fulls de càlcul, s'extreu amb estratègies d'ETL per a ser carregada a l'estructura del magatzem de dades prèviament creada.

Finalment es van desenvolupar els informes identificats a la fase de presa de requeriments.

Índex de continguts

Dedicatòria	1
Resum	2
Índex de continguts	3
Índex de Figures	5
Memòria	6
1. Introducció	6
1.1. Justificació	6
1.2. Objectius	6
1.3. Enfocament i mètode seguit	7
1.4. Planificació del projecte	8
1.5. Productes obtinguts	12
1.6. Breu descripció dels altres capítols de la memòria	13
1.6.1. Anàlisi	13
1.6.2. Disseny	13
1.6.3. Conclusions	13
1.6.4. Línies d'evolució futur	13
1.6.5. Glossari	13
1.6.6. Bibliografia	13
1.6.7. Annexos	13
2. Anàlisi	14
2.1. Casos d'ús	14
2.2. Model conceptual	15
3. Disseny	19
3.1. Diagrama de l'arquitectura del software	19
3.2. Diagrama de l'arquitectura del hardware	20
3.3. Disseny de la base de dades	21
3.4. Disseny i descripció dels informes creats	24
3.4.1. Premis per artista, any i festival	24
3.4.2. Premis per artista, any, festival i categoria	26
3.4.3. Premis a un director per any, festival i film	27
3.4.4. Premis per any, festival, categoria i pel·lícula	28
3.4.5. Premis per any, pel·lícula i festival	29
3.4.6. Premis obtinguts per països i categories	30
3.4.7. Premis per any i festival	31
3.4.8. Èxit o fracàs d'una pel·lícula	33
3.4.9. Relació entre premis	34
4. Conclusions	35

5. Línies d'evolució futura.....	35
Glossari	36
Bibliografia.....	37
Annexos.....	38
A. Creació d'estructures.....	38
B. Càrrega de dades.....	41
C. Transformació i càrrega de les taules auxiliars	43
D. Càrrega de les taules de fets	45
E. Accés a la Base de Dades	47
F. Oracle BI Discoverer Administrator	50
G. Oracle BI Discoverer Desktop	52

Índex de Figures

Figura 1 – Cicle de vida Kimball.....	7
Figura 2 – Digrama de Gannt I.....	10
Figura 3 – Digrama de Gannt II.....	11
Figura 4 – Cas d'ús Administrador IT.....	14
Figura 5 – Cas d'ús Administrador Funcional.....	14
Figura 6 – Cas d'ús Usuari.....	15
Figura 7 – Digrama de Fets Fet_Artista.....	16
Figura 8 – Digrama de Fets Fet_Film.....	17
Figura 9 – Digrama de Fets Fact_Festival.....	17
Figura 10 – Digrama de Fets Fact_Director.....	18
Figura 11 – Diagrama de l'arquitectura del software.....	19
Figura 12 – Arquitectura del hardware.....	20
Figura 13 – Arquitectura del hardware virtualitzat.....	21
Figura 14 – Premis per actor, any i festival.....	24
Figura 15 – Premis per actriu, any i festival.....	25
Figura 16 – Premis per director, any i festival.....	25
Figura 17 – Premis per artista, any, festival i categoria.....	26
Figura 18 – Premis i nominacions dels directors.....	27
Figura 19 – Premis per any, festival, categoria i film.....	28
Figura 20 – Premis per any, film i festival.....	29
Figura 21 – Països ordenats per premis obtinguts.....	30
Figura 22 – Desglossament de les categories del país amb més premis.....	31
Figura 23 – Petició de l'any a consultar.....	32
Figura 24 – Actors, actrius i films nominats per festival segons l'any demanat.....	32
Figura 25 – Films amb èxit o fracàs.....	33
Figura 26 – Relació entre premis.....	34
Figura 27 – Creació de les taules.....	39
Figura 28 – Creació de les seqüències.....	39
Figura 29 – Creació dels disparadors.....	40
Figura 30 – Fitxer Excel per a carregar.....	42
Figura 31 – Càrrega de les dades.....	42
Figura 32 – Fitxer de control de càrrega de les dades dels Òscars.....	43
Figura 33 – Accés a la BB.DD.....	47
Figura 34 – Entrada a la BB.DD.....	48
Figura 35 – Direcció IP de la MV.....	48
Figura 36 – Connexió SQL Developer.....	49
Figura 37 – Detall de les taules des de SQL Developer.....	49
Figura 38 – Accés a Discoverer.....	50
Figura 39 – Creació de l'EUL.....	50
Figura 40 – Seleccionar usuari EUL.....	51
Figura 41 – Creació final de l'EUL.....	51
Figura 42 – Objectes creats per Discoverer.....	52
Figura 43 – Accés a Discoverer Desktop.....	52
Figura 44 – Obrir un llibre de treball.....	53
Figura 45 – Llibre de treball.....	53

Memòria

1. Introducció

1.1. Justificació

Aquest projecte és l'encàrrec de l'Acadèmia de Cinema Andorrà (ACA) consistent en la creació d'un magatzem de dades amb informació de les nominacions i premis de diversos festivals internacionals de cine al llarg del temps.

La informació ha estat recopilada per l'ACA en fulls de càlcul Excel. Aquesta informació és difícil de explotar, consisteix en una tasca molt manual i no ofereix la consistència i fiabilitat necessària.

Es demana la càrrega de la informació dels fulls de càlcul a un repositori de dades multidimensional. La informació necessitarà un procés de depuració i unificació previ ja que les fulls de càlcul venen en diferents idiomes. Per exemple s'hauran d'unificar els noms dels països i de les categories dels premis. Finalment, les dades estaran en disposició de ser creuades i explotades en els diferents informes, demanats per l'ACA, mostrant la informació de forma agregada.

1.2. Objectius

L'objectiu fonamental d'aquest treball és l'adquisició d'experiència en el disseny, construcció i explotació de dades d'un magatzem de dades (Data Warehouse - DW) multidimensional partint d'informació disponible en una base de dades transaccional.

Hi ha, però, uns altres objectius inherents al projecte:

- Gestió i planificació d'un projecte informàtic.
- Anàlisi d'un problema complex.
- Assolir un major coneixement del llenguatge SQL.
- Coneixement d'estratègies d'ETL.
- Explotació de la informació mitjançant l'elaboració d'informes d'usuari final fent servir l'eina de reporting OLAP Oracle BI Discoverer.
- Presentació de la memòria final del projecte.
- Creació d'una presentació final que mostri el treball realitzat.

1.3. Enfocament i mètode seguit

La metodologia que s'ha fet servir per a portar a terme aquest treball ha estat basada en el mètode Kimball gràcies a la seva versatilitat i visió ascendent de la construcció del magatzem de dades de forma incremental.

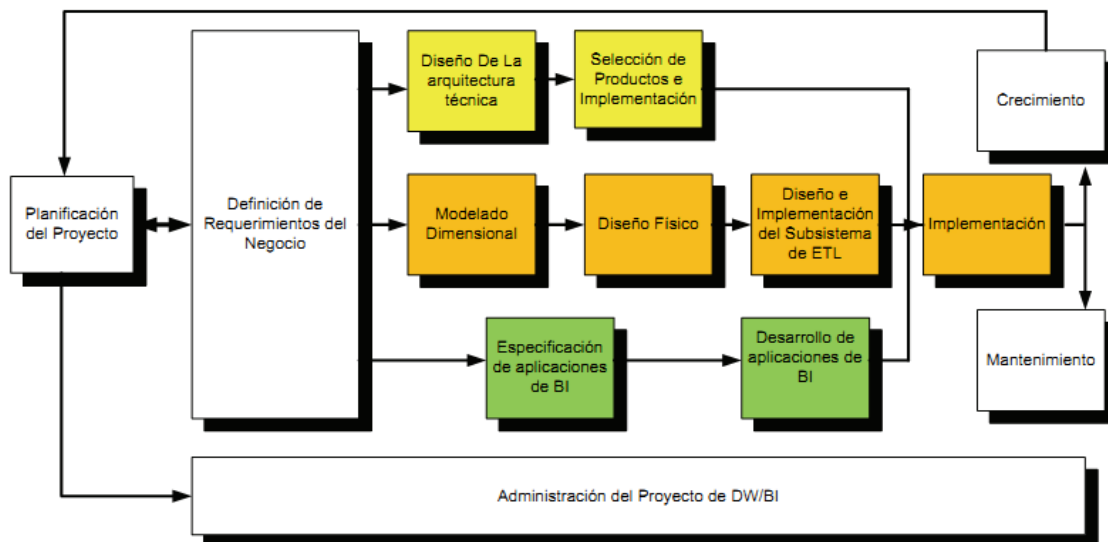


Figura 1 – Cicle de vida Kimball

A primera vista es pot observar la part la rellevància que se li atorga a la planificació del projecte i sobre tot a la definició dels requeriments del magatzem de dades. Posteriorment es veuen clarament definits tres camins. El superior té a veure amb la tecnologia que es farà servir. En aquest cas ens ve donada i és el SGBD Oracle Express 10g. El camí inferior està relacionat amb les eines d'exploració de la informació. També es defineix al principi del projecte que es farà servir l'Oracle BI Discoverer i es demanen els informes que s'hauran de crear. El camí del mig és el que s'ha fet servir, hi està especificat el disseny i implementació del model multidimensional i del sistema d'extracció, transformació i càrrega de dades (ETL).

- Planificació: es defineix l'abast del projecte. Es van identificar i planificar les tasques que es van dur partint d'un anàlisi previ dels requisits del projecte. Finalment es va presentar un document amb el pla del projecte.
- Anàlisi i disseny: enteniment del problema plantejat i disseny de la solució. És la part més densa del projecte on s'estableixen les bases per a la construcció del magatzem de dades. A la vegada ha estat subdividit:
 - Anàlisi de requeriments: es va tractar d'entendre el problema a partir de la documentació entregada on queden plasmats els requisits del client.
 - Model multidimensional: aquest és un procés iteratiu. Partint d'un model d'alt nivell, es va establir que el nivell de detall seria al màxim que permetessin els fulls de càlcul entregats, es van escollir les dimensions adients i es van identificar els seus atributs i els indicadors de les taules de fets. Al final d'aquesta fase es van obtenir els diagrames de les taules de fets identificades, les dimensions i els atributs necessaris.

- Disseny del subsistema d'ETL: definició i disseny del processos necessaris per tal d'extreure la informació dels fitxers Excel, depurar-la i carregar-la a les taules identificades a la fase de disseny.
- Implementació: es porta a terme el disseny, creant els objectes i processos definits a les fases anteriors. És una fase totalment incremental i iterativa. Començant per una part, es van anar desenvolupant pilots per a provar la implementació. En aquesta fase es poden trobar errades de disseny que faran que s'hagi de canviar per tal de donar solució al problema inicial.

1.4. Planificació del projecte

Aquest projecte s'ha planificat tenint en compte les dates establertes per les entregues de les fites.

Es van identificar les tasques, les dependències entre elles i els dies necessaris per a dur-les a terme que es mostren a la taula:

Id	Nom Tasca	Dies	Inici	Fi	Prec
0	0. TFC Magatzem de Dades	93	02/03/2011	23/06/2011	
1	1. TFC Magatzem de Dades	3	02/03/2011	04/03/2011	
2	1.1. Inici del Semestre	1	02/03/2011	02/03/2011	
3	1.2. Lectura del pla docent	2	03/03/2011	04/03/2011	2
4	2. Preparació del entorn informàtic	2	04/03/2011	05/03/2011	
5	2.1. Instal·lació de software necessari	1	04/03/2011	04/03/2011	
6	2.2. Preparació de les eines de backup	1	05/03/2011	05/03/2011	5
7	3. PAC1 - Pla de treball	11	05/03/2011	16/03/2011	
8	3.1. Lectura i comprensió del TFC	2	05/03/2011	06/03/2011	
9	3.2. Descàrrega i instal·lació màquina virtual	2	07/03/2011	08/03/2011	8
10	3.3. Lectures relacionades amb els DW	4	07/03/2011	10/03/2011	8
11	3.4. Començament de l'elaboració del pla de treball	1	11/03/2011	11/03/2011	10
12	3.5. Assistència a la trobada a la seu de la UOC	1	12/03/2011	12/03/2011	11
13	3.6. Finalització del pla de treball	1	13/03/2011	13/03/2011	11
14	3.7. Creació del document de la PAC1	1	15/03/2011	15/03/2011	13
15	3.8. Revisió i lliurament de la PAC1	0	16/03/2011	16/03/2011	
16	4. PAC2 - Requeriments i disseny	28	17/03/2011	20/04/2011	
17	4.1. Lectura detallada de l'enunciat	1	17/03/2011	17/03/2011	
18	4.2. Recerca d'informació sobre el tema exposat	2	18/03/2011	20/03/2011	17
19	4.3. Anàlisi de les dades a carregar	2	21/03/2011	22/03/2011	18
20	4.4. Anàlisi dels requeriments	3	23/03/2011	25/03/2011	19
21	4.5. Creació del disseny tècnic	8	26/03/2011	05/04/2011	20

Id	Nom Tasca	Dies	Inici	Fi	Prec
22	4.6. Elaboració de l'esborrany	11	28/03/2011	10/04/2011	20
23	4.7. Entrega de l'esborrany	1	11/04/2011	11/04/2011	22
24	4.8. Revisió de les correccions del tutor	6	12/04/2011	19/04/2011	23
25	4.9. Creació del document final de la PAC2	6	13/04/2011	20/04/2011	23
26	4.10. Entrega de la PAC2	0	20/04/2011	20/04/2011	25
27	5. PAC3 - Construcció del DW	28	21/04/2011	25/05/2011	
28	5.1. Lectura detallada de l'enunciat	1	21/04/2011	21/04/2011	
29	5.2. Revisió del document de disseny	1	22/04/2011	22/04/2011	28
30	5.3. Preparació del SGBD	1	23/04/2011	23/04/2011	29
31	5.4. Creació de les taules necessàries	2	24/04/2011	25/04/2011	30
32	5.5. Processos ETL: neteja de dades i scripts de càrrega.	3	26/04/2011	28/04/2011	31
33	5.6. Entrenament amb Discoverer	2	29/04/2011	02/05/2011	32
34	5.7. Creació d'informes amb Oracle Discoverer	3	03/05/2011	05/05/2011	33
35	5.8. Elaboració de l'esborrany	7	06/05/2011	13/05/2011	34
36	5.9. Entrega de l'esborrany	1	18/05/2011	18/05/2011	35
37	5.10. Revisió de les correccions del tutor	5	19/05/2011	25/05/2011	36
38	5.11. Creació del document final de la PAC2	4	20/05/2011	25/05/2011	36
39	5.12. Entrega de la PAC3	0	25/05/2011	25/05/2011	38
40	6. Memòria del projecte	14	26/05/2011	13/06/2011	39
41	6.1. Revisió de la documentació lliurada durant el semestre	2	26/05/2011	27/05/2011	
42	6.2. Creació del document de memòria	4	30/05/2011	02/06/2011	41
43	6.3. Creació de la presentació	6	03/06/2011	10/06/2011	42
44	6.4. Revisió dels documents finals	2	11/06/2011	13/06/2011	43
45	6.5. Entrega de la memòria	0	13/06/2011	13/06/2011	44
46	7. Debat Virtual	4	20/06/2011	23/06/2011	
47	7.1. Participació al debat virtual	4	20/06/2011	23/06/2011	

El diagrama de Gantt resultant és el següent:

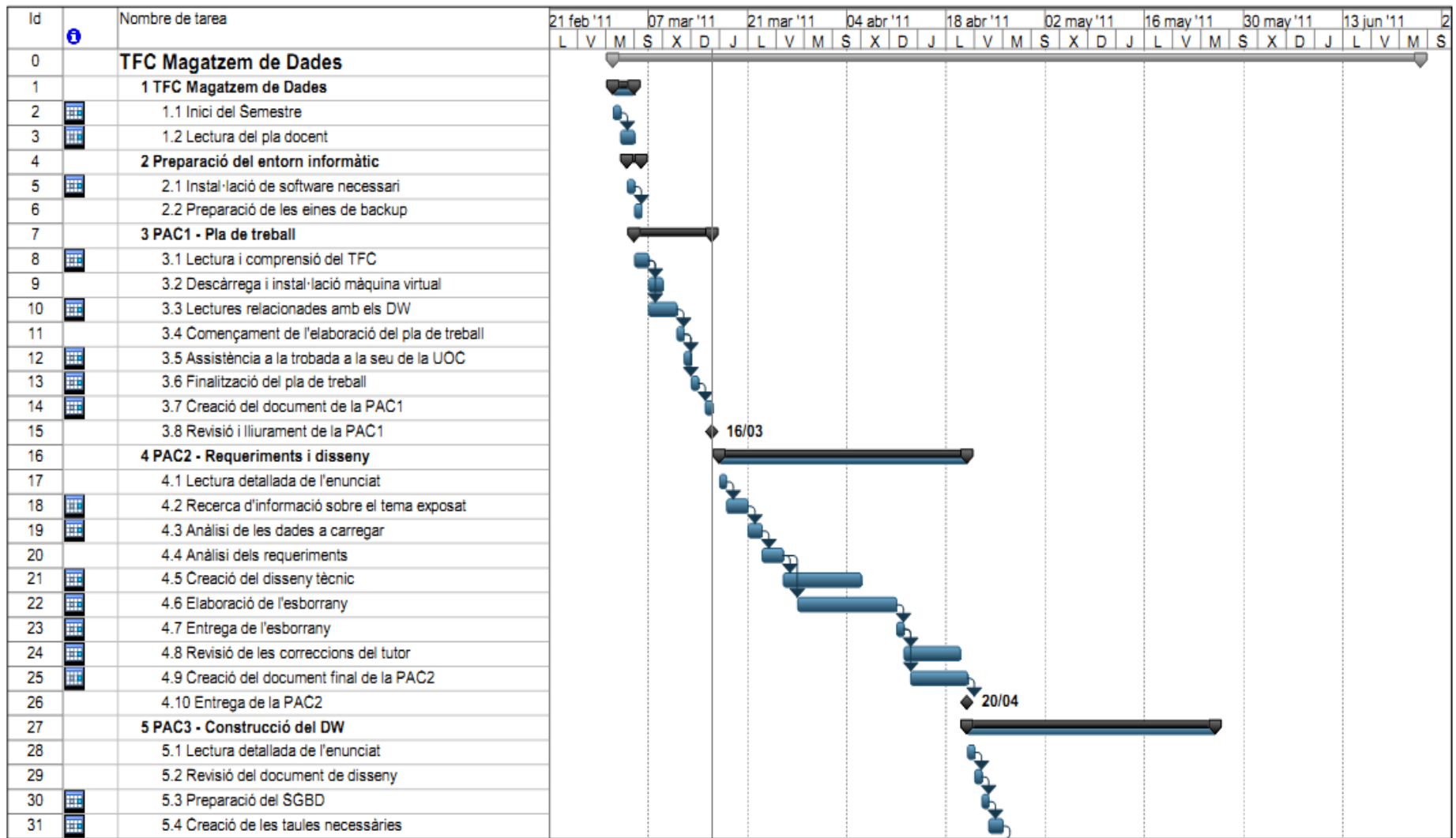


Figura 2 – Digrama de Gannt I

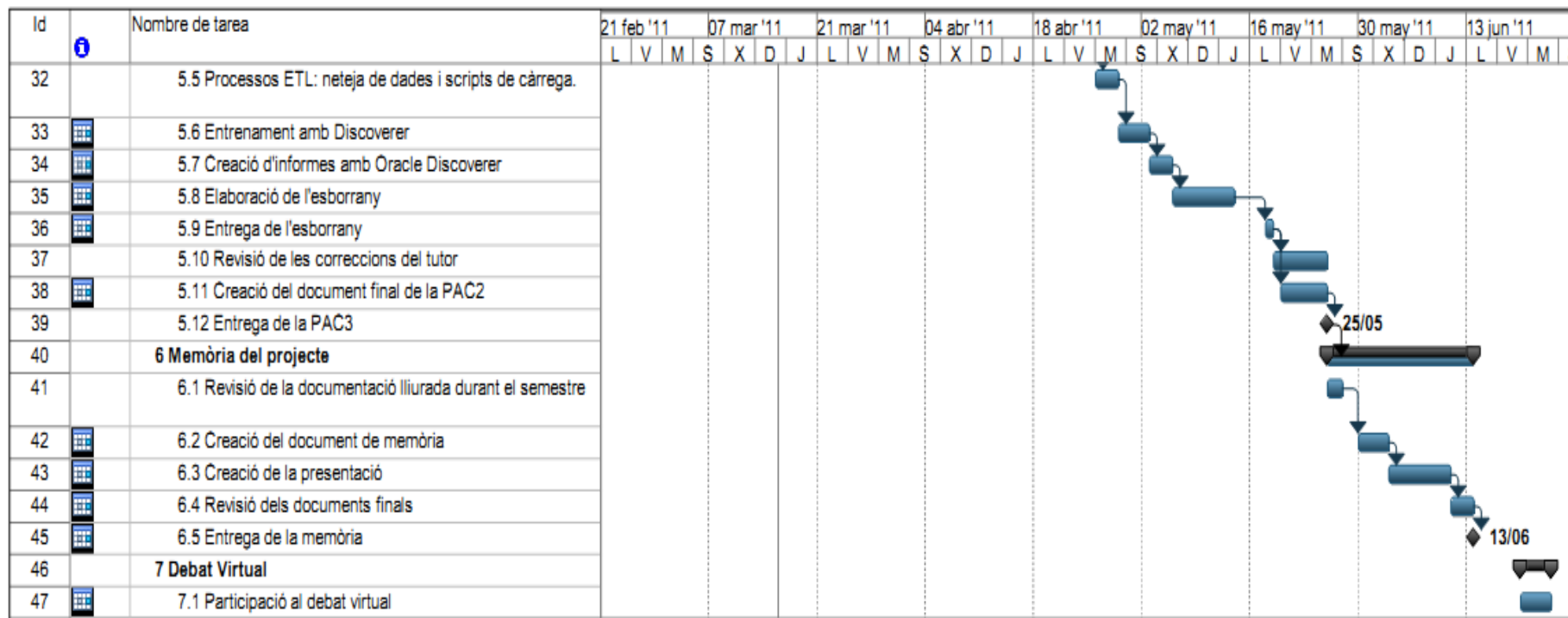


Figura 3 – Digrama de Gannt II

1.5. Productes obtinguts

A continuació es presentaran els productes que s'han obtingut en el desenvolupament d'aquest treball final de carrera. Aquests productes són processos de creació de taules, de càrrega de dades, informes, documents entregats a cada fita, una imatge d'un disc dur d'una màquina virtual, una presentació i un video on es fa la presentació del treball.

Producte	Descripció
Pla de treball	Establiment del pla de treball del TFC entregat a la PAC1.
Anàlisi de requeriments i disseny conceptual	Estudi dels requeriments, anàlisi i disseny conceptual del projecte, entregat a la PAC2.
Procés de càrrega de dades	Explicació de la creació i càrrega de la base de dades entregada a la PAC3.
Procés d'accés a la BB.DD.	Explicació de l'accés a la BB.DD. entregada a la PAC3
Procés Oracle Discoverer	Explicació de l'accés a l'Oracle Discoverer, descripció i mostra dels informes creats. També a la PAC3.
ACA	La base de dades que s'ha fet servir en el projecte.
crea_estructura.bat	Executable que crida els processos de creació de les taules i atributs a la base de dades.
carrega_dades.bat	Executable que crida la càrrega de les dades dels fulls de càlcul al repositori central.
01_tipus_artista.bat	Crea els diferents tipus d'artista
02_carrega_artista.bat	Carrega els noms dels actors, actrius i directius a la taula Artista.
03_carrega_pais.bat	Carrega els països a la taula País.
04_insert_categoria.bat	Carrega les categories a la taula Categoria.
05_insert_festival.bat	Carrega els noms dels festivals a la taula Festival.
06_insert_film.bat	Carrega els títols originals i d'ordre dels films
07_insert_any.bat	Carrega els anys en els que s'han celebrat els festivals.
08_fact_artista.bat	Carrega la taula de fets Fact_Artista.
09_fact_film.bat	Carrega la taula de fets Fact_Film.
10_fact_festival.bat	Carrega la taula de fets Fact_Festival.
11_fact_director.bat	Carrega la taula de fets Fact_Director.
insert_categoria.bat	Assigna les categories als artistes.
ACA.DIS	Llibre de treball d'Oracle BI Discoverer que conté els informes creat en aquest projecte.
dguerrerom_memoria.pdf	Memòria d'aquest TFC.
dguerrerom_mv.zip	Imatge de disc dur d'una màquina virtual Oracle MV VirtualBox amb la base de dades, fitxers de càrrega i informes generats.
dguerrerom_memoria.mp4	Vídeo amb la presentació virtual del TFC.

1.6. Breu descripció dels altres capítols de la memòria

Seguidament es passarà a descriure breument els següents capítols que es troben a la memòria.

1.6.1. Anàlisi

Es mostrarà l'anàlisi del problema plantejat, una descripció dels casos d'ús i el diagrama del model conceptual.

1.6.2. Disseny

Prenent l'anàlisi com a punt de partida, es presentarà el disseny de la solució proposada, amb els diagrames de les arquitectures de software i hardware, el disseny de la base de dades, el diagrama del model i la descripcions del informes creats.

1.6.3. Conclusions

Es farà l'explicació de les conclusions a les que s'ha arribat una vegada finalitzat el treball.

1.6.4. Línies d'evolució futur

Es faran unes pinzellades de l'evolució que pot tenir el magatzem de dades creat.

1.6.5. Glossari

Un petit glossari de termes que s'han fet servir en aquesta memòria.

1.6.6. Bibliografia

Es llistarà la bibliografia utilitzada en aquest treball.

1.6.7. Annexos

Es mostraran els processos de creació de les estructures a la base de dades, els d'ETL efectuats al projecte i els accessos a la base de dades i a Oracle Discoverer.

2. Anàlisi

2.1. Casos d'ús

Un diagrama de casos d'ús mostra la relació que hi ha entre les funcionalitats que ofereix un sistema i els actors que el fan servir.

Un actor normalment és una persona que fa servir el sistema, encara que també podria ser una màquina o un altre sistema interactuant. Un cas d'ús descriu les interaccions que hi han entre el sistema i els actors a partir de les operacions que poden fer.

S'han identificat tres actors:

- **Administrador IT:** administra la base de dades, fa el manteniment de l'estructura del sistema i dels processos d'ETL i dels usuaris, carrega les dades, garanteix l'accessibilitat al sistema i fa els backups necessaris per una eventual pèrdua d'informació per, per exemple, una corrupció de les dades.

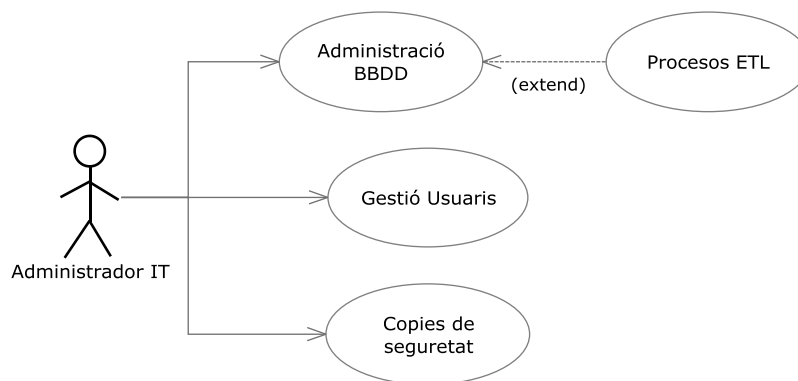


Figura 4 – Cas d'ús Administrador IT

- **Administrador Funcional:** s'encarrega del manteniment dels informes a l'Oracle BI Discoverer, vetlla per a que la informació mostrada sigui correcta i per la consistència de les estructures. Genera nous informes a petició dels usuaris.

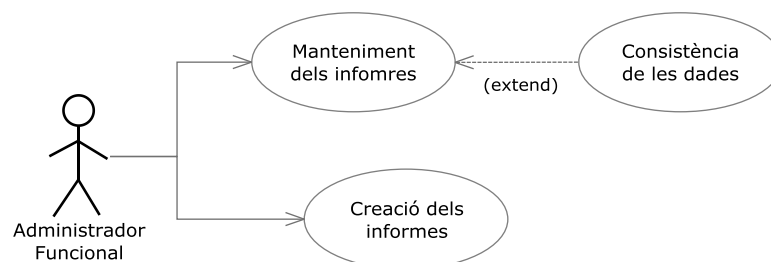


Figura 5 – Cas d'ús Administrador Funcional

- **Usuari:** és l'actor que analitza les dades a partir dels informes generats al Discoverer. Si convé, pot demanar la creació de nous informes a l'administrador funcional.

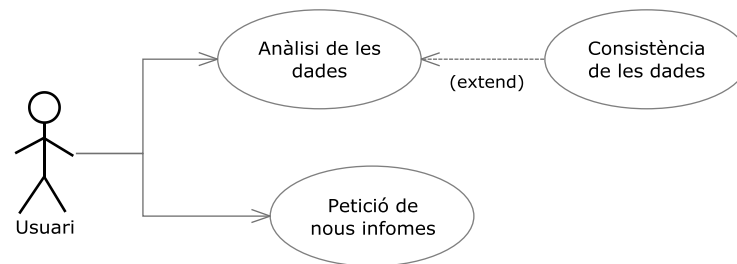


Figura 6 – Cas d'úsUsuari

2.2. Model conceptual

El model multidimensional està orientat al fet que serà l'objecte d'anàlisi. Consta de dimensions i fets. Es situen les dades en un espai n-dimensional on cada dimensió representa un punt de vista per analitzar la informació. Els fets són les dades que es volen analitzar. Les dimensions són formades per objectes homogenis entre sí. Per exemple, una dimensió País contindrà els països, una dimensió Any contindrà els números que representin un any. Aquestes dimensions es podran jerarquitzar al model multidimensional per a poder totalitzar les dades, navegant de menys a més detall (drill down) o al revés (drill up). Els fets que s'analitzen es troben al creuament dels objectes de les diferents dimensions.

Les set dimensions creades han he estat identificades a partir l'estructura dels fulls de càlcul entregats i són necessàries per a construir els indicadors requerits, Les dimensions són:

- Artista
- Tipus d'artista
- Categoria del premi
- Festival
- Film
- País
- Temps (Any)

Aquestes dimensions es combinaran per a formar les quatre taules de fets creades, una pels fets que tenen a veure amb els artistes, altra amb els corresponents de les pel·lícules, la tercera darrera amb els relacionats amb els festivals i la darrera amb informacions relatives als directors de pel·lícules.

A la taula de fets Fet_Artista es recolliran els indicadors nominacions i premis necessaris per aconseguir els nombres de nominacions i premis aconseguits per actors, actrius i directors. Les dimensions necessàries són:

- Artista, amb una relació amb tipus d'artista que marcarà si és un actor, actriu o director.
- Categoria del premi.
- Festival

- Any

El diagrama que el representa és el següent:

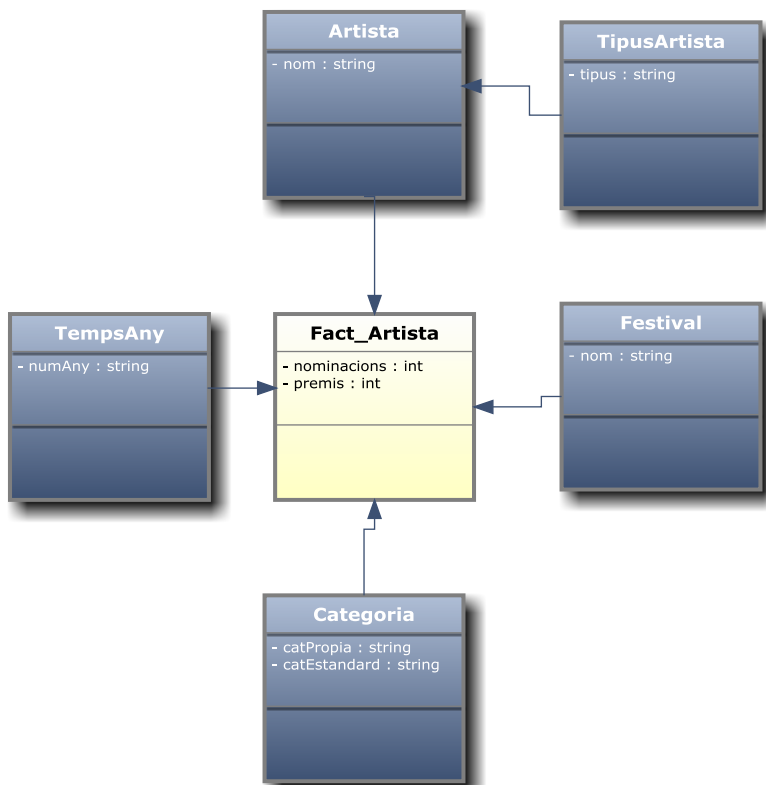


Figura 7 – Diagrama de Fets Fet_Artista

A la taula de fets Fet_Film es recolliran els indicadors nominacions i premis per aconseguir els nombres de premis i nominacions d'un film i el país que més premis aconsegueix ja que un film és produït per un país .Les dimensions necessàries són:

- Film
- Festival
- Any
- Pais
- Categoria

El diagrama que el representa és el següent:

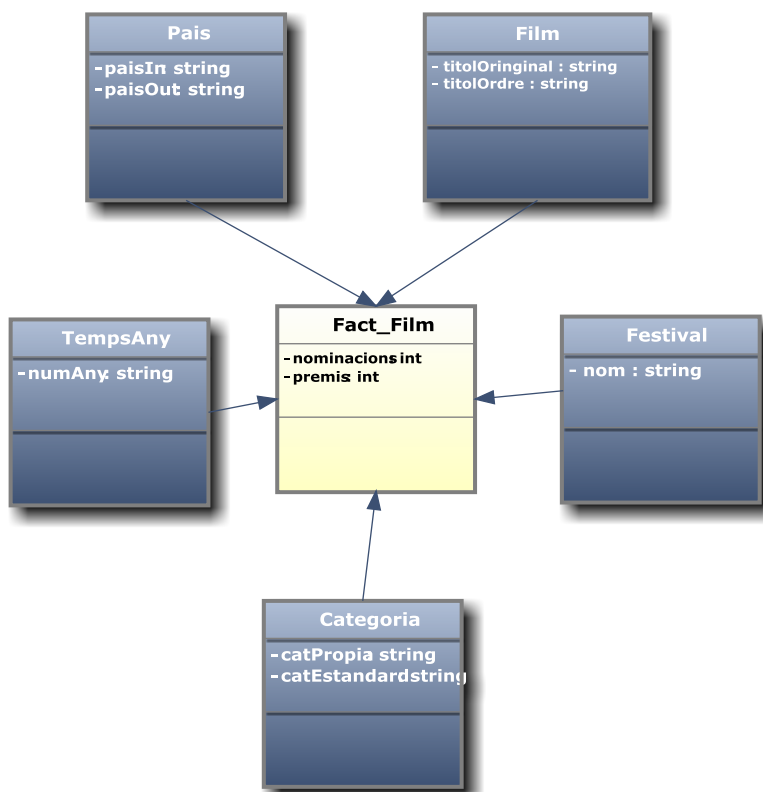


Figura 8 – Digrama de Fets Fet_Film

Partint d'aquest esquema es podrà identificar una pel·lícula exitosa o fracassada segons la relació percentual entre les nominacions i els premis aconseguits a partir d'un mínim de nominacions. Es pot considerar una pel·lícula exitosa el film que tenint 10 nominacions o més aconsegueix al menys el 70% dels premis als que opta i fracassada si aconsegueix menys del 50%.

A la taula de fets Fet_Festival es recolliran els indicadors necessaris per aconseguir els nombres d'actors, actrius i films nominats i el número d'edicions realitzades per una data .Les dimensions necessàries són:

- Festival
- Any

Per a cada any i festival es buscaran els valors que hauran de tenir els atributs que representen la quantitat d'actors, actrius i films nominats en aquell creuament.

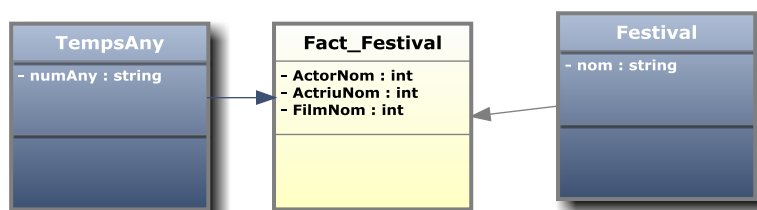


Figura 9 – Digrama de Fets Fact_Festival

Finalment, la taula Fact_Director emmagatzemarà aquella informació que ens sigui útil per a donar resposta a l'informe requerit sobre els directors

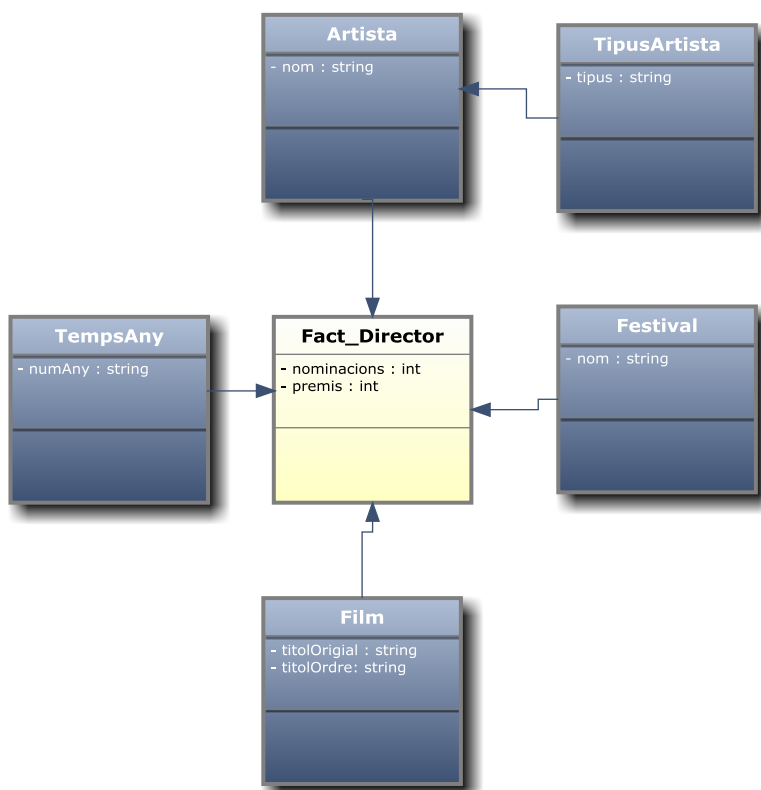


Figura 10 – Digrana de Fets Fact_Director

Les dimensions escollides són:

- Artista i Tipus d'artista: per a poder seleccionar els directors.
- TempsAny: representarà la data en la va tenir lloc el festival.
- Festival: és el festival en el que es fan les nominacions.
- Film: és la pel·lícula que ha fet el director.

3. Disseny

3.1. Diagrama de l'arquitectura del software

Aquest treball segueix l'esquema de l'arquitectura client-servidor en la que bàsicament la part client d'un software fa peticions a la part servidora que li dona les respostes. Aquesta arquitectura és profitosa en un sistema multiusuari a una xarxa d'ordinadors. Aporta l'avantatge de la centralització de les dades i els processos.

S'ha fet servir una màquina virtual Oracle VirtualBox on hi era instal·lat el sistema gestor de base de dades Oracle Database 10g Express Edition Release 10.2.0.1.0, i l'eina de creació d'informes Oracle BI Discoverer. També s'ha treballat des de la màquina amfitriona amb el programari Oracle SQL Developer connectant-se mitjançant la xarxa amb la màquina virtual.

El procés comença amb la càrrega dels fitxers Excels a un repositori central dins la base de dades. Aquestes dades alimenten les taules que seran les dimensions de les taules de fets. Després es completen els mapejos necessaris per a unificar les dades de les dimensions, per exemple els països són en diferents idiomes i s'han unificat en català per a que les dades siguin consistents i que Spain sigui igual a Espagne. Finalment s'executen el processos que omplen les taules de fets.

L'Oracle BI Discoverer es connecta a la base de dades mitjançant l'End User Layer. Discoverer crea una sèrie de taules a la base de dades que farà servir pels seus processos interns. Amb el mòdul Desktop del Discoverer es creen els informes, que es connectaran a la base de dades quan s'executen per primera vegada i quan es refresquen.,

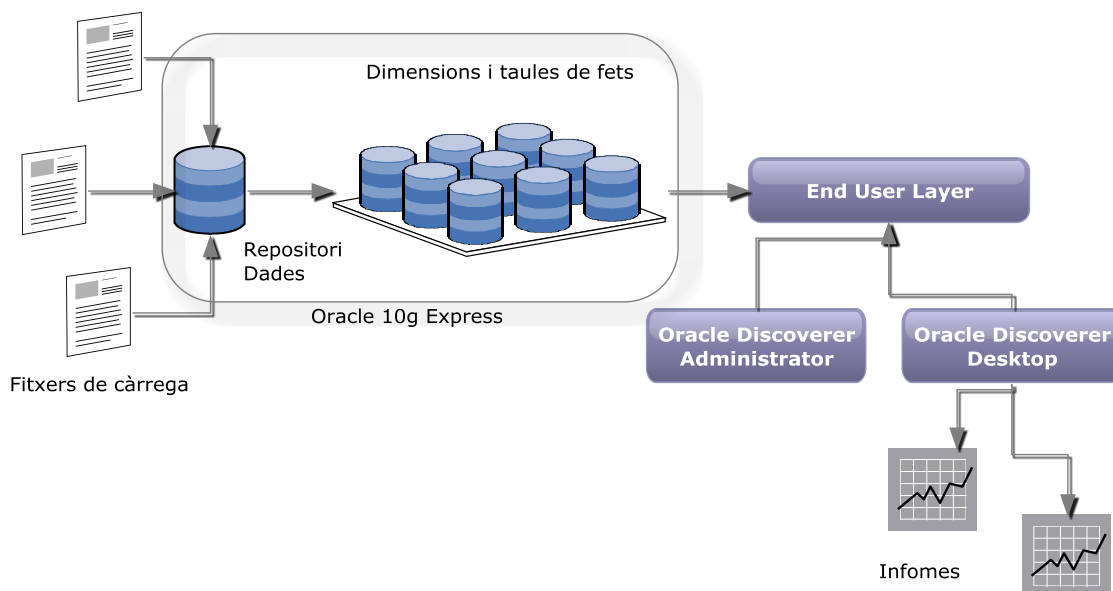


Figura 11 – Diagrama de l'arquitectura del software

3.2. Diagrama de l'arquitectura del hardware

Un Data Warehouse (DW) sol ser un sistema que no fa servir concurrentment un gran nombre d'usuaris a la vegada al contrari dels sistemes transaccionals, però es fonamental que la generació dels informes OLAP sigui e més ràpid i àgil possible ja que el perfil general d'usuari d'un DW és el d'un analista qualificat que no vol ni pot perdre massa temps esperant que un informe es generi. Per tant, els components de hardware a nivell de servidors i de xarxes serà important en qualsevol projecte d'aquestes característiques. El sistema segueix l'esquema client-servidor en el que el client fa peticions que són processades i retornades pel servidor.

El TFC s'ha realitzat íntegrament a una sola màquina amb una màquina virtual instal·lada que contenia el servidor de base de dades i l'Oracle BI Discoverer. Es feien connexions des de la màquina amfitriona a la base de dades mitjançant l'eina SQL Developer simulant dues màquines.

La recomanació per l'ACA es que es disposi de servidors escalables per tal de donar resposta a la possible ampliacions de funcionalitat del sistema i/o l'augment d'usuaris que es connectin concurrentment. Per reduir la càrrega dels servidors amb els càlculs, es recomana estudiar separar els servidors en un servidor de base dades, al que es connectaria directament l'usuari administrador IT, que contingués la base de dades i un servidor d'aplicacions amb l'Oracle BI Discoverer, al que es connectarien l'administrador funcional i els usuaris. Aquest servidor hauria de ser força escalable. L'avantatge d'aquesta arquitectura és l'alliberació del temps de procés dels servidors i la negativa és que la xarxa pot patir una sobrecàrrega.

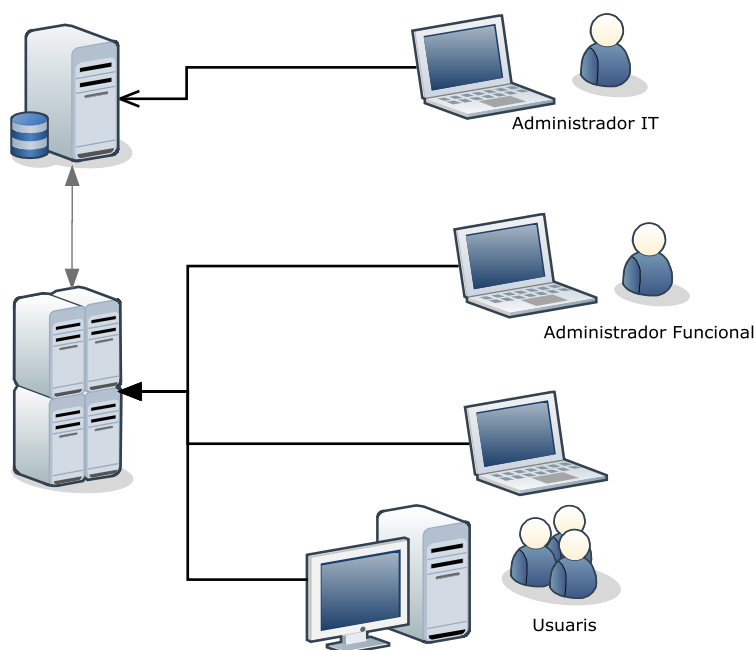


Figura 12 – Arquitectura del hardware

Gràcies als avenços de la tecnologia, es fan servir cada vegada més els servidors virtuals. Una màquina pot simular diversos servidors, això comporta una reducció evident dels costos d'adquisició i manteniment de hardware. És totalment escalable i davant una fallada del hardware es pot solucionar ràpidament traspasant les màquines virtuals a una altra màquina física. Tot això és transparent pel usuari, que no nota la diferència entre connectar-se a una màquina física d'una virtual. Així, en una màquina física es simularien els servidors de base de dades i d'aplicacions, separats amb els seus propis recursos. En cas de necessitat d'escalar el

servidor, una vegada instal·lat el nou hardware, només caldria canviar els paràmetres necessaris del software de virtualització.

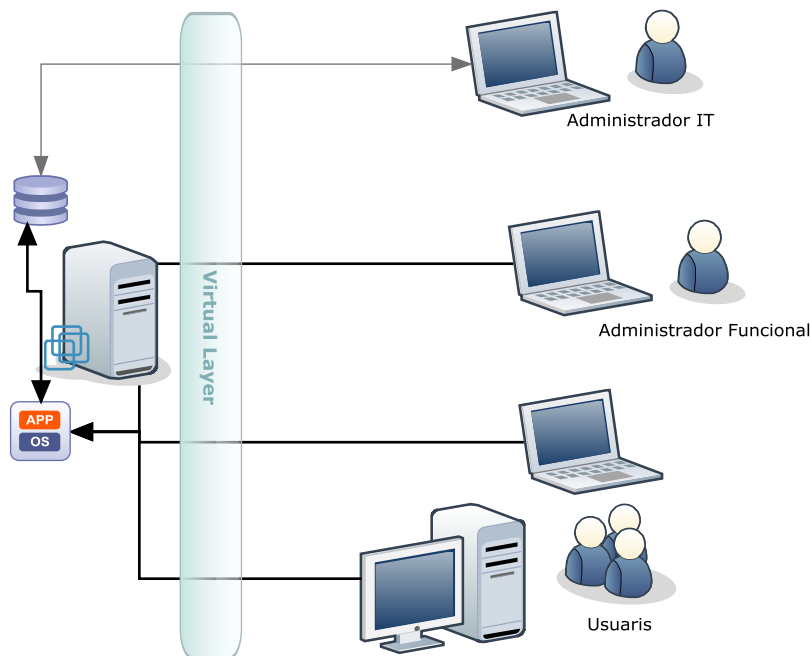


Figura 13 – Arquitectura del hardware virtualitzat

3.3. Disseny de la base de dades

Les dimensions són els punts de vista que es podran fer servir com a perspectiva dels anàlisis realitzats a l'eina de reporting. Són les que omplen de significat els indicadors de les taules de fets.

Les dimensions identificades, amb els seus atributs són:

Tipus_Artista		
Camps	Tipus	Descripció
Id_tipus	Number	Clau primària
Tipus	Varchar2(20)	Actor, actriu o director.

Artista		
Camps	Tipus	Descripció
Id_artista	Number	Clau primària
Nom_artista	Varchar2(255)	Nom de l'artista
Tipus	Varchar2(255)	Actor, actriu o director

Categoria		
Camps	Tipus	Descripció
Id_Categoria	Number	Clau primària
Cat_propia	Varchar2(255)	Categoria pròpia de cada festival
Cat_estandard	Varchar2(255)	Categoria estàndard mapejada per a unificar-les.

Festival		
Camps	Tipus	Descripció
Id_festival	Number	Clau primària
Nom_festival	Varchar2(255)	Nom del festival

Film		
Camps	Tipus	Descripció
Id_film	Number	Clau primària
Titol_original	Varchar2(255)	Títol original del film
Titol_ordre	Varchar2(255)	Títol d'ordre del film

Pais		
Camps	Tipus	Descripció
Id_Pais	Number	Clau primària
Pais_in	Varchar2(255)	Pais carregat del fitxer original
Pais_out	Varchar2(255)	Pais mapejat per unificar les dades.

Temps_any		
Camps	Tipus	Descripció
Id_temps_any	Number	Clau primària
Num_any	Varchar2(255)	Any de producció del film

Les taules de fets amb els indicadors, eixos centrals de l'anàlisi de les dades, construïdes són:

Fact_Artista		
Camps	Tipus	Descripció
Id_Artista	Number	Clau de l'artista
Id_Festival	Number	Clau del festival
Id_Categoria	Number	Clau de la categoria
Num_any	Number	Clau de l'any
Nominacions	Number	Número de nominacions de l'artista per festival, categoria i any.
Premis	Number	Número de premis de l'artista per festival, categoria i any.

Fact_Director		
Camps	Tipus	Descripció
Id_Director	Number	Clau del director a la taula Artista
Id_Festival	Number	Clau del festival
Num_any	Number	Clau de l'any
Id_Film	Number	Clau de la pel·lícula
Nominacions	Number	Número de nominacions del director per festival, any i film.
Premis	Number	Número de premis del director per festival, any i film.

Fact_Festival		
Camps	Tipus	Descripció
Id_Festival	Number	Clau del festival
Num_any	Number	Clau de l'any
Actor_Nom	Number	Número d'actors nominats per festival i any
Actriu_Nom	Number	Número d'actrius nominats per festival i any
Film_Nom	Number	Número de films nominats per festival i any

Fact_Film		
Camps	Tipus	Descripció
Id_Film	Number	Clau del film
Id_Festival	Number	Clau del festival
Id_Categoria	Number	Clau de la categoria
Id_Pais	Number	Clau del film
Num_any	Number	Clau de l'any
Nominacions	Number	Número de nominacions per film, festival, categoria, país i any
Premis	Number	Número de premis per film, festival, categoria, país i any

La taula que fa de repositori central de dades té la següent estructura:

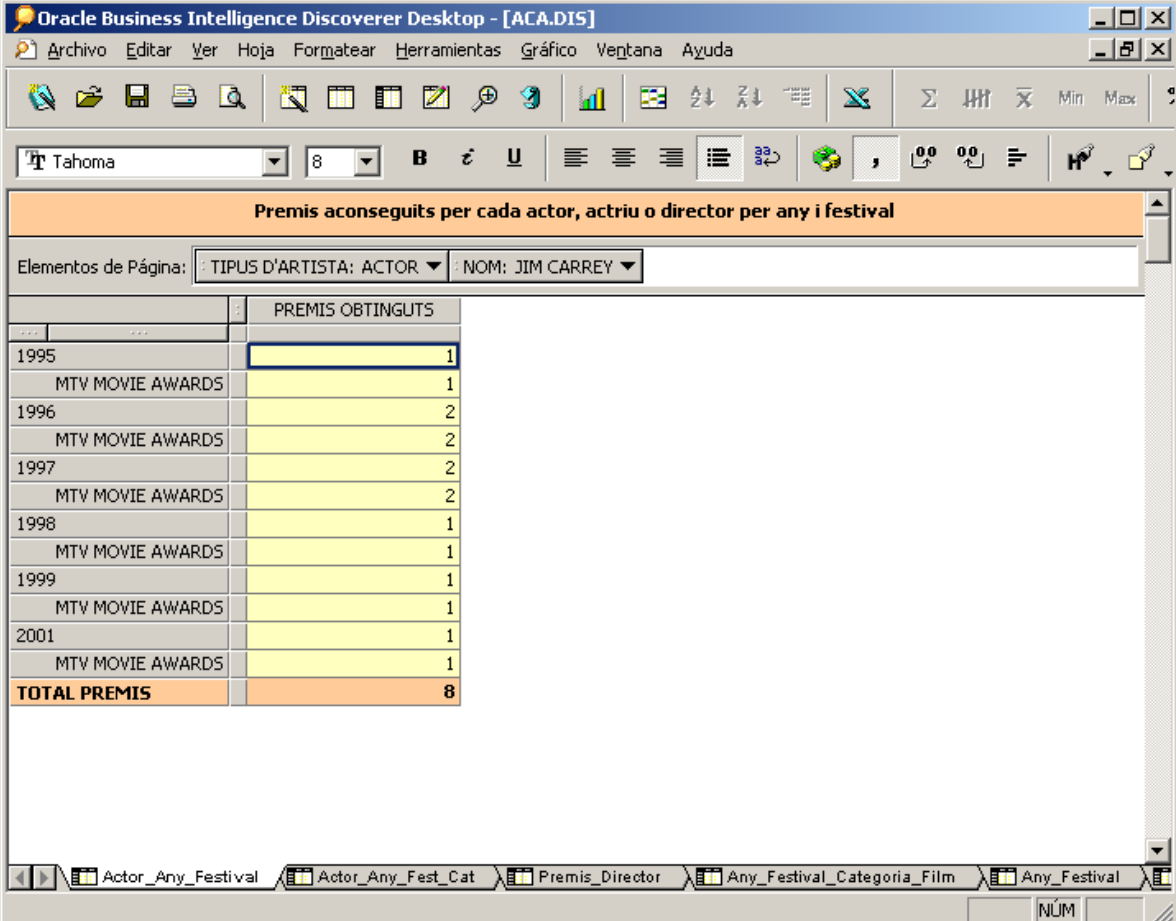
Dades		
Camps	Tipus	Descripció
Id_dades	Number	Clau primària
Titol_original	Varchar2(255)	Títol original del film
Titol_angles	Varchar2(255)	Títol en anglès del film, no es carregarà a les taules de fets.
Titol_ordre	Varchar2(255)	Títol estàndard del film
Temps_any	Varchar2(10)	Any de producció del film.
Pais	Varchar2(255)	País que produeix el film
Festival	Varchar2(255)	Festival al que participa el film
Categoria	Varchar2(255)	Categoria del premi
Guanyador	Varchar2(255)	Guanyador del premi
Nominat	Varchar2(255)	Nominat al premi

3.4. Disseny i descripció dels informes creats

En aquest capítol es descriuran els informes creats amb Oracle BI Discoverer que satisfan les peticions de l'ACA. S'ha creat un únic fitxer d'informes, guardat a la base de dades, amb diferents fulles que contenen els informes demanats.

3.4.1. Premis per artista, any i festival

Aquest informe mostra el nombre de premis aconseguits per un actor/actriu (ampliat a director gràcies al tipus d'artista identificat) per any i festival, totalitzat per a saber el nombre total de premis. Es parteix de la taula de fets Fact_Artista, com atributs es van escollir tipus i nom d'artista, l'any, el festival i l'indicador total de premis.



Oracle Business Intelligence Discoverer Desktop - [ACA.DIS]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 8 B i U

Premis aconseguits per cada actor, actriu o director per any i festival

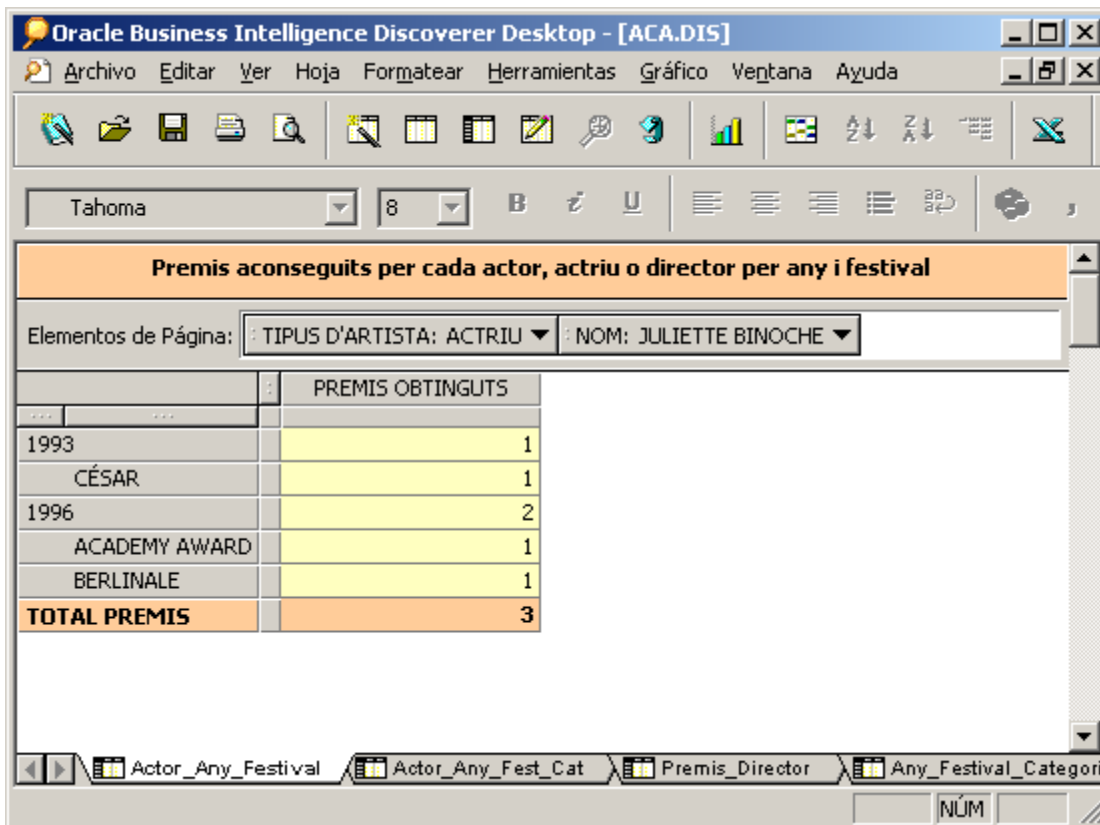
Elementos de Página: TIPIUS D'ARTISTA: ACTOR NOM: JIM CARREY

	PREMIS OBTINGUTS
1995	1
MTV MOVIE AWARDS	1
1996	2
MTV MOVIE AWARDS	2
1997	2
MTV MOVIE AWARDS	2
1998	1
MTV MOVIE AWARDS	1
1999	1
MTV MOVIE AWARDS	1
2001	1
MTV MOVIE AWARDS	1
TOTAL PREMIS	8

Actor_Any_Festival Actor_Any_Fest_Cat Premis_Director Any_Festival_Categoria_Film Any_Festival

NUM

Figura 14 – Premis per actor, any i festival



Oracle Business Intelligence Discoverer Desktop - [ACA.DIS]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 8 B U

Premis aconseguits per cada actor, actriu o director per any i festival

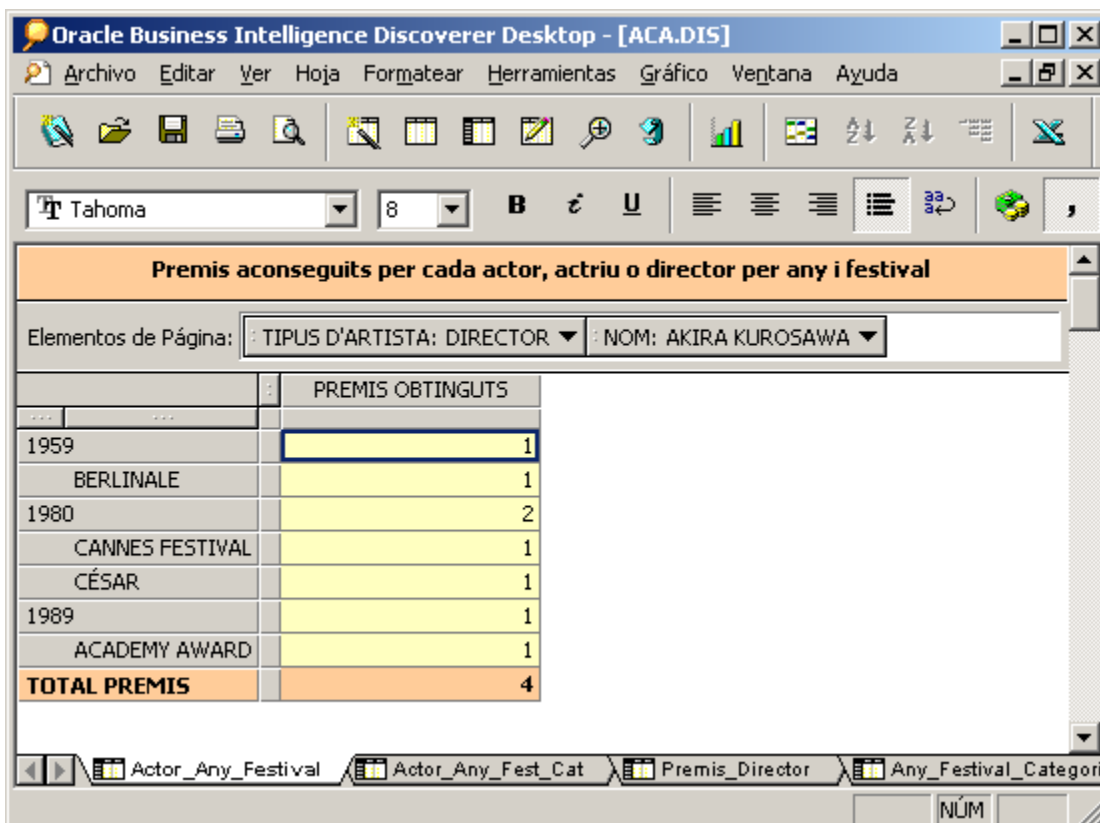
Elementos de Página: TIPUS D'ARTISTA: ACTRIU NOM: JULIETTE BINOCHÉ

	PREMIS OBTINGUTS
1993	1
CÉSAR	1
1996	2
ACADEMY AWARD	1
BERLINALE	1
TOTAL PREMIS	3

Actor_Any_Festival Actor_Any_Fest_Cat Premis_Director Any_Festival_Categori

NÚM

Figura 15 – Premis per actriu, any i festival



Oracle Business Intelligence Discoverer Desktop - [ACA.DIS]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 8 B U

Premis aconseguits per cada actor, actriu o director per any i festival

Elementos de Página: TIPUS D'ARTISTA: DIRECTOR NOM: AKIRA KUROSAWA

	PREMIS OBTINGUTS
1959	1
BERLINALE	1
1980	2
CANNES FESTIVAL	1
CÉSAR	1
1989	1
ACADEMY AWARD	1
TOTAL PREMIS	4

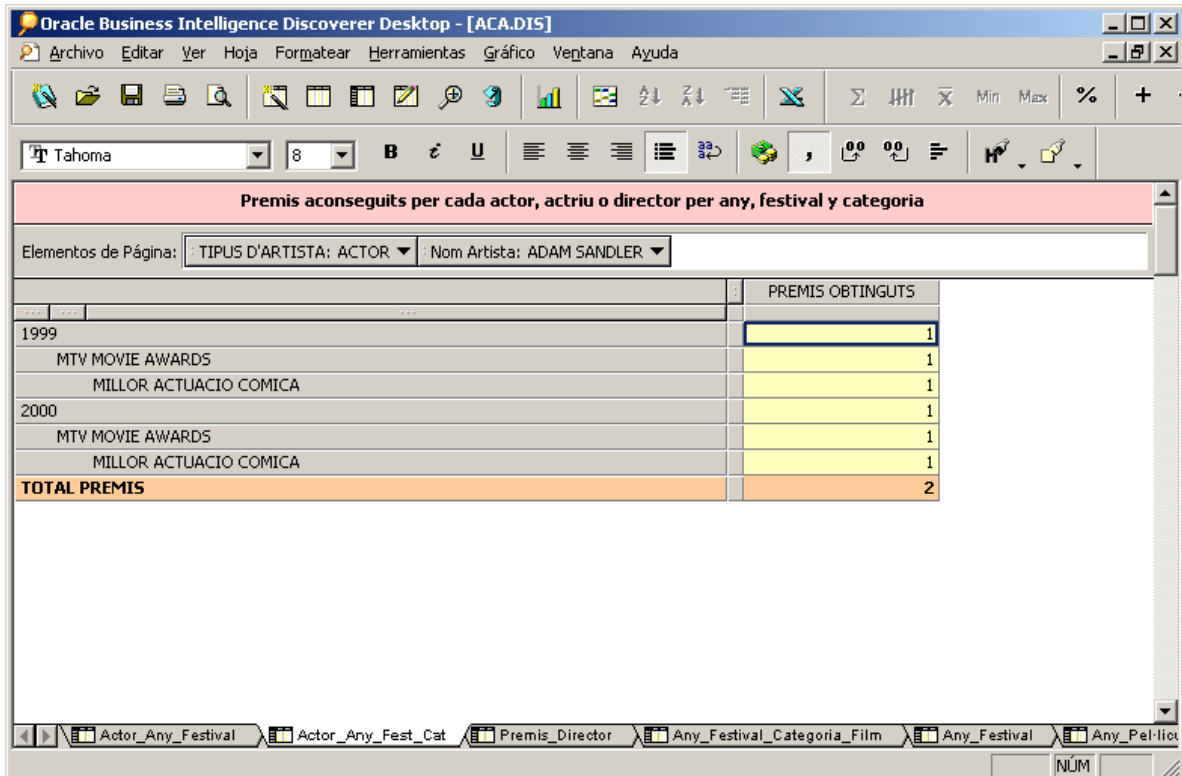
Actor_Any_Festival Actor_Any_Fest_Cat Premis_Director Any_Festival_Categori

NÚM

Figura 16 – Premis per director, any i festival

3.4.2. Premis per artista, any, festival i categoria

Aquest informe mostra el nombre de premis aconseguits per un actor/actriu (ampliat a director gràcies al tipus d'artista identificat) per any i festival i categoria. La taula de fets escollida és Fact_Artista amb els atributs tipus i nom d'artista, any, festival, categoria i el total de premis obtinguts. Es pot escollir si es vol veure actors, actrius o directors i el nom segons el tipus d'artista triat.



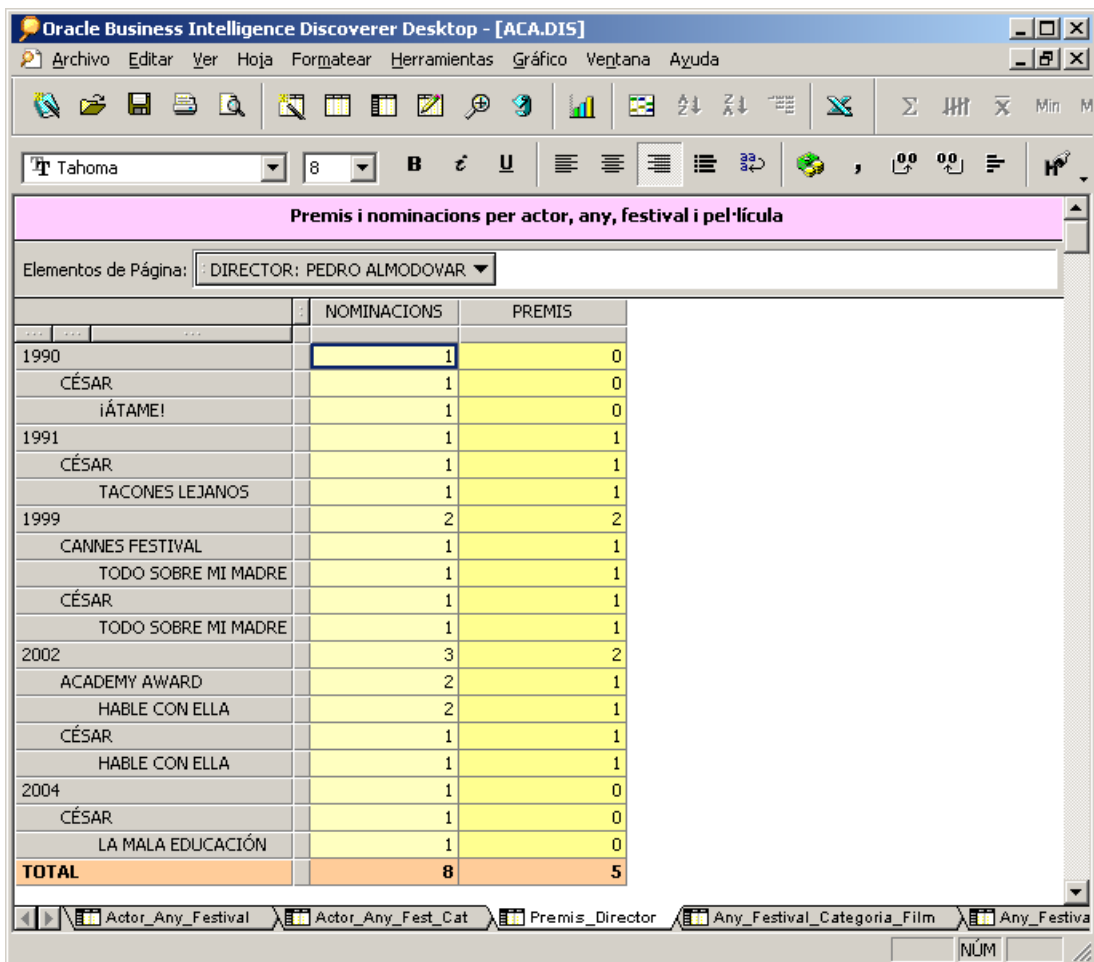
The screenshot shows the Oracle Business Intelligence Discoverer Desktop interface. The report title is "Premis aconseguits per cada actor, actriu o director per any, festival y categoria". The filters are set to "TIPUS D'ARTISTA: ACTOR" and "Nom Artista: ADAM SANDLER". The report displays a table with the following data:

	PREMIS OBTINGUTS
1999	1
MTV MOVIE AWARDS	1
MILLOR ACTUACIO COMICA	1
2000	1
MTV MOVIE AWARDS	1
MILLOR ACTUACIO COMICA	1
TOTAL PREMIS	2

Figura 17 – Premis per artista, any, festival i categoria.

3.4.3. Premis a un director per any, festival i film

Aquest informe mostra el nombre nominacions i de premis aconseguits per un director per any i festival i film. La taula de fets triada és Fact_Director amb els atributs nom d'artista, any, festival, film i el total de nominacions i premis obtinguts.



Oracle Business Intelligence Discoverer Desktop - [ACA.DIS]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 8 B U

Premis i nominacions per actor, any, festival i pel·lícula

Elementos de Página: DIRECTOR: PEDRO ALMODOVAR

	NOMINACIONES	PREMIS
1990	1	0
CÉSAR	1	0
¡ÁTAME!	1	0
1991	1	1
CÉSAR	1	1
TACONES LEJANOS	1	1
1999	2	2
CANNES FESTIVAL	1	1
TODO SOBRE MI MADRE	1	1
CÉSAR	1	1
TODO SOBRE MI MADRE	1	1
2002	3	2
ACADEMY AWARD	2	1
HABLE CON ELLA	2	1
CÉSAR	1	1
HABLE CON ELLA	1	1
2004	1	0
CÉSAR	1	0
LA MALA EDUCACIÓN	1	0
TOTAL	8	5

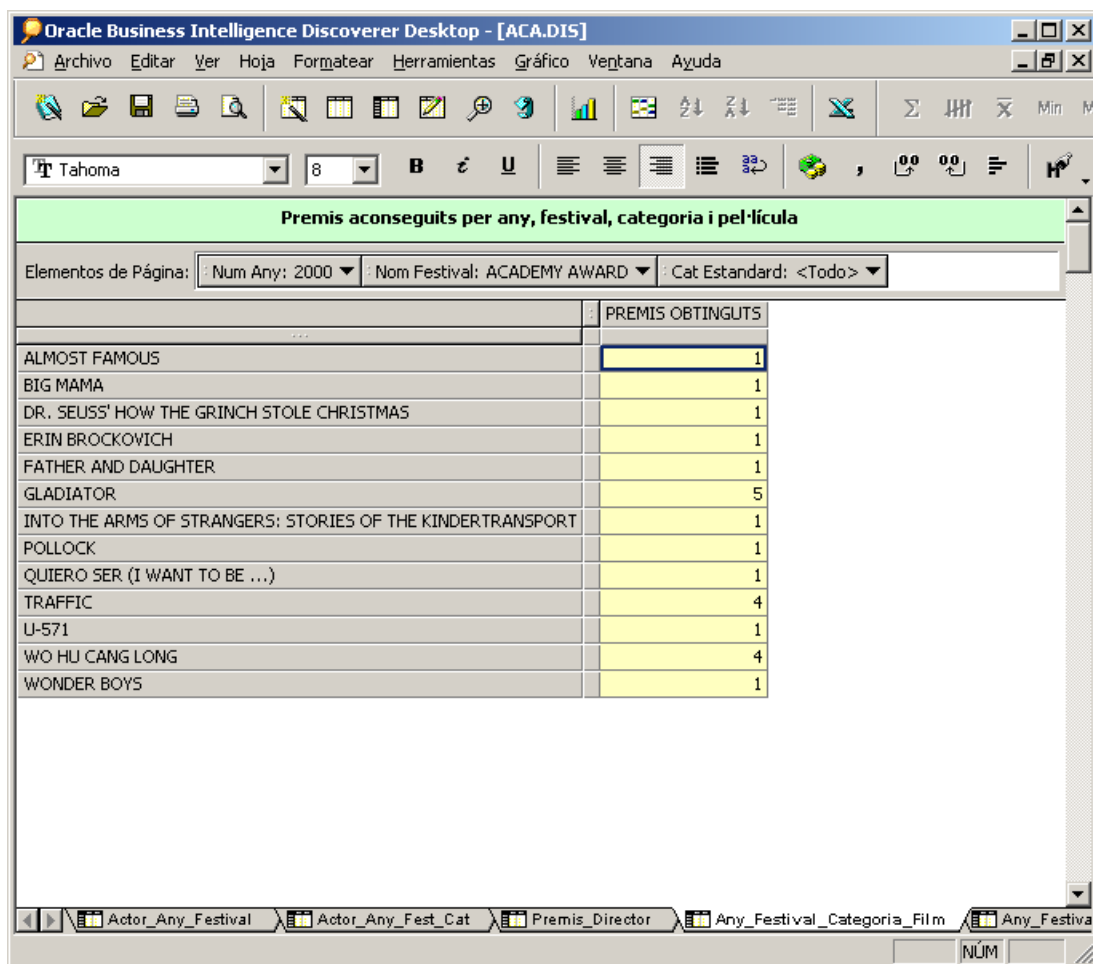
Acto... Actor_Any_Festival Actor_Any_Fest_Cat Premis_Director Any_Festival_Categoria_Film Any_Festiva

NUM

Figura 18 – Premis i nominacions dels directors

3.4.4. Premis per any, festival, categoria i pel·lícula

La fulla Any_Festival_Categoria_Film conté l'informe dels premis guanyats per any, festival categoria i pel·lícula, gràcies a la taula de fets Fact_Film i els atributs any, nom de festival, categoria, títol de la pel·lícula i com a fet el nombre de premis obtinguts.



Oracle Business Intelligence Discoverer Desktop - [ACA.DIS]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 8 B U

Premis aconseguits per any, festival, categoria i pel·lícula

Elementos de Página: Num Any: 2000 Nom Festival: ACADEMY AWARD Cat Estandard: <Todo>

	PREMIS OBTINGUTS
ALMOST FAMOUS	1
BIG MAMA	1
DR. SEUSS' HOW THE GRINCH STOLE CHRISTMAS	1
ERIN BROCKOVICH	1
FATHER AND DAUGHTER	1
GLADIATOR	5
INTO THE ARMS OF STRANGERS: STORIES OF THE KINDERTRANSPORT	1
POLLOCK	1
QUIERO SER (I WANT TO BE ...)	1
TRAFFIC	4
U-571	1
WO HU CANG LONG	4
WONDER BOYS	1

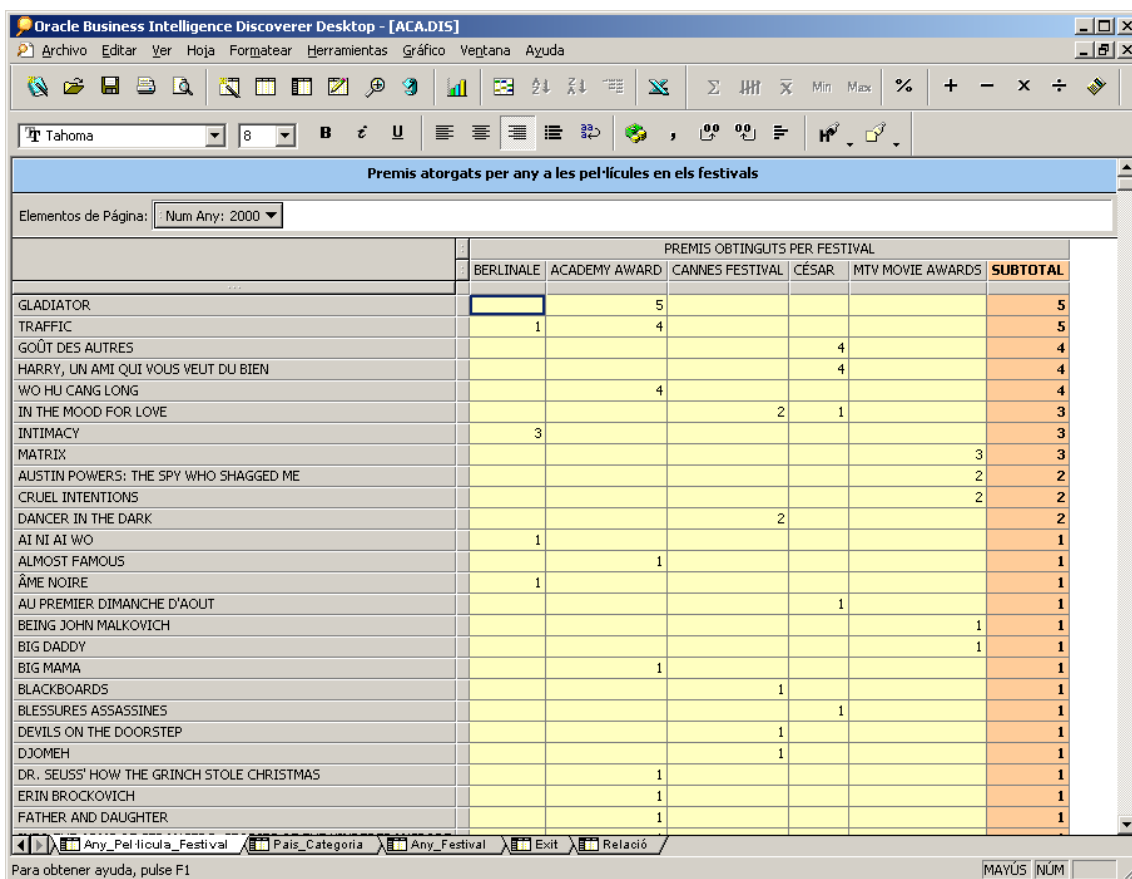
Actors: Actor_Any_Festival Actor_Any_Fest_Cat Premis_Director Any_Festival_Categoria_Film Any_Festiva

NUM

Figura 19 – Premis per any, festival, categoria i film

3.4.5. Premis per any, pel·lícula i festival

A la fulla Any_Pel·licula_Festival és l'informe que mostra per cada any, festival i categoria els premis que ha aconseguit cada film. La taula de fets és Fact_Film i els atributs triats aquesta ocasió són l'any, el festival i el títol de la pel·lícula.



	PREMIS OBTINGUTS PER FESTIVAL					SUBTOTAL
	BERLINALE	ACADEMY AWARD	CANNES FESTIVAL	CÉSAR	MTV MOVIE AWARDS	
GLADIATOR		5				5
TRAFFIC	1	4				5
GOÛT DES AUTRES				4		4
HARRY, UN AMI QUI VOUS VEUT DU BIEN				4		4
WO HU CANG LONG		4				4
IN THE MOOD FOR LOVE			2	1		3
INTIMACY	3					3
MATRIX					3	3
AUSTIN POWERS: THE SPY WHO SHAGGED ME					2	2
CRUEL INTENTIONS					2	2
DANCER IN THE DARK			2			2
AI NI AI WO	1					1
ALMOST FAMOUS		1				1
ÂME NOIRE	1					1
AU PREMIER DIMANCHE D'AOUT				1		1
BEING JOHN MALKOVICH					1	1
BIG DADDY					1	1
BIG MAMA		1				1
BLACKBOARDS			1			1
BLESSURES ASSASSINES				1		1
DEVILS ON THE DOORSTEP			1			1
DJOMEH			1			1
DR. SEUSS' HOW THE GRINCH STOLE CHRISTMAS		1				1
ERIN BROCKOVICH		1				1
FATHER AND DAUGHTER		1				1

Figura 20 – Premis per any, film i festival

3.4.6. Premis obtinguts per països i categories

La fulla Pais_Categoria mostra l'informe dels premis guanyats per a cada país, excloent els EUA, desglossat per la categoria del premi i ordenats per a mostrar el que més premis ha guanyat.

Gràcies a la jerarquia creada entre els països i les categories, es pot augmentar el detall de la dimensió país per a obtenir el detall dels premis per categoria per a cada país.

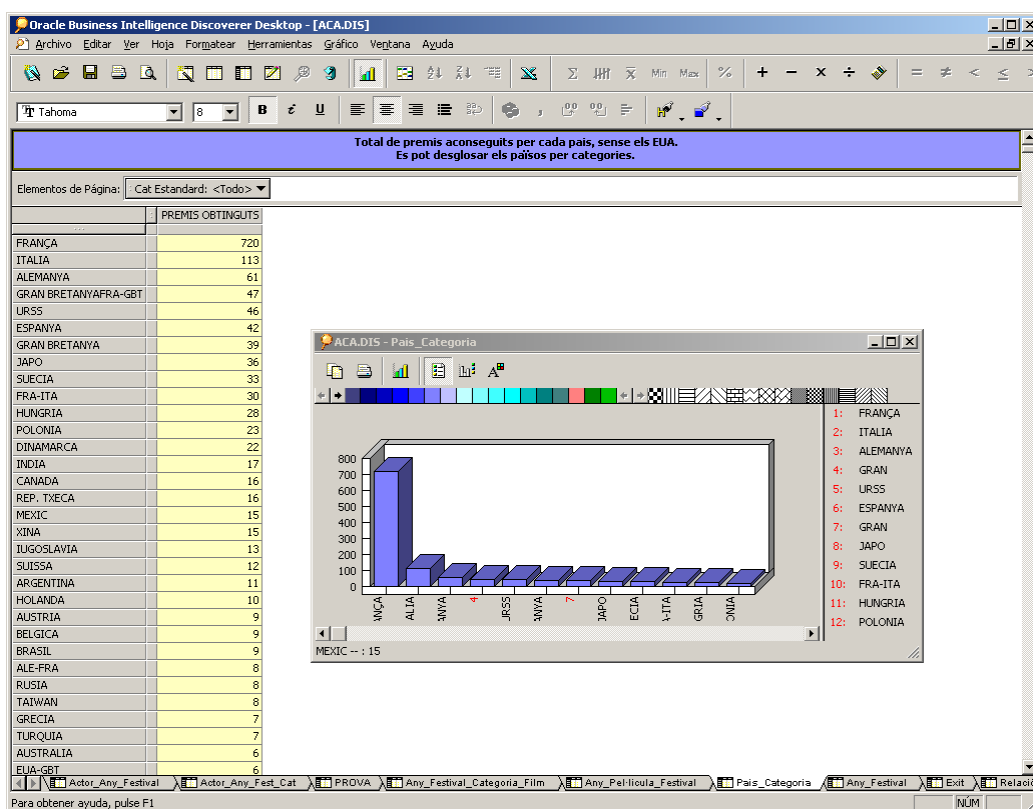
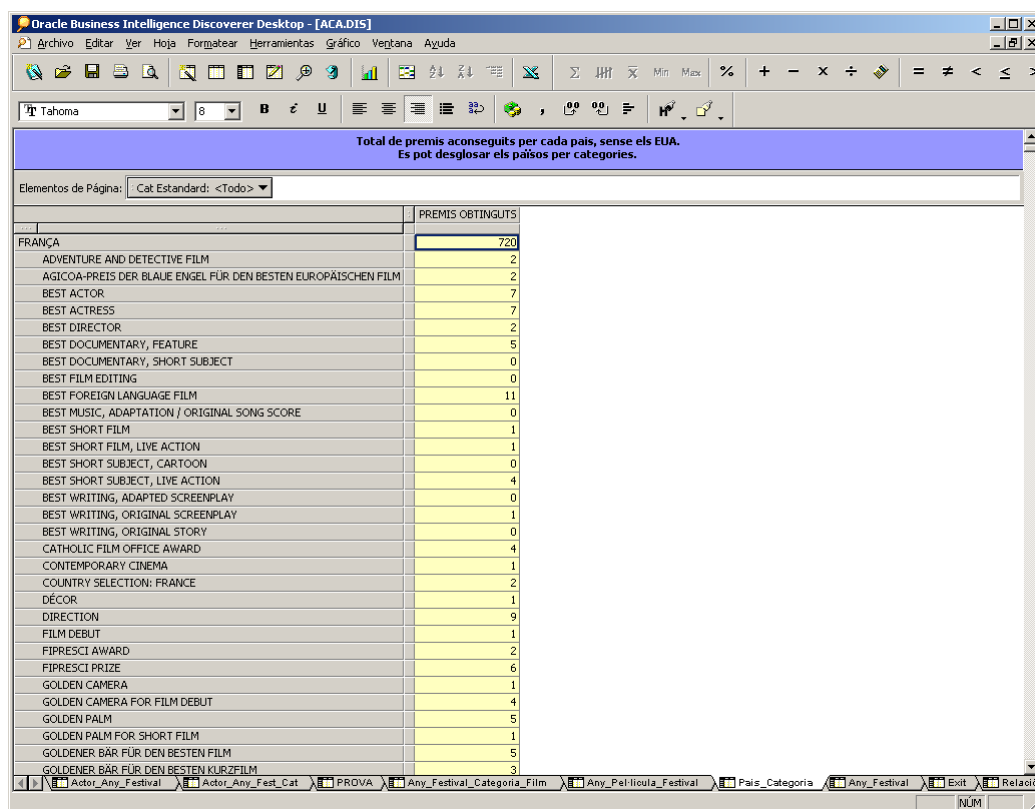


Figura 21 – Països ordenats per premis obtinguts



Oracle Business Intelligence Discoverer Desktop - [ACA.D15]

 Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

 Tahoma 8 B U

Total de premis aconseguits per cada país, sense els EUA.

Es pot desglossar els països per categories.

 Elementos de Página: Cat Estandar: <Todo>

	PREMIS OBTINGUTS
FRANÇA	720
ADVENTURE AND DETECTIVE FILM	2
AGICOA-PREIS DER BLAUE ENGEL FÜR DEN BESTEN EUROPAISCHEN FILM	2
BEST ACTOR	7
BEST ACTRESS	7
BEST DIRECTOR	2
BEST DOCUMENTARY, FEATURE	5
BEST DOCUMENTARY, SHORT SUBJECT	0
BEST FILM EDITING	0
BEST FOREIGN LANGUAGE FILM	11
BEST MUSIC, ADAPTATION / ORIGINAL SONG SCORE	0
BEST SHORT FILM	1
BEST SHORT FILM, LIVE ACTION	1
BEST SHORT SUBJECT, CARTOON	0
BEST SHORT SUBJECT, LIVE ACTION	4
BEST WRITING, ADAPTED SCREENPLAY	0
BEST WRITING, ORIGINAL SCREENPLAY	1
BEST WRITING, ORIGINAL STORY	0
CATHOLIC FILM OFFICE AWARD	4
CONTEMPORARY CINEMA	1
COUNTRY SELECTION: FRANCE	2
DÉCOR	1
DIRECTION	9
FILM DEBUT	1
FIPRESCI AWARD	2
FIPRESCI PRIZE	6
GOLDEN CAMERA	1
GOLDEN CAMERA FOR FILM DEBUT	4
GOLDEN PALM	5
GOLDEN PALM FOR SHORT FILM	1
GOLDENER BÄR FÜR DEN BESTEN FILM	5
GOLDENER BÄR FÜR DEN BESTEN KURZFILM	3

Figura 22 – Desglossament de les categories del país amb més premis

3.4.7. Premis per any i festival

La fulla Any_Festival conté un informe que demana la introducció d'un any i mostra els actors i films nominats aquell any per a cada festival. La primera vegada que s'executa demana directament l'any, per defecte està posat l'any 2002. Es posa l'any desitjat i es clica a Terminar:

Per a tornar a posar una data hem de refrescar l'informe, bé amb el menú "Hoja – Refrescar Hoja" o amb la icona amb forma de fletxa de la barra d'eines.

Aquest informe és possible gràcies a la taula de fets Fact_Festival, als atributs any i festival i als indicadors que guarden la quantitat d'actors, actrius i films nominats.

Primer es tria l'any desitjat:

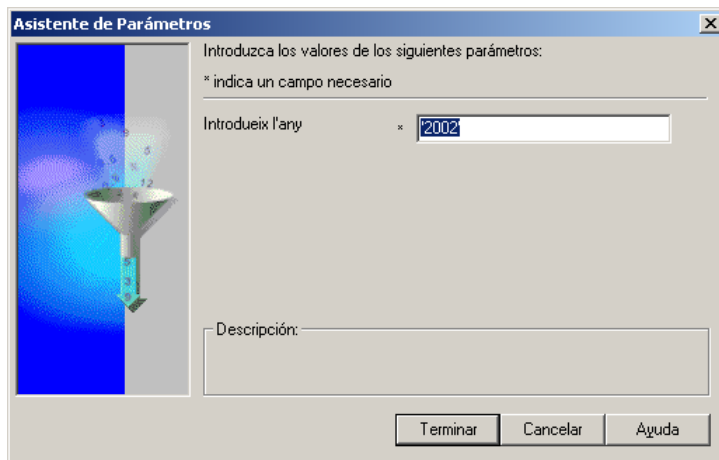


Figura 23 – Petició de l'any a consultar

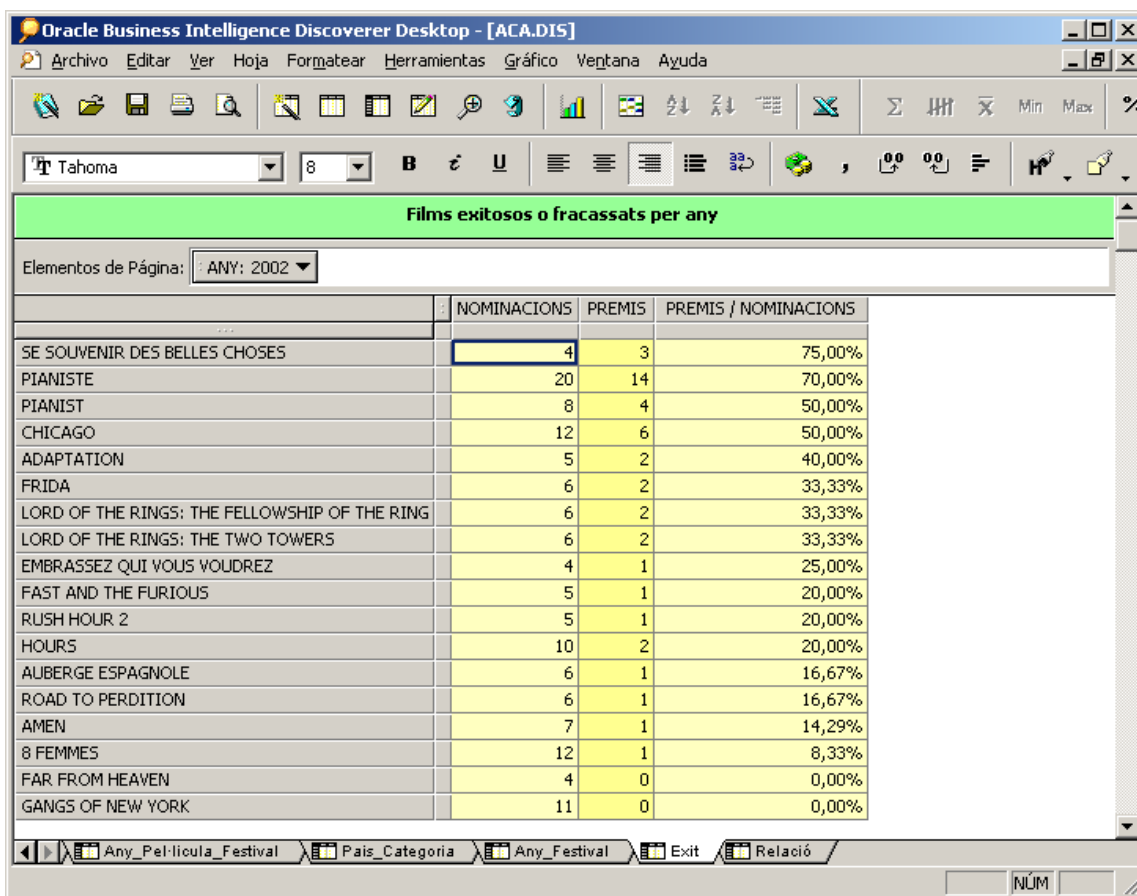
Es mostra l'informe:

	ACTORS NOMINATS	ACTRIUS NOMINADES	FILMS NOMINATS
ACADEMY AWARD	11	11	115
BERLINALE	1	1	11
CANNES FESTIVAL	1	1	12
CÉSAR	23	16	91
MTV MOVIE AWARDS	15	15	58

Figura 24 – Actors, actrius i films nominats per festival segons l'any demanat

3.4.8. Èxit o fracàs d'una pel·lícula

A la fulla Exit es pot consultar un informe que mostra per a cada any escollit a l'element de pàgina les nominacions i els premis obtinguts per un film aquell any amb una columna que calcula el percentatge. S'ha decidit mostrar solament les pel·lícules que tenen quatre o més nominacions i s'ordena per la columna del percentatge. Aquesta és una manera de veure la pel·lícula que ha tingut més èxit, tenint un percentatge alt de premis assolits respecte a les nominacions obtingudes, i les que han fracassat perquè tenint al menys quatre nominacions potser no han guanyat cap premi o el seu percentatge és baix.



Oracle Business Intelligence Discoverer Desktop - [ACA.DIS]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 8 B U

Films exitosos o fracassats per any

Elementos de Página: ANY: 2002

	NOMINACIONES	PREMIOS	PREMIOS / NOMINACIONES
SE SOUVENIR DES BELLES CHOSES	4	3	75,00%
PIANISTE	20	14	70,00%
PIANIST	8	4	50,00%
CHICAGO	12	6	50,00%
ADAPTATION	5	2	40,00%
FRIDA	6	2	33,33%
LORD OF THE RINGS: THE FELLOWSHIP OF THE RING	6	2	33,33%
LORD OF THE RINGS: THE TWO TOWERS	6	2	33,33%
EMBRASSEZ QUI VOUS VOUDREZ	4	1	25,00%
FAST AND THE FURIOUS	5	1	20,00%
RUSH HOUR 2	5	1	20,00%
HOURS	10	2	20,00%
AUBERGE ESPAGNOLE	6	1	16,67%
ROAD TO PERDITION	6	1	16,67%
AMEN	7	1	14,29%
8 FEMMES	12	1	8,33%
FAR FROM HEAVEN	4	0	0,00%
GANGS OF NEW YORK	11	0	0,00%

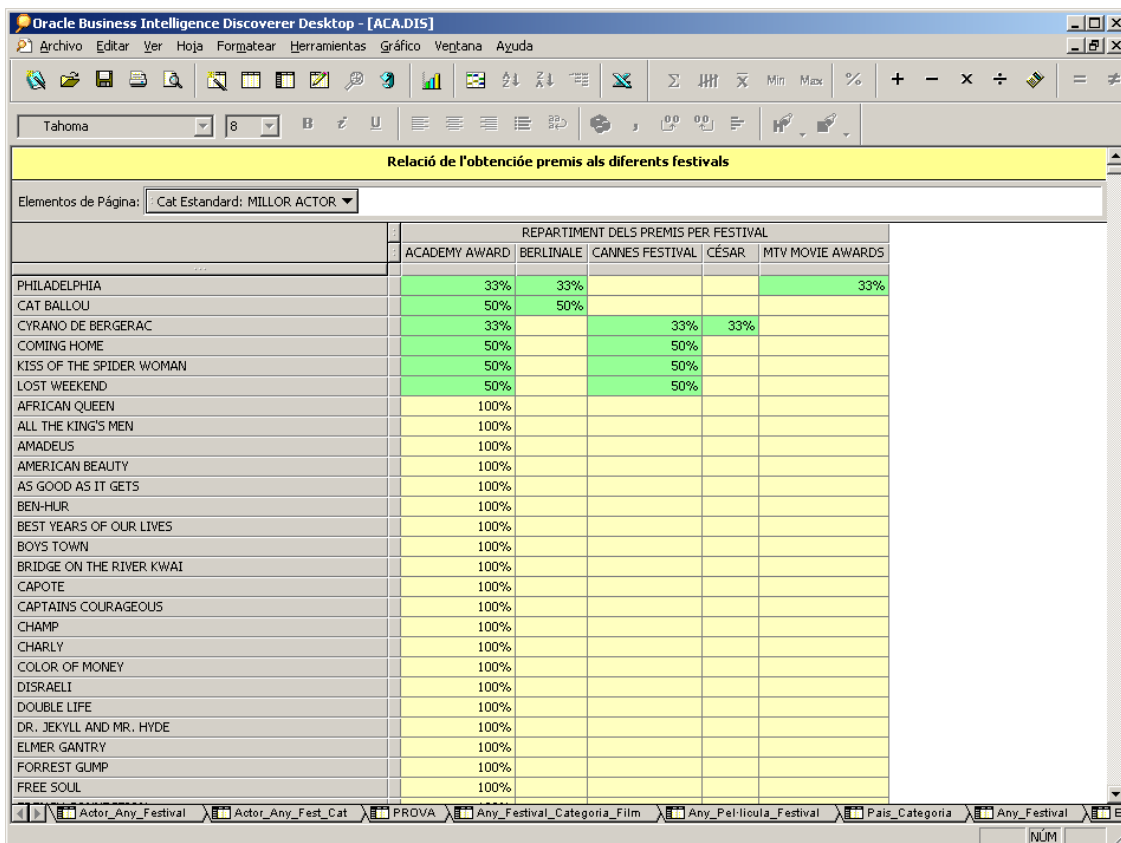
Any_Pel·lícula_Festival Pais_Categoria Any_Festival Exit Relació

NÚM

Figura 25 – Films amb èxit o fracàs

3.4.9. Relació entre premis

La darrera fulla mostra gràficament quan un film ha estat premiat a diferents festivals per categoria segons la taula Fact_Festival. Per tant es pot deduir la relació que hi ha entre premis dels diferents festivals.



The screenshot shows a software interface with a table titled "Relació de l'obtenció premis als diferents festivals". The table has a header row "REPARTIMENT DELS PREMIS PER FESTIVAL" and five columns: "ACADEMY AWARD", "BERLINALE", "CANNES FESTIVAL", "CÉSAR", and "MTV MOVIE AWARDS". The rows list various movies and their award percentages across these categories. For example, "PHILADELPHIA" has 33% in Academy Award, 33% in Berlinale, and 33% in MTV Movie Awards. "CAT BALLOU" has 50% in Academy Award and 50% in Berlinale. "CYRANO DE BERGERAC" has 33% in Academy Award, 33% in Cannes Festival, and 33% in César. "COMING HOME" has 50% in Academy Award and 50% in Cannes Festival. "KISS OF THE SPIDER WOMAN" has 50% in Academy Award and 50% in Cannes Festival. "LOST WEEKEND" has 50% in Academy Award and 50% in Cannes Festival. "AFRICAN QUEEN" through "FREE SOUL" all have 100% in the Academy Award category.

	REPARTIMENT DELS PREMIS PER FESTIVAL				
	ACADEMY AWARD	BERLINALE	CANNES FESTIVAL	CÉSAR	MTV MOVIE AWARDS
PHILADELPHIA	33%	33%			33%
CAT BALLOU	50%	50%			
CYRANO DE BERGERAC	33%		33%	33%	
COMING HOME	50%		50%		
KISS OF THE SPIDER WOMAN	50%		50%		
LOST WEEKEND	50%		50%		
AFRICAN QUEEN	100%				
ALL THE KING'S MEN	100%				
AMADEUS	100%				
AMERICAN BEAUTY	100%				
AS GOOD AS IT GETS	100%				
BEN-HUR	100%				
BEST YEARS OF OUR LIVES	100%				
BOYS TOWN	100%				
BRIDGE ON THE RIVER KWAI	100%				
CAPOTE	100%				
CAPTAINS COURAGEOUS	100%				
CHAMP	100%				
CHARLY	100%				
COLOR OF MONEY	100%				
DISRAELI	100%				
DOUBLE LIFE	100%				
DR. JEKYLL AND MR. HYDE	100%				
ELMER GANTRY	100%				
FORREST GUMP	100%				
FREE SOUL	100%				

Figura 26 – Relació entre premis

4. Conclusions

Aquest treball ha suposat una gran oportunitat d'aprofundir en el món del Data Warehouse i constatar les diferències entre un DW ROLAP, que és el que s'ha construït en aquest projecte, i un DW MOLAP, que és en el que tinc experiència laboral com a administrador funcional d'Oracle Financial Analyzer i Hyperion Planning (Essbase).

També ha estat una bona ocasió per a refrescar i aprendre coses noves sobre els conceptes d'un SGBD relacional com Oracle Express, e llenguatge SQL amb les particularitats que té Oracle i l'ús intensiu de l'eina SQL Developer.

Molt important ha estat agafar experiència en la planificació i gestió d'un projecte com aquest, la identificació i divisió de les tasques a realitzar i la planificació temporal.

Com Analista funcional sempre he tingut contacte amb els administradors de la les BB.DD. relacionals com a client. D'aquesta forma he pogut experimentar les dificultats que comporta el procés de neteja de les dades, i l'ETL en general, quan les dades no venen netes i provenen de fulls de càlculs sense disciplina.

5. Línies d'evolució futura

Un punt millorable per l'ACA de cara al futur és la millora de la qualitat de la informació d'origen. Els fitxers Excel són molt habituals però són molt manuals. En un full de càlcul hi cap tot. Una eina de Data Quality ben configurada ajudaria a garantir la consistència de les dades gràcies a l'eliminació de dades duplicades i la correcció de dades amb anomalies o eliminació de les dades errònies. En resum seria útil per a complir amb la integritat, consistència de les dades.

Si es combina amb una eina d'ETL, es podria automatitzar el procés de càrrega. En els processos sensibles, quant major grau d'automatització tinguin, millor, encara que sempre es necessiti la intervenció humana per a la comprovació de les dades. Seria recomanable comptar amb un analista amb el perfil tècnic necessari però amb uns coneixements amplis del món del cinema i dels festivals per a detectar irregularitats que un neòfit no veuria.

La recomanació seria l'adquisició de les eines Oracle Data Integrator (ODI) com a ETL i Oracle Data Quality que es combina amb ODI per a aconseguir la màxima confiança en les dades.

Glossari

Data quality: conjunt d'eines i estratègies destinades a garantir la integritat i consistència de les dades.

Diagrama de Gantt: gràfic que mostra el temps de dedicació a cada tasca d'un projecte.

Dimensió: perspectiva d'anàlisi d'un cub multidimensional. Per exemple temps, pel·lícula o festival. Les dimensions es combinen entre sí per a formar cubs multidimensionals.

ETL: acrònim d'Extract, Transform and Load (extreure, transformar i carregar). Procés que permet extreure les dades de diferents orígens, netejar-les i preparar-les per a finalment executar la càrrega al sistema receptor.

Fet: és l'objecte d'anàlisi. És la raó de ser del cub al que pertany. Conté les dades que s'analitzaran en funció de les dimensions amb les que està relacionat.

Magatzem de dades o Data Warehouse: sistema integrat de consolidació de dades orientades a un tema, no volàtil i que reflexa el canvi de les dades a través del temps.

Model conceptual: model que mostra l'esquema conceptual de la base de dades a partir dels seus objectes i descripcions.

OLAP: acrònim d'On-line Analytical Processing. Dona resposta ràpida a les consultes demanades gràcies a la seva constitució multidimensional. Si la base de dades en la que es troben les dades és relacional es tracta d'un sistema ROLAP (Relacional OLAP) com és el cas d'aquest treball.

Oracle BI Discoverer: sistema de generació d'informes OLAP.

Oracle Express: sistema gestor de base de dades relacional de l'empresa nord-americana Oracle.

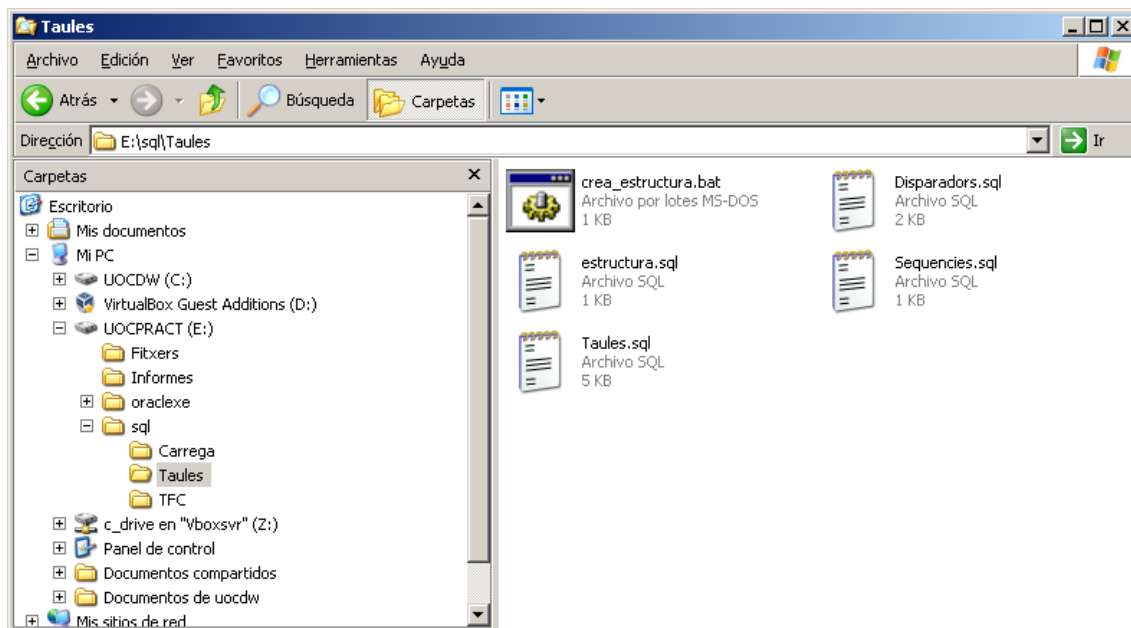
Bibliografia

- Oracle Database Data Warehousing Guide 10g Release 2 (10.2)
- The Data Warehouse ETL Toolkit. Ralph Kimball, Wiley Publishing, Inc.
- La metodología de Kimball para el diseño de almacenes de datos (Data warehouses) - Gustavo R. Rivadera
- http://www.orafaq.com/wiki/Main_Page
- <http://www.kimballgroup.com>
- <http://www.softcatala.org/traductor>

Annexos

A. Creació d'estructures

A la carpeta E:\sql de la màquina virtual hi ha les diferents carpetes amb els scripts de creació de les estructures, la càrrega i transformació de les dades. La carpeta taules conté els scripts de creació de les taules, seqüències i disparadors. S'assumeix que prèviament s'hagi creat l'esquema aca/aca a la base de dades per poder crear l'estructura a la base de dades d'aquest usuari.



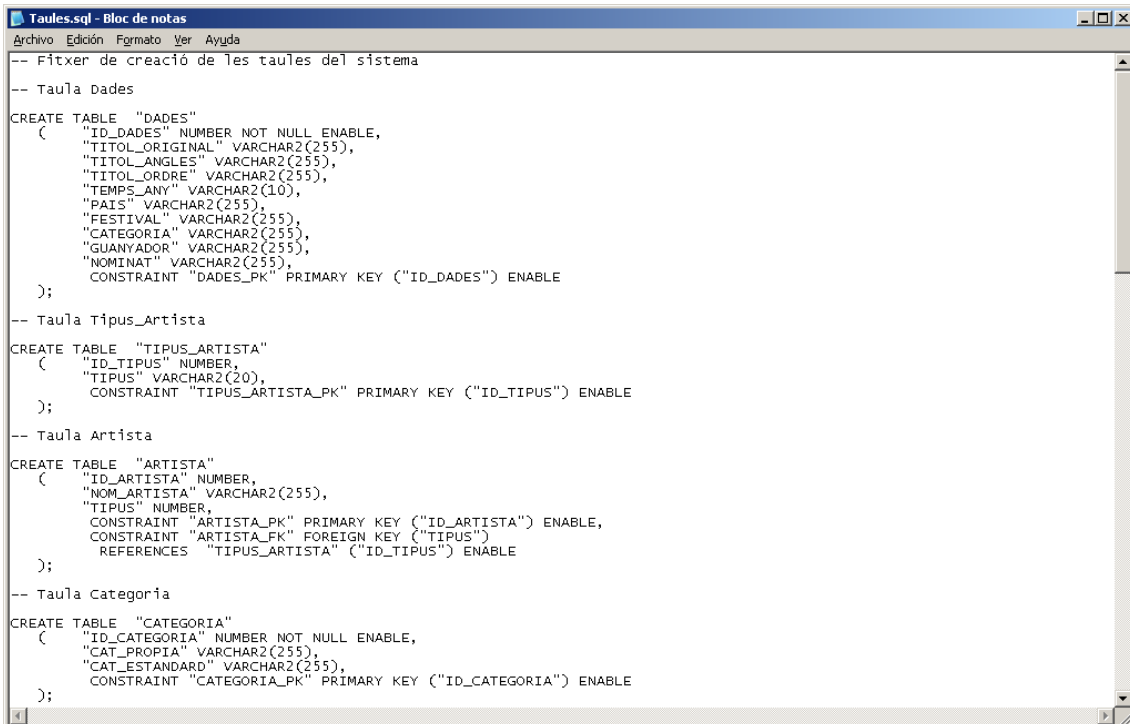
El fitxer crea_estructura.bat és el fitxer que crida els diferents fitxers sql per a crear tota l'estructura de taules de la base de dades. Aquest fitxer conté la línia:

```
sqlplus aca/aca@xe @estructura.sql
```

per a connectar-se a la base de dades mitjançant sqlplus passant com paràmetres l'usuari i el password i el fitxer a carregar. El contingut d'aquest fitxer és:

```
@taules.sql  
@sequencies.sql  
@disparadors.sql  
  
exit
```

Així que realment fa una crida als diferents fitxers .sql de càrrega. El primer crea les taules, el següent crea les seqüències necessàries i el tercer crea els disparadors que fan augmentar les seqüències.

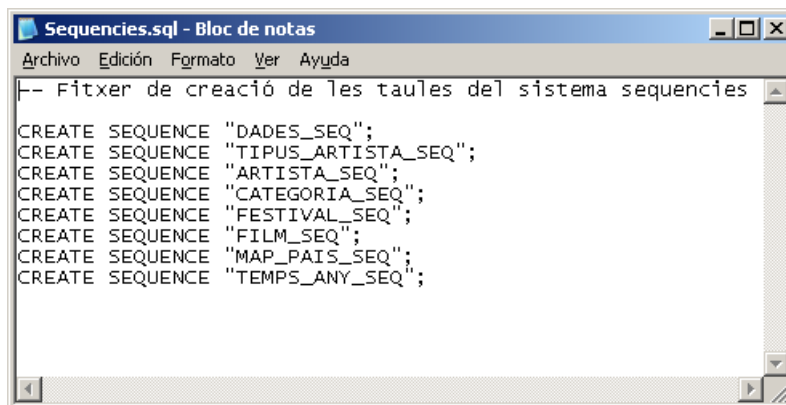


```

Taulas.sql - Bloc de notes
Archivo Edición Formato Ver Ayuda
-- Fitxer de creació de les taules del sistema
-- Taula Dades
CREATE TABLE "DADES"
(
  "ID_DADES" NUMBER NOT NULL ENABLE,
  "TITOL_ORIGINAL" VARCHAR2(255),
  "TITOL_ANGLES" VARCHAR2(255),
  "TITOL_ORDRE" VARCHAR2(255),
  "TEMPS_ANY" VARCHAR2(10),
  "PAIS" VARCHAR2(255),
  "FESTIVAL" VARCHAR2(255),
  "CATEGORIA" VARCHAR2(255),
  "GUANYADOR" VARCHAR2(255),
  "NOMINAT" VARCHAR2(255),
  CONSTRAINT "DADES_PK" PRIMARY KEY ("ID_DADES") ENABLE
);
-- Taula Tipus_Artista
CREATE TABLE "TIPUS_ARTISTA"
(
  "ID_TIPUS" NUMBER,
  "TIPUS" VARCHAR2(20),
  CONSTRAINT "TIPUS_ARTISTA_PK" PRIMARY KEY ("ID_TIPUS") ENABLE
);
-- Taula Artista
CREATE TABLE "ARTISTA"
(
  "ID_ARTISTA" NUMBER,
  "NOM_ARTISTA" VARCHAR2(255),
  "TIPUS" NUMBER,
  CONSTRAINT "ARTISTA_PK" PRIMARY KEY ("ID_ARTISTA") ENABLE,
  CONSTRAINT "ARTISTA_FK" FOREIGN KEY ("TIPUS")
  REFERENCES "TIPUS_ARTISTA" ("ID_TIPUS") ENABLE
);
-- Taula Categoria
CREATE TABLE "CATEGORIA"
(
  "ID_CATEGORIA" NUMBER NOT NULL ENABLE,
  "CAT_PROPIA" VARCHAR2(255),
  "CAT_ESTANDARD" VARCHAR2(255),
  CONSTRAINT "CATEGORIA_PK" PRIMARY KEY ("ID_CATEGORIA") ENABLE
);

```

Figura 27 – Creació de les taules

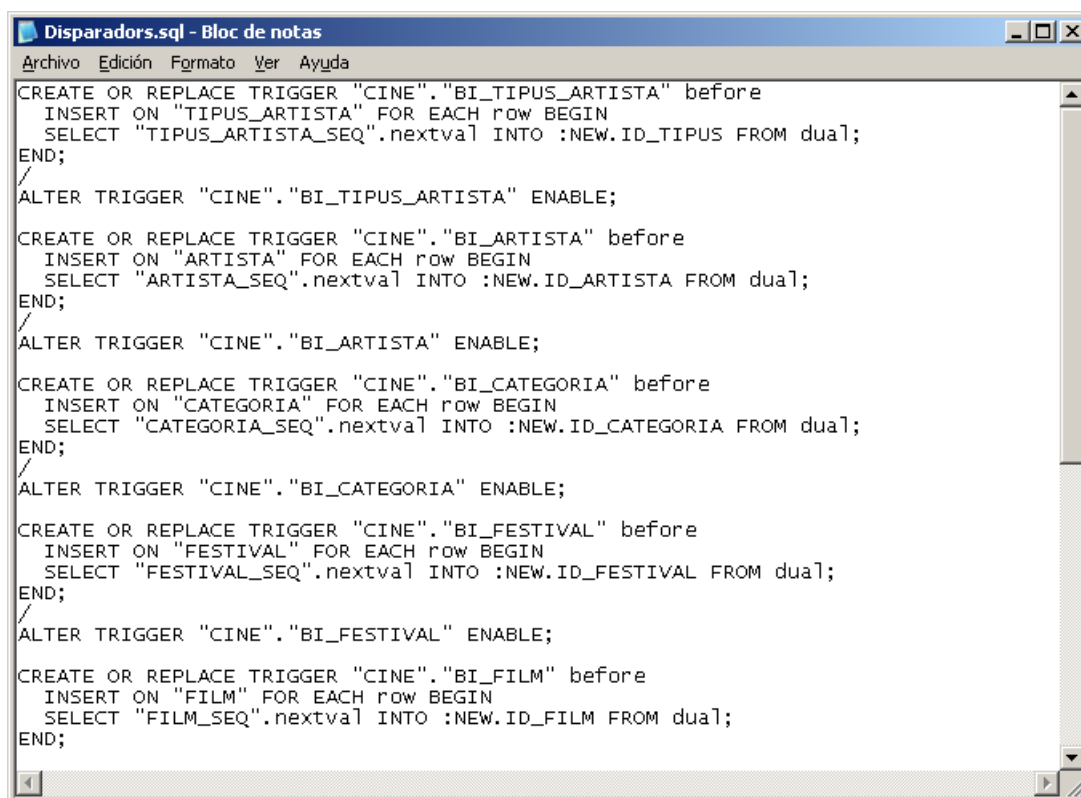


```

Sequencies.sql - Bloc de notes
Archivo Edición Formato Ver Ayuda
-- Fitxer de creació de les taules del sistema sequencies
CREATE SEQUENCE "DADES_SEQ";
CREATE SEQUENCE "TIPUS_ARTISTA_SEQ";
CREATE SEQUENCE "ARTISTA_SEQ";
CREATE SEQUENCE "CATEGORIA_SEQ";
CREATE SEQUENCE "FESTIVAL_SEQ";
CREATE SEQUENCE "FILM_SEQ";
CREATE SEQUENCE "MAP_PAIS_SEQ";
CREATE SEQUENCE "TEMPS_ANY_SEQ";

```

Figura 28 – Creació de les seqüències



```

CREATE OR REPLACE TRIGGER "CINE"."BI_TIPUS_ARTISTA" before
  INSERT ON "TIPUS_ARTISTA" FOR EACH row BEGIN
  SELECT "TIPUS_ARTISTA_SEQ".nextval INTO :NEW.ID_TIPUS FROM dual;
END;
/
ALTER TRIGGER "CINE"."BI_TIPUS_ARTISTA" ENABLE;

CREATE OR REPLACE TRIGGER "CINE"."BI_ARTISTA" before
  INSERT ON "ARTISTA" FOR EACH row BEGIN
  SELECT "ARTISTA_SEQ".nextval INTO :NEW.ID_ARTISTA FROM dual;
END;
/
ALTER TRIGGER "CINE"."BI_ARTISTA" ENABLE;

CREATE OR REPLACE TRIGGER "CINE"."BI_CATEGORIA" before
  INSERT ON "CATEGORIA" FOR EACH row BEGIN
  SELECT "CATEGORIA_SEQ".nextval INTO :NEW.ID_CATEGORIA FROM dual;
END;
/
ALTER TRIGGER "CINE"."BI_CATEGORIA" ENABLE;

CREATE OR REPLACE TRIGGER "CINE"."BI_FESTIVAL" before
  INSERT ON "FESTIVAL" FOR EACH row BEGIN
  SELECT "FESTIVAL_SEQ".nextval INTO :NEW.ID_FESTIVAL FROM dual;
END;
/
ALTER TRIGGER "CINE"."BI_FESTIVAL" ENABLE;

CREATE OR REPLACE TRIGGER "CINE"."BI_FILM" before
  INSERT ON "FILM" FOR EACH row BEGIN
  SELECT "FILM_SEQ".nextval INTO :NEW.ID_FILM FROM dual;
END;
  
```

Figura 29 – Creació dels disparadors

Hi ha un disparador especial que actualitza les dades de la taula de fets Fact_Festival, les quantitats que representen les nominacions d'actors, actrius i directors, cada vegada que s'insereixen els registres que identifiquen a un festival i l'any que s'ha celebrat el festival. Es fa servir una taula temporal per a evitar l'error de les taules mutants que sorgeix quan es vol fer una consulta a una taula que està sent actualitzada.

```

CREATE OR REPLACE TRIGGER "ACA"."FACT_FESTIVAL_ACTORNOM"
AFTER INSERT ON FACT_FESTIVAL_TEMP
REFERENCING NEW AS N

FOR EACH ROW

BEGIN

INSERT INTO FACT_FESTIVAL(ID_FESTIVAL, NUM_ANY)
VALUES (:N.ID_FESTIVAL, :N.NUM_ANY);

UPDATE FACT_FESTIVAL SET ACTOR_NOM =
(SELECT COUNT(DADES.NOMINAT)
FROM DADES, FESTIVAL, TEMPS_ANY, ARTISTA, TIPUS_ARTISTA
WHERE FESTIVAL.ID_FESTIVAL = :N.ID_FESTIVAL
AND FESTIVAL.NOM_FESTIVAL = DADES.FESTIVAL
AND TEMPS_ANY.ID_TEMPS_ANY = :N.NUM_ANY
AND TEMPS_ANY.NUM_ANY = DADES.TEMPS_ANY
AND ARTISTA.NOM_ARTISTA = DADES.NOMINAT
AND ARTISTA.TIPUS = TIPUS_ARTISTA.ID_TIPUS
AND TIPUS_ARTISTA.TIPUS = 'ACTOR')
WHERE FACT_FESTIVAL.ID_FESTIVAL = :N.ID_FESTIVAL
AND FACT_FESTIVAL.NUM_ANY = :N.NUM_ANY;
  
```

```
UPDATE FACT_FESTIVAL SET ACTRIU_NOM =
(SELECT COUNT(DADES.NOMINAT)
FROM DADES, FESTIVAL, TEMPS_ANY, ARTISTA, TIPUS_ARTISTA
WHERE FESTIVAL.ID_FESTIVAL = :N.ID_FESTIVAL
AND FESTIVAL.NOM_FESTIVAL = DADES.FESTIVAL
AND TEMPS_ANY.ID_TEMPS_ANY = :N.NUM_ANY
AND TEMPS_ANY.NUM_ANY = DADES.TEMPS_ANY
AND ARTISTA.NOM_ARTISTA = DADES.NOMINAT
AND ARTISTA.TIPUS = TIPUS_ARTISTA.ID_TIPUS
AND TIPUS_ARTISTA.TIPUS = 'ACTRIU')
WHERE FACT_FESTIVAL.ID_FESTIVAL = :N.ID_FESTIVAL
AND FACT_FESTIVAL.NUM_ANY = :N.NUM_ANY;

UPDATE FACT_FESTIVAL SET FILM_NOM =
(SELECT COUNT(DADES.TITOL_ORIGINAL)
FROM DADES, FESTIVAL, TEMPS_ANY
WHERE FESTIVAL.ID_FESTIVAL = :N.ID_FESTIVAL
AND FESTIVAL.NOM_FESTIVAL = DADES.FESTIVAL
AND TEMPS_ANY.ID_TEMPS_ANY = :N.NUM_ANY
AND TEMPS_ANY.NUM_ANY = DADES.TEMPS_ANY)
WHERE FACT_FESTIVAL.ID_FESTIVAL = :N.ID_FESTIVAL
AND FACT_FESTIVAL.NUM_ANY = :N.NUM_ANY;

END;
/
ALTER TRIGGER "ACA"."FACT_FESTIVAL_ACTORNOM" ENABLE;
```

B. Càrrega de dades

Disposem de cinc fitxers excel per a fer la càrrega de dades, un per a cada festival. Tots els fitxers tenen el mateix ordre i format de les columnes:

- Títol del film.
- Títol local: aquesta columna no es carregarà.
- Títol d'ordre.
- Any de producció del film.
- País o grup de països que produeixen la pel·lícula
- Festival: els Oscars, Cannes, César, Berlin i MTV.
- Categoria del premi: millor film, millor actor, etc.
- Guanyador: marca si l'entrada es refereix a un guanyador del premi referit.
- Nominat: marca si es tracta d'un nominat.
- Número de certamen: no es carrega.
- Ítem: no es carrega.

El primer que es va fer va ser transformar-los a format .csv. Com que la màquina virtual no te instal·lada una versió d'Excel, es va fer a la versió que hi ha a la màquina amfitriona, carregant els fitxers i guardant-los com .csv que separa les dades dels camps amb ';'. Finalment es traspassen els fitxers a la màquina virtual.

	A	B	C	D	E	F
	Film Title	Localized Title	Sort Title	Year	Country	Award
1						
2	Die Vier im Jeep	Die Vier im Jeep	Vier im Jeep	1951	Schweiz	Berlinale
3	Il cammino della speranza	Weg der Hoffnung	Il cammino della speranza	1951	Italien	Berlinale
4	The Blowing Version	Konflikt des Herzens	Blowing Version	1951	UK	Berlinale
5	Sans laisser d'adresse	Ohne Angabe der Adresse	Sans laisser d'adresse	1951	Frankreich	Berlinale
6	L'espoire fait vivre	L'espoire fait vivre	L'espoire fait vivre	1951	Schweden	Berlinale
7	The Mating season	SOS - Zwei Schwiegermütter	Mating season	1951	USA	Berlinale
8	Beaver Valley	Beaver Valley	Beaver Valley	1951	USA	Berlinale
9	The Undefeated	The Undefeated	Undefeated	1951	USA	Berlinale
10	Justice est faite	Schurgericht	Justice est faite	1951	Frankreich	Berlinale
11	Destination Moon	Endstation Mond	Destination Moon	1951	USA	Berlinale
12	Cinderella	Cinderella	Cinderella	1951	USA	Berlinale
13	Tales Of Hoffmann	Hoffmanns Erzählungen	Tales Of Hoffmann	1951	UK	Berlinale
14	Il Cristo proibito	Il Cristo proibito	Il Cristo proibito	1951	Italien	Berlinale
15	Dieu a besoin des hommes	Gott brauchst Menschen	Dieu a besoin des hommes	1951	Frankreich	Berlinale
16	Hon dansade en sommar	Sie tanzte nur einen Sommer	Hon dansade en sommar	1952	Schweden	Berlinale
17	Fanfan, la tulipe	Fanfan, der Husar	Fanfan, la tulipe	1952	Frankreich	Berlinale
18	Cry The Beloved Country	Denn sie sollen getröstet werden	Cry The Beloved Country	1952	UK	Berlinale
19	Le salaire de la peur	Lohn der Angst	salare de la peur	1953	Frankreich/Italien	Berlinale
20	Magia verde	Das grüne Geheimnis	Magia verde	1953	Italien	Berlinale
21	Das Pestalozzidorf	Das Pestalozzidorf	Pestalozzidorf	1953	Schweiz/UK	Berlinale
22	Hobson's Choice	Herr im Haus bin ich	Hobson's Choice	1954	UK	Berlinale
23	Pane, amore e fantasia	Brot, Liebe und Fantasia	Pane, amore e fantasia	1954	Italien	Berlinale
24	Le détroqué	Der Abtrünnige	détroqué	1954	Frankreich	Berlinale
25	Die Ratten	Die Ratten	Ratten	1955	Deutschland	Berlinale
26	Marcellino pan y vino	Das große Geheimnis des Marcelino	Marcellino pan y vino	1955	Spanien	Berlinale
27	Carmen Jones	Carmen Jones	Carmen Jones	1955	USA	Berlinale
28	Zimmerleute des Waldes	Zimmerleute des Waldes	Zimmerleute des Waldes	1955	Deutschland	Berlinale
29	Siam - Land und Leute	Siam - Land und Leute	Siam - Land und Leute	1955	USA	Berlinale
30	Pantomimes	Pantomimes	Pantomimes	1955	Frankreich	Berlinale
31	Invitation To The Dance	Einladung zum Tanz	Invitation To The Dance	1956	USA	Berlinale
32	Autumn Leaves	Herbststürme	Autumn Leaves	1956	USA	Berlinale
33	Donatella	Junge Liebe in Rom	Donatella	1956	Italien	Berlinale
34	Trapeze	Trapeze	Trapeze	1956	USA	Berlinale
35	La societe	Das Mädchen aus dem Wald	La societe	1956	Frankreich	Berlinale
36	Richard III	Richard III	Richard III	1956	UK	Berlinale
37	12 Angry Men	Die 12 Geschorenen	12 Angry Men	1957	USA	Berlinale
38	Padri e figli	Väter und Söhne	Padri e figli	1957	Italien	Berlinale
39	Woman In A Dressing Gown	Die Frau im Morgenrock	Woman In A Dressing Gown	1957	UK	Berlinale
40	Tizoo	Tizoo	Tizoo	1957	Mexiko	Berlinale
41	Ameneecer en puerta oscura	Der Tag der Verdammten (Wem Gott vergibt)	Ameneecer en puerta oscura	1957	Spanien	Berlinale

Figura 30 – Fixter Excel per a carregar

La carrega dels fitxers obtinguts es fa mitjançant la utilitat sqldr. Tots es carreguen a la taula Dades que és el repositori de dades. D'aquesta taula partiran les dades netejades a les taules de fet.

S'ha creat un fitxer .bat per a automatitzar la càrrega. Aquest script crida la càrrega de cadascun dels fitxers segons els fitxers de configuració amb extensió .ctl i deixant el log de les operacions.

```

carrega_dades.bat - Bloc de notas
Archivo Edición Formato Ver Ayuda
sqldr aca/aca control='dades_awards.ctl' log='dades_awards.log'
sqldr aca/aca control='dades_berlinale.ctl' log='dades_berlinale.log'
sqldr aca/aca control='dades_cannes.ctl' log='dades_cannes.log'
sqldr aca/aca control='dades_cesar.ctl' log='dades_cesar.log'
sqldr aca/aca control='dades_mtv.ctl' log='dades_mtv.log'
  
```

Figura 31 – Càrrega de les dades

Als fitxers de control es defineixen els fitxers que es carregaran, el fitxer que recollirà els possibles errors, la taula que es carregarà, el format del fitxer que es carrega i l'ordre de les columnes que es carregaran i a quin camp de la taula anirà la dada llegida. També es fan modificacions per eliminar els espais en blanc, s'agafen només les quatre últimes xifres de l'any per unificar la dada, s'hi pot trobar anys amb format yyy1-yyy2 i yyyy2 però ens quedem amb el segon format, i si es carregaran les dades en majúscules.

```

dades_awards.ctl - Bloc de notas
Archivo Edición Formato Ver Ayuda
OPTIONS (SKIP=1)
LOAD DATA
CHARACTERSET 'WE8ISO8859P1'
INFILE 'E:\sql\Carrega\Academy_Awards_2006.csv'
BADFILE 'E:\sql\Carrega\Academy_Awards_2006.err'
INTO TABLE DADES
APPEND
FIELDS TERMINATED BY ";"
TRAILING NULLCOLS
(
  TITOL_ORIGINAL "ltrim(rtrim(upper(:TITOL_ORIGINAL)))",
  TITOL_ANGLES "ltrim(rtrim(upper(:TITOL_ANGLES)))",
  TITOL_ORDRE "ltrim(rtrim(upper(:TITOL_ORDRE)))",
  TEMPS_ANY "substr(:TEMPS_ANY, -4, 4)",
  PAIS "ltrim(rtrim(upper(:PAIS)))",
  FESTIVAL "ltrim(rtrim(upper(:FESTIVAL)))",
  CATEGORIA "ltrim(rtrim(upper(:CATEGORIA)))",
  GUANYADOR NULLIF GUANYADOR='0',
  NOMINAT "ltrim(rtrim(upper(:NOMINAT)))",
  ID_DADES "DADES_SEQ.NEXTVAL"
)
  
```

Figura 32 – Fitxer de control de càrrega de les dades dels Òscars.

C. Transformació i càrrega de les taules auxiliars

Una vegada carregat el repositori s'han de netejar les dades per a passar-les al seu destí que són les taules de fets Fact_Artista, Fact_Festival i Fact_Film a partir de les quals es crearan els informes demanats. Es camps nuls que no són nominat ni guanyadors es transformen a 'NO INFORMAT' per a que no hi hagi problemes perquè 'count' no compta els valors nuls. Es fa una inspecció visual dels títols per tal d'identificar errades. Es troben errades com, per exemple, el film 'Cría Cuervos' se'n diu també 'Cría'. Es transforma el valor a 'Cría cuervos' després de consultar el títol correcte a www.imdb.com.

Taula Tipus_Artista

S'omple la taula Tipus_Artista amb els valors 'Actor', 'Actriu' i 'Director' executant el fitxer '01_tipus_artista.bat' que crida la inserció directa dels valors:

```

INSERT INTO "TIPUS_ARTISTA" (TIPUS) VALUES ('ACTOR');
INSERT INTO "TIPUS_ARTISTA" (TIPUS) VALUES ('ACTRIU');
INSERT INTO "TIPUS_ARTISTA" (TIPUS) VALUES ('DIRECTOR');
  
```

Degut a que són molt pocs valors també s'hagués pogut introduir els valors directament a la taula gràcies a la facilitat que ofereix l'eina SQL Developer.

Taula Artista

La càrrega dels artistes es van fer per tipus d'artista, identificant els que havien obtingut un premi corresponent a un actor, actriu o director per a poder seleccionar-los. S'assoleix mitjançant l'executable '02_carrega_artista.bat'.

```

INSERT INTO ARTISTA(NOM_ARTISTA)
SELECT DISTINCT DADES.NOMINAT
FROM DADES, CATEGORIA
WHERE CATEGORIA.CAT_ESTANDARD LIKE '%ACTOR%'
AND CATEGORIA.CAT_PROPIA = DADES.CATEGORIA;
  
```

A cada càrrega d'artistes els hi corresponia la posterior assignació del seu tipus, que es pot saber perquè el camp 'Tipus_artista' es nul:

```
UPDATE ARTISTA
SET ARTISTA.TIPUS =
(SELECT TIPUS_ARTISTA.ID_TIPUS
FROM TIPUS_ARTISTA
WHERE TIPUS_ARTISTA.TIPUS = 'ACTOR')
WHERE ARTISTA.TIPUS IS NULL;
```

Es va obrar de la mateixa forma per a les actrius i els directors.

Taula Pais

El fitxer '03_carrega_pais.bat' crida la càrrega de la taula Pais es de la informació del la taula 'DADES'. Es carrega el camp 'PAIS_IN'.

```
INSERT INTO PAIS(PAIS_IN)
(SELECT DISTINCT DADES.PAIS
FROM DADES
ORDER BY DADES.PAIS);
```

Es va mapejar manualment la taula Pais per tal d'unificar els valors i obtenir dades homogènies als informes. Per minimitzar la possible errada per una entrada manual, es classifica alfabèticament els països carregats al camp 'pais_in' ja que normalment es trobaran a prop els noms de països en anglès, francès i alemany. S'utilitza la tècnica de copiar i enganxar per a minimitzar la quantitat de dades picades manualment que és una font d'errors. D'aquesta manera s'omple el camp 'pais_out' per a completar el mapejat dels països.

S'assumeix que els valors '0' i 'null' fan referència als països fan referència als Estats Units, per tant aquest mapeig sí que està automatitzat a la mateixa càrrega de països:

```
UPDATE PAIS SET PAIS_OUT = 'EUA'
WHERE PAIS_IN = '0'
OR PAIS_IN IS NULL;
```

Taula Categoria

Primer es van carregar les categories des de la taula Dades al camp 'Categoria_propia' seleccionant les distintes categories mitjançant l'execució del fitxer bat '04_insert_categoria.bat':

```
INSERT INTO CATEGORIA(CAT_ORIGINAL)
SELECT DISTINCT DADES.CATEGORIA
FROM DADES
ORDER BY CATEGORIA;
```

Per al mapejat de la categoria es seleccionen les categories que porten la paraula 'Actor' per a mapejar-les com 'Millor Actor', 'Actriu' per a 'Millor Actriu' i 'Director' com a 'Millor Director', tenint en compte les diferents maneres de d'anomenar els premis a cada festival. Les categories que romanen sense mapejar es fan de manera manual, directament a la taula gràcies a les facilitats que ofereix l'SQL Developer al camp 'Categoria_Estandard'.

Taula Festival

S'omple la taula festival amb els noms dels festivals carregats al repositori:

```
INSERT INTO FESTIVAL (NOM_FESTIVAL)
SELECT DISTINCT DADES.FESTIVAL
FROM DADES;
```

Taula Film

S'omple la taula film amb els títols originals i títols d'ordre dels films que són a la taula dades. S'executa el fitxer 06_insert_film.bat.

```
INSERT INTO FILM (TITOL_ORIGINAL, TITOL_ORDRE)
SELECT DISTINCT D.TITOL_ORIGINAL, D.TITOL_ORDRE
FROM DADES D;
```

Taula Temps_Any

S'omple la taula Temps_Any amb els anys de la taula dades. S'executa el fitxer 07_insert_any.bat.

```
INSERT INTO TEMPS_ANY (NUM_ANY);
SELECT DISTINCT D.TEMPS_ANY
FROM DADES D;
```

D. Càrrega de les taules de fets

A continuació s'explicarà el procés de càrrega de les tres taules de fets identificades. Aquesta càrrega ha estat automatitzada mitjançant un script .bat que crida l'sqlplus i que passa el fitxer amb les ordres sql que carreguen les taules de fets.

Fact_Artista

S'omple la taula de fets relatiu als artistes amb les claus primàries de les taules amb les que estan relacionades. Els indicadors nominat (quantitat de nominacions) i guanyador (quantitat de premis obtinguts) s'assoleixen comptant les vegades que apareixen al repositori. Les taules que estan relacionades són artista per a saber de quin artista estem parlant, festival que indica a quin festival està nominat o premiat, categoria indica la categoria de premi al que opta l'artista i any per a saber l'any de la nominació.

```
INSERT INTO FACT_ARTISTA (ID_ARTISTA, ID_FESTIVAL, ID_CATEGORIA,
NUM_ANY, NOMINACIONS, PREMIS)
SELECT ARTISTA.ID_ARTISTA, FESTIVAL.ID_FESTIVAL,
CATEGORIA.ID_CATEGORIA,
TEMPS_ANY.ID_TEMPS_ANY, COUNT (DADES.NOMINAT), COUNT (DADES.GUANYADOR)
FROM ARTISTA, FESTIVAL, CATEGORIA, TEMPS_ANY, DADES
WHERE DADES.NOMINAT = ARTISTA.NOM_ARTISTA
AND DADES.FESTIVAL = FESTIVAL.NOM_FESTIVAL
AND DADES.CATEGORIA = CATEGORIA.CAT_PROPIA
AND DADES.TEMPS_ANY = TEMPS_ANY.NUM_ANY
GROUP BY ARTISTA.ID_ARTISTA, FESTIVAL.ID_FESTIVAL,
CATEGORIA.ID_CATEGORIA, TEMPS_ANY.ID_TEMPS_ANY;
```

Fact_Film

Aquesta és la taula de fets amb els indicadors relatius a les pel·lícules. Les claus foranes provenen de les taules film, òbviament per a saber de quin film estam parlant, festival, indica el festival al que hi participa, categoria, premi al que està nominat, país, país o grup de països que produeixen la pel·lícula necessaris pels informes relatius als països i finalment num_any indica l'any del festival al que es presenta la pel·lícula.

Els indicadors són la quantitat de nominacions i de premis obtinguts per a cada conjunt de claus anteriorment explicat.

```
INSERT INTO FACT_FILM(ID_FILM, ID_FESTIVAL, ID_CATEGORIA, ID_PAIS,
NUM_ANY,
NOMINACIONS, PREMIS)
SELECT FILM.ID_FILM, FESTIVAL.ID_FESTIVAL, CATEGORIA.ID_CATEGORIA,
PAIS.ID_PAIS, TEMPS_ANY.ID_TEMPS_ANY, COUNT(DADES.NOMINAT),
COUNT(DADES.GUANYADOR)
FROM FILM, FESTIVAL, CATEGORIA, PAIS, TEMPS_ANY, DADES
WHERE DADES.TITOL_ORDRE = FILM.TITOL_ORDRE
AND DADES.FESTIVAL = FESTIVAL.NOM_FESTIVAL
AND DADES.CATEGORIA = CATEGORIA.CAT_PROPIA
AND DADES.PAIS = PAIS.PAIS_IN
AND DADES.TEMPS_ANY = TEMPS_ANY.NUM_ANY
GROUP BY FILM.ID_FILM, FESTIVAL.ID_FESTIVAL, CATEGORIA.ID_CATEGORIA,
PAIS.ID_PAIS, TEMPS_ANY.ID_TEMPS_ANY;
```

Fact_Festival

Conté els indicadors referents als festivals que s'han celebrat, tenint en compte els festivals celebrats i els anys en el que s'han celebrat. Es carrega una taula auxiliar i es llença un disparador a cada inserció que es carrega els festivals, els anys i calcula els indicadors sobre el nombre de nominacions dels actors, actrius i pel·lícules.

```
INSERT INTO FACT_FESTIVAL_TEMP(ID_FESTIVAL, NUM_ANY)
SELECT DISTINCT FESTIVAL.ID_FESTIVAL, TEMPS_ANY.ID_TEMPS_ANY
FROM DADES, FESTIVAL, TEMPS_ANY
WHERE DADES.FESTIVAL = FESTIVAL.NOM_FESTIVAL
AND DADES.TEMPS_ANY = TEMPS_ANY.NUM_ANY;
```

Fact_Director

Aquesta taula de fets guarda la informació respecte als directors

```
INSERT INTO FACT_DIRECTOR(ID_DIRECTOR, ID_FESTIVAL, NUM_ANY, ID_FILM,  
NOMINACIONES, PREMIS)  
SELECT ARTISTA.ID_ARTISTA, FESTIVAL.ID_FESTIVAL,  
TEMPS_ANY.ID_TEMPS_ANY,  
FILM.ID_FILM, COUNT(DADES.NOMINAT), COUNT(DADES.GUANYADOR)  
FROM ARTISTA, TIPUS_ARTISTA, FESTIVAL, TEMPS_ANY, FILM, DADES  
WHERE DADES.NOMINAT = ARTISTA.NOM_ARTISTA  
AND ARTISTA.TIPUS = TIPUS_ARTISTA.ID_TIPUS  
AND TIPUS_ARTISTA.TIPUS = 'DIRECTOR'  
AND DADES.FESTIVAL = FESTIVAL.NOM_FESTIVAL  
AND DADES.TEMPS_ANY = TEMPS_ANY.NUM_ANY  
AND DADES.TITOL_ORDRE = FILM.TITOL_ORDRE  
GROUP BY ARTISTA.ID_ARTISTA, FESTIVAL.ID_FESTIVAL,  
TEMPS_ANY.ID_TEMPS_ANY,  
FILM.ID_FILM;
```

E. Accés a la Base de Dades

Es va crear un usuari a la base de dades amb les dades següents:

User: aca

Password: aca

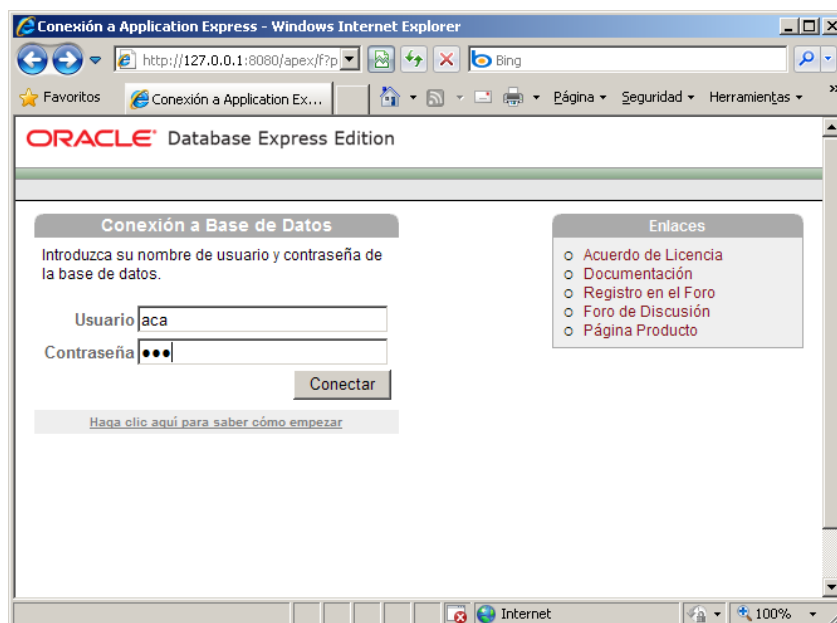


Figura 33 – Accés a la BB.DD.

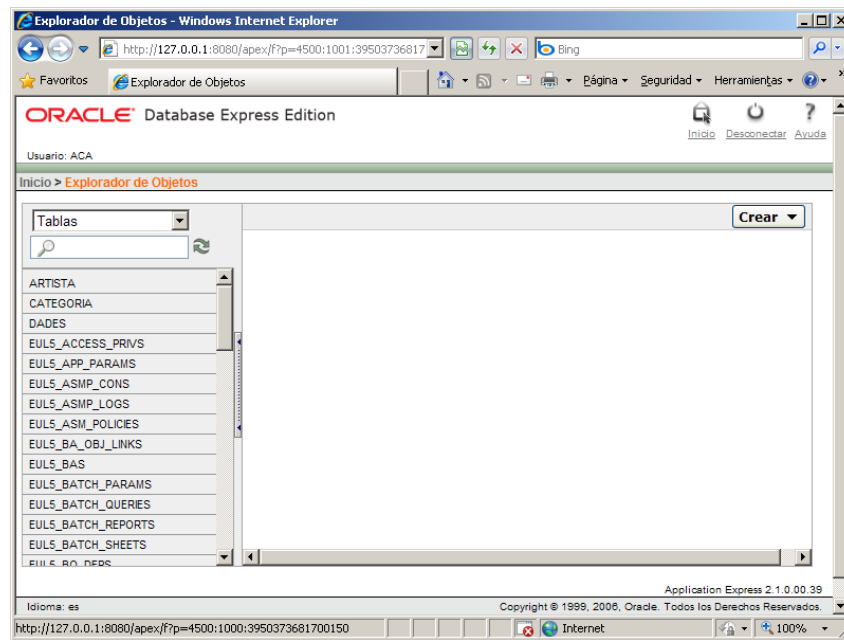


Figura 34 – Entrada a la BB.DD.

O mitjançant el programari Oracle SQL Developer executat desde la màquina amfitriona sobre la que corria la màquina virtual canviant a la màquina virtual la seva connexió a la xarxa a solament amfitrió. Una vegada aconseguida la direcció ip, en aquest cas 192.168.56.101

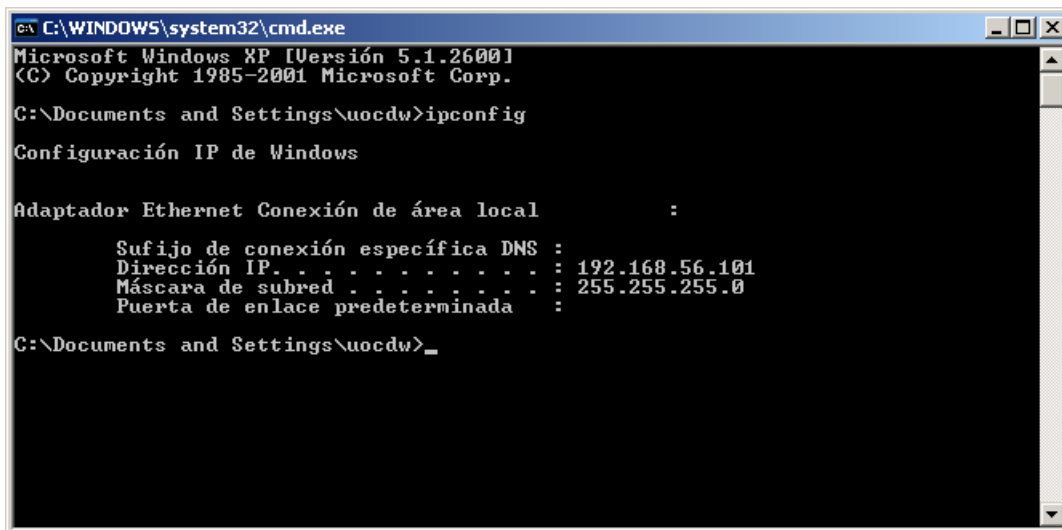


Figura 35 – Direcció IP de la MV

Es va crear la connexió a l'SQL Developer:

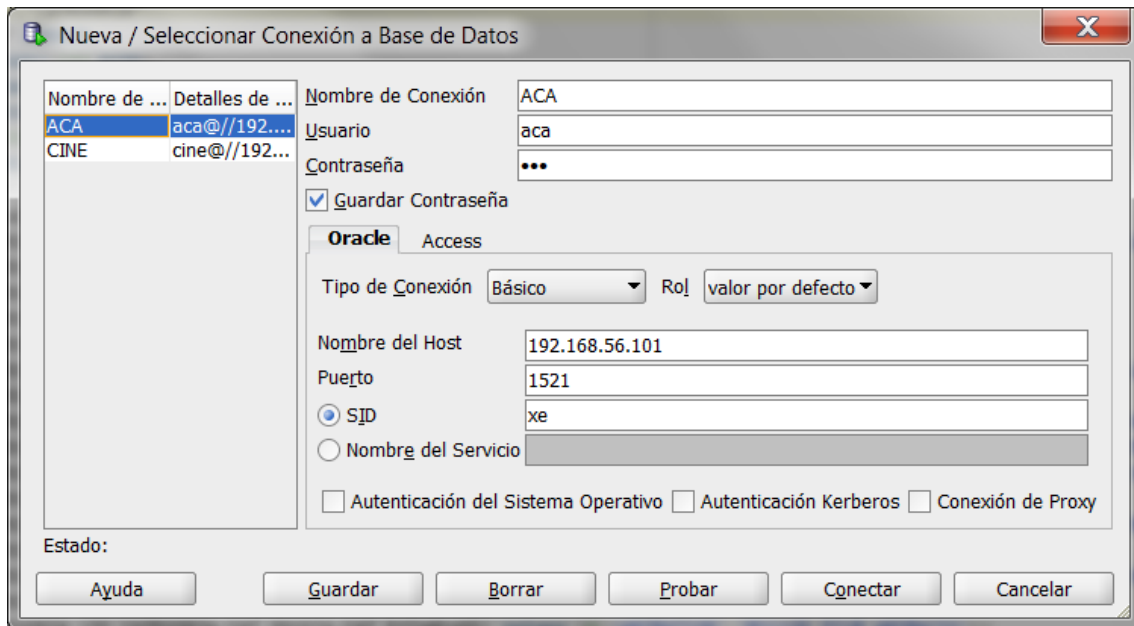


Figura 36 – Connexió SQL Developer

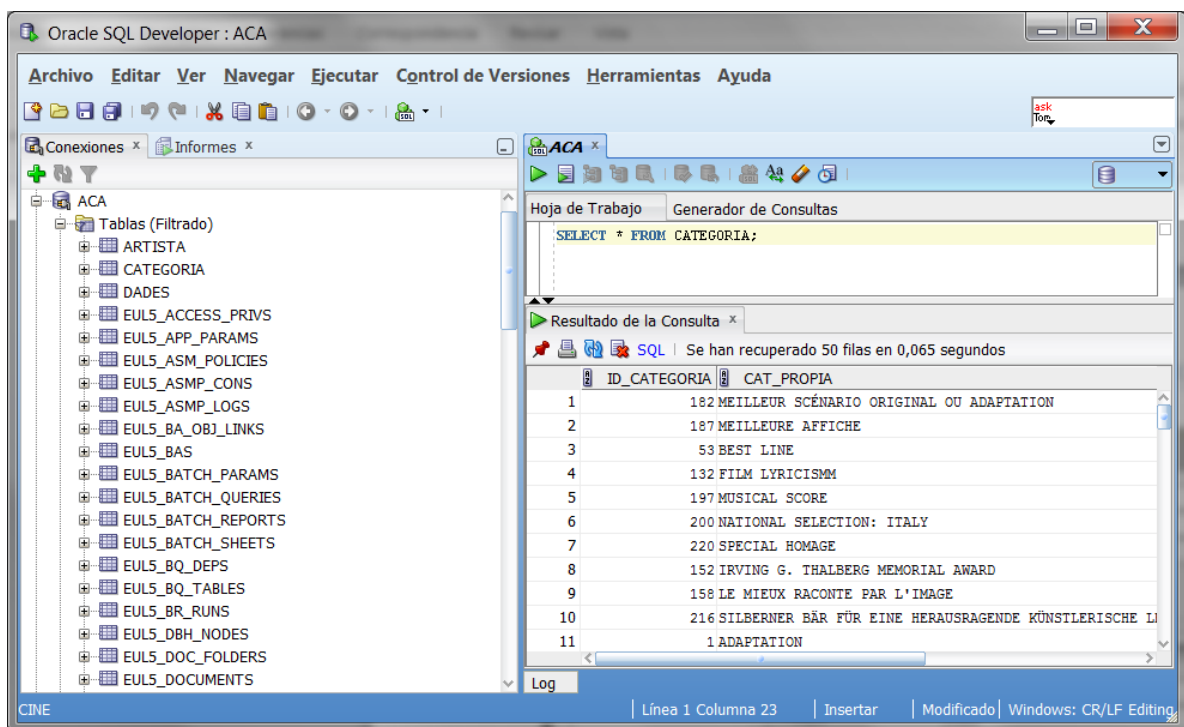


Figura 37 – Detall de les taules des de SQL Developer

F. Oracle BI Discoverer Administrator

Amb el mòdul d'administració del Discoverer vaig crear un End User Layer (EUL) que contindrà l'àrea de negocis i informes associats a la base de dades. Quan es crea un EUL es creen automàticament una sèrie de taules i vistes a l'esquema de l'usuari amb les metadades de les àrees de negoci i objectes de l'EUL.

La connexió es va fer amb l'usuari creat aca amb password aca.

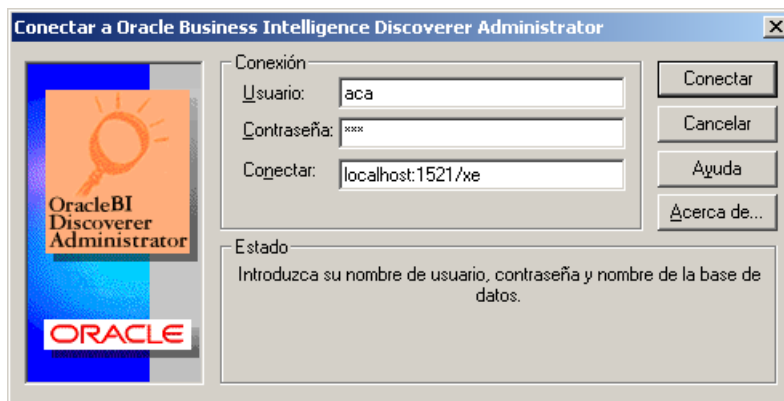


Figura 38 – Accés a Discoverer

L'aplicació va demanar crear un EUL.

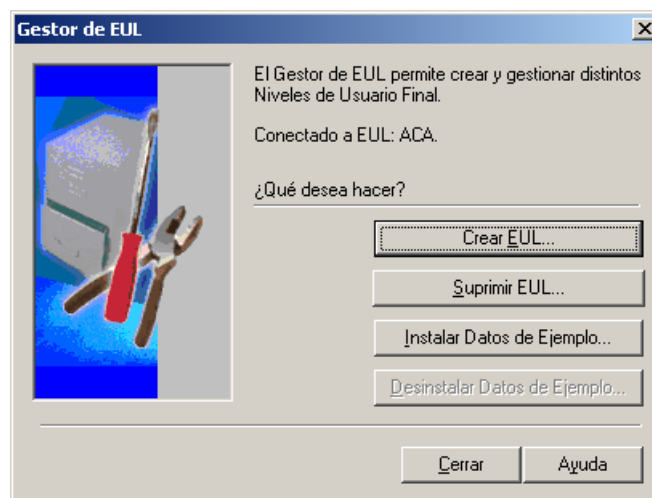


Figura 39 – Creació de l'EUL

Es va seleccionar l'usuari aca.

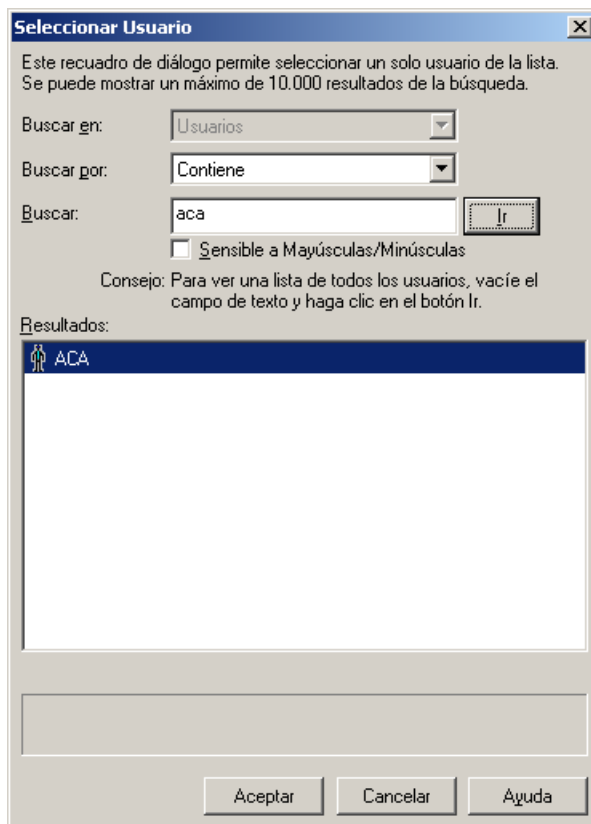


Figura 40 – Seleccionar usuari EUL

I finalment es va crear l'EUL.

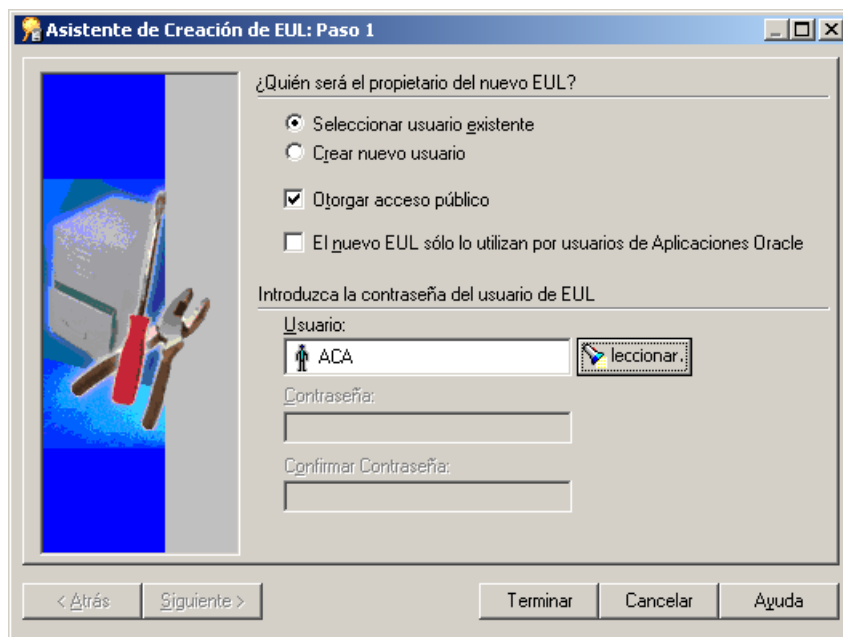


Figura 41 – Creació final de l'EUL

Discoverer crea per defecte una sèrie d'objectes:

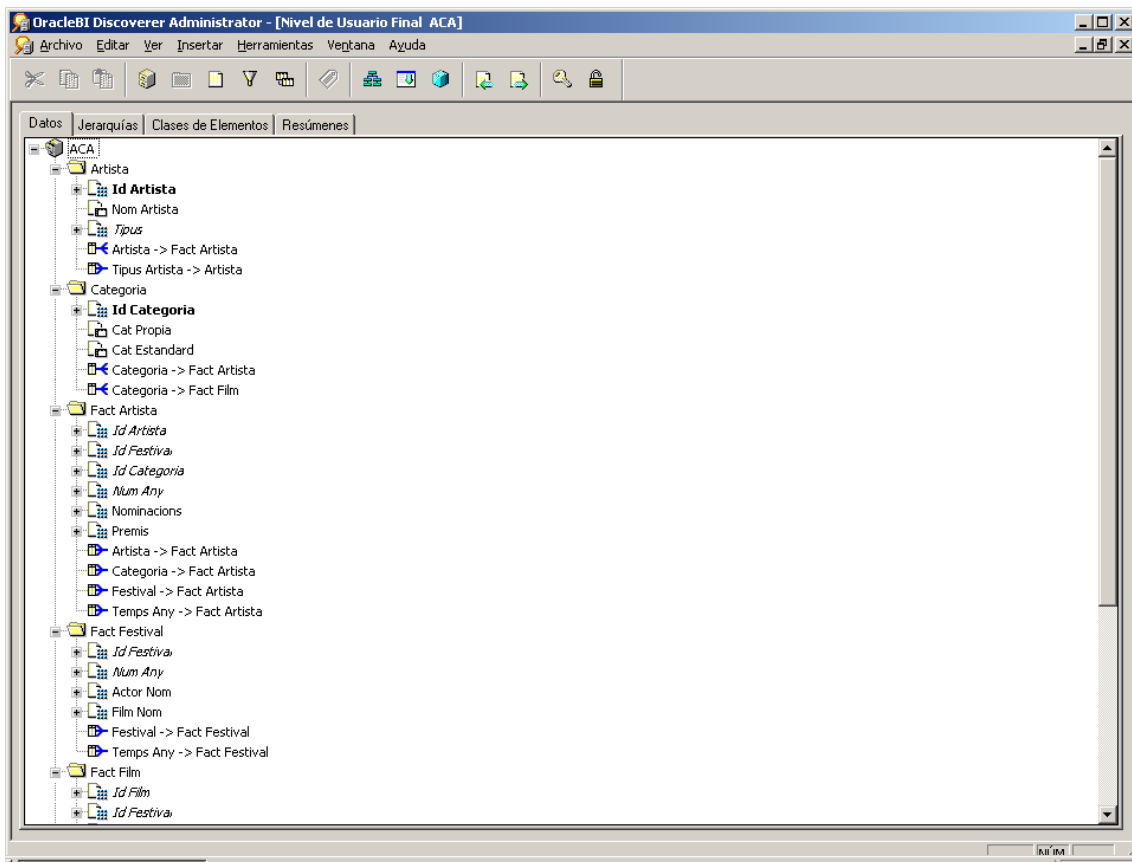


Figura 42 – Objectes creats per Discoverer

G. Oracle BI Discoverer Desktop

Es pot accedir a aquest mòdul com un usuari per a la creació i explotació dels informes. Per a accedir a l'eina Oracle BI Discoverer Desktop es farà servir l'usuari **aca** amb el password **aca** també. La següent imatge mostra els paràmetres que s'han d'introduir per accedir-hi:

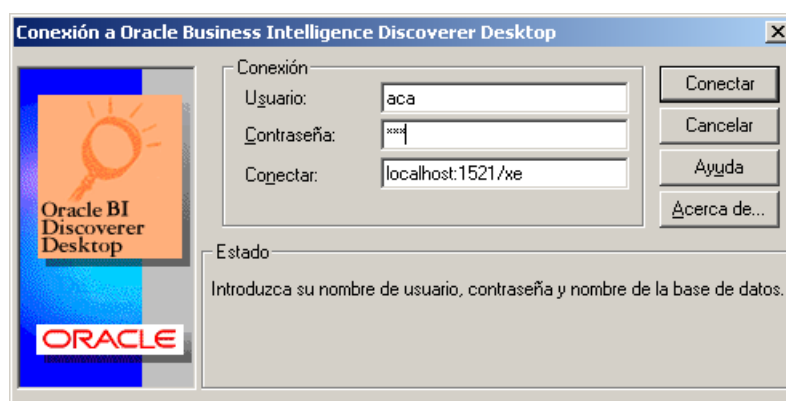


Figura 43 – Accés a Discoverer Desktop

A la pantalla següent es demana crear o obrir un llibre de treball, clicarem a obrir un llibre de treball per poder obrir el que tenim:

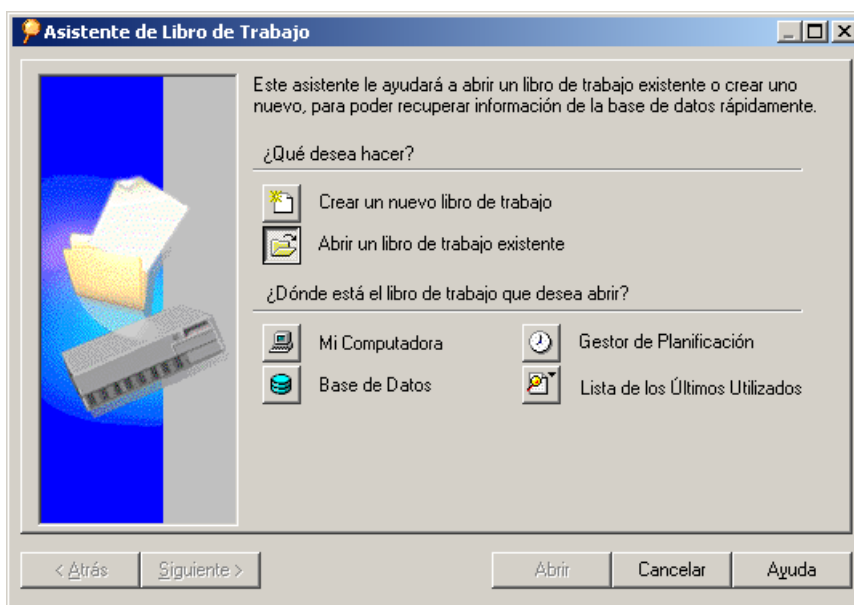


Figura 44 – Obrir un llibre de treball

L'informe està creat a la base de dades, per tant s'ha de clicar al botó 'Base de Datos' i seleccionar-lo, el seu nom es ACA.DIS

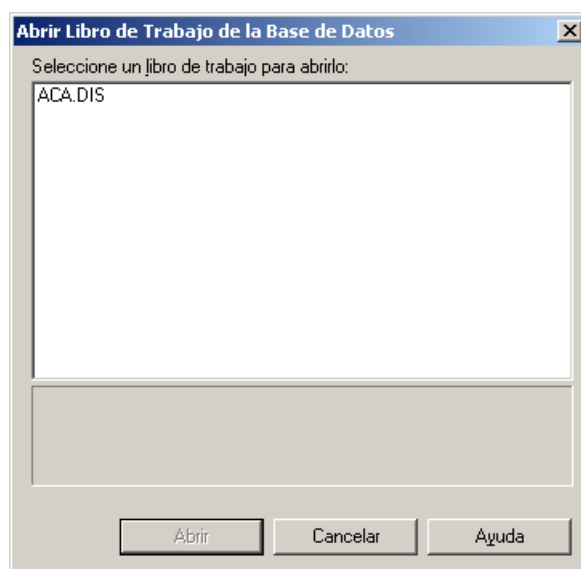


Figura 45 – Llibre de treball

I finalment s'obrirà la darrera fulla que va ser oberta.