Projecte final de carrera

MEMORIA

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

Bartomeu Antich Luque

E.T.I.G.

Consultor: José Ángel Martín Carballo

Curs 2010/11 - Segon Semestre

ÍNDEX

1.- Dedicatòria i agraïments

A na Sonia, sense la qual mai hagués sigut possible arribar fins aquí.

2.- Resum i paraules claus

2.1.- Resum

Actualment la implementació de la informàtica en tots els àmbits de la societat genera una quantitat d'informació que és difícilment assimilable. Passant pel món empresarial, l'educatiu o la sanitat, qualsevol entitat actual disposa d'una gran quantitat de dades procedent dels seus sistemes d'informació. Un dels objectius finals de tenir tota aquesta informació és, o hauria de ser en la majoria de casos, poder-la analitzar d'una forma eficient: la gestió de la informació.

En aquest punt és on intervenen els Magatzems de Dades o Data Warehouse (DW), recopilant, resumint i tractant eficientment el gran volum de dades existent, per poder facilitar l'anàlisi de la informació i oferint suport a la presa de decisions.

En el projecte proposat ens trobem com a client l'Acadèmia de Cinema Andorrà (ACA), clar exemple que els magatzems de dades es poden implementar en qualsevol tipus d'organització que disposi d'una quantitat de dades significativa.

L'ACA ens planteja unificar, creuar i explotar un seguit de dades procedents de diferents arxius, que contenen la informació relativa a les nominacions i premis de diferents festivals de cinema. El seu principal objectiu és poder analitzar tota aquesta informació àgilment.

2.2.- Paraules clau

Data Warehouse, Oracle, SQL, Discoverer, Magatzem de dades, informació cinematogràfica, Treball de final de carrera.

2.3.- Índex de figures

3.- Introducció

El present document és la memòria del Treball de Final de Carrera de la Enginyeria Tècnica de Gestió de l'àrea de Magatzem de Dades.

Hi trobarem el resum del procés de creació complerta d'un magatzem de dades amb eines Oracle, des de l'anàlisi preliminar del problema plantejat fins a la creació dels informes per a la explotació de les dades.

La base de dades que s'ha utilitzat correspon a informació cinematogràfica refent a premis aconseguits.

3.1.- Objectius del TFC

3.1.1.- Objectius generals

El Treball de Fi de Carrera té com objectiu general sintetitzar els coneixements adquirits durant els estudis de l'Enginyeria Tècnica, sent un treball pràctic i vinculat a l'exercici professional. Tenim el repte de poder desenvolupar un cas pràctic real des del començament fins el final i així adquirir experiència en la realització de projectes.

Concretament s'han d'adquirir competències en:

- L'anàlisi d'un cas real pràctic, i la realització d' un projecte.
- Planificar i estructurar el projecte.
- Creació d'una memòria i la presentació del projecte

En l'àrea de Treball final de carrera que ens ocupa, Magatzems de Dades, els objectius de síntesi es centren especialment en les assignatures de bases de dades i estadística.

El nostre objectiu principal ha de ser adquirir experiència en el camp dels Magatzem de Dades, en el seu disseny, construcció i explotació.

3.1.2.- Objectius específics

En la realització d'un Treball final de carrera de l'àrea dels Magatzems de dades, els objectius específics han de ser:

- Adquirir experiència en el disseny d'un magatzem de dades.
- Ampliar coneixement de bases de dades.

- Adquirir experiència en les eines usades .
- Aplicar coneixements d'estadística.
- Creació d'un magatzem de dades.
- Creació d'informes i consultes dirigits a l'explotació de les dades i a la presa de decisions per part de les entitats.

3.1.3.- Informes realitzats

- Nombre de premis aconseguits per un actor/actriu per any i festival. L'informe ha d'estar totalitzat per saber el nombre total de premis.
- Nombre de premis aconseguits per un actor/actriu per any, festival i categoria.
- Nombre de premis aconseguits per un director per any, festival i pel·lícula. L'informe ha d'estar totalitzat per saber el nombre total de premis.
- Nombre de premis aconseguits per any, festival, categoria (s'ha de poder triar si es vol veure l'estàndard o la pròpia) i pel·lícula.
- Donada una data a seleccionar per l'ACA, per cada tipus de festival mostrar el nombre d'edicions realitzades del festival, nombre total d'actors que han estat nominats, nombre total d'actrius que han estat nominades i nombre total de pel·lícules nominades.
- Per a cada any, nombre de premis guanyats per cada pel·lícula i festival (i subtotal per any/pel·lícula).

- Una definició de "pel·lícula exitosa" i "pel·lícula fracassada". A partir d'aquesta definició, i per a cada any, hem de llistar la pel·lícula més exitosa i més fracassada.
- Hem de proporcionar una manera de veure el país (Estats Units a banda) que més premis ha guanyat. Per aquest país, també voldrem veure el nombre de premis per categoria guanyats.
- Explicació si hi ha una relació entre guanyar un premi a un festival i guanyar-ne el mateix a un altre. Detallar també quins criteris s'han seguit per arribar al resultat donat.
- Donat un actor/actriu concret veure el premis aconseguits

3.2.- Enfocament i mètode seguit

Per a realitzar el projecte s'ha seguit la següent metodologia:

- Anàlisi:

- ✓ Anàlisi previ del projecte.
- ✓ Creació del pla de treball.
- ✓ Anàlisi de les dades
- ✓ Anàlisi acurat de requeriments del projecte

- Disseny:
  - ✓ Disseny conceptual del model multidimensional
  - ✓ Disseny físic del model
  - ✓ Disseny del procés ETL
- Implementació:
  - ✓ Construcció del magatzem de dades
  - ✓ Procés ETL
  - ✓ Creació d'informes

3.3.- Planificació

3.3.1.- Tasques

Les tasques realitzades dins cada fase són les següents:

Fase 1: Realització del pla de treball i anàlisi preliminar.

- Realització del pla de treball
- Anàlisi preliminar de dimensions, atributs i informes.
- Anàlisis de les dades operacionals proporcionades.

Fase 2: Anàlisi de requeriments i disseny conceptual i tècnic.

- Anàlisi detallada de requeriments.
- Disseny conceptual del model dimensional.
- Disseny del procediment d'extracció de dades a alt nivell.

Fase 3: Implementació.

- Construcció del magatzem de dades.
- Instal·lació de l'eina d'exportació de dades.
- Construcció dels informes.

- Anàlisi de la informació.

Fase Final: Memòria i presentació.

- Elaboració de la memòria del projecte.
- Presentació del projecte.

3.3.2.- Temporització i Diagrama de Gantt

FASE	TASCA	INICI	FI	DURADA
1	Instal·lació del programari	05/03/2011	10/03/2011	6
	Pla de Treball	09/03/2011	12/03/2011	4
	Anàlisi preliminar	12/03/2011	16/03/2011	5
	Lliurament PAC1	**16/03/2011**		**fita**
2	Anàlisi de requeriments	16/03/2011	01/04/2011	17
	Disseny del model	01/04/2011	20/04/2011	20
	Lliurament PAC2	**20/04/2011**		**fita**
3	Construcció del Magatzem de dades	21/04/2011	09/05/2011	19
	Construcció Informes	09/05/2011	17/05/2011	9
	Anàlisi de la Informació	17/05/2011	25/05/2011	9
	Lliurament PAC3	**25/05/2011**		**fita**
FINAL	Memòria	27/05/2011	07/06/2011	12
	Presentació	07/06/2011	13/06/2011	7
	Lliurament final	**13/06/2011**		**fita**

Diagrama de Gantt:

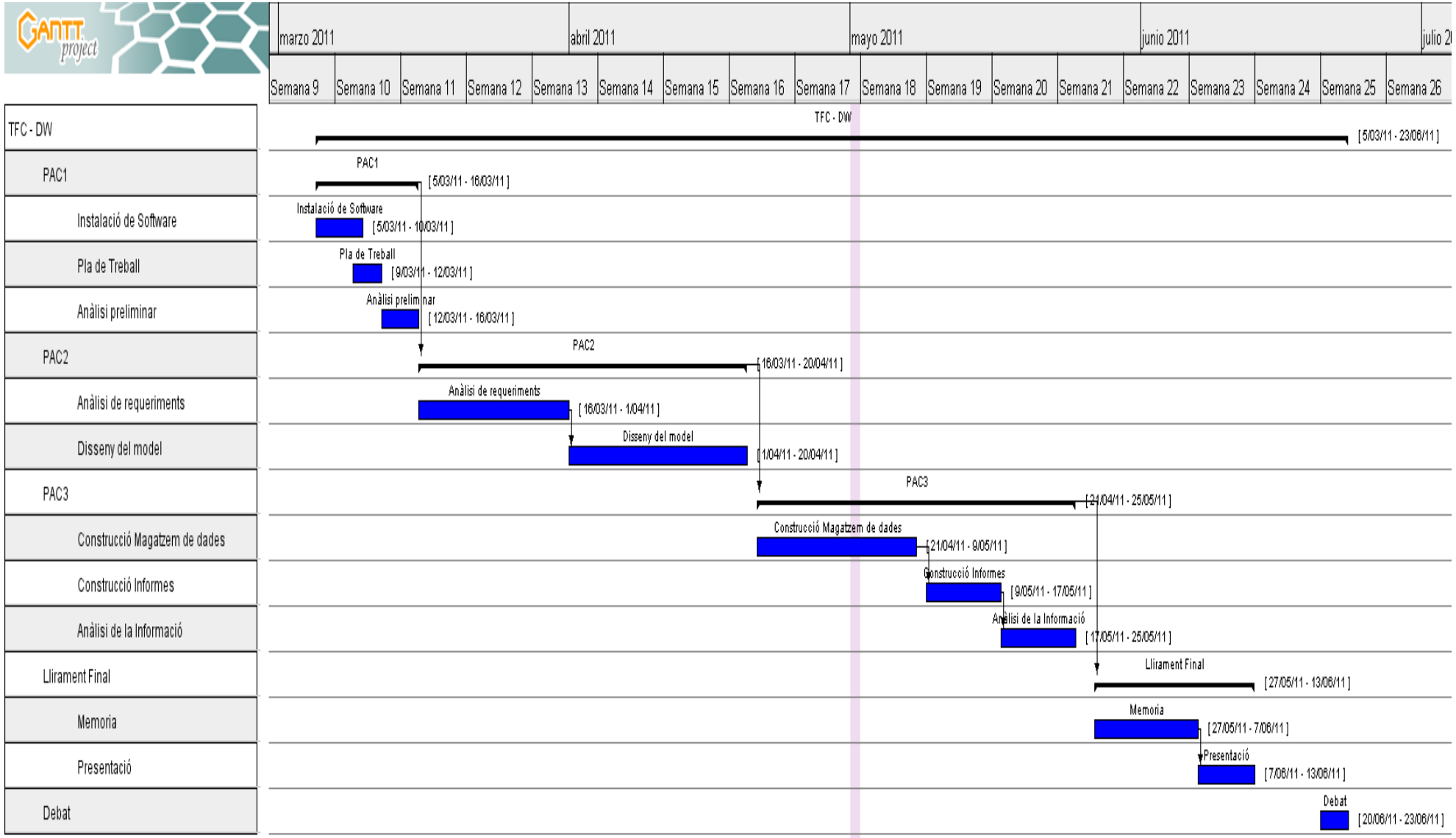


Figura 1.

3.4.- Productes obtinguts

En la realització del present projecte hem obtingut els següent productes:

- Document d'anàlisi preliminar
- Document del pla de treball
- Document d'Anàlisi de requeriments
- Disseny conceptual del magatzem de dades
- Disseny físic del magatzem de dades
- Creació del magatzem de dades
- Procés ETL
- Creació de l'Àrea de Negoci
- Creació dels informes sol·licitats

3.5.- Descripció dels altres capítols de la Memòria.

Anàlisi

En aquest apartat s'analitzen les dades de partida del projecte, els requeriments i les demandes del projecte, així com els fets i les dimensions necessàries a tal efecte.

Disseny

En aquest apartat es presenten tots els dissenys referents al projecte: disseny de programari, maquinari, conceptual i físic.

Captures de pantalla

En aquest apartat es presenten les captures de pantalla del desenvolupament del projecte, la creació de la base de dades, el model multidimensional, i els informes realitzats.

Conclusions

En aquest apartat es presenten les conclusions a la finalització del projecte.

4.- Anàlisi

De l'anàlisi realitzat dels requeriments d'informes del client i de les fonts de dades s'ha obtingut la informació referent a les dimensions, amb els seus atributs, i els fets, amb els indicadors i les dimensions relacionades, que composen el model conceptual del projecte. Aquest model es presenta en els següents apartats.

4.1.- Diagrames de casos d'ús

Segons el projecte que se'ns ha platejat podem distingir dos actors diferenciats, l'administrador i l'usuari final, i entenem que les tasques i casos d'us es distribueixen segons els diagrames següents:

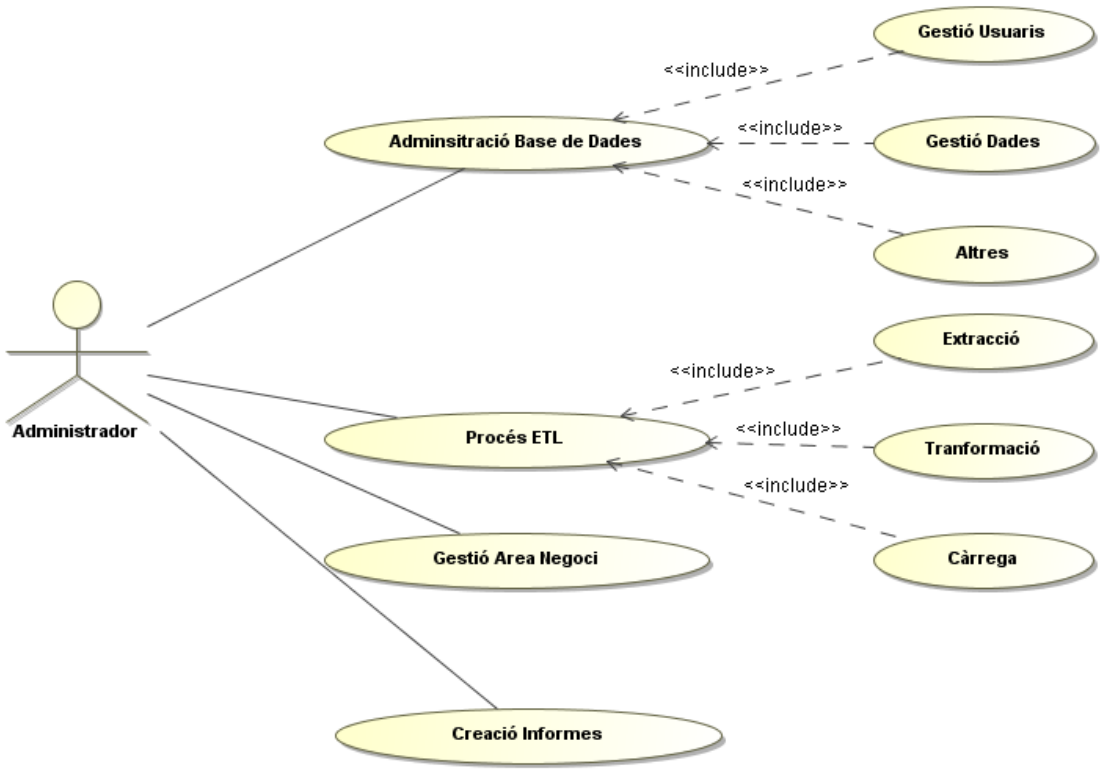


Figura 2.

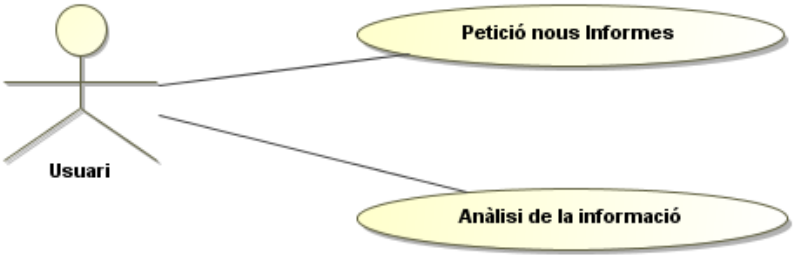


Figura 3.

El procés ETL es podrà realitzar manualment o amb una automatització que permetrà una actualització periòdica de les dades i per tant pot passar a formar part o bé d'un actor rellotge o de l'actor usuari.

4.2.- Fonts de dades

Les fonts de dades proporcionades per l'Acadèmia de Cinema Andorrà són 5 arxius en format Excel, algun d'ells amb més d'un full. Cada arxiu correspon a les dades històriques de nominacions i premis de 5 festivals de cinema diferents:

- Premis de l'Acadèmia, Academy Awards, els Oscars.
- Festival Internacional de Cinema de Berlín, Berlinale.
- Festival Internacional de Cinema de Canes
- Premis César.
- MTV Movie Awards.

Ens interessen concretament aquells fulls on hi apareixen les dades de nominacions i premis.

Les dades que podem extreure de cada arxiu són:

- **Títol Original de la pel·lícula**, és un valor alfanumèric, s'haurà de comprovar que no hi ha errades i unificar els títols.
- **Títol traduït segons la nacionalitat del festival de cinema**, es un valor alfanumèric.

- **Any del festival o de la pel·lícula**, pot presentar dos tipus de format, yyyy o yyyy-xxxx, en l'enunciat del projecte s'ha especificat que en el segon cas caldrà agafar com a referència xxxx.
- **País d'origen de la pel·lícula**, caldrà unificar el nom dels països, per exemple apareix "Germany" i "Allemagne", a banda tenim la problemàtica de la participació de diferents nacionalitats en una mateixa pel·lícula que es veu reflectida en aquest camp. També tenim el cas on no s'especifica país, que en la majoria de casos és quan es refereix a Estats Units.
- **Nom del festival de cinema**, en aquest cas només podem tenir 5 entrades diferents, una per a cada festival.
- **Categoria del premi**, caldrà crear categories unificades, a més es mantindrà l'original per petició del client.
- **Identificador de guanyador/nominat**, una X indica guanyadors, 0 o *espai* indica no guanyador
- **Nominat**, nom del nominat. En el cas dels actor, actriu i director haurem d'unificar els noms. Per exemple, s'ha detectat que apareix algun comentari juntament amb el nom de l'actor ("(refused the award)"). A més poden aparèixer dos noms conjuntament en la mateixa nominació ("Warren Beatty, Buck Henry").
- **Ordre del festival**, valor alfanumèric
- **Número identificador**, valor numèric.

Alguns dels camps tenen noms diferents segons l'arxiu, per aquest motiu s'ha estandarditzat per una major simplificació en aquest anàlisi.

4.3.- Diagrama del model conceptual

De l'anàlisi realitzat dels requeriments d'informes del client i de les fonts de dades s'ha obtingut la informació referent a les dimensions, amb els seus atributs, i els fets, amb els indicadors i les dimensions relacionades, que composen el model conceptual del projecte. Aquest model es presenta en els següents apartats.

4.3.1.- Dimensions

En l'anàlisi s'han observat 6 dimensions diferents: festival, any, categoria, títol, país i nominats per a poder realitzar els informes sol·licitats pel client. Aquestes dimensions es descriuen a continuació.

AWARD

La dimensió Festival, conté els festivals, els atributs són el nom del festival i el codi identificador del festival de cinema. Internament l'anomenarem AWARD per conservar la nomenclatura de les dades del client.

Tenim 5 festivals enumerats: *MTV Movie Awards, César, Cannes Festival, Berlinale* i *Academy Award.*

YEAR

LA dimensió Any, és una dimensió de temps sense jerarquies, en el cas que ens ocupa no hi apareixen mesos o dies que s'haurien de jerarquitzar. Els seus atributs seran el codi identificador i l'any corresponent. L'anomenarem YEAR per conservar la nomenclatura de les dades proporcionades pel client.

Les dades que tenim corresponen a l'interval 1928 a 2006.

CATEGORY

La dimensió Categoria, anomenada CATEGORY pels motius abans explicats, contindrà la informació referent a les categories dels premis atorgats als certàmens de cinema, els seus atributs seran un codi identificador, la denominació de la categoria pròpia del Festival, i una denominació estandarditzada per a algunes de les categories que siguin comunes o ens interessi comparar, com pot ser Millor Actor (*Best Actor*, *Meiller Acteur*).

En les dades que ens han facilitat tenim 258 categories sense estandarditzar.

FILM

La dimensió del Títol conté la informació referent als títols de les pel·lícules nominades en els diferents Festivals. El seus atributs seran un codi identificador únic, el títol original de la pel·lícula, es podria afegir el títol traduït de la pel·lícula com a atribut, però sent el cas que cada festival el té traduït en el seu idioma, no ens resulta de gran

utilitat. Encara que existeixen premis no relacionats amb cap film, l'atribut del nom no usarà el valor NULL per utilitat.

COUNTRY

La dimensió país conté la informació referent al país d'origen de la pel·lícula, els seus atributs són el codi identificador i el nom del país en anglès, aquesta dada com s'ha dit a l'apartat referent a fonts dades s'haurà de transformar. Si el client ho sol·licités es podrien augmentar amb atributs amb traduccions a altres idiomes.

NOMINEES

La dimensió Nominats, anomenada Nominees per conservar la nomenclatura de les dades subministrades pel client, conté les dades dels nominats en els Festivals, els seus atributs són un codi identificador, i el nom o noms dels nominats. En els casos en que hi hagi varis nominats conjuntament no s'ha separat, pel fet de considerar que es podria donar duplicitat de premis, i es considera premi o nominació compartida, no pertanyerà individualment als nominats.

NOMINATION_FOR

S'ha creat una dimensió auxiliar per donar resposta als requeriments dels informes sol·licitats, diferenciant els nominats segons si són actor, actrius o directors, segons un anàlisi realitzat a les dades. Aquesta dimensió no surt reflectida en el disseny ja que només s'utilitza com a filtre.

4.3.2.- Fets

Com a fet principal tenim el nombre de premis aconseguits encreuat amb diverses combinacions de dimensions, però amés tenim un informe que ens demana unes totalitzacions: edicions de festivals, nominacions d'actors, actrius i pel·lícules nominades. Per aquest motiu els agruparem de la següent manera:

- Nombre de premis

- Nombre d'edicions, nominacions d'actors, nominacions d'actrius i nominacions de pel·lícules.

En la definició del fets següents no tenim en compte les propostes d'informes realitzades que no ha sol·licitat el client.

Nombre de premis i nominacions

De manera general el fet principal tindria associades totes les dimensions abans enumerades:

- AWARD
- CATEGORY
- YEAR
- TITLE
- COUNTRY
- NOMINEES

I l'indicador seria el nombre de premis.

El seu esquema seria:

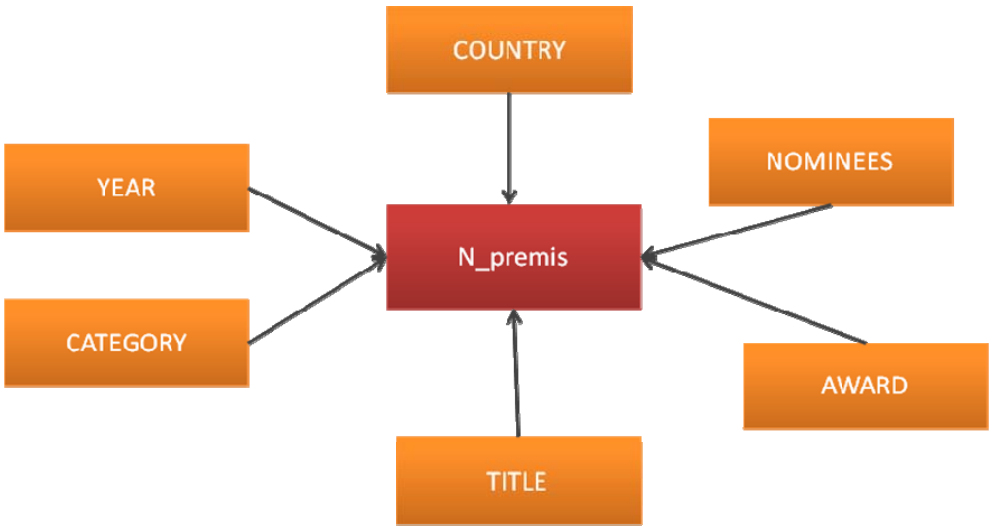


Figura 4.

Però entrant més en detall en cada un dels informes sol·licitats tenim els següents esquemes, on surten reflectides les dimensions que hi intervenen:

- Nombre de premis aconseguits per un actor/actriu per any i festival. Com no podem diferenciar si els nominats són actors o directors per exemple, la dimensió CATEGORY ens servirà per escollir aquelles categories que van destinades només a actors i actrius.

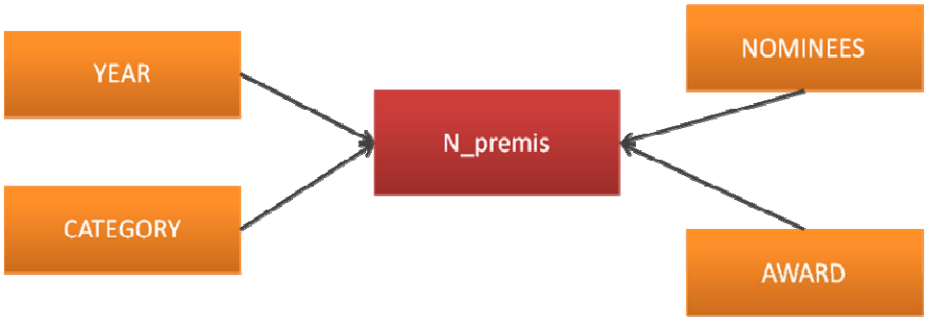


Figura 5.

- Nombre de premis aconseguits per un actor/actriu per any, festival i categoria. Amb el mateix esquema que l'anterior.

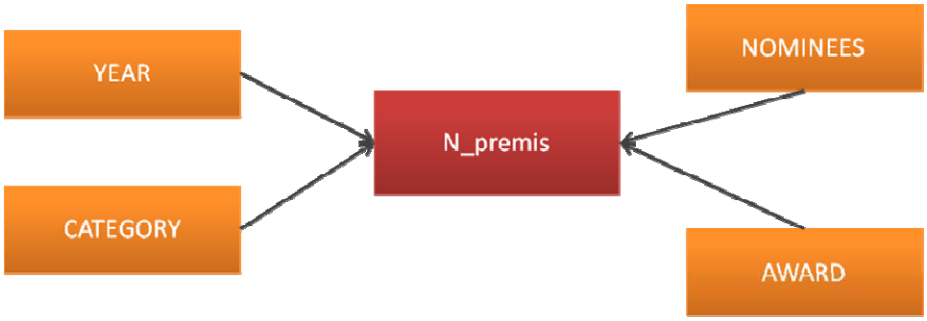


Figura 6.

- Nombre de premis aconseguits per un director per any, festival i pel·lícula. Pel mateix motiu que anteriorment necessitem la dimensió CATEGORY per poder diferenciar entre actors o directors, per exemple.

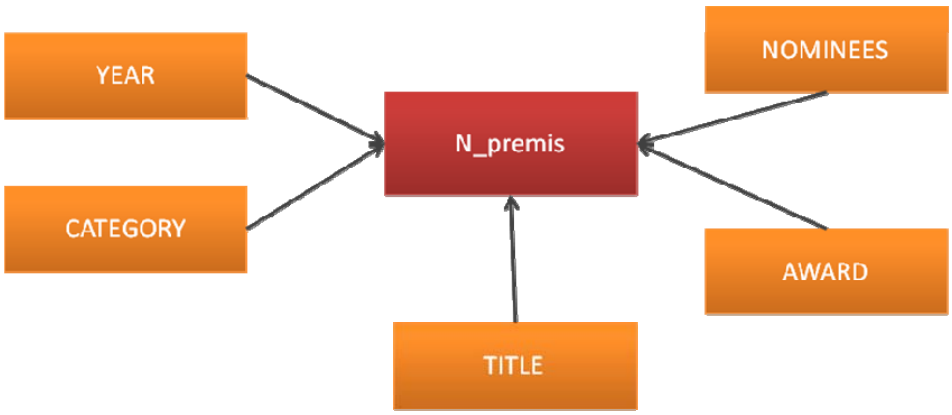


Figura 7.

- Nombre de premis aconseguits per any, festival, categoria i pel·lícula.

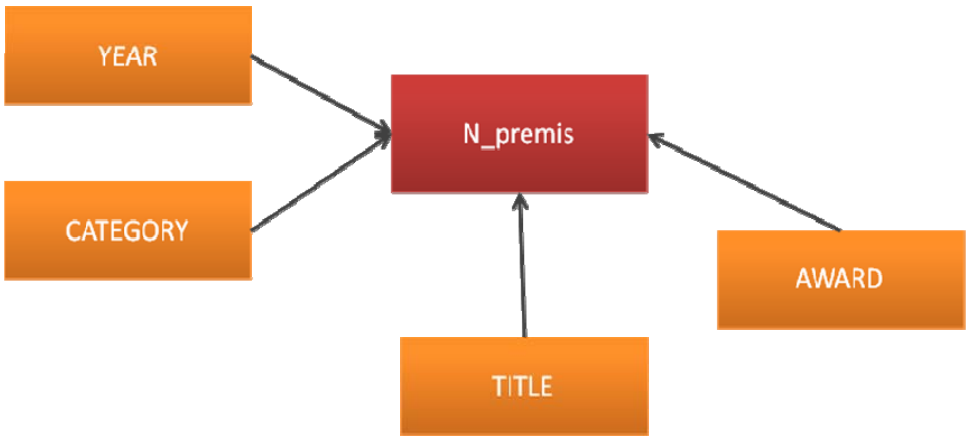


Figura 8.

- Per a cada any, nombre de premis guanyats per cada pel·lícula i festival.

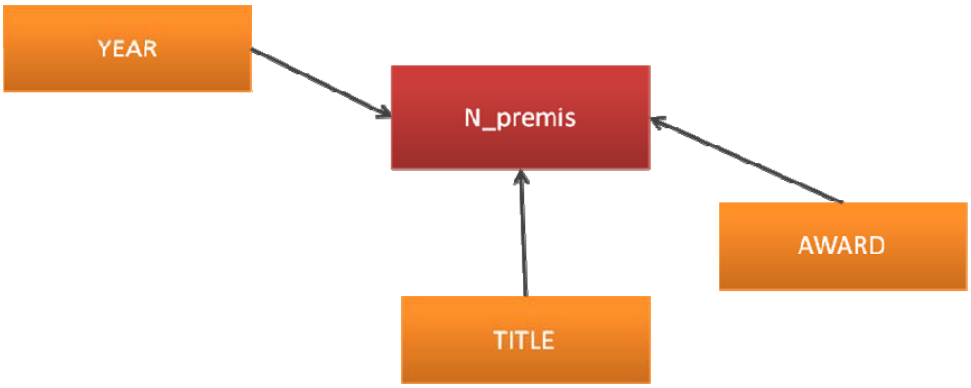


Figura 9.

- País (Estats Units a banda) que més premis ha guanyat. Per aquest país, també voldrem veure el nombre de premis per categoria guanyats.

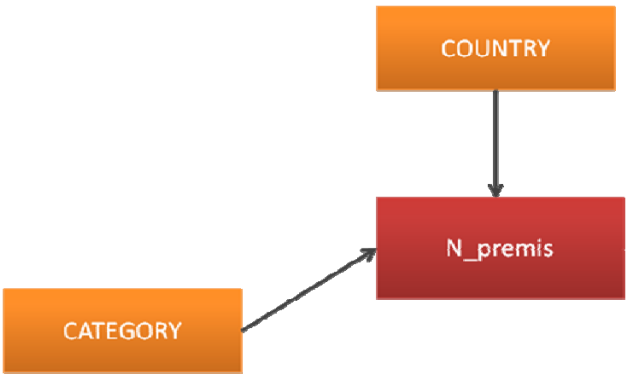


Figura 10.

Totalitzacions

Aquest fet ve donat per la següent sol·licitud:

- Donada una data a seleccionar per l'ACA, per cada tipus de festival mostrar el nombre d'edicions realitzades del festival, nombre total d'actors que han estat nominats, nombre total d'actrius que han estat nominades i nombre total de pel·lícules nominades.

Intervindrien les dimensions:

- AWARD
- CATEGORY
- YEAR

Com en casos anteriors per poder diferenciar, entre les professions que es poden nominar ens cal la dimensió CATEGORY.

Els indicadors d'aquest fet són: el nombre d'edicions del festival, el nombre d'actors nominats, el nombre d'actrius nominades i el nombre de pel·lícules nominades

L'esquema del fet seria:

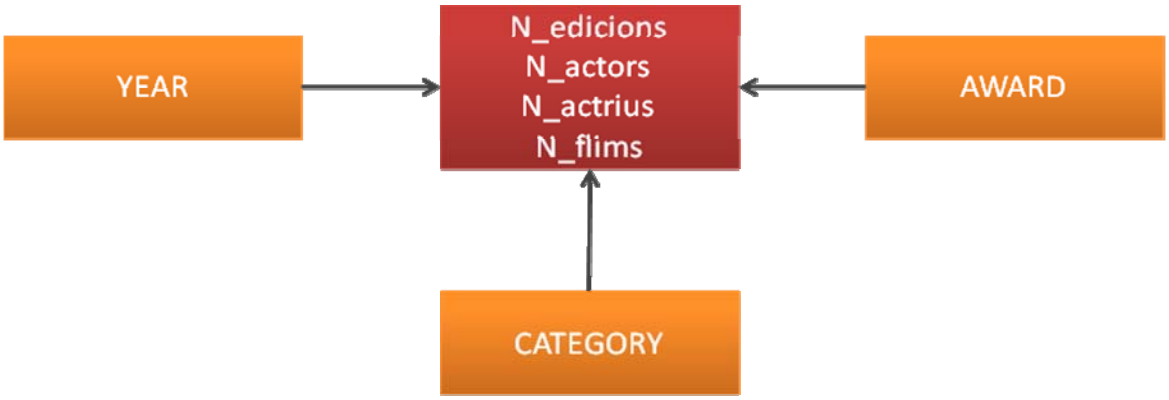


Figura 11.

5.- Disseny

En aquesta part del projecte passarem del disseny conceptual al disseny tècnic. Les parts del model conceptual, dimensions i fets es construiran mitjançant taules amb relacions entre elles per les claus primàries de les dimensions, i creant l'esquema d'estrella.

Així doncs cal crear una taula per a cada dimensió, la taula del fet i les relacions entre elles.

Primerament presentarem també el diagrama de programari i de maquinari requerits.

5.1.- Diagrama de l'arquitectura de programari

Pel desenvolupament del projecte s'ha usat el següent programari:

- Oracle 10g Express Edition
- Oracle Discoverer
- Oracle SQL Developer, versió 1.0.0.15.17
- Microsoft Excel

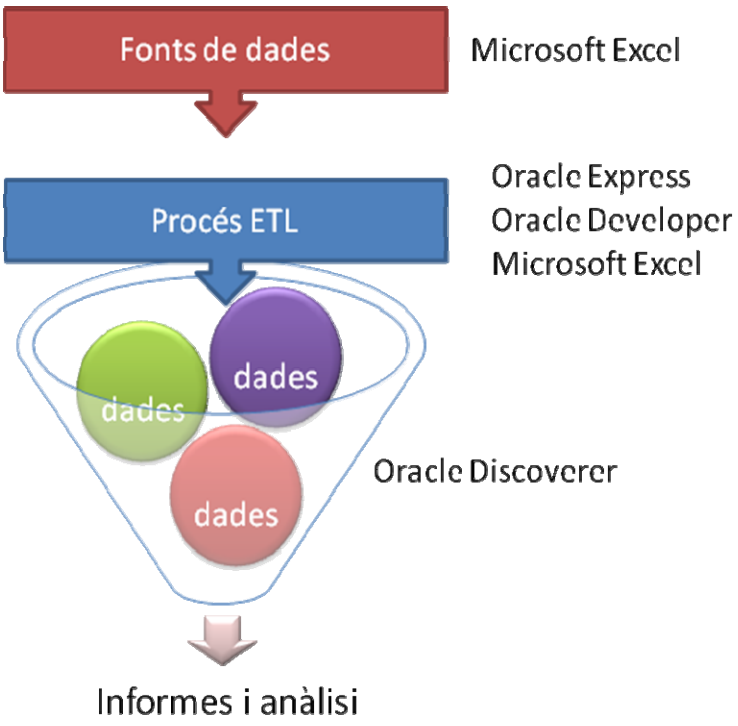


Figura 12.

5.2.- Diagrama de l'arquitectura de maquinari

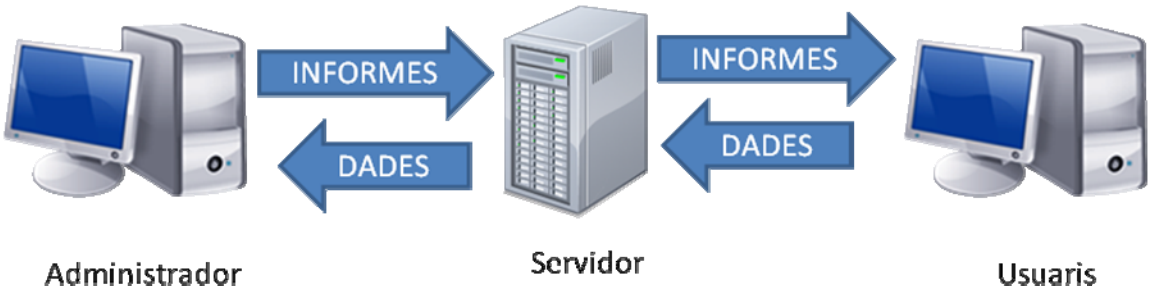


Figura 13.

5.3.- Diagrama de la base de dades i model físic

En aquesta part del projecte passarem del disseny conceptual al disseny tècnic. Les parts del model conceptual, dimensions i fets es construiran mitjançant taules amb relacions entre elles per les claus primàries de les dimensions, i creant l'esquema d'estrella.

Així doncs, cal crear una taula per a cada dimensió, la taula del fet i les relacions entre elles.

5.3.1.- Dimensions

AWARD

La taula AWARD conté els atributs del nom del festival i el codi identificador del festival de cinema, aquest darrer serà la clau primària, i tindrà la relació de clau forania dins la taula del fet. Ambdós atributs no nuls.

Figura 14.

YEAR

La Taula YEAR té per atributs un codi identificador i l'any, el codi és la clau primària.

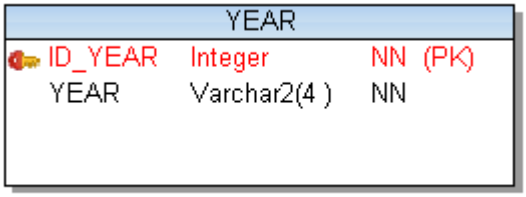


Figura 15.

CATEGORY

La Taula CATEGORY tindrà per atributs un codi identificador, la denominació de la categoria pròpia del Festival (DO_CATEGORY), i una denominació estandarditzada (D_CATEGORY). El codi identificador és la clau primària de la taula. S'ha considerat que D_CATEGORY pot ser NULL per que no totes les categories de premis tenen perquè estar estandarditzades, existeixen categories de premis només a un Festival.

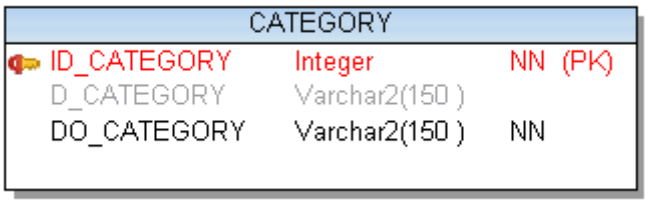


Figura16.

FILM

La taula FILM tindrà per atributs un codi identificador únic i el títol original de la pel·lícula. El codi identificador serà la clau primària de la taula.

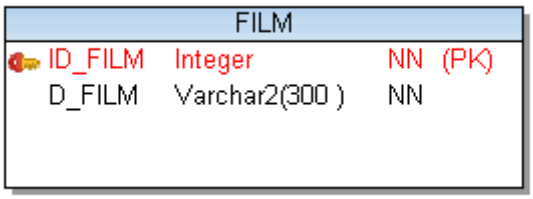


Figura 17.

COUNTRY

La taula COUNTRY té per atributs el codi identificador i el nom del país en anglès, el codi identificador del país serà la clau primària de la taula

Figura 18.

NOMINEES

La taula NOMINEES té per atributs el codi identificador i l'atribut que conté els nominats/nominat.

Figura 19.

5.3.2.- Fets

Premis i nominacions

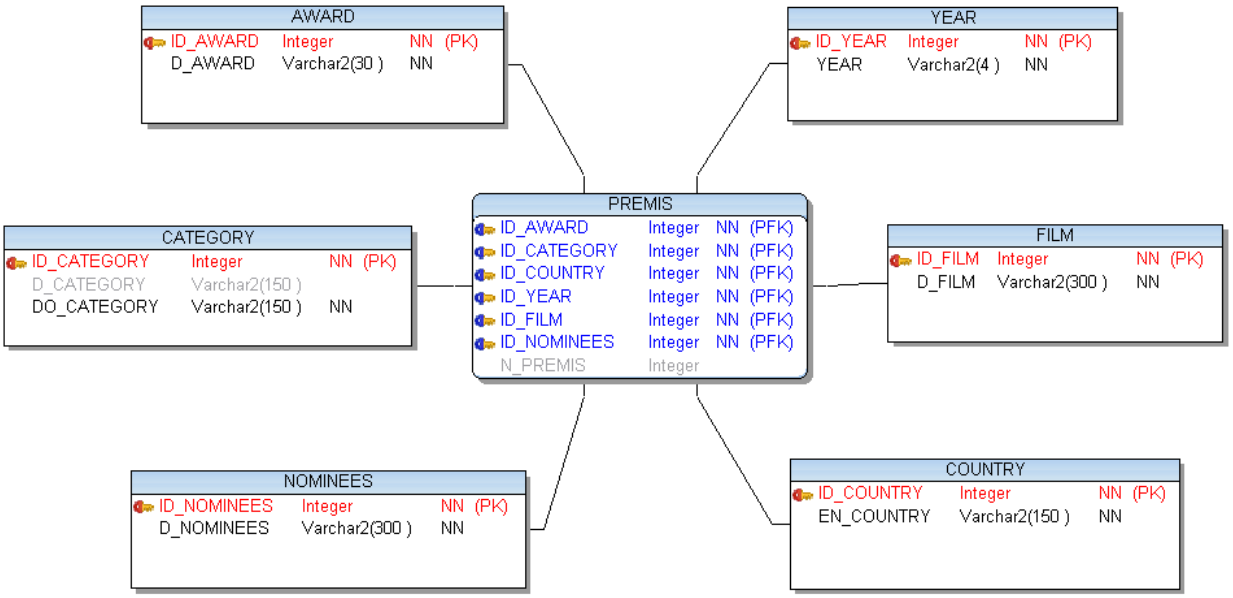


Figura 20.

Totalitzacions

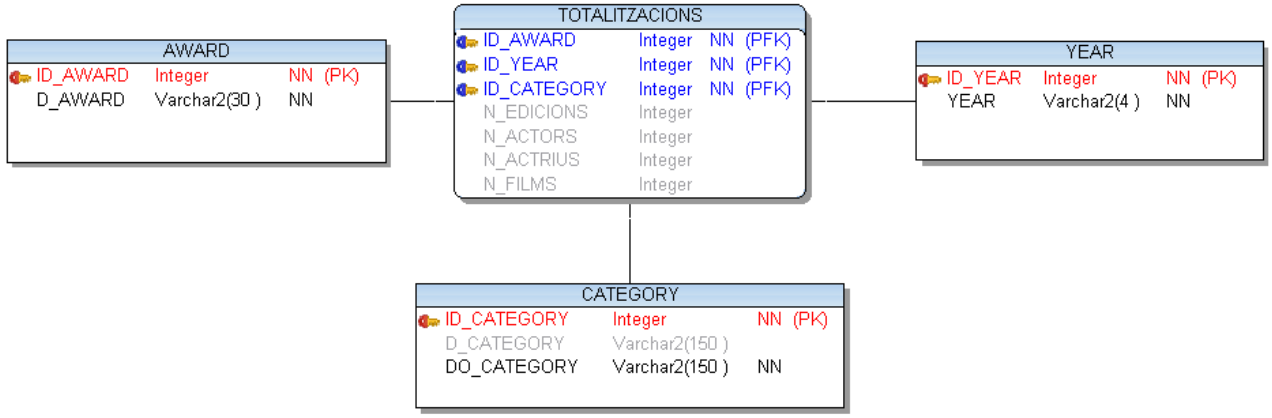


Figura 21.

A la implementació, per a alguns atributs, s'ha decidit canviar el nom per una major agilitat i comprensió dels informes, aquests canvis només afecten majoritàriament a prefixos (D_ i DO_).

5.4.- Procés ETL

El procés ETL (Extract, transform and load) és el procés que realitza les operacions d'extraure, transformar i carregar les dades per al seu ús.

Tots els script's que s'han utilitzat en aquest procés que no han sigut guardats en l'aplicació Oracle Express, amb un indicador de l'ordre a seguir per a excutar-los:

Propietario	Nombre	Descripción
ACA	1_CREACIO_DADES_TMP	Crea la taula de copia de seguretat de Dades
ACA	10_CANVI_NULL_0	Realitza el canvi dels nulls per 0s en winner
ACA	11_INSERT_AWARD	-
ACA	12_INSERT_CATEGORY	-
ACA	13_INSERT_COUNTRY	-
ACA	14_INSERT_EDITION	-
ACA	15_INSERT_FILM	-
ACA	16_INSERT_NOMINEES	-

Figura 22.

Extracció

En la fase d'extracció el client ens ha proporcionat 5 fitxer Excel amb la informació que cal carregar, amb les mateixes columnes de dades i tipus de contingut a pesar de tenir noms diferents.

Per poder passar a la fase de transformació, i donat que la grandària dels fitxers ho permetia, s'han agrupat els 5 en un de sol, per poder fer l'anàlisi preliminar.

Aquest anàlisi s'ha fet mitjançat el propi programari Excel usant filtres i taules dinàmiques i amb la càrrega a una taula temporal usant el programari Oracle 10g Express i el SQL Developer.

Taula de dades creada i carregada amb Oracle Express:

```
CREATE TABLE  "DADES"
   (     "FILM" VARCHAR2(255),
         "YEAR" NUMBER,
         "AWARD" VARCHAR2(255),
         "NOMINEES" VARCHAR2(765),
         "EDITION_ORIG" VARCHAR2(255),
         "COUNTRY_ORIG" VARCHAR2(255),
         "CATEGORY_ORIG" VARCHAR2(255),
         "EDITION_UNIF" VARCHAR2(255),
         "COUNTRY_UNIF" VARCHAR2(255),
         "CATEGORY_UNIF" VARCHAR2(255),
         "NOMINATION_FOR" VARCHAR2(255),
         "ITEM" NUMBER,
         "ID_DADES" NUMBER,
         "WINNER" VARCHAR2(1),
          CONSTRAINT "DADES_PK" PRIMARY KEY ("ID_DADES") ENABLE
   )
/

CREATE OR REPLACE TRIGGER  "BI_DADES"
  before insert on "DADES"
  for each row
begin
    select "DADES_SEQ".nextval into :NEW.ID_DADES from dual;
end;

/
ALTER TRIGGER  "BI_DADES" ENABLE
/
```

Nombre De Columna	Tipo De Dato	Nulo	Valor Por Defecto	Clave Primaria
ID_DADES	NUMBER	No	-	1
FILM	VARCHAR2(255)	Yes	-	-
YEAR	NUMBER	Yes	-	-
AWARD	VARCHAR2(255)	Yes	-	-
NOMINEES	VARCHAR2(765)	Yes	-	-
EDITION_ORIG	VARCHAR2(255)	Yes	-	-
COUNTRY_ORIG	VARCHAR2(255)	Yes	-	-
CATEGORY_ORIG	VARCHAR2(255)	Yes	-	-
EDITION_UNIF	VARCHAR2(255)	Yes	-	-
COUNTRY_UNIF	VARCHAR2(255)	Yes	-	-
CATEGORY_UNIF	VARCHAR2(255)	Yes	-	-
NOMINATION_FOR	VARCHAR2(255)	Yes	-	-
ITEM	NUMBER	Yes	-	-
WINNER	VARCHAR2(1)	Yes	-	-
				1 - 14

Figura 23.

Càrrega a la taula dades mitjançat les utilitats d'Oracle Express:

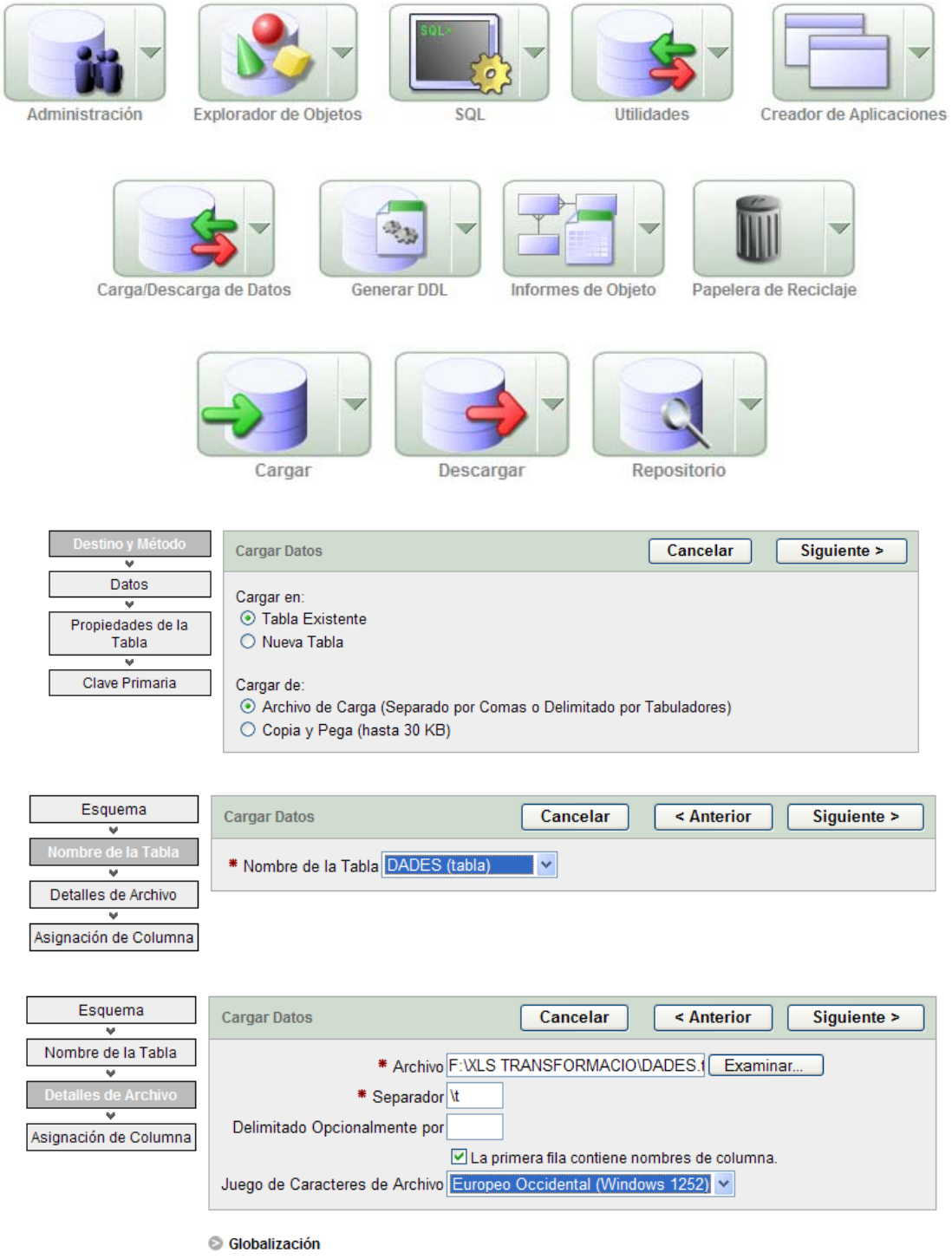


Figura 24.

Transformació

El procés de transformació és la part del ETL encarregada de netejar i estandarditzar les dades abans de la càrrega definitiva. A més, en aquesta part es creen les claus i és calculen dades derivades si escau.
Podem dividir les transformacions que calen fer a les dades en dos grups: errades i estandarditzacions:

- **Errades:** A les dades d'origen, a l'anàlisi preliminar ja es detectaren diverses errades:
  - Falta de títol original a alguna pel·lícula.
  - Noms de nominats diferents.
  - Títol del mateix film diferent.
  - Espais en les denominacions de les categories
  - Mateixa nominació (totes les dades iguals repetides)
- **Estandarditzacions:**
  - Noms dels països en diferents idiomes segons la font de les dades.
  - Any en diferent format.

Les transformacions necessàries per solucionar els errors puntuals s'han realitzat en primera instància en el fitxer Excel agrupat.

En el cas dels nominats, per exemple, s'han trobat 108 noms amb errades, podem veure els següents exemples:

Nominats erròns
Agnieszka Holland
Agniezka Holland
Andre Cayatte
André Cayatte
Anouk Aimee
Anouk Aimeé
Anouk Aimée
Burt Lancaster
Burt Lancester
Catalina Sandino Moreno
Catalina Sanino Moreno
Cécile De France
Cécile de France

Les transformacions més generals s'han realitzat mitjançant SQL.

Les claus primàries del model és crearan en el procés de càrrega a les taules.

Càrrega

En el procés de Càrrega, després de la transformació de les dades, seguirem les següents passes:

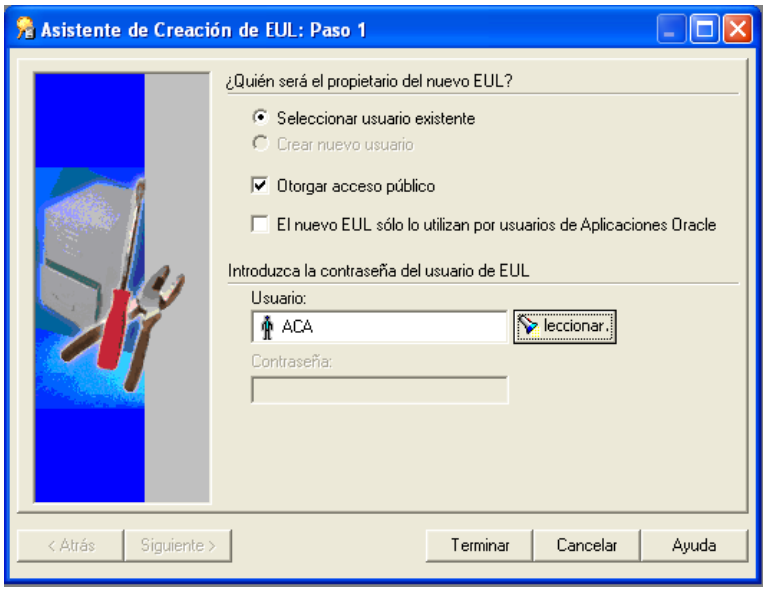
- Creació de les taules necessàries per al model multidimesional, taules de dimensions i de fets, amb les seves corresponents seqüències i triggers, segons el disseny tècnic desenvolupat.
- Càrrega de les dades de la taula temporal a les taules definitives.
- Eliminació de la taula temporal.

6.- Captures de pantalla i explicacions d'informes.

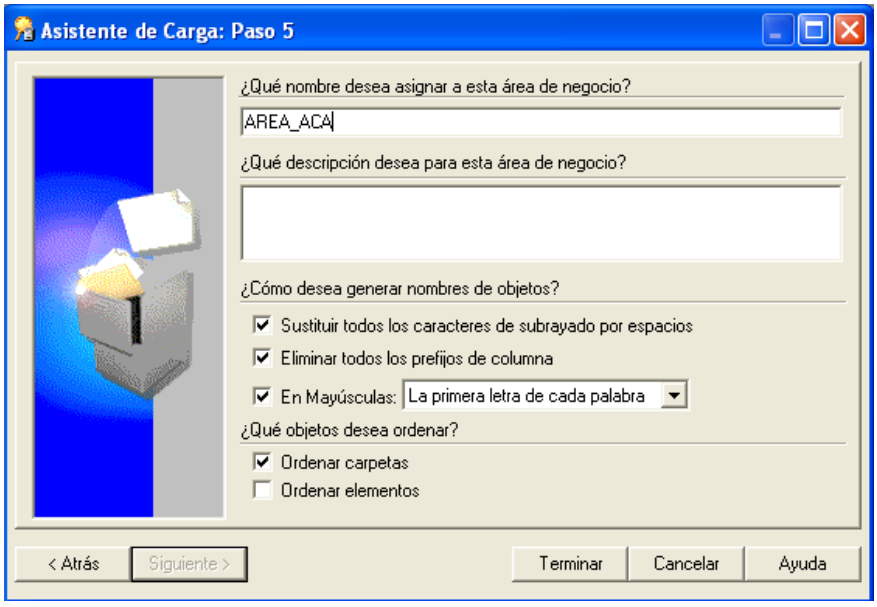
Ara es presenten les diferents captures de pantalla del producte que s'ha obtingut amb el projecte, així com les explicacions oportunes, si escau, de cada una d'elles.

6.1.- Creació del magatzem de dades

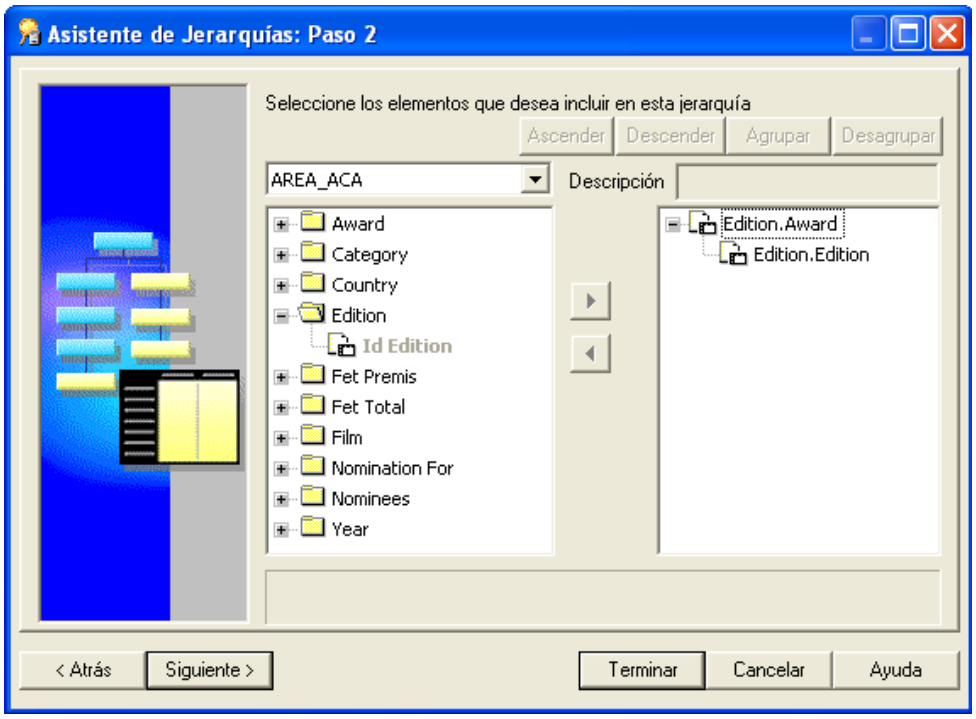
Creació de usuari EUL



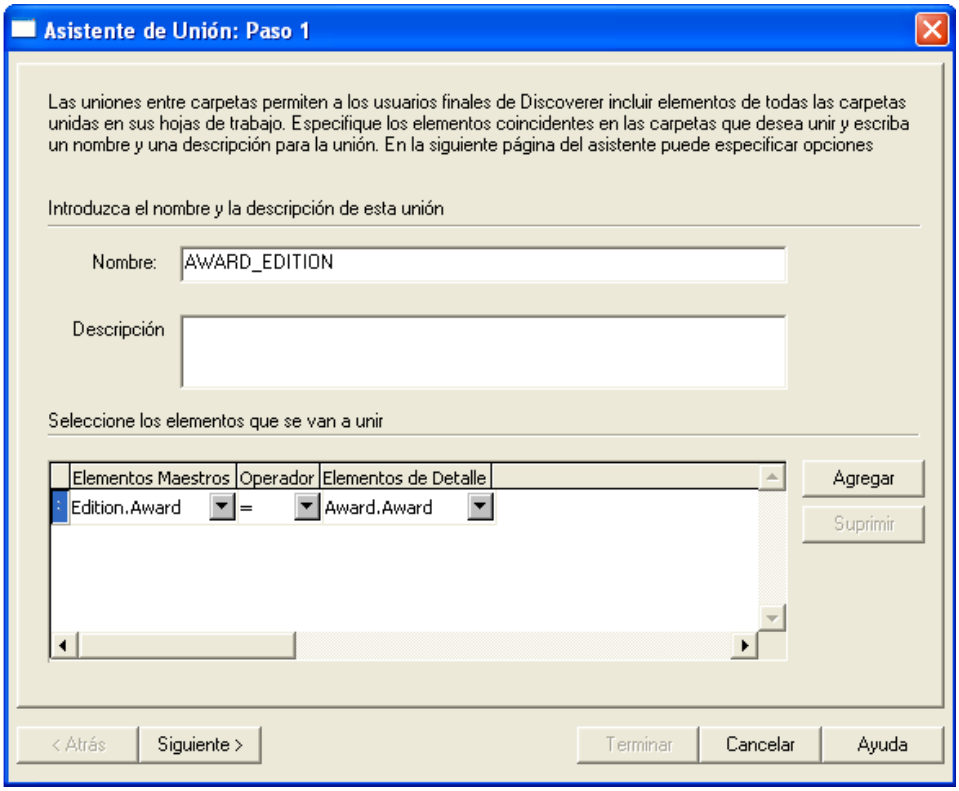
Creació de l'àrea de negoci.



Creació de noves jerarquies.



Creació de noves unions.



6.2.- Informes realitzats

Informe 1.

A l'informe 1 podem veure els premis aconseguits per cada any i festivals dels diferents actors i actrius, només s'han considerat premis directament relacionats amb la professió, no es tenen en compte premis honorífics.

UOC Universitat Oberta de Catalunya
www.uoc.edu

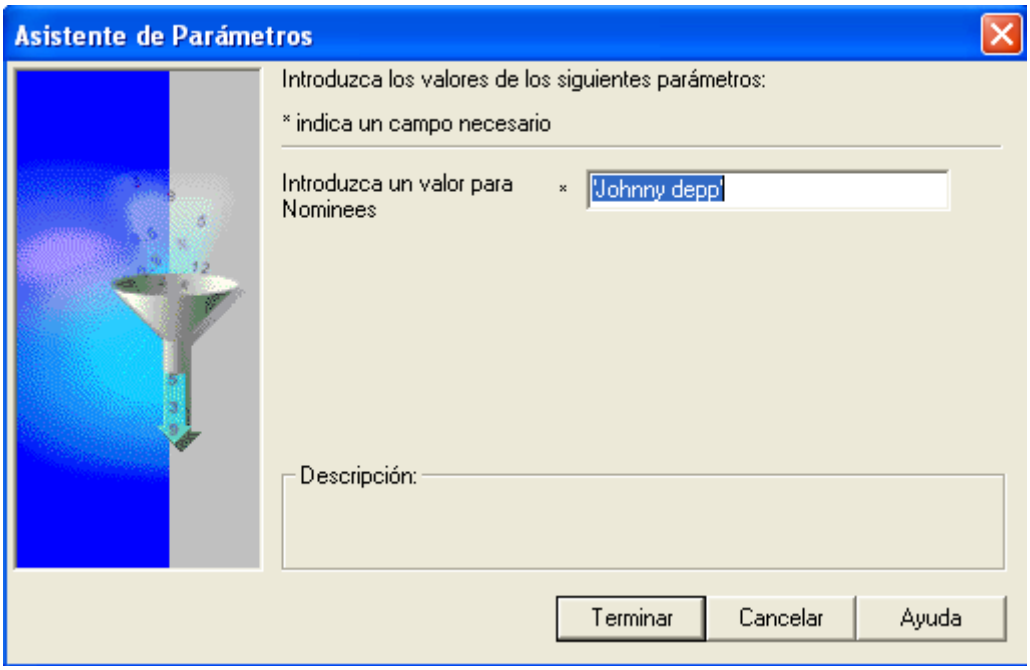
Premis aconseguits per actors/actrius per any i festival

Elementos de Página: Nomination For: <Todo>

	PREMIS					
	Academy Award	Berlinale	Cannes Festival	César	MTV Movie Awards	TOTAL
Jim Carrey					8	**8**
1996					2	**2**
1997					2	**2**
1995					1	**1**
1998					1	**1**
1999					1	**1**
2001					1	**1**
Isabelle Adjani		1	2	4		**7**
1981			2	1		**3**
1988		1		1		**2**
1983				1		**1**
1994				1		**1**
Denzel Washington	2	2			2	**6**
1989	1					**1**
1992		1				**1**
1993					1	**1**
1999		1				**1**
2001	1					**1**
2002					1	**1**
Jack Lemmon	2	1	2			**5**
1955	1					**1**
1973	1					**1**
1979			1			**1**

Informe 1.1

S'ha realitzat una versió de l'informe 1 on podem veure els premis aconseguits per cada any i festivals d'un actor o actriu, que prèviament passem com a paràmetre.



UOC Universitat Oberta de Catalunya www.uoc.edu

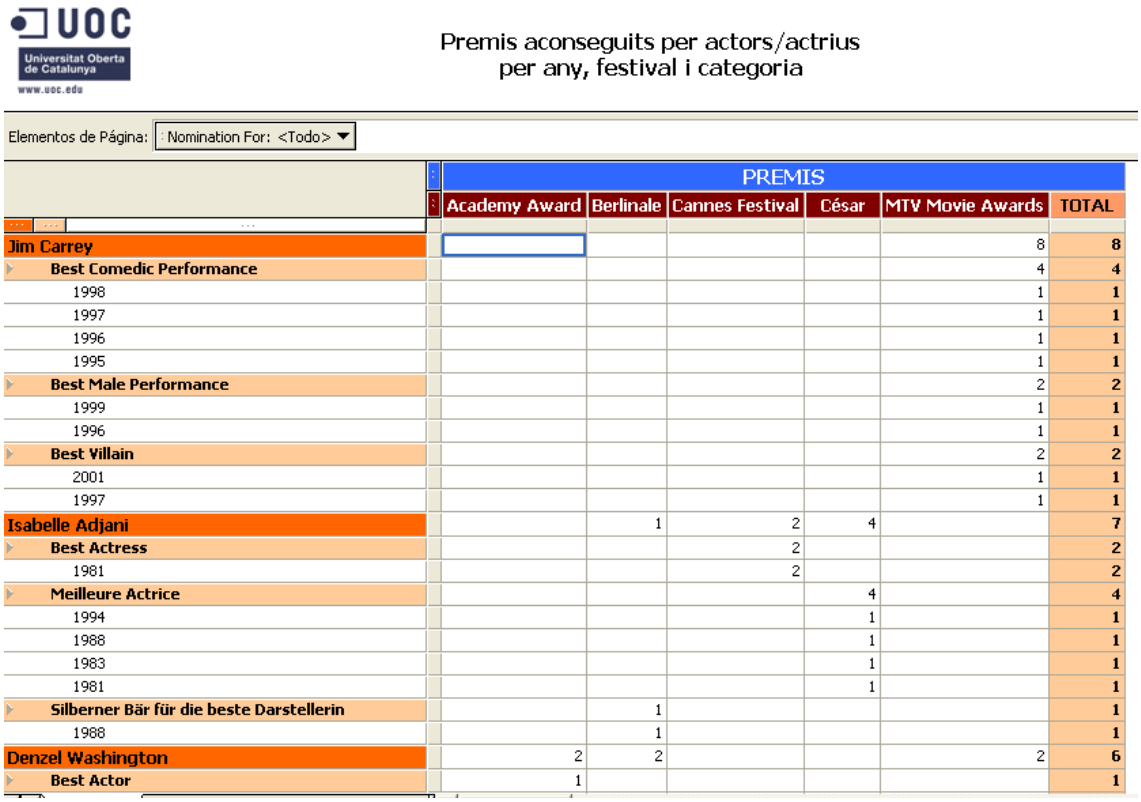
Premis aconseguits per Actor/Actriu: 'Johnny depp' per any i festival

Elementos de Página: Nomination For: ACTOR

	Academy Award		MTV Movie Awards		Total Nominacions	Total Premis
	Nominacions	Premis	Nominacions	Premis		
Johnny Depp	2	0	3	1	**5**	**1**
2004	1	0	2	1	**3**	**1**
2003	1	0			**1**	**0**
1994			1	0	**1**	**0**

Informe 2

A l'informe 2 podem veure els premis aconseguits per cada any, categoria i festivals dels diferents actors i actrius, només s'han considerat premis directament relacionats amb la professió, no es tenen en compte premis honorífics.



UOC Universitat Oberta de Catalunya www.uoc.edu

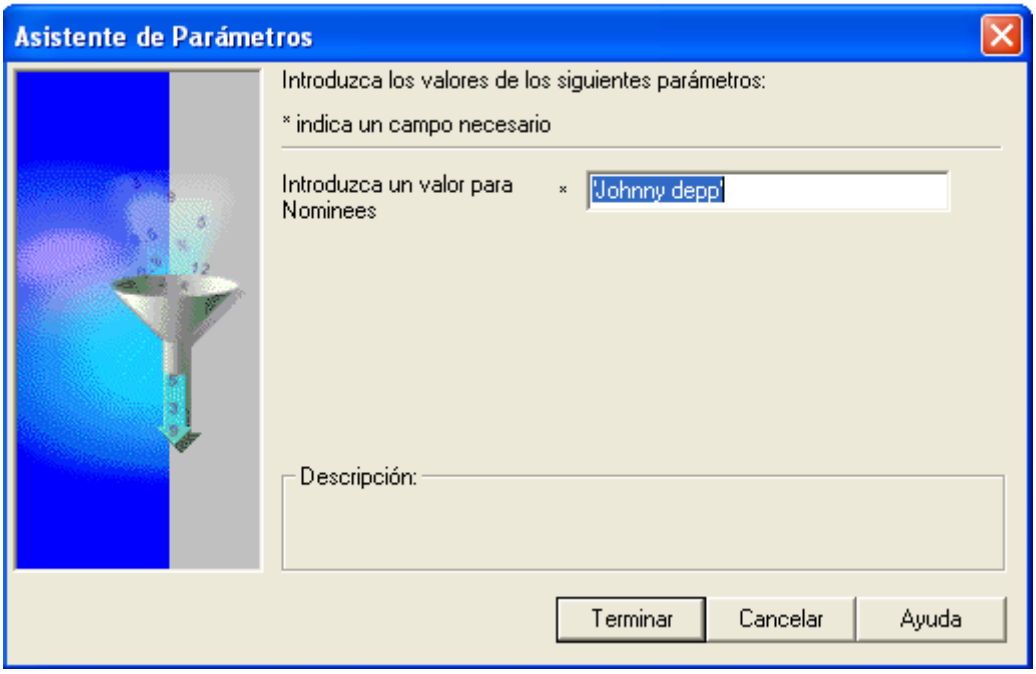
Premis aconseguits per actors/actrius per any, festival i categoria

Elementos de Página: Nomination For: <Todo>

	PREMIS					
	Academy Award	Berlinale	Cannes Festival	César	MTV Movie Awards	TOTAL
Jim Carrey					8	**8**
Best Comedic Performance					4	**4**
1998					1	**1**
1997					1	**1**
1996					1	**1**
1995					1	**1**
Best Male Performance					2	**2**
1999					1	**1**
1996					1	**1**
Best Villain					2	**2**
2001					1	**1**
1997					1	**1**
Isabelle Adjani		1	2	4		**7**
Best Actress			2			**2**
1981			2			**2**
Meilleure Actrice				4		**4**
1994				1		**1**
1988				1		**1**
1983				1		**1**
1981				1		**1**
Silberner Bär für die beste Darstellerin		1				**1**
1988		1				**1**
Denzel Washington	2	2			2	**6**
Best Actor	1					**1**

Informe 2.1

S'ha realitzat una versió de l'informe 2 on podem veure els premis aconseguits per cada any, categoria i festivals d'un actor o actriu, que prèviament passem com a paràmetre.



UOC Universitat Oberta de Catalunya www.uoc.edu

Premis aconseguits per Actor/Actriu: 'Johnny depp' per any, festival i categoria

Elementos de Página: Nomination For: ACTOR

	Academy Award		MTV Movie Awards		Total Nominacions	Total Premis
	Nominacions	Premis	Nominacions	Premis		
Johnny Depp	2	0	3	1	**5**	**1**
Best Actor	2	0			**2**	**0**
2004	1	0			**1**	**0**
2003	1	0			**1**	**0**
Best Comedic Performance			2	0	**2**	**0**
2004			1	0	**1**	**0**
1994			1	0	**1**	**0**
Best Male Performance			1	1	**1**	**1**
2004			1	1	**1**	**1**

Informe 3

A l'informe 3 podem veure els premis aconseguits per cada any, pel·lícula i festivals de cada director, només s'han considerat premis directament relacionats amb la professió, no es tenen en compte premis honorífics.

UOC Universitat Oberta de Catalunya
www.uoc.edu

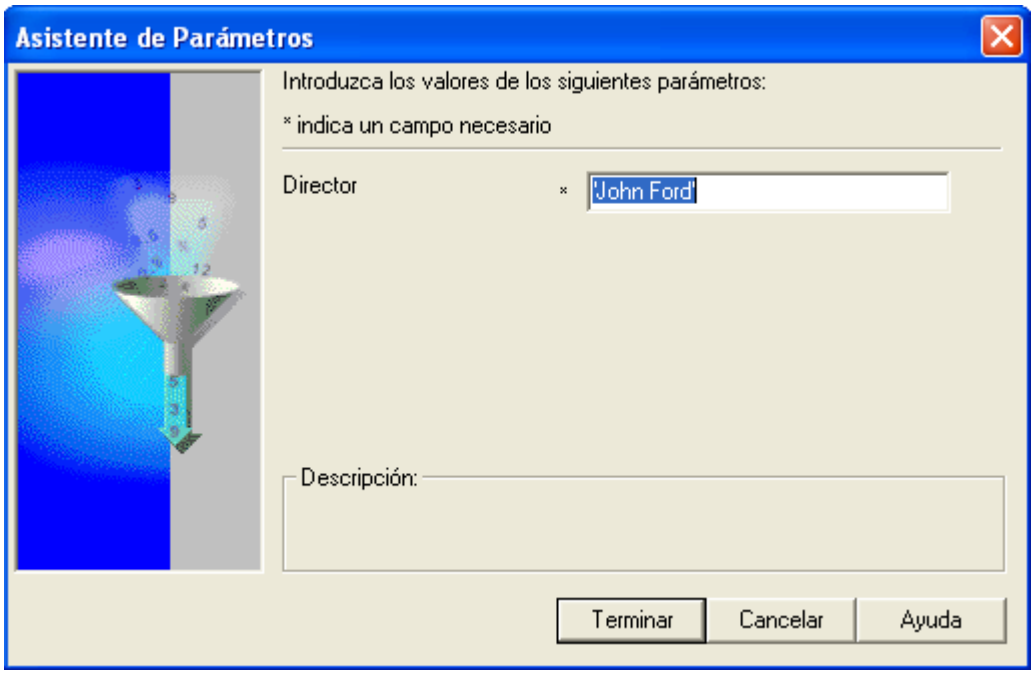
Premis aconseguits per directors
per any, festival i película

Elementos de Página: Nomination For: DIRECTOR

	PREMIS					
	Academy Award	Berlinale	Cannes Festival	César	MTV Movie Awards	TOTAL
John Ford	4					4
1935	1					1
The Informer	1					1
1940	1					1
The Grapes of Wrath	1					1
1941	1					1
How Green Was My Valley	1					1
1952	1					1
The Quiet Man	1					1
Bertrand Tavernier			1	2		3
1974				1		1
Que la fête commence				1		1
1984			1			1
Un dimanche à la campagne			1			1
1996				1		1
Capitaine Conan				1		1
Ettore Scola		1	1	1		3
1976			1			1
Brutti sporchi e cattivi			1			1
1983				1		1
Le Bal				1		1
1984		1				1
Le bal		1				1
Frank Capra	3					3

Informe 3.1

S'ha realitzat una versió de l'informe 3 on podem veure els premis aconseguits per un director que prèviament passem com a paràmetre.



UOC Universitat Oberta de Catalunya www.uoc.edu

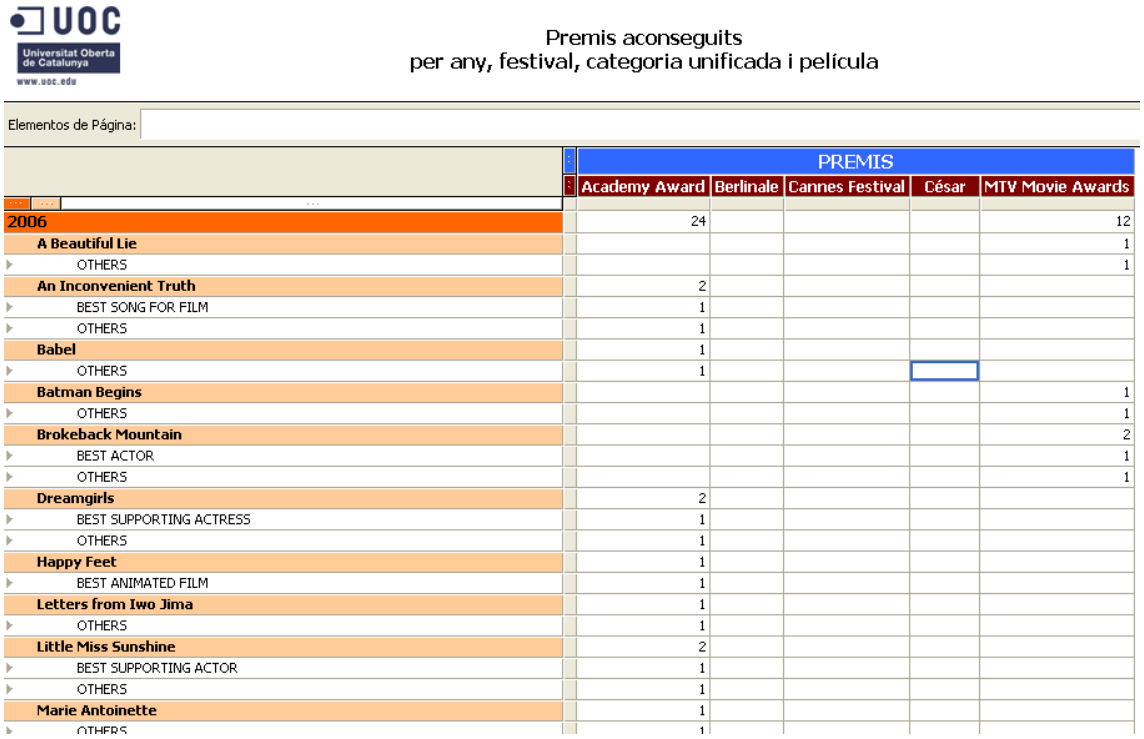
Premis aconseguits per Director: 'John Ford' per any, festival i película

Elementos de Página: Nomination For: DIRECTOR

	Academy Award		Total Nominacions	Total Premis
	Nominacions	Premis		
John Ford	5	4	**5**	**4**
1952	1	1	**1**	**1**
The Quiet Man	1	1	**1**	**1**
1941	1	1	**1**	**1**
How Green Was My Valley	1	1	**1**	**1**
1940	1	1	**1**	**1**
The Grapes of Wrath	1	1	**1**	**1**
1935	1	1	**1**	**1**
The Informer	1	1	**1**	**1**
1939	1	0	**1**	**0**
Stagecoach	1	0	**1**	**0**

Informe 4

A l'informe 4 podem veure els premis aconseguits per cada any, categoria i festivals de cada pel·lícula.



UOC Universitat Oberta de Catalunya www.uoc.edu

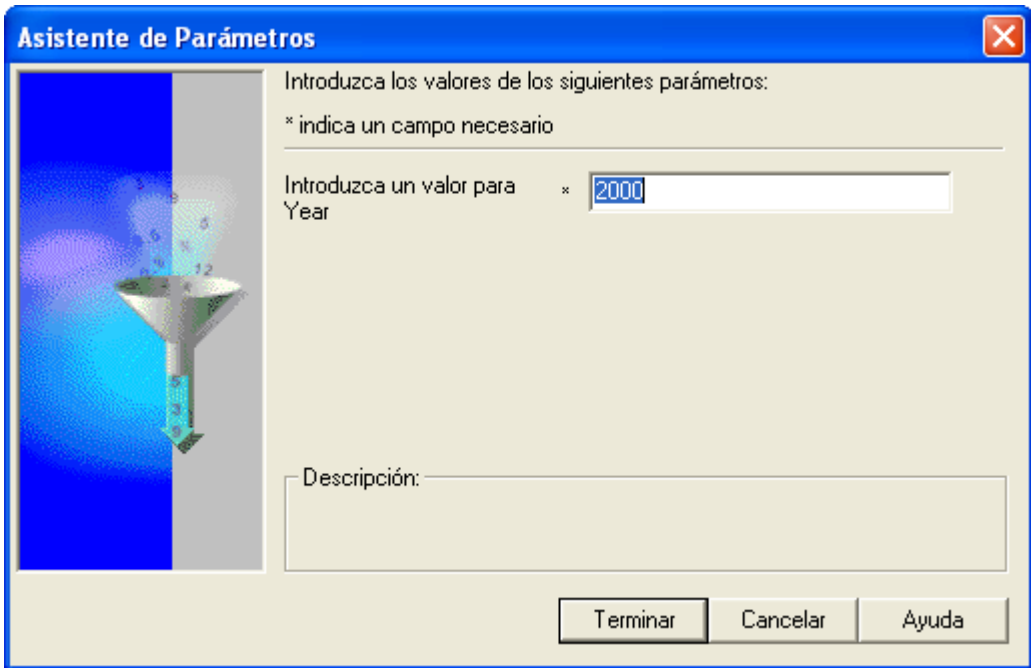
Premis aconseguits
per any, festival, categoria unificada i película

Elementos de Página:

	PREMIS				
	Academy Award	Berlinale	Cannes Festival	César	MTV Movie Awards
2006	24				12
A Beautiful Lie					1
OTHERS					1
An Inconvenient Truth	2				
BEST SONG FOR FILM	1				
OTHERS	1				
Babel	1				
OTHERS	1				
Batman Begins					1
OTHERS					1
Brokeback Mountain					2
BEST ACTOR					1
OTHERS					1
Dreamgirls	2				
BEST SUPPORTING ACTRESS	1				
OTHERS	1				
Happy Feet	1				
BEST ANIMATED FILM	1				
Letters from Iwo Jima	1				
OTHERS	1				
Little Miss Sunshine	2				
BEST SUPPORTING ACTOR	1				
OTHERS	1				
Marie Antoinette	1				
OTHERS	1				

Informe 5

A l'informe 5 podem veure les edicions, nombre d'actors i actrius nominats i pel·lícules de cada festival fins a l'any passat com a paràmetre.



UOC Universitat Oberta de Catalunya www.uoc.edu

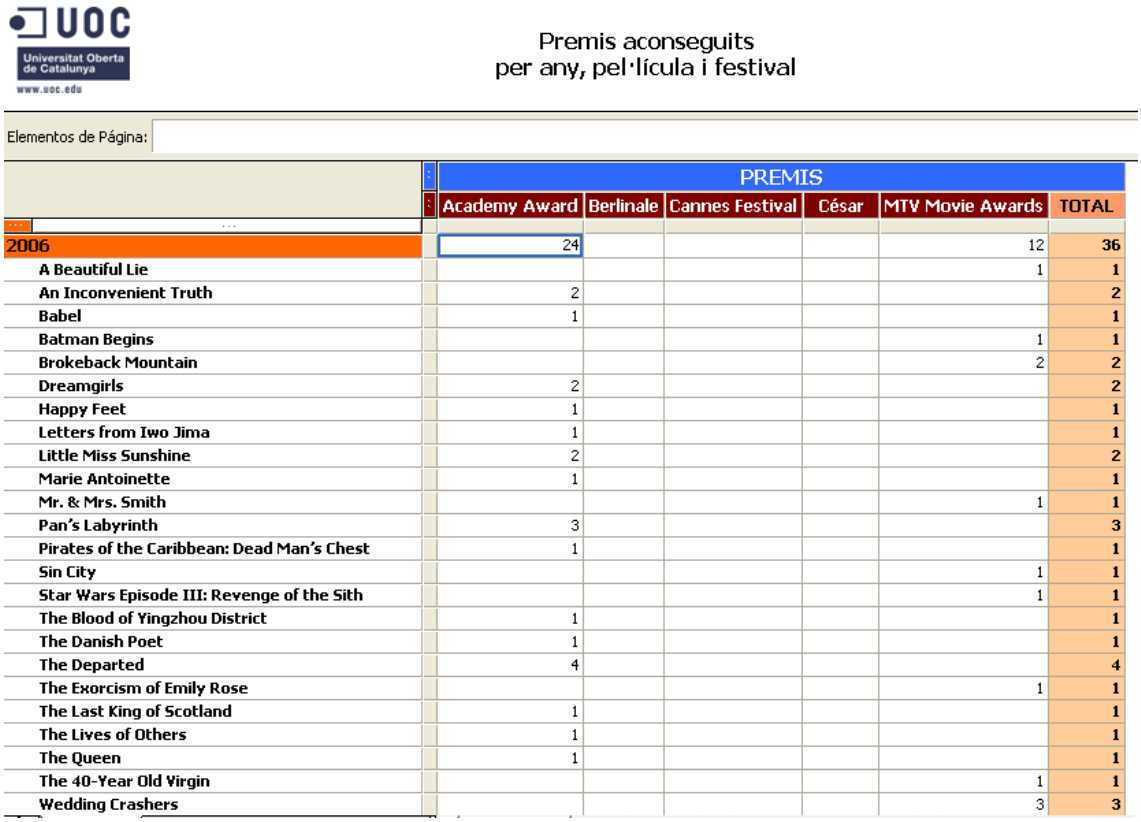
Edicions, nombre d'actors, actrius i pel·lícules nominades
Fins l'any: '2000'

Elementos de Página: Year: <Todo>

	Edicions	Actors	Actrius	Pel·lícules
Academy Award	73	686	689	917
Berlinale	43	44	45	88
Cannes Festival	49	54	61	112
César	33	338	339	387
MTV Movie Awards	9	164	106	178

Informe 6

A l'informe 6 podem veure els premis per any, pel·lícula i festival.



UOC Universitat Oberta de Catalunya www.uoc.edu

Premis aconseguits
per any, pel·lícula i festival

Elementos de Página:

	PREMIS					
	Academy Award	Berlinale	Cannes Festival	César	MTV Movie Awards	TOTAL
2006	24				12	36
A Beautiful Lie					1	1
An Inconvenient Truth	2					2
Babel	1					1
Batman Begins					1	1
Brokeback Mountain					2	2
Dreamgirls	2					2
Happy Feet	1					1
Letters from Iwo Jima	1					1
Little Miss Sunshine	2					2
Marie Antoinette	1					1
Mr. & Mrs. Smith					1	1
Pan's Labyrinth	3					3
Pirates of the Caribbean: Dead Man's Chest	1					1
Sin City					1	1
Star Wars Episode III: Revenge of the Sith					1	1
The Blood of Yingzhou District	1					1
The Danish Poet	1					1
The Departed	4					4
The Exorcism of Emily Rose					1	1
The Last King of Scotland	1					1
The Lives of Others	1					1
The Queen	1					1
The 40-Year Old Virgin					1	1
Wedding Crashers					3	3

Informe 7

Per avaluar l'èxit o el fracàs de les pel·lícules s'ha creat un sistema de puntuació:

(premis x 3) – (nominacions)

Així podem mesurar l'èxit no només pel nombre de premis, sinó per la relació en les nominacions i els premis aconseguits.

En l'informe podem veure el llistat segons l'orde de puntuació per a cada any.

UOC Universitat Oberta de Catalunya www.uoc.edu

Exit de les pel·licules
(premi * 3 - nominació)

Elementos de Página:

	Nominacions	Premis	Puntuacio
2006	174	36	-66,00
The Departed	5	4	7,00
An Inconvenient Truth	2	2	4,00
Brokeback Mountain	2	2	4,00
Wedding Crashers	5	3	4,00
Pan's Labyrinth	6	3	3,00
A Beautiful Lie	1	1	2,00
Happy Feet	1	1	2,00
Little Miss Sunshine	4	2	2,00
Marie Antoinette	1	1	2,00
The Blood of Yingzhou District	1	1	2,00
The Danish Poet	1	1	2,00
The Last King of Scotland	1	1	2,00
The Lives of Others	1	1	2,00
West Bank Story	1	1	2,00
Mr. & Mrs. Smith	2	1	1,00
The Exorcism of Emily Rose	2	1	1,00
Batman Begins	3	1	0,00
Sin City	3	1	0,00
Star Wars Episode III: Revenge of the Sith	3	1	0,00
After the Wedding	1	0	-1,00
Binta Y La Gran Idea	1	0	-1,00
Borat Cultural Learnings of America for Make Benefit Glorious Na	1	0	-1,00
Click	1	0	-1,00
Curse of the Golden Flower	1	0	-1,00
Deliver Us from Evil	1	0	-1,00

Informe 8

A l'informe 8 podem veure els premis aconseguits pels diferents països (amb noms unificats) per ordre de número de premis, sense tenir en compte EUA, amb les categories unificades i les pròpies.

UOC Universitat Oberta de Catalunya www.uoc.edu

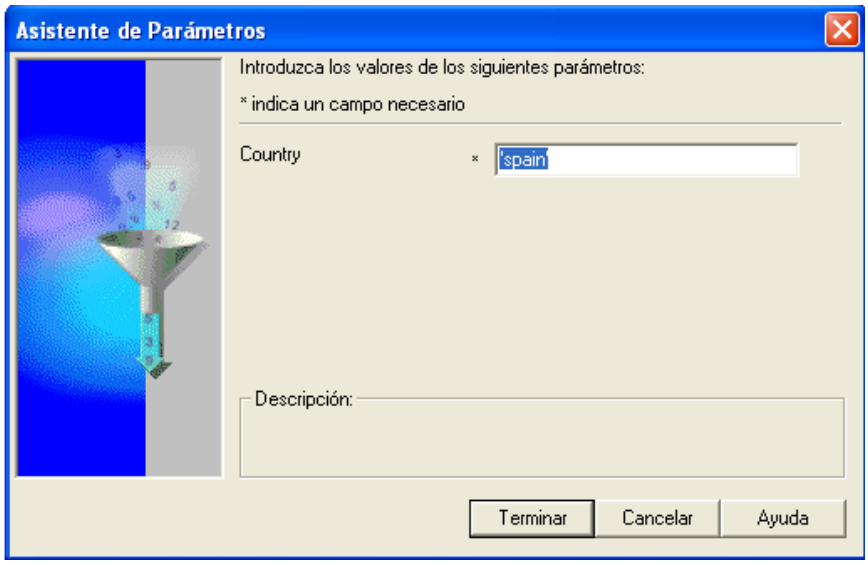
Premis aconseguits per països (sense EUA o desconeguts)

Elementos de Página: Award: <Todo>

	Premis aconseguits
France	699
BEST ACTOR	42
Best Actor	7
Meilleur Acteur	31
Silberner Bär für den besten Darsteller	4
BEST ACTRESS	42
Best Actress	7
Meilleure Actrice	31
Silberner Bär für die beste Darstellerin	4
BEST BREAKTHROUGH ACTOR	24
Meilleur Espoir masculin	24
BEST BREAKTHROUGH ACTRESS	24
Meilleur Espoir féminin	24
BEST DIRECTOR	47
Best Director	2
Direction	9
Golden Camera	1
Meilleur Réalisateur	31
Silberner Bär für die beste Regie	4
BEST FILM	41
Golden Palm	4
Goldener Bär für den besten Film	5
Meilleur Film	32
BEST SHORT FILM	19
Best Short Film	1
Best Short Film, Live Action	1

Informe 8.1

S'ha modificat l'informe anterior per a poder veure directament els premis aconseguits per un país concret, passat prèviament com a paràmetre.



UOC Universitat Oberta de Catalunya
www.uoc.edu

Premis aconseguits per Country: 'spain'

	Premis aconseguits				
	Academy Award	Berlinale	Cannes Festival	César	TOTAL
Spain	5	15	18	3	**41**
BEST ACTOR		1	3		**4**
Best Actor			3		**3**
Silberner Bär für den besten Darsteller		1			**1**
BEST ACTRESS		1			**1**
Silberner Bär für die beste Darstellerin		1			**1**
BEST DIRECTOR		3	1		**4**
Direction			1		**1**
Silberner Bär für die beste Regie		3			**3**
BEST FILM		6	1		**7**
Golden Palm			1		**1**
Goldener Bär für den besten Film		6			**6**
OTHERS	5	4	13	3	**25**
Artistic Contribution			1		**1**
Best Foreign Language Film	4				**4**
Best Writing, Original Screenplay	1				**1**
Fipresci Award			1		**1**
Fipresci Prize			1		**1**
Jury Prize			2		**2**
Meilleur Film de l'Union Européenne				1	**1**
Meilleur Film Etranger				2	**2**
OCIC Award			1		**1**
Prize for Good Humor			1		**1**
Screenplay			1		**1**
Silberner Bär für eine herausragende künstlerische Leistung		4			**4**
Special Homage			1		**1**
Special Jury Mention			1		**1**

7.- Conclusions

Després de la realització del projecte, puc considerar que el producte obtingut és prou bo donades les dades que ens han presentat, no sempre en la realització d'un magatzem de dades tenim les dades unificades, i aquest n'ha sigut un bon exemple, de fet ha sigut un bon exercici de transformació, i probablement aquesta part ha resultat la més costosa de tot el projecte.

La informació obtinguda mitjançant els informes, considero que, s'ajusta als requeriments, si més no, amb algunes variacions que poden ser considerades ampliacions dels informes sol·licitats.

L'objecte d'anàlisi ha suposat un repte però no l'habitual que es sol donar en la utilització dels DW, normalment molt més econòmics o numèrics que aquest cas.

8.- Glossari

Data Warehouse: magatzem de dades, col·lecció de dades orientada a un domini, integrat, no volàtil, i variant en el temps que permet la presa de decisions a les empreses o organitzacions.

OLAP: processament analític en línea, Solució utilitzada en la Intel·ligència de Negocis per a realitzar consultes a estructures multidemsionals.

ETL: Extract, Tranform and Load. Procés d'extracció, transformació i càrrega de les dades.

Mesura: Atribut numèric.

Fet: Objecte d'anàlisi.

Model d'estrella: Esquema clàssic del model multidimesional.

Dimensió: punt de vista de l'anàlisi del fet.

9.- Bibliografia

Relació de la bibliografia que s'ha utilitzat durant el desenvolupament del projecte:

- Pla docent del TFC (documentació UOC)

- Pla docent del TFC-DW (documentació UOC)

- Treball Final de Carrera (documentació UOC)

- Competències comunicatives (documentació UOC)