



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

*Miquel Luis Peñacerrada*

Enginyeria Tècnica Informàtica de Gestió

*Consultor: José Ángel Martín Carballo*

*13/06/2011*

## 2 Agraïments

Els meu principal agraïment es per a la meva família per els ànims i suport que m'han transmès en els moments difícils, sense ells no hagués pogut arribar fins aquí.

## 3 Resum

Els Sistemes de Magatzem de dades representen un tipus de bases de dades on estan integrades les dades de diferents fonts de dades d'una organització. El Magatzem de dades esta orientat a donar suport a la presa de decisions dins una organització per lo que esta orientat a un determinat àmbit depenent de la organització.

Com es veurà en aquest Projecte Final de carrera el principal objectiu d'un Sistema de Magatzem de dades es treure el màxim rendiment de la informació emmagatzemada, aquest rendiment es realitza amb l'extracció de dades per un anàlisi que ajudi a prendre decisions dins l'entorn de la organització.

En aquest treball es realitzarà l'estudi dels Magatzems de dades amb un cas pràctic en l'àmbit de la industria cinematogràfica concretament estarà centrat en els premis atorgats per a diferents festivals. La informació dels premis s'ha lliurat en format Excel y representa els següents festival:

- Premis de l'Acadèmia.
- Festival Internacional de Cinema de Berlín.
- Festival Internacional de Cinema de Canes.
- Premis Cèsar.
- MTV Movie Awards.

El cos de la memòria s'estructurara en tres seccions ben diferenciats, Introducció, Anàlisi i Disseny.

Paraules Clau: Magatzem de Dades, Data Warehouse, Oracle, Discoverer, SQL, ACA,ETL.

## 4 Índex

|  |           |
|--|-----------|
| <b>2 AGRAIMENTS .....</b>                                  | <b>2</b>  |
| <b>3 RESUM .....</b>                                       | <b>2</b>  |
| <b>4 INDEX.....</b>  | <b>3</b>  |
| <b>5.1 CAPÍTOL 1: INTRODUCCIO .....</b>                    | <b>4</b>  |
| 5.1.1 JUSTIFICACIO .....                                   | 6         |
| 5.1.1 JUSTIFICACIO .....                                   | 6         |
| 5.1.2 OBJECTIUS .....                                      | 7         |
| <i>Objectius de l'assignatura.....</i>                     | <i>7</i>  |
| <i>Objectius del Projecte Final de Carrera .....</i>       | <i>7</i>  |
| 5.1.3 ENFOCAMENT I METODE .....                            | 8         |
| 5.1.4 PLANIFICACIO DEL PROJECTE .....                      | 9         |
| 5.1.5 PRODUCTES OBTINGUTS .....                            | 10        |
| 5.1.6 CAPÍTOLS .....                                       | 10        |
| <b>5.2 CAPÍTOL 2 ANALISI .....</b>                         | <b>11</b> |
| 5.2.1 REQUERIMENTS.....                                    | 11        |
| 5.2.2 FONTS DE DADES .....                                 | 12        |
| 5.2.3 MODEL CONCEPTUAL.....                                | 13        |
| <i>Triar el Fet.....</i>                                   | <i>13</i> |
| <i>Grànul escaient .....</i>                               | <i>13</i> |
| <i>Dimensions.....</i>                                     | <i>14</i> |
| <i>Atributs .....</i>                                      | <i>14</i> |
| <i>Distingir Descriptors i Jerarquies d agregació.....</i> | <i>15</i> |
| <i>Mesures.....</i>  | <i>15</i> |
| <i>Definir Celles .....</i>                                | <i>15</i> |
| <i>Restriccions d Integritat .....</i>                     | <i>15</i> |
| <i>Viabilitat.....</i>                                     | <i>15</i> |
| <i>Diagrama Model conceptual.....</i>                      | <i>16</i> |
| <b>5.3 CAPÍTOL 3: DISSENY .....</b>                        | <b>18</b> |
| 5.3.1 ARQUITECTURA DE SOFTWARE .....                       | 19        |
| 5.3.2 ARQUITECTURA DE HARDWARE .....                       | 20        |
| 5.3.3 DISSENY DE LA BASE DE DADES .....                    | 21        |
| 5.3.4 PROCES DE CARREGA DE INFORMACIO .....                | 29        |
| 5.3.5 DISSENY DELS INFORMES.....                           | 39        |
| <b>5.4 CONCLUSIONS.....</b>                                | <b>49</b> |
| <b>5.6 LINIES D EVOLUCIO .....</b>                         | <b>49</b> |
| <b>6 GLOSSARI .....</b>                                    | <b>50</b> |
| <b>7 BIBLIOGRAFIA .....</b>                                | <b>51</b> |

## 5.1 Capítol 1: Introducció

Els Sistemes de Magatzem de dades representen un tipus de bases de dades on estan integrades les dades de tots els sistemes d'una organització, la dades son de tipus no volàtil i variables en el temps. A continuació es detallaran les principals característiques d'un Magatzem de dades.

- Orientant al Tema
- Degut a que normalment no es poden conèixer els requeriments del usuaris en el moment en que es construeix un Magatzem de dades la informació no s'estructura segons la seva funcionalitat sinó dividida per temes d'interès.
- Integració de dades
- Degut a que un Magatzem de dades es la unió de la informació de diversos sistemes operacionals, les dades d'aquests s'han d'integrar per tal de que la informació aparenti que prové d'una única font. Un component que ajuda a la integració son els Metadatas.
- Informació Històrica i no volàtil

Aquestes dues característiques fan referència al temps, la historicitat indica quan ocorre un determinat esdeveniment al temps i la no volatilitat indica quan tenim constància de que ha ocorregut aquest fet a la Base de dades.

La historicitat es important per analitzar quan han ocorregut els fets, es molt important que qualsevol dada del magatzem de dades estigui acompanyada del seu període de validesa.

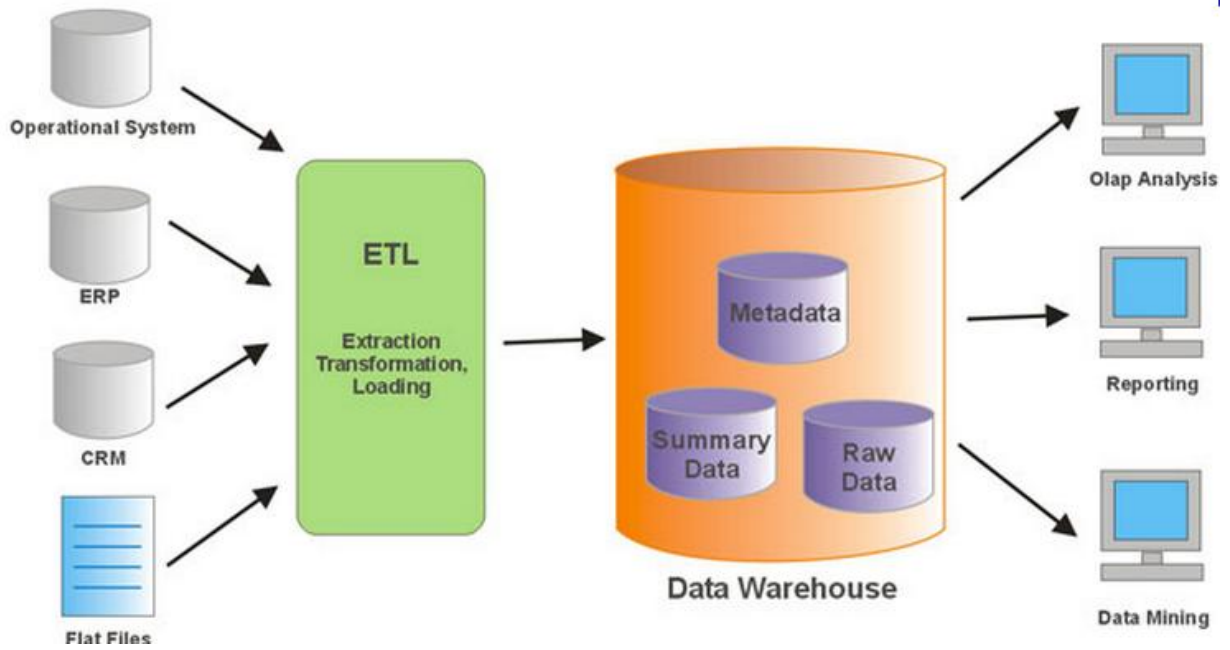
La no volatilitat indica en quin moment ens adona'm d'un fet, aquesta característica implica que les operacions de modificar i esborrar no siguin necessaris sinó que s'insereixen les correccions i la data en que s'han realitzat.

Els magatzems de dades son també coneguts con sistemes analítics representant per les sigles OLAP y es emprat per usuaris que no tenen perquè ser informàtics o tenir coneixements de programació.

Un procés important dins l'entorn dels magatzems de dades son els processos ETL, aquets processos son els encarregats de la extracció , transformació i carrega de informació dins el Magatzem de dades.

Al següent diagrama es pot veure d'esquerra a dreta com les dades provenen dels diferents sistemes d'una organització de treball. Aquestes dades per poder introduir-se i alimentar correctament el Magatzem de dades han de passar per el proses ETL.

A continuació es mostra un diagrama de flux sobre un magatzem de dades típic.



### 5.1.1 Justificació

Actualment la acadèmia de cinema Andorra té diversos problemes per a explotar i donar utilitat a la informació de la que disposa. Per aquest motiu se'ns ha encomanat la implementació d'un magatzem de dades per a assolir els següents beneficis:

- Ajudar a la presa de decisions

Es tracta del principal objectiu d'un magatzem de dades, per poder complir aquest objectiu és molt important que el magatzem de dades contengui la informació necessària sobre el negoci o àmbit en el que s'aplica.

- Segmentar les dades de negoci

La segmentació de dades ajuda a que la consulta d'informació sobre la base de dades es realitzi de forma més àgil i ràpida, i es puguin fer anàlisis fàcilment. La segmentació també pot ajudar a preveure on hi pot haver futurs problemes que no s'han detectat en una primera instància.

- Gestionar el coneixement de l'empresa

Una visió diferent de un magatzem de dades és pensar que es tracta d'un contenidor de dades de l'empresa.

- Depurar les dades

Degut a que durant la implementació d'un magatzem de dades s'ha de realitzar una primera migració i unificació de dades dels diferents sistemes, és una bona oportunitat per a realitzar una depuració de dades, tant per a eliminar dades redundants com per a unificar criteris.

## 5.1.2 Objectius

### Objectius de l'assignatura

El principal objectiu de l'assignatura Projecte Final de Carrera cursada a la Universitat Oberta de Catalunya es treballar les aptituds adquirides a diverses assignatures dins el Pla d' Estudis de l' Enginyeria Tècnica d'informàtica de Gestió. Aquest objectiu general es pot concretar en els següents:

- Analitzar un problema complex de tipus pràctic.
- Planificar i estructurar el desenvolupament del projecte.
- Treballar aspectes formals del desenvolupament de projectes.
- Sintetitzar una solució viable i realista al problema.
- Elaborar una memòria de projecte.
- Elabora una presentació del desenvolupament i resultats finals del projecte.

### Objectius del Projecte Final de Carrera

Els objectius d'aquest projecte Final de Carrera estan basats en la temàtica en la que esta basada, els magatzems de dades. A continuació es detallen els principals objectius que es volen assolir relacionats amb el tema:

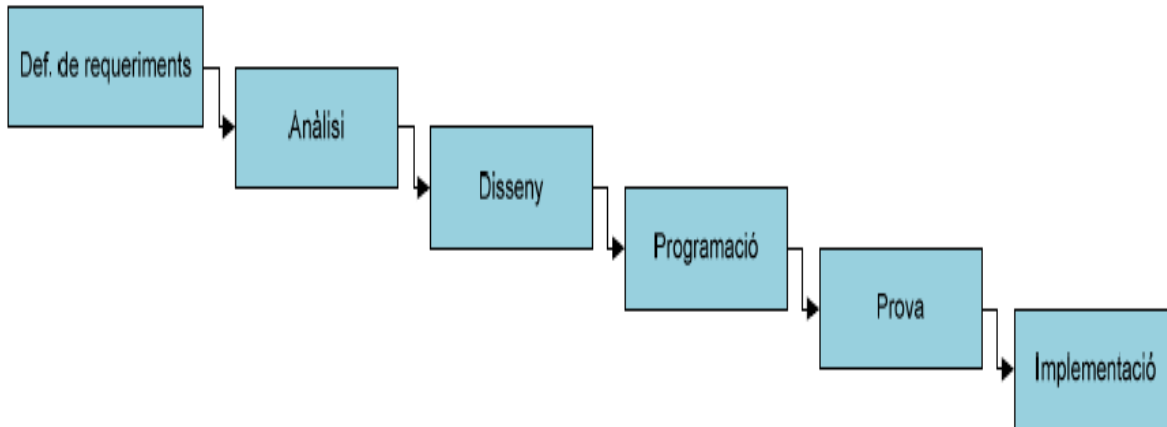
- Conèixer el propòsit dels Magatzems de dades.
- Conèixer l'arquitectura general d'un Magatzem de dades.
- Dissenyar un Magatzem de dades.
- Aprofundir sobre les tècniques de tractament de dades.
- Aprofundir sobre els llenguatges de bases de dades.
- Conèixer les eines d'explotació de dades i les eines OLAP.

Es podrà treballar amb bases de dades Oracle i utilitzar el llenguatge SQL tant per a la creació del magatzem de dades com per al tractament i extracció de dades.

S'haurà de desenvolupar un cas pràctic sobre Magatzems de dades dins l'àmbit de l'empresa cinematogràfica.

### 5.1.3 Enfocament i mètode

L'enfocament i el mètode seguit per a la realització d'aquest projecte es correspon amb el cicle de vida clàssic d'un projecte, també anomenat en cascada.



Els requeriments del projecte son extrets del enunciat del projecte, per lo que en aquest projecte no es s'inclou la fase de definició de requeriments pròpiament dita.

A la fase de anàlisi de requeriments es defineix detalladament les necessitats d'informació que haurà de resoldre el magatzem de dades a implementar. Durant la etapa de disseny s'especifica com es dona resposta a les necessitats definides a l'anterior etapa d'anàlisi.

A la següent etapa, corresponent a la Programació es realitzarà la codificació dels processos necessaris per a la incorporació de dades al magatzem de dades (migració) així com les tasques de creació de base de dades e informes.

La fase de prova no es realitzarà directament amb el client, sinó que es farà en base al feedback rebut per part del consultor durant el projecte.

En la última fase s'entregarà la implementació de tot el projecte juntament amb la memòria i la presentació virtual.



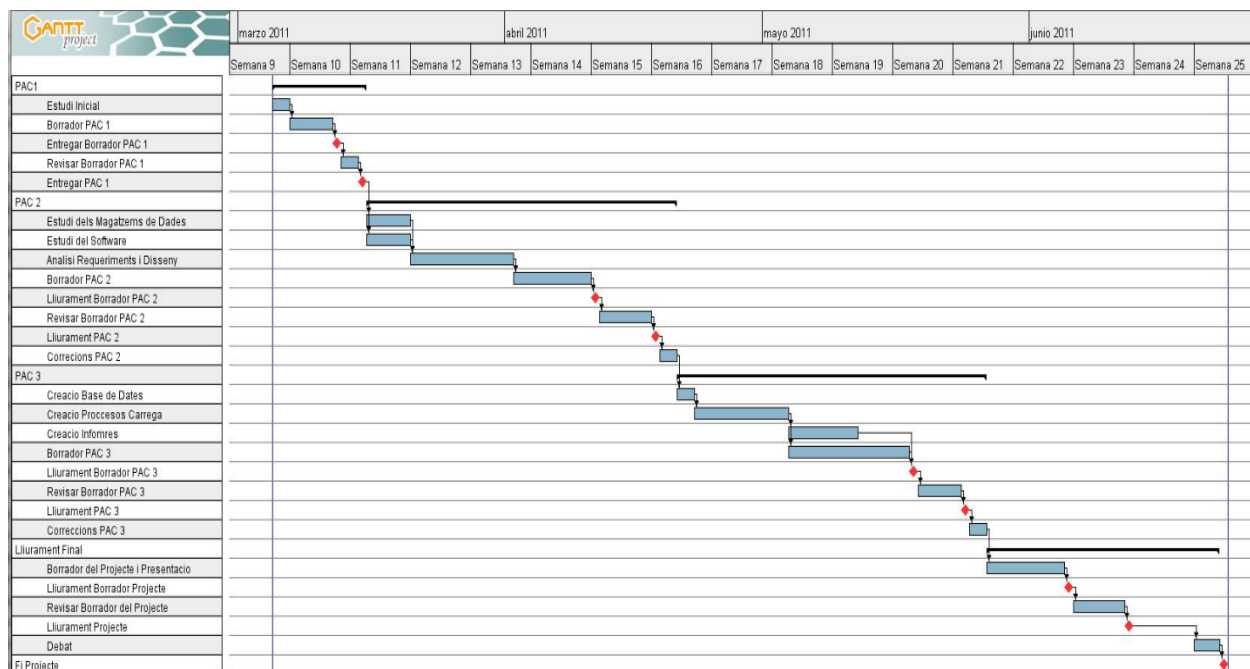
### 5.1.4 Planificació del Projecte

La planificació del projecte recull les diferents tasques de les que es compona i la planificació estimada per a cada una d'aquestes tasques, la planificació s'ha basat amb les diferents entregues de les Pacs establertes per el pla docent d'aquesta assignatura així com l'entrega de la memora i debat final.

Taula amb les dates importants durant el projecte:

| Fita                           | Data          |
|--------------------------------|---------------|
| Entregar Borrador PAC 1        | 13/03/2011    |
| Entregar PAC 1                 | 16/03/2011    |
| Entregar Borrador PAC 2        | 12/04/2011    |
| Entregar PAC 2                 | 19/04/2011    |
| Entregar Borrador PAC 3        | 19/05/2011    |
| Entregar PAC 3                 | 25/05/2011    |
| Entregar Borrador Memòria      | 06/06/2011    |
| Entregar Memòria i Presentació | 13/06/2011    |
| Debat                          | 20-23/06/2011 |

A Continuació es mostra el diagrama de Gantt amb la planificació detallada



### 5.1.5 Productes Obtinguts

A continuació es detallen els productes obtinguts durant aquest projecte:

- Document de Planificació corresponent a la PAC 1.
- Document de Anàlisi de Requeriments i Disseny corresponent a la PAC 2.
- Implementació provisional del producte corresponent a la PAC 3.
- Implementació definitiva del producte incloent la següent informació:
  - Fitxers de mapeig de dades mestres.
  - Script de Creació del Magatzem de Dades.
  - Script de Carrega de Informació al Magatzem de Dades.
  - Informes requerits per el client.
- Memòria del projecte.
  - Presentació virtual del projecte.

### 5.1.6 Capítols

Els principals capítols de aquesta memòria son el capítol 2 dedicat a l'anàlisi del cas pràctic i el capítol 3 dedicat al disseny .

Al segon capítol corresponent a l'anàlisi s'explicarà el model conceptual emprat per a resoldre el problema plantejat a n'aquest projecte, mentre que al tercer capítol es donarà una solució al problema plantetjat indicant l 'arquitectura de Software i Hardware emprada, també es mostrarà la base de dades implementada així com els informes creats per donar una solució a les necessitats de l 'ACA

## 5.2 Capítol 2 Anàlisi

### 5.2.1 Requeriments

A continuació es detallen els requeriments exposats per l'Acadèmia de cinema Andorrà:

- El magatzem de dades ha d'incorporar la informació dels diferents Excel i permetre creuar aquesta informació.
- Les dades dels diferents Excels s'han d'unificar, a continuació es detallen les dades que son necessaris unificar:
  - Nom dels Actors i Actrius
  - Noms dels Directors
  - Nom de les Categories dels Premis
- L'ACA vol mantenir la nomenclatura original, per lo que s'ha de mantenir dins el magatzem de dades.
- S'ha de proporcionar una definició de pel·lícula exitosa i pel·lícula fracassada.
- Indicar la relació entre guanyar una premi a diversos festivals i indicar els criteris d'avaluació.
- Definició dels Informes:
  - Informe de Premis aconseguits per actor/actriu per any i festival. A l' informe s'ha d'indicar el total de premis per l actor/actriu al festival i a tots els festivals.
  - Informe Premis aconseguits per actor/actriu per any, festival i categoria.  
Al contingut d'aquest informe s'indicarà el total de premis i es mostrarà la categoria original i la categoria amb el nom comú.
  - Informe Premis aconseguits per un director per any, festival i pel·lícula.  
A l' informe s'ha d'indicar el total de premis per el director al festival i a tots els festivals
  - Informe Premis aconseguits per una pel·lícula per any, festival i categoria.  
Al contingut d'aquest informe es mostrarà la categoria original i la categoria amb el nom comú.
  - Informe que permetrà sobre una determinada data (edició) indicar la següent informació des de la primera edició del festival fins a la data seleccionada:
    - Total edicions de festival.
    - Total actors nominats.
    - Total actrius nominades.
    - Total pel·lícules nominades

- Informe global sobre els premis per cada pel·lícula i festival, indicant el total de premis guanyats per pel·lícula i el subtotal per any.
- Informe del País (sense tenir en compte els Estats Units) amb mes premis que ha guanyat per any, la informació a llistar per aquest país es la següent, l'informe ha de mostrar els premis per categoria i un apartat amb els premis totals.

### 5.2.2 Fonts de dades

Les fonts de dades del magatzem de dades són els Excel·lents proporcionats per l'ACA.

Sobre les dades d'aquests Excel·lents s'haurà de realitzar una unificació de dades. A continuació s'indica la principal informació dels Excel·lents:

- Any d'edició del Festival.
- País d'origen de la pel·lícula.
- Títol original de la pel·lícula.
- Títol traduït a l'anglès.
- Nom del Festival
- Categoria del premi.
- Guanyador del premi a la categoria corresponent.
- Nominats al premi a la categoria corresponent.

Per saber si un Actor o Director és guanyador o tan sols ha quedat com a nominat utilitzarem una columna on té un valor lògic per indicar si s'ha premiat o no.

A continuació detalla'm les fonts de dades mestres que s'han de registrar:

- Festivals
- Categories  
Al Magatzem de dades s'assignarà un identificador per cada categoria, per a les descripcions s'utilitzarà una descripció unificada (per totes els festivals) i una descripció per cada un dels Festivals on la categoria tinguí sentit.
- Pel·lícules  
El magatzem de dades assignarà un identificador a cada pel·lícula.  
Per unificar les pel·lícules s'emprarà per a cada un dels registres de pel·lícules el títol original.
- Actors  
Per identificar unívocament els Actors s'utilitzarà la columna Nominats de cada un dels Excel·lents proporcionats i es realitzarà una unificació d'aquestes dades, per als actors el magatzem de dades assignarà un identificador únic per cada registre.  
Com que a la columna nominats podem tenir actors i directors, identificarem que es tracta d'un actor per el tipus de categoria del premi.
- Directors

Es realitzarà la mateixa lògica que els Actors.

- Països  
Els països es troben associats a cada una de les pel·lícules en alguna ocasió una pel·lícula pot tenir mes d'un país.  
Cada Festival te la seva nomenclatura de països (amb diferent idioma) es realitzarà un agrupament i es guardarà el nom en angles. El magatzem de dades assignarà un identificador únic per a cada País que serà el que s'associarà a les pel·lícules.

### 5.2.3 Model Conceptual

El magatzem de dades es dissenyarà segons el model conceptual, aquest contindrà una sèrie de cubs de dades per explotar la informació àgilment. A continuació es detallen els passos seguits per a realitzar el disseny conceptual.

#### Triar el Fet

Segons el disseny del model conceptual s'ha d'escollir l'objecte d'anàlisi. Podem anomenar el fet com un conjunt d'esdeveniments amb dades associades. Normalment son els processos de negoci, on un processos de negoci es pot definir com els processos operacionals de l'organització.

Segons els requeriments establerts al nostre cas podem dir que el nostre fets estan relacionats amb les el tipus de premi que s'atorga a un Festival:

Fet de que un Actor guanya un Premi a un Any determinat a un Festival per una Categoria amb una determinada Pel·lícula.

Fet de que una Director guanya un Premi a un Any determinat a un Festival amb una determinada Pel·lícula.

Fet de que una Pel·lícula guanya un Premi a un Any determinat a un Festival.

També hem de tenir en compte els fets d'estar nominats:

Fet de que un Actor esta nominat a un Festival a un Any determinat.

Fet de que una Pel·lícula esta nominada a un Festival a un Any determinat.

#### Grànul escaient

El grànul escaient tracta la ultima part de la informació que volem analitzar, es tractaria de la cella mes petita que volem disposar. En el nostre cas els nostres grànuls estarien formants per la següent informació:

- Premis aconseguits per un Actor o Actriu a una determinada categoria i festival y any.
- Premis aconseguits per una Pel·lícula per any, festival i categoria.
- Premis aconseguits per un Director per any, festival i categoria.
- Actors Nominats per Categoria i Festival i Any.
- Pel·lícules Nominades per Categoria i Festival i Any.
- Numero de Nominacions d'un determinat Festival.

## Dimensions

Amb l'ajuda dels requeriments establerts anteriorment, a continuació s'identificaran les dimensions que complementa les dades de les taules de fets:

- Festival
- Actor/Actriu
- Director
- Pel·lícula
- Temps
- Categoria

## Atributs

A continuació es detallen els atributs necessaris identificats en cada una de les dimensions nombrades a l'apartat anterior:

- Festival  
ID Festival, Nom Festival.
- Actor/Actriu  
ID Actor/Actriu, Nom Actriu.
- Director  
ID Director, Nom Director.
- Pel·lícula  
ID Pel·lícula, Nom Pel·lícula, País.
- Temps  
Any
- Categoria  
ID de la categoria, Nom de la Categoria unificada, Nom de la Categoria al Festival Acadèmia, Nom de la Categoria al Festival Berlinale, Nom de la Categoria al Festival Cannes, Nom de la Categoria al Festival Cesar, Nom de la Categoria al Festival MTV,

### Distingir Descriptors i Jerarquies d agregació

D'entre els atributs que hi ha en una Dimensió, n'hem de distingir dos tipus els que utilitzarem per a agrupar i els que serviran simplement per a seleccionar. En el moment de realització del projecte es detecta que s'ha de definir una jerarquia d'agregació entre País i Pel·lícula.

### Mesures

A continuació es detallen les mesures que seran mes útils per a la implementació del magatzem de dades:

Per el Fet de guanyar un Premi:

- Premis aconseguits per actor/actriu per any i festival.
- Premis aconseguits per actor/actriu per any, festival i categoria.
- Premis aconseguits per un director per any, festival i pel·lícula.
- Premis aconseguits per una pel·lícula per any, festival i categoria.
- Premis per cada pel·lícula i festival.
- Nominacions actors/actrius per Festival y any.
- Nominacions Pel·lícula per Festival y any
- Edicions per Festival

### Definir Celles

Cal definir les Cel·les que contenen Mesures interessants per analitzar les dades. Les Cel·les que contenen les mesures anteriorment anomenades son les corresponents a guanyar un premi y estar nominat.

### Restriccions d Integritat

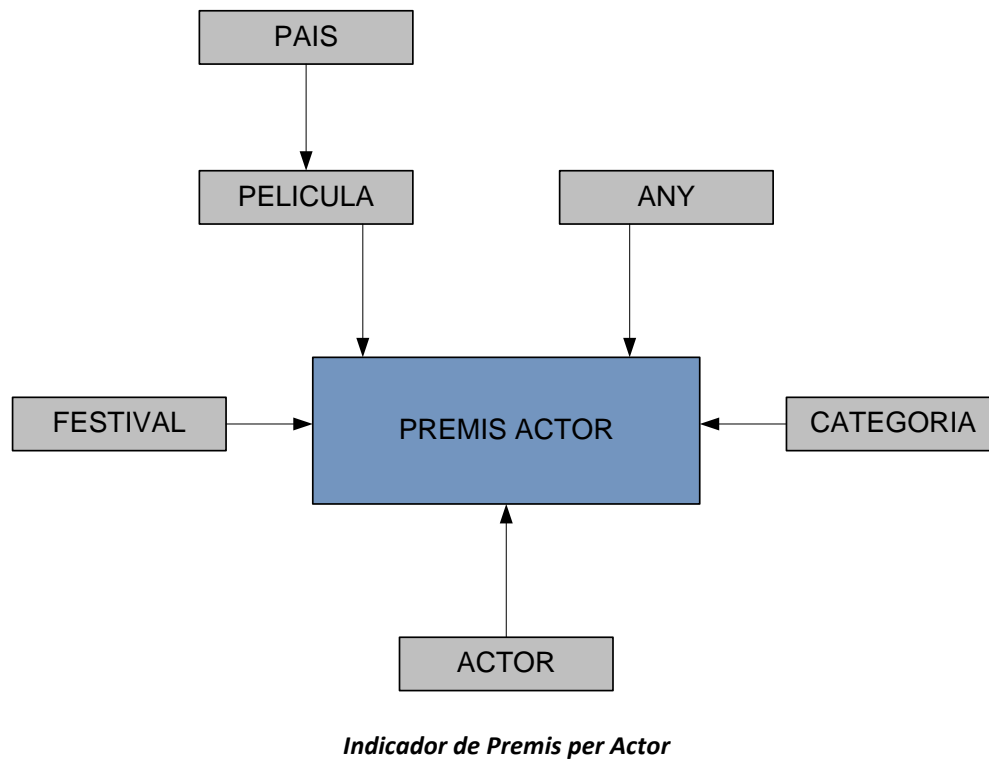
S'han de tenir en compte totes les restriccions de integritat amb les dimensions associades a les taules de fets, així com la restricció d'integritat entre pel·lícula i país.

### Viabilitat

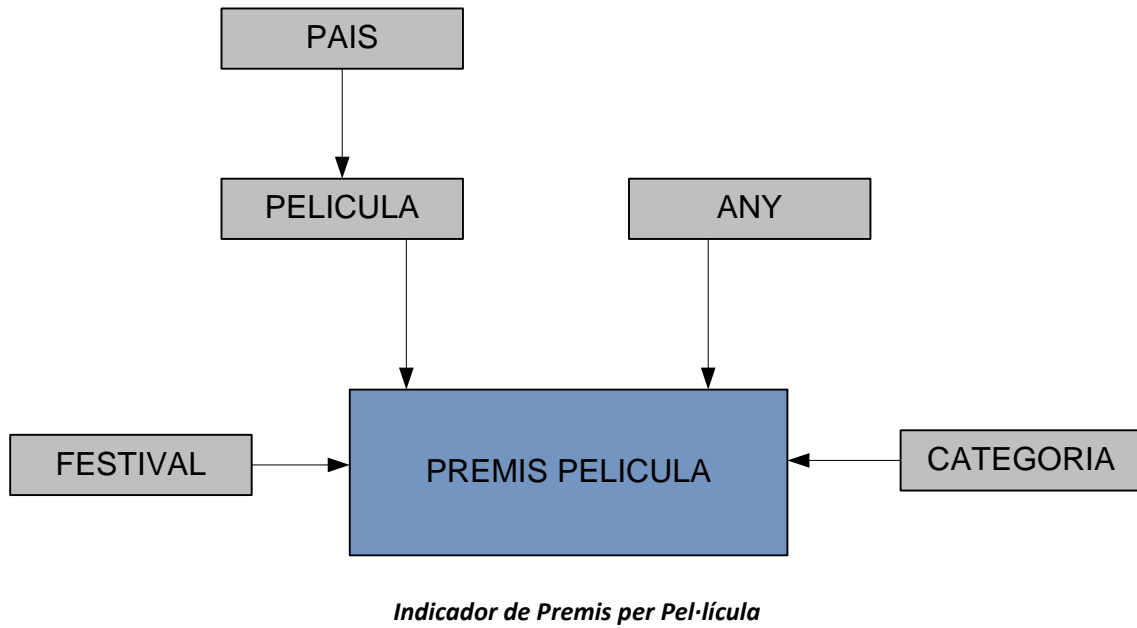
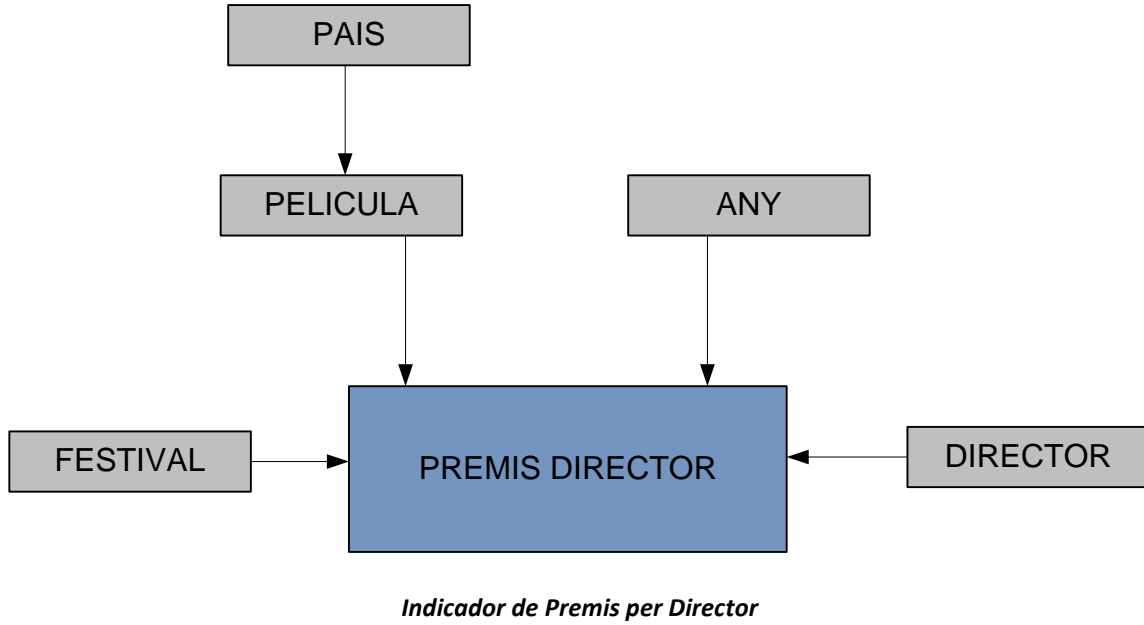
Considerant el volum de dades que representa els Excels analitzats i tenguent present la tecnologia actual no tindrem cap impediment tecnològic per a desenvolupar el model de Magatzem de dades.

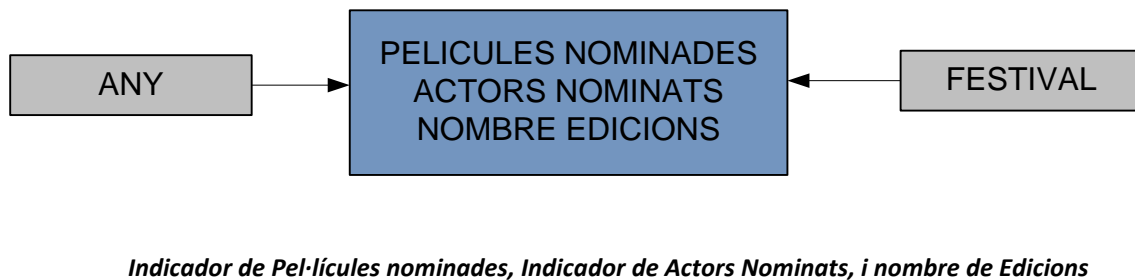
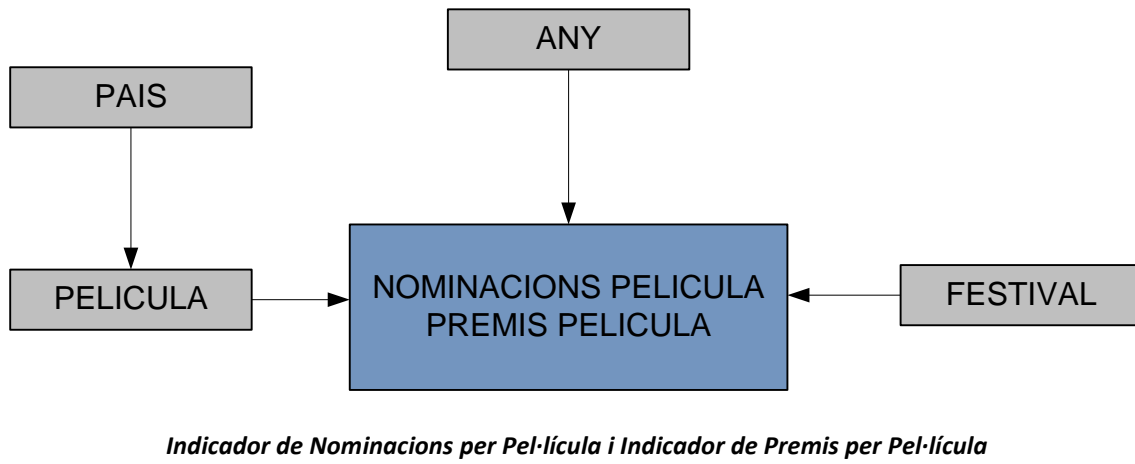
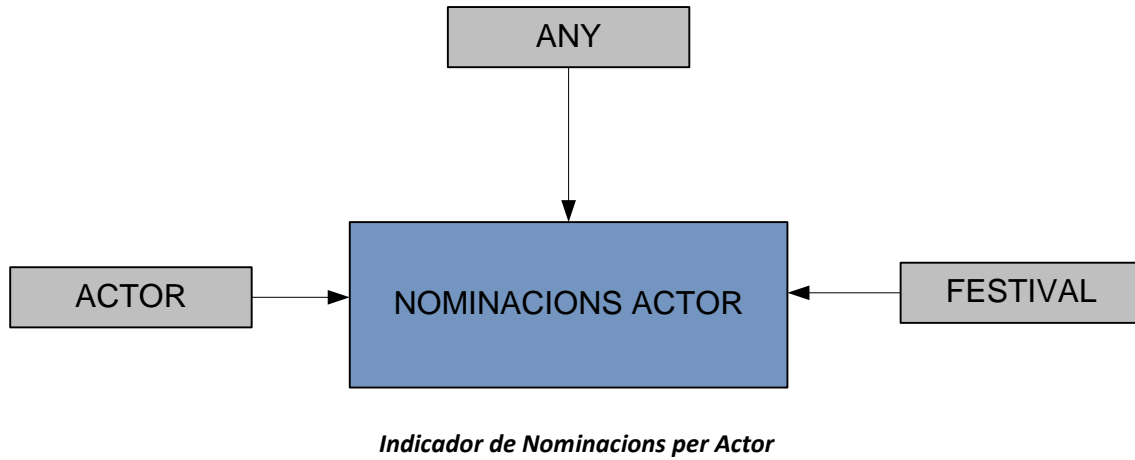
### Diagrama Model conceptual

A continuació es detallen les estrelles implementades al magatzem de dades, totes elles corresponents als fets anomenats anteriorment.









## 5.3 Capítol 3: Disseny

### 5.3.1 Arquitectura de Software

El software elegit per a la implementació del Magatzem de dades es distribuït per l'empresa Oracle amb un reconeixement mundial molt important.

Les Solucions utilitzades son la Base de Dades Oracle 10 g com a software d'emmagatzament i l'eina Oracle Discoverer per a l'explotació de dades. Per a gestionar la Base de Dades Oracle 10g s'utilitzarà l'entorno web proporcionat anomenat **ORACLE Database Express Edition**.

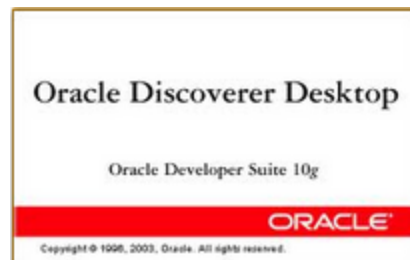


L'eina Oracle Discoverer estarà dividit en dos mòduls:

- Discoverer Administrator, sistema emprat per definir el magatzem de dades, dimensions, relacions, mòdul de permisos d'usuari, etc.



- Discoverer Desktop sistema emprat per els usuaris finals per analitzar la informació del magatzem de dades, es un entorno fàcil de utilitzar i dirigit a usuaris sense coneixements tècnic.



### 5.3.2 Arquitectura de Hardware

A continuació es presenta el model d'arquitectura del Hardware utilitzada per el projecte.

Inicialment s'han definit tres entorns diferenciats:

- Entorn de Desenvolupament

Entorn utilitzat per el desenvolupament i testeig basic dels processos programats. Contindrà un conjunt de dades mínim anonimitzades, per complir amb la llei orgànica de protecció de dades (LOPD), que permeti realitzar el desenvolupament i test necessaris.

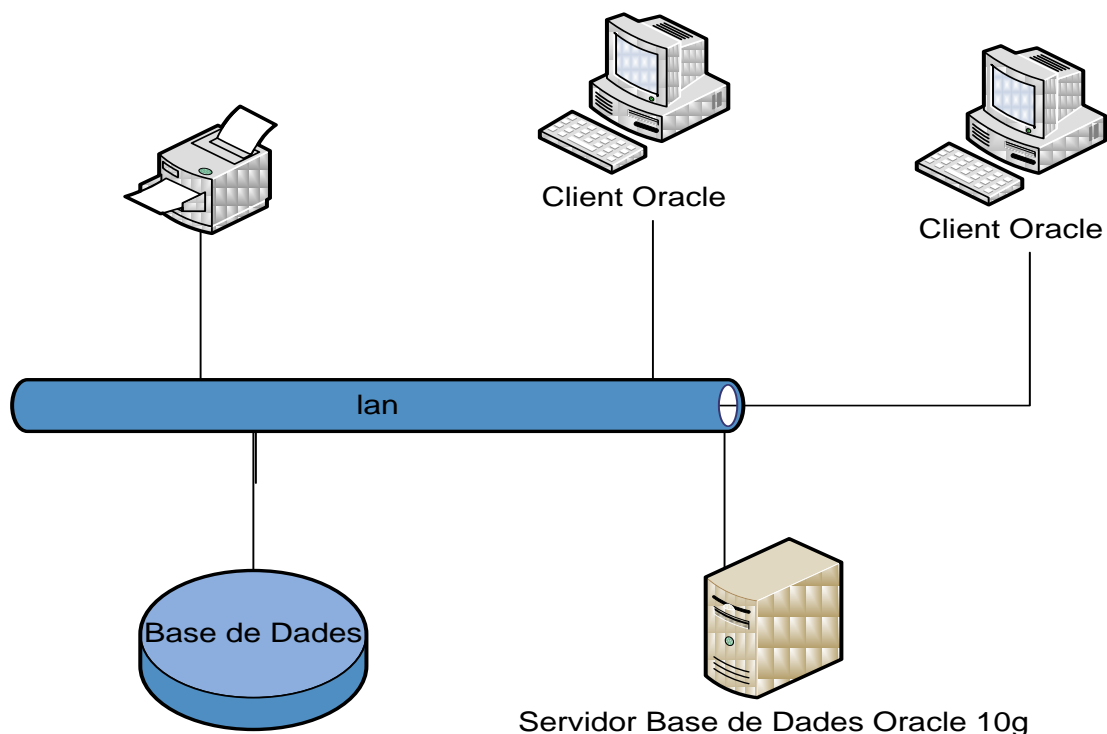
- Entorn de Certificació

E entorn on l'usuari podrà realitzar un guions de test complets i on es certificarà que el desenvolupament compleix els requeriments i que es pot aplicar a l'entorn de producció.

Aquest Entorn contindrà un conjunt de dades major que l'entorn de Desenvolupament representant com a màxim un 50 % del volum de producció, en aquest entorn les dades també estaran anonimitzades.

- Entorn de Producció

Entorn de producció amb les dades reals de DataWareHouse on els usuaris finals podran extreure els informes per realitzar els processos d'anàlisi.



### 5.3.3 Disseny de la Base de Dades

A continuació es detallarà el disseny de la base de dades indicant cada una de les taules que formaran el magatzem de dades i quina relació tindran entre elles, així com indicar les claus primàries i tipus de dades.

- Taula País

Taula mestre on es registren els països del sistema. La clau primària de la taula es ID\_PAIS y es tracta de un valor numèric generat automàticament mitjançant una seqüència. La columna NOM\_STD conte el nom del país unificat entre tots els festivals y sempre ha de contenir un valor. La resta de columnes NOM\_ACADEMY, NOM\_CESAR, NOM\_CANNES, NOM\_MTV i NOM\_BERLINALE es corresponen als nom del país per el Festival Academy, Cesar, Cannes, MTV y Berlinale respectivament, aquestes columnes poden ser nul·les ja que no tots els països estan presents a tots els festivals.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| ID_PAIS           | NUMBER        | No   | -                 | 1              |
| NOM_STD           | VARCHAR2(200) | No   | -                 | -              |
| NOM_ACADEMY       | VARCHAR2(200) | Yes  | -                 | -              |
| NOM_CESAR         | VARCHAR2(200) | Yes  | -                 | -              |
| NOM_CANNES        | VARCHAR2(200) | Yes  | -                 | -              |
| NOM_MTV           | VARCHAR2(200) | Yes  | -                 | -              |
| NOM_BERLINALE     | VARCHAR2(200) | Yes  | -                 | -              |
|                   |               |      |                   | 1-7            |

*taula Pais*

- Taula Festival

Taula mestre on es registraran els Festivals. La clau primària de la taula es ID\_FESTIVAL y es tracta de un valor numèric generat automàticament mitjançant una seqüència. La columna NOM conte la el nom del festival y no pot ser nul.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| ID_FESTIVAL       | NUMBER        | No   | -                 | 1              |
| NOM               | VARCHAR2(200) | Yes  | -                 | -              |
|                   |               |      |                   | 1-2            |

***taula Festival***

- Taula Categoria

Taula mestre on es registraran les Categories. La clau primària de la taula es ID\_Categoria y es tracta de un valor numèric generat automàticament mitjançant una seqüència .La columna NOM\_STD conte el nom del país unificat entre tots els festivals y sempre ha de contenir un valor. La resta de columnes NOM\_ACADEMY, NOM\_CESAR, NOM\_CANNES,NOM\_MTV i NOM\_BERLINEN es corresponen als nom de la categoria corresponent per el Festival Academy,Cesar,Cannes, MTV y Berlinale respectivament, aquestes columnes poden ser nul·les ja que no totes les categories es troben ma petjades.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| ID_CATEGORIA      | NUMBER        | No   | -                 | 1              |
| NOM_STD           | VARCHAR2(200) | No   | -                 | -              |
| NOM_ACADEMY       | VARCHAR2(200) | Yes  | -                 | -              |
| NOM_CESAR         | VARCHAR2(200) | Yes  | -                 | -              |
| NOM_BERLINEN      | VARCHAR2(200) | Yes  | -                 | -              |
| NOM_CANNES        | VARCHAR2(200) | Yes  | -                 | -              |
| NOM_MTV           | VARCHAR2(200) | Yes  | -                 | -              |
|                   |               |      |                   | 1-7            |

***taula Categoria***

- Taula Actor

Taula mestre on es registraran els actors. La clau primària de la taula es ID\_ACTOR y es tracta de un valor numèric generat automàticament mitjançant una seqüència. La columna NOM conte el nom unificat del Actor y no pot ser nul.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| ID_ACTOR          | NUMBER        | No   | -                 | 1              |
| NOM_ACTOR         | VARCHAR2(200) | No   | -                 | -              |
|                   |               |      |                   | 1-2            |

*taula Actor*

- Taula Director

Taula mestre on es registraran els Directors. La clau primària de la taula es ID\_DIRECTOR y es tracta de un valor numèric generat automàticament mitjançant una seqüència. La columna NOM conte la el nom unificat del Director y no pot ser nul.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| ID_DIRECTOR       | NUMBER        | No   | -                 | 1              |
| NOM               | VARCHAR2(200) | No   | -                 | -              |
|                   |               |      |                   | 1-2            |

*taula Director*

- Taula Any

Taula de la data (solament Any) en que es celebra un Festival. La clau primària de la taula es ID\_ANY y es tracta de un valor numèric generat automàticament mitjançant una seqüència. La columna ANYO conte l'any en format YYYY.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| ID_ANY            | NUMBER        | No   | -                 | 1              |
| ANYO              | VARCHAR2(200) | No   | -                 | -              |
|                   |               |      |                   | 1-2            |

*taula Any*

- Taula Pel·lícula

Taula mestre on es registraran les pel·lícules. La clau primària de la taula es ID\_PELICULA y es tracta de un valor numèric generat automàticament mitjançant una seqüència. La columna NOM conte la el nom unificat de la pel·lícula y no pot ser nul. La columna PAIS conte una clau forana al país al qual pertany la pel·lícula.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| ID_PELICULA       | NUMBER        | No   | -                 | 1              |
| NOM               | VARCHAR2(300) | Yes  | -                 | -              |
| PAIS              | NUMBER        | Yes  | -                 | -              |
|                   |               |      |                   | 1-3            |

*taula Pelicula*



- Taula Premis\_Actor

Taula on es registraran els Nombre de Premis per els Actors per Any,Festival, Categoria i Pel·lícula. La columna ID\_ACTOR conte una clau forana a la taula mestre de actors, ID\_PELICULA conte una clau forana a la taula mestre de pel·lícules, ID\_CATEGORIA conte una clau forana a la taula mestre de categories, ID\_ANY conte una clau forana a la taula mestre de anys, ID\_FESTIVAL conte una clau forana a la taula mestre de festivals. La columna N\_PREMIS conte el indicador de nombre de premis.

La clau primària esta formada por els valors de les columnes ID\_ACTOR,ID\_PELICULA,ID\_CATEGORIA,ID\_ANY, i ID\_FESTIVAL.

| Nombre De Columna | Tipo De Dato | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|--------------|------|-------------------|----------------|
| ID_ACTOR          | NUMBER       | No   | -                 | 1              |
| ID_PELICULA       | NUMBER       | No   | -                 | 2              |
| ID_CATEGORIA      | NUMBER       | No   | -                 | 3              |
| ID_ANY            | NUMBER       | No   | -                 | 4              |
| ID_FESTIVAL       | NUMBER       | No   | -                 | 5              |
| N_PREMIS          | NUMBER       | Yes  | -                 | -              |
|                   |              |      |                   | 1 - 6          |

**Taula de Fets Premis\_Actor**

- Taula Premis\_Director

Taula on es registraran el nombre de Premis per els Directors per Any,Festival i Pel·lícula. La columna ID\_DIRECTOR conte una clau forana a la taula mestre de directors, ID\_PELICULA conte una clau forana a la taula mestre de pel·lícules ,ID\_ANY conte una clau forana a la taula mestre de anys, ID\_FESTIVAL conte una clau forana a la taula mestre de festivals. La columna N\_PREMIS conte el indicador de nombre de premis.

La clau primària esta formada por els valors de les columnes ID\_DIRECTOR,ID\_PEL·LÍCULA, ,ID\_ANY, i ID\_FESTIVAL.

| Nombre De Columna | Tipo De Dato | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|--------------|------|-------------------|----------------|
| ID_DIRECTOR       | NUMBER       | No   | -                 | 1              |
| ID_FESTIVAL       | NUMBER       | No   | -                 | 2              |
| ID_ANY            | NUMBER       | No   | -                 | 3              |
| ID_PELICULA       | NUMBER       | No   | -                 | 4              |
| N_PREMIS          | NUMBER       | No   | -                 | -              |
|                   |              |      |                   | 1 - 5          |

**taula de Fets Premis\_Director**

- Taula Premis\_Pel·lícules

Taula on es registraran el nombre de Premis per Pel·lícula, any, festival i categoria. La columna ID\_PELICULA conte una clau forana a la taula mestre de pel·lícules ,ID\_ANY conte una clau forana a la taula mestre de anys, ID\_FESTIVAL conte una clau forana a la taula mestre de festivals, La columna ID\_Categoria conte una clau forana a la taula mestre de categories. La columna N\_PREMIS conte el indicador de nombre de premis.

La clau primària esta formada por els valors de les columnes ID\_CATEGORIA, ID\_PEL·LÍCULA, ,ID\_ANY, i ID\_FESTIVAL.

| Nombre De Columna | Tipo De Dato | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|--------------|------|-------------------|----------------|
| ID_PELICULA       | NUMBER       | No   | -                 | 1              |
| ID_ANY            | NUMBER       | No   | -                 | 2              |
| ID_FESTIVAL       | NUMBER       | No   | -                 | 3              |
| ID_CATEGORIA      | NUMBER       | No   | -                 | 4              |
| N_PREMIS          | NUMBER       | No   | -                 | -              |
|                   |              |      |                   | 1 - 5          |

***taula de Fets Premis\_Pelicules***

- Taula Nominacions\_Actors

Taula on es registraran el nombre de nominacions dels actors, per festival y any. La columna ID\_ACTOR conte una clau forana a la taula mestre de actors, ID\_ANY conte una clau forana a la taula mestre de anys, ID\_FESTIVAL conte una clau forana a la taula mestre de festivals. La columna N\_PREMIS conte el indicador de nombre de premis.

La clau primària esta formada por els valors de les columnes ID\_ACTOR,,ID\_ANY, i ID\_FESTIVAL.

| Nombre De Columna | Tipo De Dato | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|--------------|------|-------------------|----------------|
| ID_ACTOR          | NUMBER       | No   | -                 | 1              |
| ID_FESTIVAL       | NUMBER       | No   | -                 | 2              |
| ID_ANY            | NUMBER       | No   | -                 | 3              |
| N_NOMINACIONES    | NUMBER       | Yes  | -                 | -              |
|                   |              |      |                   | 1 - 4          |

***taula de Fets Nominacions\_Actor***

- Taula Premis\_Pel·lícula\_AF

Taula on es registraran el nombre de nominacions y premis de les pel·lícules per festival y any. La columna ID\_PELICULA conte una clau forana a la taula mestre de pel·lícules, ID\_ANY conte una clau forana a la taula mestre de anys, ID\_FESTIVAL conte una clau forana a la taula mestre de festivals. La columna N\_PREMIS conte el indicador de nombre de premis, La columna N\_NOMIACIONS conte el indicador de nombre de premis

La clau primària esta formada por els valors de les columnes ID\_PELICULA, ID\_ANY, i ID\_FESTIVAL.

| Nombre De Columna | Tipo De Dato | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|--------------|------|-------------------|----------------|
| ID_PELICULA       | NUMBER       | No   | -                 | 1              |
| ID_FESTIVAL       | NUMBER       | No   | -                 | 2              |
| ID_ANY            | NUMBER       | No   | -                 | 3              |
| N_PREMIS          | NUMBER       | No   | -                 | -              |
| N_NOMIACIONS      | NUMBER       | Yes  | -                 | -              |
|                   |              |      |                   | 1 - 5          |

*taula de Fets Premis\_Pelicula\_AF*

- Taula NOMINATS\_EDICIONS\_FESTIVAL

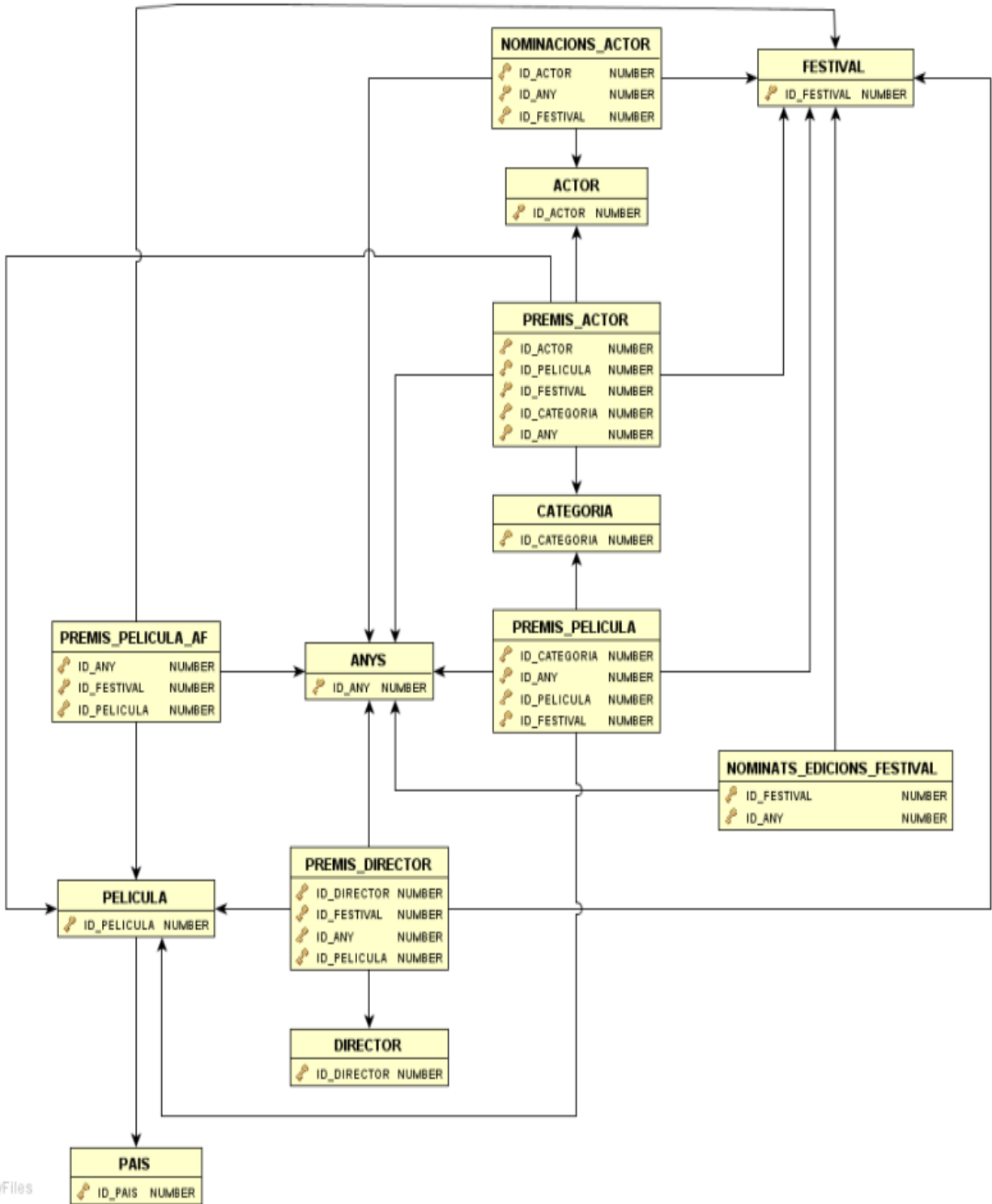
Taula on es registraran el nombre de nominacions totals de actors y pel·lícules per festival y any, això com un indicador amb el nombre de edicions. La columna ID\_ANY conte una clau forana a la taula mestre de anys, ID\_FESTIVAL conte una clau forana a la taula mestre de festivals. La columna N\_ACTORS\_NOM conte el indicador de nombre de actors nominats, La columna N\_PELICULES\_NOM conte el indicador de nombre de pel·lícules nominades, La columna N\_EDICIONS conte el nombre edicions.

La clau primària esta formada por els valors de les columnes ID\_ANY, i ID\_FESTIVAL.

| Nombre De Columna | Tipo De Dato | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|--------------|------|-------------------|----------------|
| ID_FESTIVAL       | NUMBER       | No   | -                 | 1              |
| ID_ANY            | NUMBER       | No   | -                 | 2              |
| N_EDICIONS        | NUMBER       | Yes  | -                 | -              |
| N_ACTORS_NOM      | NUMBER       | Yes  | -                 | -              |
| N_PELICULES_NOM   | NUMBER       | Yes  | -                 | -              |
|                   |              |      |                   | 1 - 5          |

*taula de Fets Nominats\_edicions\_festival*

Diagrama de Base de Dades



### 5.3.4 Procés de Carrega de Informació

A continuació es defineixen els processos de carrega (ETL) per tal d'incorporar les dades dels diferents Excels al magatzem de dades. Així com hem explicat a l'apartat de Anàlisi s'ha de realitzar una forta tasca manual d'unificació de dades:

- Països

Els Països es troben definits en diferents idiomes segons el Festival, es decideix que els registres es guardaran unificats i es realitzarà una tasca prèvia de traducció al anglès per tal de guardar un valor unificat per a tots els països. S'ha identificat que una mateixa pel·lícula pot estar assignada a diversos països, en aquest cas es crearà un país fictici amb la descripció de tots els països implicats. Es crearà un fitxer de mapeig per cada Festival.

- Categories

Les Categories tenen diverses denominacions segons el Festival sense deixar el mateix tipus de premi. En aquest cas s'han de guardar ambdues definicions de la Categoria i crear una unificada, la Categoria unificada serà la de l Festival Academy per renom. Per tal de realitzar aquest mapeig es crearan unes taules externes on es relacionaran les mateixes Categories als diferents Festivals. En els casos de que una Categoria tan sols sigui d'un Festival no tindrà relació amb altres. Es crearà un fitxer de mapeig per cada Festival.

- Actors

Per tal de unificar el nom dels actors es guardarà el nom amb majúscules per tal d'evitar duplicitats per diferències entre minúscules i majúscules. Als Excels disposa'm de registres on per una mateixa pel·lícula hi ha diversos actors nominats separats per comes, aquets actors es crearan per separat utilitzat el caràcter “,” com a separador. En aquest cas s'ha detectat que la coma també delimita els actors anomenats NOM, JR, aquest cas s'ha de tenir en conte ja que es tracta de un sol actor.

- Directors

S'aplicarà la mateixa lògica que per els actors.

- Pel·lícules

Per tal de unificar el nom de les pel·lícules es guardarà el nom amb majúscules per tal d'evitar duplicitats per diferències entre minúscules i majúscules. S'ha de tenir en compte que les Pel·lícules tenen diversos noms, en ocasions en anglès, s'ha decidit que primer de tot es guardarà el nom original (present a tots els festivals). Les pel·lícules que no tinguin un país associat quedaran lligades al país Estats Units.

- Anys

Per a la carrega d'anys s'ha detectat que hi ha valors registrats amb el següent format yyyy1-yyyy2 en aquets casos es guardarà únicament el any definit per la cadena yyyy2.

Per a fer la carrega dels Excels es guardaran els fitxers en format csv (separats) per comes a, a continuació s'explica el contingut de cada un dels fitxers:

| CURSOR                                | Propòsit   |
|---------------------------------------|--|
| <b>ACADEMY_PAISES</b>                 | Mapeig dels països per festival Academy            |
| <b>ACADEMY_CATEGORIAS</b>             | Mapeig de les categories per el festival Academy   |
| <b>ACADEMY_PREMIOS_NOMINACIONES</b>   | Excel corresponent al festival Academy             |
| <b>CESAR_PAISES</b>                   | Mapeig dels països per festival Cesar              |
| <b>CESAR_CATEGORIAS</b>               | Mapeig de les categories per el festival Cesar     |
| <b>CESAR_PREMIOS_NOMINACIONES</b>     | Excel corresponent al festival Cesar               |
| <b>BERLINALE_PAISES</b>               | Mapeig dels països per festival Berlinale          |
| <b>BERLINALE_CATEGORIAS</b>           | Mapeig de les categories per el festival Berlinale |
| <b>BERLINALE_PREMIOS_NOMINACIONES</b> | Excel corresponent al festival Berlinale           |
| <b>CANNES_PAISES</b>                  | Mapeig dels països per festival Cannes             |
| <b>CANNES_CATEGORIAS</b>              | Mapeig de les categories per el festival Cannes    |
| <b>CANNES_PREMIOS_NOMINACIONES</b>    | Excel corresponent al festival Cannes              |

El procés per a poder tractar els fitxers anomenats en el punt anterior serà mitjançant la creació de taules externes. Es crea una taula externa per a cada fitxer. A continuació es mostren les taules externes mostrades des de el gestor de oracle.

- Academy\_categories\_ext.

Taula corresponent al fitxer de categories per el Festival Academy.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| NOM_ORG           | VARCHAR2(150) | Yes  | -                 | -              |
| NOM_STD           | VARCHAR2(150) | Yes  | -                 | -              |
|                   |               |      |                   | 1 - 2          |

- Academy\_paises\_ext

Taula corresponent al fitxer mapetjat de països per el Festival Academy.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| NOM_ORG           | VARCHAR2(100) | Yes  | -                 | -              |
| NOM_STD           | VARCHAR2(100) | Yes  | -                 | -              |
|                   |               |      |                   | 1 - 2          |

- Academy\_premios\_ext

Taula corresponent al fitxer Excel per el Festival Academy. El nom de cada columna s'ha respectat el del fitxer Excel per tal de depurar les dades mes fàcilment.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| ORIGINAL_TITLE    | VARCHAR2(500) | Yes  | -                 | -              |
| ENGLISH_TITLE     | VARCHAR2(500) | Yes  | -                 | -              |
| SORT_TITLE        | VARCHAR2(500) | Yes  | -                 | -              |
| YEAR              | VARCHAR2(500) | Yes  | -                 | -              |
| COUNTRY           | VARCHAR2(500) | Yes  | -                 | -              |
| AWARD             | VARCHAR2(500) | Yes  | -                 | -              |
| CATEGORY          | VARCHAR2(500) | Yes  | -                 | -              |
| WINNER            | VARCHAR2(500) | Yes  | -                 | -              |
| NOMINEES          | VARCHAR2(500) | Yes  | -                 | -              |
| EDITION           | VARCHAR2(500) | Yes  | -                 | -              |
| ITEM              | VARCHAR2(500) | Yes  | -                 | -              |
|                   |               |      |                   | 1 - 11         |

- MTV\_categorias\_ext.

Taula corresponent al fitxer de categories per el Festival MTV.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| NOM_ORG           | VARCHAR2(150) | Yes  | -                 | -              |
| NOM_STD           | VARCHAR2(150) | Yes  | -                 | -              |
|                   |               |      |                   | 1 - 2          |

- MTV\_paises\_ext

Taula corresponent al fitxer mapetjat de països per el Festival MTV.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| NOM_ORG           | VARCHAR2(100) | Yes  | -                 | -              |
| NOM_STD           | VARCHAR2(100) | Yes  | -                 | -              |
|                   |               |      |                   | 1 - 2          |

- MTV\_premios\_ext

Taula corresponent al fitxer Excel per el Festival MTV. El nom de cada columna s'ha respectat el del fitxer Excel per tal de depurar les dades mes fàcilment.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| ORIGINAL_TITLE    | VARCHAR2(500) | Yes  | -                 | -              |
| ENGLISH_TITLE     | VARCHAR2(500) | Yes  | -                 | -              |
| SORT_TITLE        | VARCHAR2(500) | Yes  | -                 | -              |
| YEAR              | VARCHAR2(500) | Yes  | -                 | -              |
| COUNTRY           | VARCHAR2(500) | Yes  | -                 | -              |
| AWARD             | VARCHAR2(500) | Yes  | -                 | -              |
| CATEGORY          | VARCHAR2(500) | Yes  | -                 | -              |
| WINNER            | VARCHAR2(500) | Yes  | -                 | -              |
| NOMINEES          | VARCHAR2(500) | Yes  | -                 | -              |
| EDITION           | VARCHAR2(500) | Yes  | -                 | -              |
| ITEM              | VARCHAR2(500) | Yes  | -                 | -              |
|                   |               |      |                   | 1 - 11         |



- CESAR\_categorias\_ext.

Taula corresponent al fitxer de categories per el Festival CESAR.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| NOM_ORG           | VARCHAR2(150) | Yes  | -                 | -              |
| NOM_STD           | VARCHAR2(150) | Yes  | -                 | -              |
|                   |               |      |                   | 1-2            |

- CESAR\_paises\_ext

Taula corresponent al fitxer mapetjat de països per el Festival CESAR.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| NOM_ORG           | VARCHAR2(100) | Yes  | -                 | -              |
| NOM_STD           | VARCHAR2(100) | Yes  | -                 | -              |
|                   |               |      |                   | 1-2            |

- CESAR\_premios\_ext

Taula corresponent al fitxer Excel per el Festival CESAR. El nom de cada columna s'ha respectat el del fitxer Excel per tal de depurar les dades mes fàcilment.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| FILM_TITLE        | VARCHAR2(500) | Yes  | -                 | -              |
| LOCALIZED_TITLE   | VARCHAR2(500) | Yes  | -                 | -              |
| SORT_TITLE        | VARCHAR2(500) | Yes  | -                 | -              |
| YEAR              | VARCHAR2(500) | Yes  | -                 | -              |
| COUNTRY           | VARCHAR2(500) | Yes  | -                 | -              |
| AWARD             | VARCHAR2(500) | Yes  | -                 | -              |
| CATEGORY          | VARCHAR2(500) | Yes  | -                 | -              |
| WINNER            | VARCHAR2(500) | Yes  | -                 | -              |
| NOMINEE           | VARCHAR2(500) | Yes  | -                 | -              |
| EDITION           | VARCHAR2(500) | Yes  | -                 | -              |
| ITEM              | VARCHAR2(500) | Yes  | -                 | -              |
|                   |               |      |                   | 1-11           |

- CANNES\_categorias\_ext.

Taula corresponent al fitxer de categories per el Festival CANNES.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| NOM_ORG           | VARCHAR2(150) | Yes  | -                 | -              |
| NOM_STD           | VARCHAR2(150) | Yes  | -                 | -              |
|                   |               |      |                   | 1 - 2          |

- CANNES\_paises\_ext

Taula corresponent al fitxer ma petjat de paisos per el Festival CANNES.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| NOM_ORG           | VARCHAR2(100) | Yes  | -                 | -              |
| NOM_STD           | VARCHAR2(100) | Yes  | -                 | -              |
|                   |               |      |                   | 1 - 2          |

- CANNES\_premios\_ext

Taula corresponent al fitxer Excel per el Festival CANNES. El nom de cada columna s'ha respectat el del fitxer Excel per tal de depurar les dades mes fàcilment.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| FILM_TITLE        | VARCHAR2(500) | Yes  | -                 | -              |
| LOCALIZED_TITLE   | VARCHAR2(500) | Yes  | -                 | -              |
| SORT_TITLE        | VARCHAR2(500) | Yes  | -                 | -              |
| YEAR              | VARCHAR2(500) | Yes  | -                 | -              |
| COUNTRY           | VARCHAR2(500) | Yes  | -                 | -              |
| AWARD             | VARCHAR2(500) | Yes  | -                 | -              |
| CATEGORY          | VARCHAR2(500) | Yes  | -                 | -              |
| WINNER            | VARCHAR2(500) | Yes  | -                 | -              |
| NOMINEE           | VARCHAR2(500) | Yes  | -                 | -              |
| EDITION           | VARCHAR2(500) | Yes  | -                 | -              |
| ITEM              | VARCHAR2(500) | Yes  | -                 | -              |
|                   |               |      |                   | 1 - 11         |

- BERLINALE\_categorias\_ext.

Taula corresponent al fitxer de categories per el Festival BERLINALE.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| NOM_ORG           | VARCHAR2(150) | Yes  | -                 | -              |
| NOM_STD           | VARCHAR2(150) | Yes  | -                 | -              |
|                   |               |      |                   | 1 - 2          |

- BERLINALE\_paises\_ext

Taula corresponent al fitxer mapetjat de països per el Festival BERLINALE.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| NOM_ORG           | VARCHAR2(100) | Yes  | -                 | -              |
| NOM_STD           | VARCHAR2(100) | Yes  | -                 | -              |
|                   |               |      |                   | 1 - 2          |

- BERLINALE\_premios\_ext

Taula corresponent al fitxer Excel per el Festival BERLINALE. El nom de cada columna s'ha respectat el del fitxer Excel per tal de depurar les dades mes fàcilment.

| Nombre De Columna | Tipo De Dato  | Nulo | Valor Por Defecto | Clave Primaria |
|-------------------|---------------|------|-------------------|----------------|
| FILM_TITLE        | VARCHAR2(500) | Yes  | -                 | -              |
| LOCALIZED_TITLE   | VARCHAR2(500) | Yes  | -                 | -              |
| SORT_TITLE        | VARCHAR2(500) | Yes  | -                 | -              |
| YEAR              | VARCHAR2(500) | Yes  | -                 | -              |
| COUNTRY           | VARCHAR2(500) | Yes  | -                 | -              |
| AWARD             | VARCHAR2(500) | Yes  | -                 | -              |
| CATEGORY          | VARCHAR2(500) | Yes  | -                 | -              |
| WINNER            | VARCHAR2(500) | Yes  | -                 | -              |
| NOMINEE           | VARCHAR2(500) | Yes  | -                 | -              |
| EDITION           | VARCHAR2(500) | Yes  | -                 | -              |
| ITEM              | VARCHAR2(500) | Yes  | -                 | -              |
|                   |               |      |                   | 1 - 11         |

Per a processar les dades des de les taules externes s'han emprat cursos per tal de poder recollir i tractar la informació necessària abans de fer les insercions i/o modificacions a base de dades. A continuació s'inclouen els cursors utilitzats.

Hi ha dos tipus de cursors els emprats per completar les taules mestres (actor, país..etc) i cursors per completar la informació dins les taules de fets (premis\_actors, premis\_director,etc...)

#### **Creació de les dades mestres:**

A continuació enumera'm els cursos involucrats a n' aquest procés, els cursors apliquen la lògica explicada anteriorment:

| CURSOR                 | Propòsit   |
|------------------------|--|
| CARGA_CESAR_PAISES     | Llegir la taula externa de mapeig de països per el festival CESAR i insertar els registres a la taula mestre de Països. Si el país ja existeix com estàndard s' actualitza la columna del festival CESAR.              |
| CARGA_ACADEMY_PAISES   | Llegir la taula externa de mapeig de països per el festival ACADEMY i insertar els registres a la taula mestre de Països. Si el país ja existeix com estàndard se actualitza la columna del festival ACADEMY.          |
| CARGA_CANNES_PAISES    | Llegir la taula externa de mapeig de països per el festival CANNES i insertar els registres a la taula mestre de Països. Si el país ja existeix com estàndard s' actualitza la columna del festival CANNES.            |
| CARGA_BERLINALE_PAISES | Llegir la taula externa de mapeig de països per el festival BERLINALE i insertar els registres a la taula mestre de Països. Si el país ja existeix com estàndard se actualitza la columna del festival BERLINALE.      |
| CARGA_CESAR_CATEGORIA  | Llegir la taula externa de mapeig de categories per el festival CESAR i insertar els registres a la taula mestre de Categories. Si la categoria ja existeix com estàndard se actualitza la columna del festival CESAR. |
| CARGA_CANNES_CATEGORIA | Llegir la taula externa de mapeig de categories per el festival CANNES i insertar els registres a la taula mestre de Categories. Si la categoria ja existeix com estàndard se actualitza la columna                    |

|                           |   |
|---------------------------|---|
| CARGA_MTV_CATEGORIA       | del festival CANNES.<br>Llegir la taula externa de mapeig de categories per el festival MTV i insertar els registres a la taula mestre de Categories. Si la categoria ja existeix com estàndard se actualitza la columna del festival MTV.  |
| CARGA_BERLINALE_CATEGORIA | Llegir la taula externa de mapeig de categories per el festival BERLINALE i insertar els registres a la taula mestre de Categories. Si la categoria ja existeix com esdentar s' actualitza la columna del festival BERLINALE.   |
| CARGA_ACADEMY_CATEGORIA   | Llegir la taula externa de mapeig de categories per el festival ACADEMY i insertar els registres a la taula mestre de Categories. Si la categoria ja existeix com estàndard s' actualitza la columna del festival ACADEMY.  |
| CARGA_ACTORES             | Cursor encarregat de crear els actors a la taula mestre de actors, es llegeixen totes les files de les taules temporals de premis de cada un dels festivals (per les categories relacionades amb els actors) i s' incerta a la taula de actors si l'actor ja existeix es descarta la transacció. Per tal de unificar els actors s'ha fet un UPPER a tots els registres .                    |
| CARGA_ANYOS               | Llegir els Anys de totes les taules externes de premis de cada una de les categories i crear un únic registre per any (distint).  |
| CARGA_DIRECTORES          | Cursor encarregat de crear els directors a la taula mestre de directors, es llegeixen totes les files de les taules temporals de premis de cada un dels festivals (per les categories relacionades amb els directors) i s' incerta a la taula de directors si el director ja existeix es descarta la transacció. Per tal de unificar els directors s'ha fet un UPPER a tots els registres . |
| CARGA_FESTIVAL            | Creació dels Festivals a les taules mestres.  |
| CARGA_PELICULAS           | Cursor encarregat de crear les pel·lícules a la taula mestre de pel·lícules, es llegeixen totes les files de les taules temporals de premis de cada un dels festivals i s'inseria a la taula de pel·lícula si la  |

pel·lícula ja existeix es descarta la transacció. Per tal de unificar les pel·lícules s'ha fet un UPPER a tots els registres .

#### Inserció de les dades de les taules de Fets:

A continuació enumera'm els cursos involucrats a n' aquest procés, els cursos apliquen la lògica explicada anteriorment:

| CURSOR                     | Propòsit   |
|----------------------------|--|
| CARGA_PREMIOS_ACTOR        | Cursor encarregat de llegir els premis de actors per a totes les taules externes de premis. Les dades es carregaran a la taula PREMIS_ACTOR  |
| CARGA_PREMIOS_DIRECTOR     | Cursor encarregat de llegir els premis de directors per a totes les taules externes de premis. Les dades es carregaran a la taula PREMIS_DIRECTOR  |
| CARGA_PREMIOS_PELICULES    | Cursor encarregat de llegir els premis de pel·lícules per a totes les taules externes de premis. Les dades es carregaran a la taula PREMIS_PELICULA  |
| CARGA_PREMIOS_PELICULES_AF | Cursor encarregat de llegir els premis y nominacions de pel·lícules premiades per Festival y Any per a totes les taules externes de premis. Les dades es carregaran a la taula PREMIS_PELICULA_AF            |
| CARGA_NOM_PELICULES_AF     | Cursor encarregat de llegir les nominacions de pel·lícules sense premiar per Festival y Any per a totes les taules externes de premis . Les dades es carregaran /actualitzaran a la taula PREMIS_PELICULA_AF |
| CARGA_NOMINACIONS_ACTOR    | Cursor encarregat de llegir les nominacions per a totes les taules externes de premis dels actors per any y Festival y guardar les dades a la taula NOMINACIONS_ACTOR.                                       |
| CARGA_PELICULAS_NOMINADAS  | Cursor encarregat de llegir les nominacions de pel·lícula per a Festival i any per a totes les taules externes de premis i carregar la informació a la taula NOMINATS_EDICIONS_FESTIVAL.                     |
| CARGA_ACTORS_NOMINATS_V2   | Cursor encarregat de llegir les nominacions de   |

**CARGA\_EDICIONS**

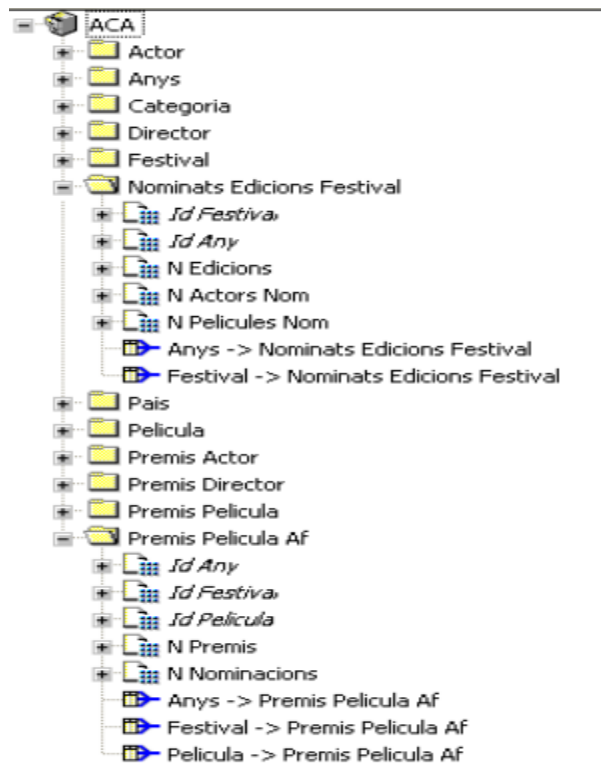
actors de la taula de fets de NOMINACIONS\_ACTOR y actualitzar la informació a la taula NOMINATS\_EDICIONS\_FESTIVAL.

Cursor encarregat d'actualitzar a la taula de NOMINATS\_EDICIONS\_FESTIVAL el nombre de edicions per any als festivals.

### 5.3.5 Disseny dels Informes

A continuació es detalla com s'ha realitzat la implementació dels informes.

El primer pas es definir el model de negoci mitjançant l'eina Oracle Discoverer Administrator, en el nostre cas el nom assignat ha estat ACA\_DWH, en aquest procés es seleccionen les taules del model de dades que s'empraran per a realitzar els informes així com definir altres elements com jerarquies, relacions entre taules. Al Oracle Discoverer Administrator també es poden crear càlculs i restriccions predefinides que després podran ser emprades per l'usuari final des de Oracle Discoverer Desktop



Finalment mitjançant Oracle Discoverer Desktop es defineixen els informes requerits per l'ACA, a continuació es mostren els informes creats:

- Informe Premis aconseguits per actor/actriu per any i festival. A l'informe s'ha d'indicar el total de premis per l'actor/actriu. Al contingut d'aquest informe s'indicarà la següent informació:
  - Nom Actor/Actriu
  - Any
  - Festival
  - Total al Festival
  - Total a tots els Festivals (en vermell a l'infome)

## Informe 1: Premis per Actor, Any i Festival

Elementos de Página:

|   | Actor        | Any  | Festival      | Total Festival-Any |
|---|--------------|------|---------------|--------------------|
| 1 | ADRIEN BRODY | 2002 | César         | 1                  |
| 2 |              | 2002 | Academy Award | 1                  |
| 3 |              |      |               | TOTAL: 2           |
| 4 | AGNÈS JAQUI  | 1997 | César         | 1                  |
| 5 |              |      |               | TOTAL: 1           |
| 6 | AL PACINO    | 1992 | Academy Award | 1                  |
| 7 |              |      |               | TOTAL: 1           |
| 8 | ALAIN DELON  | 1984 | César         | 1                  |
| 9 |              |      |               | TOTAL: 1           |



- Informe Premis aconseguits per actor/actriu per any, festival i categoria.  
Al contingut d'aquest informe s'indicarà la següent informació:
  - Nom Actor/Actriu
  - Any
  - Festival
  - Categoria Unificada
  - Total

| <b>Informe 2: Premis per Actor,<br/>Any i Festival i Categoria</b>      |            |      |               |       |
|---|------------|------|---------------|-------|
| Elementos de Página: <input type="text" value="Actor: ADRIEN BRODY ▼"/> |            |      |               |       |
|   | Categoria  | Any  | Festival      | Total |
| 1   | Best Actor | 2002 | César         | 1     |
| 2   | Best Actor | 2002 | Academy Award | 1     |

- Informe Premis aconseguits per un director per any, festival i pel·lícula.  
A l' informe s'ha d'indicar el total de premis per el director.  
Al contingut d'aquest informe s'indicarà la següent informació:
  - Nom Director
  - Pel·lícula
  - Any
  - Festival
  - Total per Festival
  - Total Global

| <b>Informe 3: Premis per Director<br/>Any Festival i Pelicula</b>           |                    |      |                    |                   |
|---|--------------------|------|--------------------|-------------------|
| Elementos de Página: <input type="text" value="Festival: Academy Award ▼"/> |                    |      |                    |                   |
|   | Director           | Any  | Pelicula           | Total             |
| 1   | ANG LEE            | 2005 | BROKEBACK MOUNTAIN | 1                 |
| 2   |                    |      |                    | Total Director: 1 |
| 3   | ANTHONY MINGHELLA  | 1996 | ENGLISH PATIENT    | 1                 |
| 4   |                    |      |                    | Total Director: 1 |
| 5   | BARRY LEVINSON     | 1988 | RAIN MAN           | 1                 |
| 6   |                    |      |                    | Total Director: 1 |
| 7   | BERNARDO BERTOLUCC | 1987 | LAST EMPEROR       | 1                 |
| 8   |                    |      |                    | Total Director: 1 |
| 9   | BILLY WILDER       | 1960 | APARTMENT          | 1                 |
| 10  |                    | 1945 | LOST WEEKEND       | 1                 |
| 11  |                    |      |                    | Total Director: 2 |

- Informe Premis aconseguits per una pel·lícula per any, festival i categoria.  
Al contingut d'aquest informe s'indicarà la següent informació
  - Pel·lícula
  - Any
  - Festival
  - Categoria Unificada
  - Categoria Original (categoria corresponent del festival).
  - Total

## Informe 4: Premis per Pelicula Any, Festival i Categoria

Elementos de Página: Pelicula: LORD OF THE RINGS: THE RETURN OF THE KING ▼

|    | Any  | Festival         | Categoria Standard               | Total              | Nom Academy                      |
|----|------|------------------|----------------------------------|--------------------|----------------------------------|
| 1  | 2003 | Academy Award    | Best Picture                     | 1                  | Best Picture                     |
| 2  | 2003 | Academy Award    | Best Makeup                      | 1                  | Best Makeup                      |
| 3  | 2003 | Academy Award    | Best Director                    | 1                  | Best Director                    |
| 4  | 2003 | Academy Award    | Best Writing, Adapted Screenplay | 1                  | Best Writing, Adapted Screenplay |
| 5  | 2003 | Academy Award    | Best Art Direction               | 1                  | Best Art Direction               |
| 6  | 2003 | Academy Award    | Best Costume Design              | 1                  | Best Costume Design              |
| 7  | 2003 | Academy Award    | Best Film Editing                | 1                  | Best Film Editing                |
| 8  | 2003 | Academy Award    | Best Music, Original Score       | 1                  | Best Music, Original Score       |
| 9  | 2003 | Academy Award    | Best Music, Song                 | 1                  | Best Music, Song                 |
| 10 | 2003 | Academy Award    | Best Sound Mixing                | 1                  | Best Sound Mixing                |
| 11 | 2003 | Academy Award    | Best Visual Effects              | 1                  | Best Visual Effects              |
| 12 | 2004 | MTV Movie Awards | Best Picture                     | 1                  | Best Picture                     |
| 13 | 2004 | MTV Movie Awards | Best Action Sequence             | 1                  |                                  |
| 14 |      |                  |                                  | Total Pelicula: 13 |                                  |

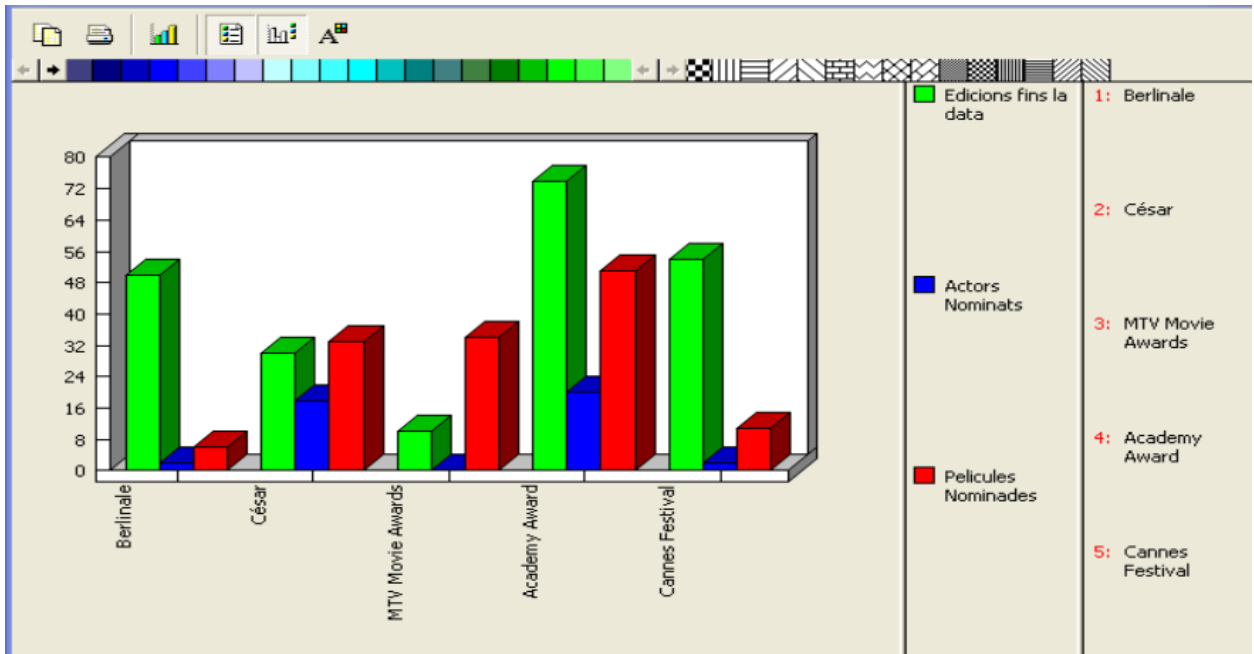
- Informe que permet sobre una determinada data (edició) indicar la següent informació des de la primera edició del festival fins a la data seleccionada:
  - Total edicions de festival fins a la data seleccionada.
  - Total actors nominats.
  - Total actrius nominades.
  - Total pel·lícules nominades

L'usuari final podrà seleccionar introduir una data per Pantalla per seleccionar la informació corresponent. En el cas del exemple la data seleccionada es el 2001.

## Informe 5: Num Edicions, Actors Nominats , Pelicules Nomiades per Festival i Any

Elementos de Página: Any: 2001 ▼

|   | Nom              | Edicions fins la data | Actors Nominats | Pelicules Nominades |
|---|------------------|-----------------------|-----------------|---------------------|
| 1 | Berlinale        | 50                    | 2               | 6                   |
| 2 | César            | 30                    | 18              | 33                  |
| 3 | MTV Movie Awards | 10                    |                 | 34                  |
| 4 | Academy Award    | 74                    | 20              | 51                  |
| 5 | Cannes Festival  | 54                    | 2               | 11                  |



- Indicar un informe global sobre els premis per cada pel·lícula i festival, indicant el total de premis guanyats per pel·lícula i el subtotal per any. La informació a incloure a l'informe es la següent:
  - Pel·lícula
  - Festival
  - Total Any
  - Total Global

## INFORME 6: Premis per Pelicula Any, i Festival

Elementos de Página:

|    | Pelicula                       | Festival        | Any  | Total Pelicula |
|----|--------------------------------|-----------------|------|----------------|
| 87 | ALL ABOUT EVE                  | Academy Award   | 1950 | 6              |
| 88 |                                | Cannes Festival | 1951 | 2              |
| 89 |                                |                 |      | <b>Suma: 8</b> |
| 90 | ALL ABOUT MY MOTHER            | Cannes Festival | 1999 | 1              |
| 91 |                                |                 |      | <b>Suma: 1</b> |
| 92 | ALL GOOD CITIZENS              | Cannes Festival | 1969 | 1              |
| 93 |                                |                 |      | <b>Suma: 1</b> |
| 94 | ALL QUIET ON THE WESTERN FRONT | Academy Award   | 1930 | 2              |
| 95 |                                |                 |      | <b>Suma: 2</b> |
| 96 | ALL THAT JAZZ                  | Academy Award   | 1979 | 4              |
| 97 |                                | Cannes Festival | 1980 | 1              |
| 98 |                                |                 |      | <b>Suma: 5</b> |

- Informe del País (sense tenir en compte els Estats units) amb mes premis que ha guanyat per any, La informació a llistar per aquest país es la següent:
  - País
  - Nombre de premis

S'extraurà la informació de la taula PremisPel·lícula, es calcularà el total de premis per Pel·lícula, a través de la Pel·lícula podem saber el País o països al que esta associada

### País amb mes premis (Sense USA)

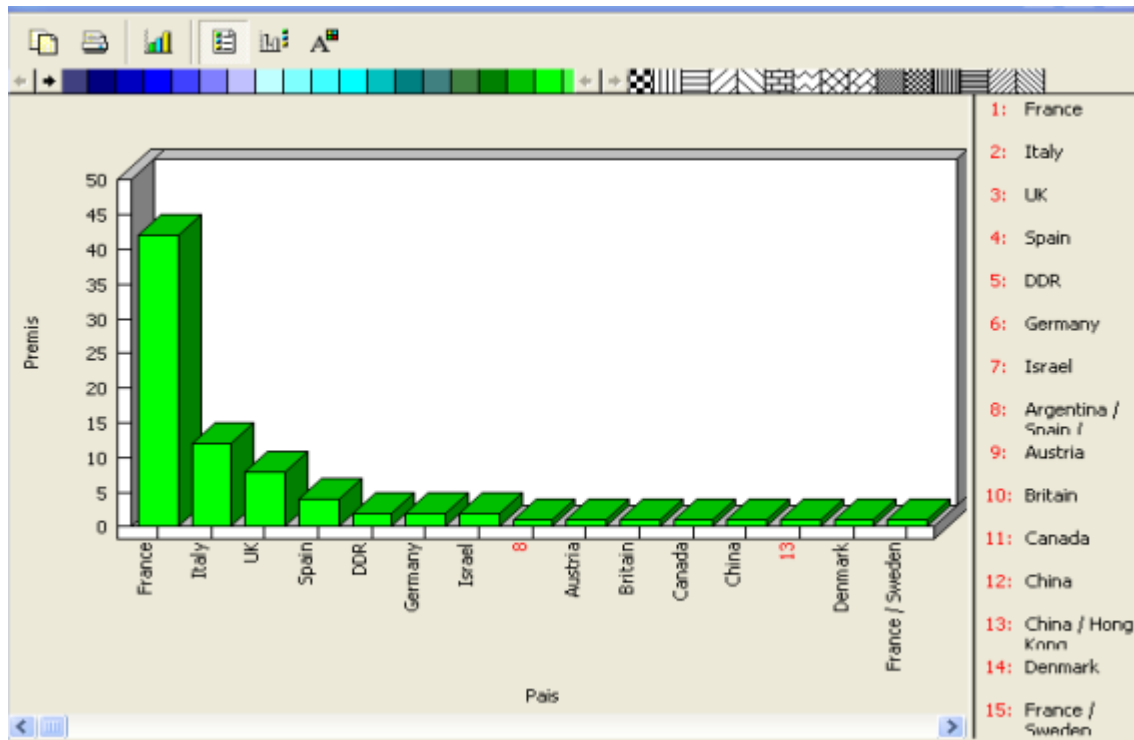
Elementos de Página:

|   | País    | Total Premis |
|---|---------|--------------|
| 1 | France  | 699          |
| 2 | Italy   | 103          |
| 3 | UK      | 79           |
| 4 | Germany | 51           |
| 5 | Spain   | 41           |
| 6 | USSR    | 39           |
| 7 | Japan   | 33           |
| 8 | Sweden  | 31           |

### País amb mes premis (Sense USA) - Categoria

Elementos de Página:  : Categoria: Best Actor ▼

|         | Total Premis |
|---------|--------------|
| ...     |              |
| France  | 42           |
| Italy   | 12           |
| UK      | 8            |
| Spain   | 4            |
| DDR     | 2            |
| Germany | 2            |
| Israel  | 2            |



- Pel·lícula Exitosa

Per extreure la pel·lícula més exitosa utilitzarem el nombre total de premis d'una pel·lícula comparat amb el nombre de nominacions. Aplicarem el següent càlcul (Numero de Premis/Numero de Nominacions) ponderat amb el nombre de premis.

| PELICULA EXITOSA FRACASA          |                    |         |
|-----------------------------------|--------------------|---------|
| Elementos de Página: Anjo: 2006 ▼ |                    |         |
|                                   | Pelicules          | EXITOSA |
| 1                                 | DEPARTED           | 4       |
| 2                                 | INCONVENIENT TRUTH | 3       |
| 3                                 | BROKEBACK MOUNTAIN | 3       |
| 4                                 | WEDDING CRASHERS   | 3       |
| 5                                 | PAN'S LABYRINTH    | 3       |
| 6                                 | WEST BANK STORY    | 2       |
| 7                                 | HAPPY FEET         | 2       |

- Pel·lícula Fracassada.

Per extreure la pel·lícula mes exitosa utilitzarem el nombre total de premis d'una pel·lícula comparat amb el nombre de nominacions. Aplicarem el següent càlcul  $\text{Numero de Premis} - \text{Numero de Nominacions}$ .

| PELICULA EXITOSA FRACASA          |                    |           |
|-----------------------------------|--------------------|-----------|
| Elementos de Página: Anjo: 2006 ▼ |                    |           |
|                                   | Pelicules          | FRACASADA |
| 1                                 | BABEL              | -6,00     |
| 2                                 | DREAMGIRLS         | -6,00     |
| 3                                 | BLOOD DIAMOND      | -5,00     |
| 4                                 | QUEEN              | -5,00     |
| 5                                 | NOTES ON A SCANDAL | -4,00     |
| 6                                 | 40-YEAR OLD VIRGIN | -4,00     |
| 7                                 | CHILDREN OF MEN    | -3,00     |

La Pel·lícula mes exitosa es la que te el valor mes alt y la mes Fracassada el valor mes baix (en negatiu) Diferenciat en dues columnes diferents.

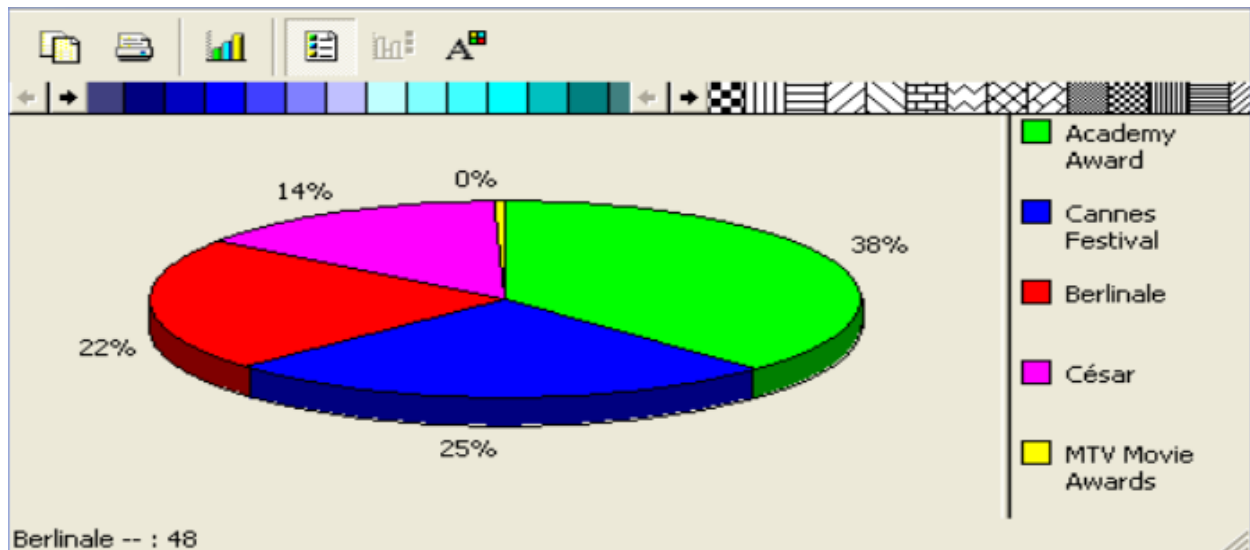
- Relació entre guanya un mateix premi a un Festival o a un altre.

Per tal de donar una relació entre guanya un premi a un Festival o un altre, s'ha creat un Informe on s'indica el nombre de premis atorgats per Festival i Categoria, per tant si un Festival ha atorgat 81 premis de Millor Actor aquest Festival tindrà més importància. També podem guiar-nos per el nombre de edicions de cada un dels festivals mostrats a l'informe .

### Premis Otorgats per Categoria y Festival

Elementos de Página: Categoría: Best Actor ▼

|   | Nom              | Total Premis |
|---|------------------|--------------|
| 1 | Academy Award    | 81           |
| 2 | Cannes Festival  | 54           |
| 3 | Berlinale        | 48           |
| 4 | César            | 31           |
| 5 | MTV Movie Awards | 1            |





## 5.4 Conclusions

El Magatzem de dades dissenyat es fàcil de fer servir, i permet als usuaris obtenir els informes requerits per l'ACA amb un bon rendiment. El projecte ajudarà a que la Acadèmia de cinema andorrà pugui disposar de les dades agrupades en una sola font des de la qual podrà realitzar els anàlisis que siguin oportuns.

A nivell personal, aquest projecte m'ha permès acabar d'adinar-me dins el món del magatzem de dades assolint una bona comprensió sobre els models multidimensionals i les diferències amb els models d'entitat relació habituals. També he pogut practicar els meus coneixements amb PL/SQL dissenyant els processos de carrega, així com aprendre a utilitzar les eines OLAP de oracle.

## 5.6 Línies d'evolució

La principal línia d'evolució en aquest projecte ha de ser millorar les dades d'origen per poder millorar els processos de carrega. Una de les tasques necessàries seria que el ACA disposa's d'un fitxer complet amb les diferents equivalències d'una categoria a diferents festivals així com la unificació dels noms dels països, ja que en la actualitat es troben en diversos idiomes.

També s'han d'unificar els separadors d'informació, en ocasions s'estan utilitzant guions (,) i d'altres barres (/).

Una evolució que permetria donar una major accessibilitat a les dades, seria la publicació dels resultats dels informes a un entorn Web on estiguin disponibles per a presentar els treballs de l'ACA.

## 6 Glossari

### Magatzem de dades

Sistema de bases de dades que agrupa la informació històrica d'una organització, el magatzem de dades té una estructura específica per a realitzar-hi consultes més eficientment

### Sistema de gestió de bases de dades

Software específic dedicat a oferir una interfície amigable per interactuar amb les bases de dades.

### SQL

Llenguatge estructurat, emprat per a crear, modificar i recuperar informació de la bases de dades.

### PL/SQL

Llenguatge de programació basant amb SQL, incorpora totes les funcions d'aquest programa afegint altres sentències i controls de variables així com estructures modulars i de control de flux (sentències for, while, if), amb aquest llenguatge també es pot fer un control de errors mitjançant l'ús de excepcions.

### ETL

Prové de les sigles de l'anglès Extract, transform and load, es el process de extracció, transformació i carrega de dades, en el qual es realitzen operacions de unificació, neteja de dades i finalment carrega a bases de dades.

### OLAP

Prové de les sigles de l'angles On-Line Analytical Processing, Es un tipus de solució emprada per a realitzar càlculs analítics sobre magatzems de dades.

### Taula de dimensions

Taula on s'emmagatzema els registres relacionats amb una taula de fets, es pot dir que es el punt de vista des de que s'analitza un fet.

### Taula de fets

Taula on es registren els fets que es volen analitzar així com les relacions establertes amb aquest fet.

### Estrella

Organització o forma en la que es disposen les dades en un magatzem de dades, en el centre es troba la taula de fets i als extrems les taules de dimensions que estan relacionades amb el fet.

### ACA

Acadèmia de cinema Andorra

## 7 Bibliografia

- Document Gestió i desenvolupament de projectes (materials UOC).
- Introducció a l'emmagatzematge de dades( materials UOC).
- Disseny multidimensional (materials UOC).
- Us dels magatzems de dades (materials UOC).
- [www.oracle.com](http://www.oracle.com)
- [http://en.wikipedia.org/wiki/Data\\_warehouse](http://en.wikipedia.org/wiki/Data_warehouse)
- Manuals Oracle Discoverer: <http://www.oracle.com/technetwork/developer-tools/discoverer/documentation/discoverer-090649.html>