

DATA WAREHOUSE



Trabajo de Final de Carrera Memoria Final

CONSTRUCCIÓN Y EXPLOTACIÓN
DE UN ALMACÉN DE DATOS
DE FESTIVALES CINEMATOGRAFICOS

Agradecimientos

Gracias. Gracias a todas aquellas personas que a lo largo de mi etapa formativa en la UOC me han apoyado, incluidos compañeros y consultores, facilitando así el poder compaginar la vida labora, la personal y la formativa.

Gracias. Gracias a mis amigos. Gracias a todos aquellos que siempre se han ocupado por encontrar un hueco y adaptarse a mis horarios.

Gracias. Gracias a mi familia. Gracias a mis padres y mi hermana por entenderme, por comprender que en muchas ocasiones no he podido estar.

Gracias. Gracias a Cristina y Coral. Gracias por haber sabido sufrir conmigo. Gracias por haber soportado los inconvenientes que mi dedicación a la carrera han podido conllevar. Gracias por esos besos de buenas noches cuando me quedaba delante del ordenador. Gracias por esos besos de buenos días cuando despertaba delante de la pantalla. Gracias. Simplemente, gracias.

TFC – Almacén de Datos

Este documento representa la memoria del trabajo de final de carrera (TFC), de los estudios de Ingeniería Técnica en Informática de Sistemas, elaborado dentro del área de Almacenes de Datos. En él se incluye una descripción detallada de los procedimientos seguidos a lo largo del TFC, así como los productos obtenidos durante las fases por las que el mismo evoluciona: estudio preliminar, planificación, análisis, diseño, implementación y explotación de un Almacén de Datos a partir del enunciado proporcionado en el aula, de la Universitat Oberta de Catalunya (UOC), correspondiente al área indicada. A lo largo del proyecto se pretende transmitir que el estudiante es capaz de utilizar los recursos y conocimientos adquiridos durante su etapa en los estudios comentados, siendo el principal objetivo obtener experiencia en el diseño, construcción y explotación de un Almacén de Datos a partir de la información disponible en una base de datos transaccional. Tras el detalle de los procedimientos seguidos para la elaboración del TFC y la presentación de los productos elaborados durante el mismo, se exponen las conclusiones obtenidas por el alumno, así como una serie de consejos que permitan la continuidad en la explotación del Almacén de Datos.

Índice

Agradecimientos	1
TFC – Almacén de Datos	1
Índice	2
Índice de figuras	3
Introducción	4
Justificación del TFC	4
Objetivos del TFC	4
Enfoque y método seguido	4
Planificación del proyecto	5
Breve descripción de los otros capítulos de la memoria	9
Análisis	10
Necesidades de la ACA	10
Datos de partida	11
Elementos del análisis multidimensional	11
Diagrama modelo conceptual	12
Diseño	15
Diseño lógico	15
Diseño físico	15
Implementación y explotación	20
Conexión y creación de tablas	20
Extracción, Transformación y Carga	24
Informes	40
Capturas de pantalla	45
Planificación	45
SQL Loader	49
Informes solicitados por la ACA	52
Conclusiones	57
Líneas de evolución futura	57
Glosario	58
Bibliografía	58

Índice de figuras

Fig.1: Tablas propias del Almacén de Datos	13
Fig.2: Tablas de apoyo para transformación y carga	14
Fig.3: Planificación del proyecto (parte1)	45
Fig.4: Planificación del proyecto (parte2)	46
Fig.5: Diagrama de Gantt (parte1)	47
Fig.6: Diagrama de Gantt (parte2)	48
Fig.7: Resultado carga fichero ACADEMY	49
Fig.8: Resultado carga fichero BERLINALE	50
Fig.9: Resultado carga fichero CANNES	50
Fig.10: Resultado carga fichero CESARS	51
Fig.11: Resultado carga fichero MTV	51
Fig.12: Resultado primer informe en primer formato	52
Fig.13: Resultado primer informe en segundo formato	52
Fig.14: Resultado segundo informe en primer formato	52
Fig.15: Resultado segundo informe en segundo formato	52
Fig.16: Resultado tercer informe en primer formato	53
Fig.17: Resultado tercer informe en segundo formato	53
Fig.18: Resultado cuarto informe	53
Fig.19: Pantalla de indicación de parámetro	54
Fig.20: Resultado quinto informe	54
Fig.21: Resultado sexto informe en ambos formatos	54
Fig.22: Resultado primer informe adicional	55
Fig.23: Resultado segundo informe adicional en segundo formato	55
Fig.24: Resultado segundo informe adicional en segundo formato	55
Fig.25: Consulta y datos obtenidos para tercer informe adicional	56
Fig.26: Análisis para tercer informe adicional	56

Introducción

Justificación del TFC

La Academia de Cine Andorrano (ACA) realiza, entre otras, tareas estadísticas y de análisis de la información sobre los festivales más importantes del sector realizados en los últimos años. Actualmente la ACA se ve obligada a realizar estas tareas de forma manual sobre varias hojas de Excel en las que almacena los datos recopilados tras cada festival.

Con la finalidad de optimizar dichas tareas, nos han encargado la creación y diseño de un nuevo sistema de información que gestione dichos datos, tanto los que ya se poseen como los que pudieran añadirse en un futuro próximo.

Objetivos del TFC

Todo Trabajo de Final de Carrera se caracteriza por una serie de objetivos generales. Los mismos, personalizados para este TFC de Almacén de Datos, son los siguientes:

- Analizar con detalle los requisitos para el diseño y uso del Almacén de Datos
- Planificar el proyecto de creación del Almacén de Datos
- Establecer las especificaciones necesarias para el diseño del Almacén de Datos
- Especificar los procesos de Extracción, Transformación y Carga de los datos iniciales a fin de poblar el Almacén de Datos
- Analizar y seleccionar las herramientas necesarias para la administración del Almacén de Datos
- Analizar y seleccionar las herramientas necesarias para la explotación del Almacén de Datos
- Analizar y especificar el contenido de los informes solicitados por la ACA
- Implementar el Almacén de Datos, permitiendo dar respuesta a las necesidades expresadas por la ACA

Alcanzando estos objetivos, se demuestra que el estudiante ha logrado los conocimientos necesarios de planificación, análisis, diseño, implementación y explotación de un Almacén de Datos durante su periodo de formación en la UOC, lo cual le permitirá aplicarlos en el mundo laboral.

Enfoque y método seguido

El enfoque que se ha pretendido dar para realizar este TFC, así como el método seguido, es el conocido como ciclo de vida clásica o en cascada, en el que se establecen las fases del mismo y cada una de ellas marca el punto de partida en el que se basa la siguiente. Las fases en las que se ha dividido este proyecto son las siguientes:

- Planificación y análisis previo de requisitos
- Análisis detallado de requisitos y diseños conceptual y técnico
- Implementación y explotación

Planificación del proyecto

Calendario laboral

Dentro de los plazos del proyecto, que tiene como fecha de inicio el 2 de Marzo de 2011 y como fecha de entrega de la documentación final el 13 de Junio de 2011, se ha establecido una jornada laboral de 20 horas semanales repartidas de la siguiente forma:

- Lunes, miércoles y viernes de 20:00 a 0:00
- Sábados de 10:00 a 13:00 y de 16:00 a 21:00

Dentro del calendario marcado se establecen las siguientes excepciones:

- Del 6 al 12 de Marzo de 2011, el horario de dedicación es de 22:00 a 0:00 debido a un viaje por motivos laborales
- Del 29 de Abril al 1 de Mayo de 2011, el horario de dedicación es nulo, debido a un viaje programado
- Domingos 5 y 12 de Junio de 2011, de 10:00 a 13:00 y de 16:00 a 21:00, debido a la cercanía de la entrega final

Recursos humanos

El equipo que ha elaborado este proyecto ha sido formado por Víctor Manuel Naranjo Ferro, estudiando de Ingeniería Técnica en Informática de Sistemas. Asimismo se ha recibido apoyo tanto del consultor asignado por la Universitat Oberta de Catalunya (UOC), Pere Juanola Juanola, como de los compañeros de las aulas de teoría y laboratorio.

Recursos tecnológicos

El equipo utilizado para la elaboración de este proyecto ha sido compuesto por un Ordenador portátil Toshiba Tecra S5, como equipo principal, un Ordenador netbook Hp Mini 210, como equipo de reserva, un dispositivo USB de 16Gb y un Disco duro externo de 1,5Tb, ambos utilizados para realizar copias de seguridad.

Asimismo se ha contado con el soporte del Campus Virtual y sus componentes, como los foros o biblioteca virtual, ofrecidos por la UOC.

Compatibilidades y especificaciones técnicas

El sistema proporcionado a la ACA es compatible, al menos, con Windows XP. Dicho Sistema Operativo es bajo el que corre la máquina virtual de VirtualBox proporcionada por la UOC y en la que se han realizado las tareas que se detallan más adelante.

Las herramientas utilizadas para la realización de dichas tareas, se basan principalmente en software de Oracle (Oracle 10g, Oracle Developer, Oracle PL/SQL y Oracle Discoverer) como SGBD, y en lenguaje PL/SQL para la elaboración de scripts de transformación y carga de datos.

Como herramientas de apoyo se hace uso de software de Microsoft, como Visio, Excel y Project.

Documentación inicial

De forma inicial se ha contado con el enunciado proporcionado por la UOC, compuesto por la redacción del mismo y una serie de archivos en los que se ubicaban los datos a tratar en el proyecto de creación y explotación del Almacén de Datos, así como la descripción de la asignatura a través de su Plan Docente. De forma adicional, durante el desarrollo del proyecto, se ha contado con información específica de cada fase en la que el mismo se encontraba.

Fechas clave

Las fechas clave indicadas por la UOC al inicio del proyecto, y sobre las que se ha trabajado a la hora de establecer la planificación del mismo, son las siguientes:

- PAC1: del 5-Mar'11 al 16-Mar'11
- PAC2: del 17-Mar'11 al 20-Abr'11
- PAC3: del 21-Abr'11 al 25-May'11
- Entrega Final: del 26-May'11 al 13-Jun'11

Cabe destacar que en la planificación seguida las fechas finales a partir de la PAC2 se han adelantado un día a fin de disponer del mismo como margen, si bien esta situación no se ha considerado necesaria en la primera entrega debido a la menor duración de la PAC1.

Incidencias y riesgos

Durante la planificación del proyecto se tuvieron en cuenta diferentes situaciones que podrían haberse dado a lo largo de la elaboración del mismo, las cuales finalmente no se han producido lo cual ha permitido seguir la planificación según lo establecido al inicio del proyecto. Las situaciones consideradas fueron las siguientes:

- Problemas por exceso de carga de trabajo en otra asignatura: por posibles coincidencias de fechas de entrega, proponiendo como solución la priorización del TFC
- Problemas de salud: por posibles enfermedades contraídas, proponiendo como solución intentar ir avanzado a la planificación prevista
- Problemas técnicos: que pudieran afectar al equipo principal, proponiendo como solución disponer de un segundo equipo
- Viajes no previstos: por motivos laborales, proponiendo como solución disponer del segundo equipo y el avance de las tareas respecto a la planificación prevista

Descripción y planificación de las tareas

Las tareas llevadas a cabo a lo largo de la elaboración del proyecto se detallan a continuación, separadas por las fases que el mismo pasa:

- Inicio del proyecto: 2 al 5 de Marzo de 2011
 - o Descarga del material de la asignatura, lectura del plan docente, presentación en el foro y confirmación de consultor
 - o Preparación hardware y software inicial
 - o Búsqueda preliminar de información web
- Planificación y análisis previo de requisitos: 5 al 16 de Marzo de 2011
 - o Descarga y lectura del enunciado del TFC
 - o Toma de contacto y comprensión de los datos proporcionados (ficheros Excel)
 - o Descarga y lectura del modelo ejemplo
 - o Elaboración del borrador del documento “Análisis de datos”
 - o Elaboración del borrador del documento “Documento de especificaciones”
 - o Entrega de los borradores elaborados hasta el momento
 - o Elaboración del borrador del documento “Planificación del proyecto”
 - o Entrega del borrador “Planificación del proyecto”
 - o Lectura de los comentarios del consultor sobre los borradores entregados
 - o Inicio de la lectura del material extraordinario suministrado (MMMD)
 - o Elaboración del documento de “Análisis de datos”
 - o Elaboración del documento de “Documento de especificaciones”
 - o Elaboración del documento de “Planificación del proyecto”
 - o Entrega de los documentos generados en esta fase
- Análisis detallado de requisitos y diseños conceptual y técnico: 18 de Marzo de 2011 al 19 de Abril de 2011

- Descarga y lectura del enunciado de la PAC2
- Instalación y puesta en marcha de la Máquina Virtual proporcionada
- Continuación de la lectura del material extraordinario suministrado (MMMD)
- Consulta bibliografía sobre “Análisis Multidimensional”
- Inicio del proceso de limpieza y depuración de los datos iniciales
- Elaboración del borrador del documento “Análisis detallado y diseño del Almacén de Datos”
- Entrega del borrador generado
- Lectura de los comentarios del consultor sobre el borrador generado
- Elaboración del documento de “Análisis detallado y diseño del Almacén de Datos”
- Entrega del documento generado en esta fase
- Implementación: 22 de Abril del 2011 al 24 de Mayo de 2011
 - Descarga y lectura del enunciado de la PAC3
 - Finalización del proceso de limpieza y depuración de los datos iniciales
 - Creación de tablas, procedimientos y disparadores
 - Carga de datos
 - Instalación de la herramienta de explotación de datos
 - Construcción de los informes detallados en el enunciado del TFC
 - Análisis de la información
 - Elaboración del boceto del documento “Implementación del Almacén de Datos”
 - Entrega del borrador generado
 - Lectura de los comentarios del consultor sobre el borrador generado
 - Pruebas sobre los informes creados
 - Resolución de posibles incidencias en los informes inciales
 - Elaboración del documento de “Implementación del Almacén de Datos”
 - Entrega del documento generado en esta fase
- Entrega final: 27 de Mayo de 2011 al 12 de Junio de 2011
 - Elaboración del boceto del documento final
 - Entrega del borrador generado
 - Lectura de los comentarios del consultor sobre el borrador generado
 - Elaboración del documento final
 - Elaboración de la presentación multimedia
 - Entrega del TFC (documento final y presentación multimedia)

- Tribunal: 15 al 18 de Junio de 2011
 - o Recepción y resolución de las cuestiones propuestas por el Tribunal de Evaluación

El nombre de cada una de las tareas descritas ha sido seleccionado con la finalidad de que resulte lo suficientemente descriptivo como para no precisar explicación alguna de su contenido. Asimismo se ha realizado, en los momentos necesarios, tarea de investigación a fin de definir o aclarar conceptos, así como procedimientos a seguir en algunas de ellas.

La planificación detallada y los diagramas de Gantt generados durante la planificación del proyecto pueden observarse en el apartado Capturas de pantalla de este documento.

Documentación y productos obtenidos

A lo largo de la vida del presente proyecto, se ha generado una serie de documentación en cada una de sus fases, la cual se detalla a continuación:

- Planificación y análisis previo de requisitos
 - o Análisis de los datos iniciales: estudio previo de los datos proporcionados inicialmente
 - o Documento de especificaciones: informe en el que se detalla las especificaciones sobre las que se desarrolla el proyecto
 - o Planificación del proyecto: detalle con la planificación prevista para la realización del proyecto
- Análisis detallado de requisitos y diseños conceptual y técnico: análisis exhaustivo de los requisitos y descripción del diseño, tanto lógico como físico, del Almacén de Datos. Asimismo se incluye una descripción de los procesos ETL y los elementos asociados a ellos.
- Implementación: detalle de la implementación de los elementos recogidos en la fase anterior y explotación del Almacén de Datos facilitando los informes solicitados por la ACA, incluyendo el producto físico integrado en el gestor Oracle.

Breve descripción de los otros capítulos de la memoria

Tras esta primera toma de contacto con el proyecto a desarrollar, en los siguientes capítulos de la presente memoria se describirán las acciones que se han realizado para la obtención de los resultados requeridos por la ACA. En un primer momento se realiza un análisis que permita hacerse una idea de lo que se tiene y lo que se espera obtener, pudiendo valorar lo que será necesario para lograrlo, lo que da como resultado el modelo conceptual del Almacén de Datos. Tras ello, y partiendo de la información analizada, se genera el diseño del Almacén de Datos, tanto lógico como físico, que

permita posteriormente la implementación y explotación del mismo, creando todos sus elementos, poblándolos de los datos previamente tratados y generando una serie de informes solicitados por la ACA. Una vez presentado por completo el proyecto en sí, se agrupan las capturas de pantalla en un apartado a fin de facilitar, y complementar, la comprensión de los conceptos expuestos a lo largo de los capítulos anteriores. Finalmente, se manifiestan una serie de conclusiones y se exponen las acciones a tomar en un futuro para ofrecer continuidad al proyecto.

Análisis

Necesidades de la ACA

La principal inquietud expresada por la ACA es la necesidad de explotar de forma eficiente los datos de los que dispone, actualmente estructurados en hojas de Excel.

El objetivo principal para cumplir con las funciones que nos ha encargado la ACA es la creación de un Almacén de Datos sobre el que se puedan obtener informes predefinidos que muestren los valores estadísticos que la misma requiere, con la posibilidad de adaptar nuevos informes que pudieran necesitarse en un futuro.

Para facilitar el alcance de dicho objetivo, el mismo puede fragmentarse en objetivos de menor complejidad, los cuales serán más fáciles de alcanzar. Estos objetivos, a priori, pueden ser los siguientes:

- Objetivo inicial: Analizar los datos de forma que puedan detectarse posibles errores en los mismos, así como necesidades de transformación sobre ellos para conseguir homogeneidad.
- Segundo objetivo: Crear un proceso automatizado que, con la ayuda de alguna herramienta ETL, permita importar la información relevante de los datos de las diferentes hojas de Excel que la ACA pone a nuestra disposición.
- Tercer objetivo: Diseñar un almacén de datos en el que puedan ser insertados los datos a través del proceso automatizado generado anteriormente y que permita aprovecharlos para realizar los informes estadísticos del siguiente objetivo.
- Cuarto objetivo: Diseñar los informes solicitados, así como proporcionar una herramienta eficiente, y las líneas generales de actuación sobre la misma, que permita generar en un futuro posibles informes que puedan precisarse más adelante.

Datos de partida

Los datos iniciales se proporcionan a partir de varios ficheros Excel entregados por la ACA. Los mismos han debido tratarse al inicio del proceso de carga, tal y como se indica más adelante en este mismo informe.

Cada fichero Excel contiene los datos correspondientes a un festival. En cada uno de ellos, se indican los premios y nominaciones para cada una de las categorías de los diferentes festivales. Los campos que en ellos figuran, junto con el tipo de cada campo, pueden observarse a continuación:

- Original Title / Film Title (título original): texto
- English Title / Localized Title (título utilizado en el festival): texto
- Sort title (ordenación del título): texto
- Year (año): número
- Country (país): país nativo de la película
- Award (festival): texto
- Category (categoría): texto
- Winner (si el título resultó agraciado en dicha categoría): texto
- Nominees/s (nominados, en caso de tratarse de premios individuales a actores/actrices/directores...): texto
- _xxx, donde "xxx" representa el nombre del festival (edición del festival): texto
- Item (identificador único del registro dentro del propio fichero): número

Elementos del análisis multidimensional

Analizando los ficheros Excel proporcionados por la ACA, así como los informes solicitados por la misma, se han podido deducir las siguientes definiciones:

- Hechos: el objeto central a analizar son las nominaciones a las diferentes categorías de los festivales
- Medidas: el hecho nominación tiene como medida la concesión del galardón, por lo que la misma representa si la nominación a conseguido el premio en su categoría o no
- Dimensiones: las dimensiones, en función a los informes solicitados, son la película, el país, la categoría, la época (tiempo/año), el festival y la persona/película nominada
- Descriptores: para cada una de las dimensiones se pueden distinguir los siguientes atributos:
 - o Película: título original, título festival, título unificado
 - o País: nombre del país
 - o Categoría: nombre original, nombre unificado, destino
 - o Tiempo: año
 - o Festival: nombre festival

- Persona/Película nominada: nombre del nominado
- Niveles: con las indicaciones recibidas por la ACA no se han estimado niveles, ya que ninguno de los descriptores es sensible a organizarse en jerarquía, si bien en el caso del descriptor Tiempo sería posible descomponerlo en Año, Mes y Día. La misma particularidad podría suceder con los países, que podrían organizarse en continentes. No obstante, según los datos de los que se dispone y los informes solicitados no se estima la aparición de niveles.

Diagrama modelo conceptual

La confección del diseño conceptual del Almacén de datos está compuesta también por diferentes fases. Las mismas, así como las conclusiones de cada una de ellas, se detallan a continuación:

- Identificar el hecho: el objeto central a analizar son las nominaciones
- Encontrar la granularidad adecuada: el detalle necesario para cumplir con los requisitos que la ACA solicita, debe ser el mismo que ofrece en los documentos Excel es decir, un registro por nominación (película, festival, año, categoría y nominado)
- Escoger las dimensiones: las dimensiones que se vinculan al hecho para dar forma al mismo son película, país, categoría, año, festival y nominado.
- Establecer los atributos de cada dimensión: para cada una de las dimensiones indicadas, ha sido preciso definir los atributos suficientes para que las mismas quedaran identificadas
 - Película: código, título original y título festival
 - País: código y nombre país
 - Categoría: código, nombre y destino
 - Año: número de año
 - Festival: código y nombre
 - Nominado: código y nombre
- Distinguir entre descriptores y jerarquías de agregación: en este caso concreto no se han estimado jerarquías, si bien el tipo de dato es susceptible a tenerlas en caso de considerarlo así la ACA
- Identificar las medidas: las medidas del hecho queda formado por el atributo “premiado”
- Definir las celdas: atributos que conforman el hecho (premiado)
- Explicitar las restricciones de integridad: además de la que forma la clave primaria del hecho (película, festival, año, categoría, nominado), la formada por la relación que vincula siempre a una película a un determinado país.
- Tablas de apoyo: como tablas de apoyo para la transformación y carga de los datos desde las hojas Excel proporcionadas por la ACA hasta el Almacén de Datos, se dispone de una tabla por fichero Excel (Festival), junto con tres tablas que establecerán la relación entre los diferentes idiomas en los que figuran

todos los países involucrados, la unificación entre las categorías de los festivales y la vinculación entre el año y la edición del festival Cesars, en el que se ha detectado durante el análisis varios errores en dicha vinculación, los cuales serán tratados más adelante durante el proceso de transformación.

- Estudio de viabilidad: se debe determinar si las conclusiones a las que se llega en cada una de las fases anteriores permite la viabilidad del proyecto, que en este caso sí es posible, tal y como se resume en la tabla incluida a continuación en la que se realiza una estimación de los registros de cada tabla y el espacio ocupado por cada uno de ellos.

Resumen espacio estimado Data Warehouse (bytes)			
Tabla	Registros	Bytes/reg.	Bytes/tabla
Premios	13279	26	345.254,00
Películas	6635	36,44	241.779,40
Nominados	8344	22,97	191.661,68
Festivales	5	15,6	78,00
Países	167	9,73	1.624,91
Año	79	4	316,00
Categorías	225	32,08	7.218,00
países_traducción	167	38,38	6.409,46
categorías_traducción	225	150,69	33.905,25
TOTAL DW			828.246,70

* No se cuentan las tablas temporales

Con ello, el diseño conceptual final del diseño del Almacén de Datos es el siguiente:

TABLAS PROPIAS DEL DATA WAREHOUSE:

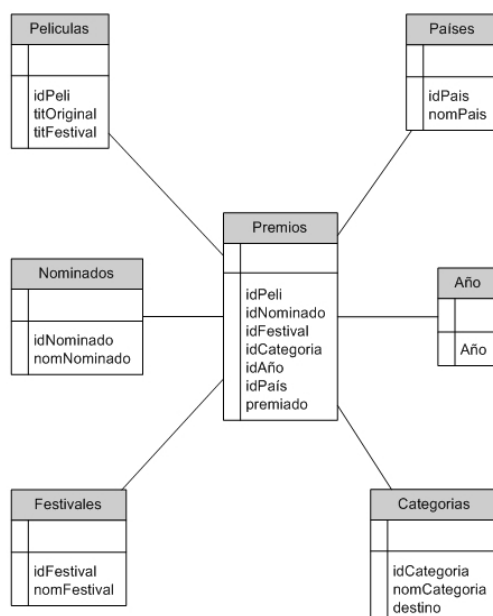


Fig.1: Tablas propias del Almacén de Datos

TABLAS DE APOYO PARA TRANSFORMACIÓN Y CARGA:

tmp_academy	tmp_berlinale	tmp_cannes	tmp_cesars	tmp_mtv
Original_Title English_Title Sort_Title Year Country Award Category Winner Nominee Academy_Awards Item	Film_Title Localized_Title Sort_Title Year Country Award Category Winner Nominee Berlinale Item	Film_Title Localized_Title Sort_Title Year Country Award Category Winner Nominee Cannes_Film_Festival Item	Film_Title Localized_Title Sort_Title Year Country Award Category Winner Nominee Cesars Item	Original_Title Localized_Title Sort_Title Year Country Award Category Winner Nominee Mtv_Movie_Awards Item

países_traduccion	categorias_unificacion	añoedicion
id academy berlinale cannes cesars mtv unificado	id academy berlinale cannes cesars mtv unificado destino	año edicion

Fig.2: Tablas de apoyo para transformación y carga

Con el diseño mostrado se garantiza poder resolver las siguientes cuestiones, solicitadas por la ACA a modo de informes:

- Número de premios conseguidos por un actor/actriz por año y festival
- Número de premios conseguidos por un actor/actriz por año, festival y categoría
- Número de premios conseguidos por un director por año, festival y película
- Número de premios conseguidos por año, festival, categoría y película
- Dada una fecha, para cada festival, mostrar el número de ediciones realizadas del festival, número total de actores que han sido nominados, número total de actrices que han sido nominadas y número total de películas nominadas
- Para cada año, número de premios ganados por cada película y festival

De forma adicional, se puede proporcionar la información necesaria para solventar las inquietudes expresadas por la ACA:

- Para cada año, indicar las películas con mayor éxito y mayor fracaso, según unos criterios que definan cada concepto
- Número de premios por categoría ganados por el país más galardonado, dejando de lado a EE.UU.
- Analizar una posible relación entre ganar un premio en un festival y ganar el mismo, o equivalente, en otro festival.

Diseño

Diseño Lógico

El diseño lógico del Almacén de Datos, obtenido a partir del diseño conceptual elaborado en el apartado anterior, es el siguiente:

Premios(idPeli, idNominado, idFestival, idCategoria, Año, idPais, premiado)
Películas(idPeli, titOriginal, titFestival)
Nominados(idNominado, nomNominado)
Festivales(idFestival, nomFestival)
Categorías(idCategoria, nomCategoria, destino)
Año(año)
Países(idPais, nomPais)

Diseño Físico

Partiendo de los resultados obtenidos anteriormente, el diseño físico del Almacén de Datos es el siguiente:

Tabla PREMIOS

Los atributos que conforman la tabla Premios, son los siguientes:

idPeli	Number(10,0)	NotNull
idFestival	Number(10,0)	NotNull
idCategoria	Number(10,0)	NotNull
idAño	Number(4,0)	NotNull
idNominado	Number(10,0)	NotNull
idPais	Number(10,0)	NotNull
premiado	Varchar2(2)	

La clave primaria está compuesta por los atributos idPeli, idFestival, idCategoria, idAño, idNominado. Además de los mismos, el atributo idPais no puede contener valores nulos.

Tabla PELICULAS

Los atributos que conforman la tabla Peliculas, son los siguientes:

idPeli	Number(10,0)	NotNull
titOriginal	Varchar2(255)	
titFestival	Varchar2(255)	

La clave primaria está compuesta por el atributo idPeli

Tabla NOMINADOS

Los atributos que conforman la tabla Nominados, son los siguientes:

idNominado	Number(10,0)	NotNull
nomNominado	Varchar2(300)	

La clave primaria está compuesta por el atributo idNominado

Tabla FESTIVALES

Los atributos que conforman la tabla Festivales, son los siguientes:

idFestival	Number(10,0)	NotNull
nomFestival	Varchar2(30)	

La clave primaria está compuesta por el atributo idFestival

Tabla CATEGORIAS

Los atributos que conforman la tabla Categorías, son los siguientes:

idCategoria	Number(10,0)	NotNull
nomCategoria	Varchar2(255)	
destino	Varchar2(10)	

La clave primaria está compuesta por el atributo idCategoria

Tabla AÑO

Los atributos que conforman la tabla Año, son los siguientes:

Año	Number(4,0)	NotNull
-----	-------------	---------

La clave primaria está compuesta por el atributo Año.

Tabla PAISES

Los atributos que conforman la tabla Países, son los siguientes:

idPais	Number(10,0)	NotNull
nomPais	Varchar2(255)	

La clave primaria está compuesta por el atributo idPais

Además de las tablas comentadas, son necesarias unas tablas de apoyo en las que se vuelcan los datos proporcionados por la ACA, vaciándolas previamente en cargas de futuras actualizaciones. En estas tablas son en las que se tratan los datos y preparan para la carga posterior en las tablas comentadas anteriormente. Las mismas se detallan a continuación.

Tabla TMP_ACADEMY

Los atributos que conforman la tabla tmp_academy, son los siguientes:

original_title	Varchar2(255)
english_title	Varchar2(255)
sort_title	Varchar2(255)
year	Number(4,0)
country	Varchar2(255)
award	Varchar2(30)
category	Varchar2(255)
winner	Varchar2(2)
nominee	Varchar2(255)
academy_awards	Varchar2(50)
item	Number(10,0)

La clave primaria está compuesta por el atributo ítem.

Tabla TMP_BERLINALE

Los atributos que conforman la tabla tmp_berlinale, son los siguientes:

film_title	Varchar2(255)
localized_title	Varchar2(255)
sort_title	Varchar2(255)
year	Number(4,0)
country	Varchar2(255)
award	Varchar2(30)
category	Varchar2(255)
winner	Varchar2(2)
nominee	Varchar2(255)
berlinale	Varchar2(35)
item	Number(10,0)

La clave primaria está compuesta por el atributo ítem.

Tabla TMP_CANNES

Los atributos que conforman la tabla tmp_cannes, son los siguientes:

film_title	Varchar2(255)
localized_title	Varchar2(255)
sort_title	Varchar2(255)
year	Number(4,0)
country	Varchar2(255)
award	Varchar2(30)
category	Varchar2(255)
winner	Varchar2(2)
nominee	Varchar2(255)
cannes_film_festival	Varchar2(35)
item	Number(10,0)

La clave primaria está compuesta por el atributo ítem.

Tabla TMP_CESARS

Los atributos que conforman la tabla tmp_cesars, son los siguientes:

film_title	Varchar2(255)
localized_title	Varchar2(255)
sort_title	Varchar2(255)
year	Number(4,0)
country	Varchar2(255)
award	Varchar2(30)
category	Varchar2(255)
winner	Varchar2(2)
nominee	Varchar2(255)
cesars	Varchar2(35)
item	Number(10,0)

La clave primaria está compuesta por el atributo ítem.

Tabla TMP_MTV

Los atributos que conforman la tabla tmp_mtv, son los siguientes:

original_title	Varchar2(255)
localized_title	Varchar2(255)
sort_title	Varchar2(255)
year	Number(4,0)
country	Varchar2(255)
award	Varchar2(30)
category	Varchar2(255)

winner	Varchar2(2)
nominee	Varchar2(255)
mtv_movie_awards	Varchar2(35)
item	Number(10,0)

La clave primaria está compuesta por el atributo ítem.

Tabla PAISES_TRADUCCION

Los atributos que conforman la tabla paises_traduccion, son los siguientes:

id	number(10,0)
academy	Varchar2(255)
berlinale	Varchar2(255)
cannes	Varchar2(255)
cesars	Varchar2(255)
mtv	Varchar2(255)
unificado	Varchar2(255)

La clave primaria está compuesta por el atributo id. En esta tabla se almacenan los datos de los países que van apareciendo en los ficheros Excel proporcionados por la ACA, junto a una denominación unificada.

Tabla CATEGORIAS_UNIFICACION

Los atributos que conforman la tabla categorías_unificacion, son los siguientes:

id	number(10,0)
academy	Varchar2(255)
berlinale	Varchar2(255)
cannes	Varchar2(255)
cesars	Varchar2(255)
mtv	Varchar2(255)
unificado	Varchar2(255)
destino	Varchar2(10)

La clave primaria está compuesta por el atributo id. En esta tabla se almacenan los datos de las categorías que van apareciendo en los ficheros Excel proporcionados por la ACA, junto con una denominación unificada y el destino al cual está dirigida la misma.

Tabla AÑOEDICION

Los atributos que finalmente conforman la tabla Añoedicion, son los siguientes:

año	number(4,0)
edicion	Varchar2(255)

En esta tabla se almacenan las relaciones entre el año y la edición que aparece en cada registro del fichero correspondiente al festival Césars, a fin de corregir posibles desviaciones detectadas en el mismo, tal y como se describe más adelante.

Implementación y explotación

Conexión y creación de tablas

Tal y como se ha descrito anteriormente el proyecto se ha llevado a cabo utilizando el motor de base de datos Oracle 10g Express Edition, en el que se ha creado un usuario para la realización del mismo. Los datos de acceso con la cuenta creada son:

- Usuario: PR
- Contraseña: pr

Siguiendo los resultados obtenidos del diseño anteriormente mostrado, las sentencias de creación de las tablas que forman el Almacén de Datos son las siguientes:

Tabla TMP_ACADEMY

```
CREATE TABLE "TMP_ACADEMY"
( "ORIGINAL_TITLE" VARCHAR2(255),
  "ENGLISH_TITLE" VARCHAR2(255),
  "SORT_TITLE" VARCHAR2(255),
  "YEAR" VARCHAR2(30),
  "COUNTRY" VARCHAR2(50),
  "AWARD" VARCHAR2(30),
  "CATEGORY" VARCHAR2(255),
  "WINNER" VARCHAR2(2) NOT NULL ENABLE,
  "NOMINEES" VARCHAR2(300),
  "ACADEMY_AWARDS" VARCHAR2(70),
  "ITEM" NUMBER(10,0),
  CONSTRAINT "PK_TMP_ACADEMY" PRIMARY KEY
( "ITEM" ) ENABLE
)
```

Tabla TMP_BERLINALE

```
CREATE TABLE "TMP_BERLINALE"
( "FILM_TITLE" VARCHAR2(255),
  "LOCALIZED_TITLE" VARCHAR2(255),
  "SORT_TITLE" VARCHAR2(255),
  "YEAR" NUMBER,
  "COUNTRY" VARCHAR2(30),
  "AWARD" VARCHAR2(30),
  "CATEGORY" VARCHAR2(255),
  "WINNER" VARCHAR2(2) NOT NULL ENABLE,
  "NOMINEES" VARCHAR2(255),
  "BERLINALE" VARCHAR2(30),
  "ITEM" NUMBER(10,0),
  CONSTRAINT "PK_TMP_BERLINALE" PRIMARY KEY
("ITEM") ENABLE
)
```

Tabla TMP_CANNES

```
CREATE TABLE "TMP_CANNES"
( "FILM_TITLE" VARCHAR2(255),
  "LOCALIZED_TITLE" VARCHAR2(255),
  "SORT_TITLE" VARCHAR2(255),
  "YEAR" NUMBER,
  "COUNTRY" VARCHAR2(255),
  "AWARD" VARCHAR2(30),
  "CATEGORY" VARCHAR2(255),
  "WINNER" VARCHAR2(2) NOT NULL ENABLE,
  "NOMINEES" VARCHAR2(255),
  "CANNES_FILM_FESTIVAL" VARCHAR2(30),
  "ITEM" NUMBER(10,0),
  CONSTRAINT "PK_TMP_CANNES" PRIMARY KEY
("ITEM") ENABLE
)
```

Tabla TMP_CESARS

```
CREATE TABLE "TMP_CESARS"
( "FILM_TITLE" VARCHAR2(255),
  "LOCALIZED_TITLE" VARCHAR2(255),
  "SORT_TITLE" VARCHAR2(255),
  "YEAR" NUMBER,
  "COUNTRY" VARCHAR2(30),
  "AWARD" VARCHAR2(30),
  "CATEGORY" VARCHAR2(255),
  "WINNER" VARCHAR2(2) NOT NULL ENABLE,
  "NOMINEES" VARCHAR2(255),
  "CESARS" VARCHAR2(30),
  "ITEM" NUMBER(10,0),
  CONSTRAINT "PK_TMP_CESAR" PRIMARY KEY ("ITEM")
ENABLE
)
```

Tabla TMP_MTV

```
CREATE TABLE "TMP_MTV"
( "ORIGINAL_TITLE" VARCHAR2(255),
  "LOCALIZED_TITLE" VARCHAR2(255) NOT NULL
  ENABLE,
  "SORT_TITLE" VARCHAR2(255),
  "YEAR" NUMBER,
  "COUNTRY" VARCHAR2(30) NOT NULL ENABLE,
  "AWARD" VARCHAR2(30),
  "CATEGORY" VARCHAR2(255),
  "WINNER" VARCHAR2(2) NOT NULL ENABLE,
  "NOMINEES" VARCHAR2(255),
  "MTV_MOVIE_AWARDS" VARCHAR2(30),
  "ITEM" NUMBER(10,0),
  CONSTRAINT "PK_TMP_MTV" PRIMARY KEY ("ITEM")
  ENABLE
)
```

Tabla PAISES_TRADUCCION

```
CREATE TABLE "PAISES_TRADUCCION"
( "ID" NUMBER(10,0),
  "ACADEMY" VARCHAR2(50),
  "BERLINALE" VARCHAR2(50),
  "CANNES" VARCHAR2(50),
  "CESARS" VARCHAR2(50),
  "MTV" VARCHAR2(50),
  "UNIFICADO" VARCHAR2(50)
)
```

Tabla CATEGORIAS_UNIFICACION

```
CREATE TABLE "CATEGORIAS_UNIFICACION"
( "ID" NUMBER(10,0),
  "ACADEMY" VARCHAR2(75),
  "BERLINALE" VARCHAR2(75),
  "CANNES" VARCHAR2(75),
  "CESARS" VARCHAR2(75),
  "MTV" VARCHAR2(75),
  "UNIFICADO" VARCHAR2(75),
  "DESTINO" VARCHAR2(10)
)
```

Tabla AÑOEDICION

```
CREATE TABLE "AÑOEDICION"
( "AÑO" NUMBER,
  "EDICION" VARCHAR2(30)
)
```

Tabla PELICULAS

```
CREATE TABLE "PELICULAS"
( "IDPELI" NUMBER(10,0) NOT NULL ENABLE,
  "TITORIGINAL" VARCHAR2(255),
  "TITINGLES" VARCHAR2(255),
  CONSTRAINT "PK_PR_PELICULAS" PRIMARY KEY
( "IDPELI" ) ENABLE
)
```

Tabla NOMINADOS

```
CREATE TABLE "NOMINADOS"
( "IDNOMINADO" NUMBER(10,0) NOT NULL ENABLE,
  "NOMNOMINADO" VARCHAR2(300),
  CONSTRAINT "PK_PR_NOMINADOS" PRIMARY KEY
( "IDNOMINADO" ) ENABLE
)
```

Tabla FESTIVALES

```
CREATE TABLE "FESTIVALES"
( "IDFESTIVAL" NUMBER(10,0) NOT NULL ENABLE,
  "NOMFESTIVAL" VARCHAR2(30),
  CONSTRAINT "PK_PR_FESTIVALES" PRIMARY KEY
( "IDFESTIVAL" ) ENABLE
)
```

Tabla CATEGORIAS

```
CREATE TABLE "CATEGORIAS"
( "IDCATEGORIA" NUMBER(10,0) NOT NULL ENABLE,
  "NOMCATEGORIA" VARCHAR2(255),
  "DESTINO" VARCHAR2(10),
  CONSTRAINT "PK_PR_CATEGORIAS" PRIMARY KEY
( "IDCATEGORIA" ) ENABLE
)
```

Tabla AÑO

```
CREATE TABLE "AÑO"
( "AÑO" NUMBER(4,0),
  CONSTRAINT "PK_PR_AÑO" PRIMARY KEY ( "AÑO" )
ENABLE
)
```


Tabla PAISES

```
CREATE TABLE "PAISES"
( "idPais" NUMBER(10,0),
  "nomPais" VARCHAR2(255),
  CONSTRAINT "PK_PR_PAISES" PRIMARY KEY
  ("idPais") ENABLE
)
```

Tabla PREMIOS

```
CREATE TABLE "PREMIOS"
( "IDPELI" NUMBER(10,0) NOT NULL ENABLE,
  "IDFESTIVAL" NUMBER(10,0) NOT NULL ENABLE,
  "IDCATEGORIA" NUMBER(10,0) NOT NULL ENABLE,
  "IDAÑO" NUMBER(4,0) NOT NULL ENABLE,
  "IDNOMINADO" NUMBER(10,0) NOT NULL ENABLE,
  "IDPAIS" NUMBER(10,0) NOT NULL ENABLE,
  "PREMIADO" VARCHAR2(2),
  CONSTRAINT "PK_PR_PREMIOS" PRIMARY KEY
  ("IDPELI", "IDFESTIVAL", "IDCATEGORIA", "IDAÑO",
  "IDNOMINADO") ENABLE,
  CONSTRAINT "FK_PR_PREMIOS_PELI" FOREIGN KEY
  ("IDPELI")
  REFERENCES "PELICULAS" ("IDPELI") ENABLE,
  CONSTRAINT "FK_PR_PREMIOS_FESTIVAL" FOREIGN
  KEY ("IDFESTIVAL")
  REFERENCES "FESTIVALES" ("IDFESTIVAL")
  ENABLE,
  CONSTRAINT "FK_PR_PREMIOS_CATEGORIA" FOREIGN
  KEY ("IDCATEGORIA")
  REFERENCES "CATEGORIAS" ("IDCATEGORIA")
  ENABLE,
  CONSTRAINT "FK_PR_PREMIOS_AÑO" FOREIGN KEY
  ("IDAÑO")
  REFERENCES "AÑO" ("AÑO") ENABLE,
  CONSTRAINT "FK_PR_PREMIOS_NOMINADO" FOREIGN
  KEY ("IDNOMINADO")
  REFERENCES "NOMINADOS" ("IDNOMINADO")
  ENABLE,
  CONSTRAINT "FK_PR_PREMIOS_PAIS" FOREIGN KEY
  ("IDPAIS")
  REFERENCES "PAISES" ("idPais") ENABLE
)
```

Extracción, Transformación y Carga

Una vez obtenido el diseño físico del Almacén de Datos y creadas las tablas que lo forman, se entra en contacto directo con el material facilitado por la ACA. En primer lugar es necesario preparar los ficheros a fin de extraer los datos a un formato que pueda ser utilizado para realizar el traslado a las tablas temporales, donde se realizará

la transformación de los mismos antes de culminar el proceso con la carga de los datos en las tablas definitivas del Almacén de Datos.

En la primera fase comentada, la extracción de los datos, se han realizado las siguientes acciones:

- En los ficheros correspondientes a los festivales Academy y Cannes se han eliminado columnas vacías a la derecha de los datos, que pudieran provocar confusiones, e incidencias en la carga, ya que se tratarían como cualquier otro campo
- En el fichero correspondiente al festival Academy, además de la acción comentada, se han sustituido los símbolos “;” por símbolos “,”, a fin de mantener la separación correcta entre columnas, puesto que el símbolo “;” será el que identifique en sucesivas fases la separación entre campos
- Se han guardado todos los ficheros Excel, una vez realizadas las dos acciones anteriores, en formato CSV
- Para la unificación de países y categorías que figuran en los ficheros iniciales se han generado nuevas tablas, utilizando el software Microsoft Excel, en las que se identifica la relación entre los diferentes países/categorías en los idiomas de cada fichero a fin de unificarlos. En cada uno de los ficheros figura una columna para cada festival, más una en la que figura el valor unificado para cada registro. En el caso de las categorías, además, se ha añadido una columna (campo) adicional en la que figura el destino de la misma.
- Para la relación entre el año y la edición del festival Cesars, en la que se han detectado errores en los ficheros proporcionados por la ACA, no se ha creado ningún fichero de precarga, puesto que la tabla que contiene dichos valores se irá poblando durante el proceso de transformación en los datos en las tablas temporales.

Una vez finalizadas estas acciones, los datos proporcionados por la ACA ya se encuentran en un formato que puede importarse sin dificultad a las tablas temporales para realizar en ellas la fase de transformación. Esta fase se ha automatizado al máximo mediante el uso de disparadores y procedimientos. Los disparadores implicados en la fase de transformación serán llamados por el SGBD antes de realizar la inserción de cada registro en cada una de las tablas. Los mismos llamarán a su vez a los procedimientos correspondientes a cada tabla (cabe observar que desde cada tabla se llaman únicamente a los procedimientos que son necesarios para la transformación de los datos que en ella se ubican) que realizarán las acciones oportunas sobre los datos, almacenándolos en las tablas temporales ya preparados para su carga posterior al Almacén de Datos. Los disparadores y procedimientos que se utilizan en esta fase son los siguientes:

- Disparador BEFORE sobre tabla TMP_ACADEMY: realiza llamadas a los procedimientos que tratan los campos year, winner, country y category. El código del mismo es el siguiente:

```
create or replace trigger DISP_ACADEMY
BEFORE
insert or update on "TMP_ACADEMY"
for each row
declare
ediciones NUMBER := 0;

begin
proc_año(:new.year);
proc_winner(:new.winner);
proc_pais('academy', :new.country);
proc_categoria('academy', :new.category);
end;
```

- Disparador BEFORE sobre tabla TMP_BERLINALE: realiza llamadas a los procedimientos que tratan los campos winner, nominees, country y category. El código del mismo es el siguiente:

```
create or replace trigger DISP_BERLINALE
BEFORE
insert or update on "TMP_BERLINALE"
for each row
begin

proc_winner(:new.winner);
proc_nominado(:new.film_title, :new.nominees);
proc_pais('berlinale', :new.country);
proc_categoria('berlinale', :new.category);
end;
```

- Disparador BEFORE sobre tabla TMP_CANNES: realiza llamadas a los procedimientos que tratan los campos winner, country, film_title, localized_title, nominees y category. El código del mismo es el siguiente:

```
create or replace trigger DISP_CANNES
BEFORE
insert or update on "TMP_CANNES"
for each row
begin
proc_winner(:new.winner);
proc_pais('cannes', :new.country);
proc_titoriginal(:new.film_title, :new.localized_title);
proc_titfestival(:new.film_title, :new.localized_title);
proc_nominado(:new.film_title, :new.nominees);
proc_categoria('cannes', :new.category);
end;
```

- Disparador BEFORE sobre tabla TMP_CESARS: realiza llamadas a los procedimientos que tratan los campos year (respecto a la relación del mismo con la edición del registro), winner, country, nominees y category. El código del mismo es el siguiente:

```
create or replace trigger DISP_CESARS
BEFORE
insert or update on "TMP_CESARS"
for each row
declare
edi NUMBER := 0;

begin
proc_edicion(:new.cesars, :new.year, edi);
proc_winner(:new.winner);
proc_pais('cesars', :new.country);
proc_nominado(:new.film_title, :new.nominees);
proc_categoria('cesars', :new.category);
end;
```

- Disparador BEFORE sobre tabla TMP_MTV: realiza llamadas a los procedimientos que tratan los campos winner, nominees, country, localized_title y category. El código del mismo es el siguiente:

```
create or replace trigger DISP_MTV
BEFORE
insert or update on "TMP_MTV"
for each row
begin
proc_winner(:new.winner);
proc_nominado(:new.original_title, :new.nominees);
proc_pais('mtv', :new.country);
proc_titfestival(:new.original_title, :new.localized_title);
proc_categoria('mtv', :new.category);
end;
```

- Procedimiento PROC_AÑO: En caso de pasar como parámetro un valor de año con el formato year1-year2, almacena únicamente el segundo año. El código del mismo es el siguiente:

```
create or replace procedure proc_año
(año IN OUT VARCHAR2)
is
begin
IF año IS NOT NULL THEN
  IF LENGTH(año) > 4 THEN
    año := substr(año,6,4);
  END IF;
END IF;
end;
```

- Procedimiento PROC_CATEGORIA: realiza acciones sobre el valor pasado como parámetro, que debe corresponder al campo 'category'. En primer lugar elimina los posibles espacios vacíos, tanto al inicio como al final de la cadena de texto. Posteriormente, selecciona el valor unificado para la categoría. El código del procedimiento es el siguiente:

```
create or replace procedure proc_categoria
(festival IN VARCHAR2, categoria IN OUT VARCHAR2)
is
begin

    IF categoria IS NOT NULL THEN

        categoria := LTRIM(categoria);
        categoria := RTRIM(categoria);

        IF festival = 'academy' THEN
            SELECT unificado INTO categoria FROM categorias_unificacion
            WHERE academy = categoria and rownum = 1;
        ELSE
            IF festival = 'berlinale' THEN
                SELECT unificado INTO categoria FROM categorias_unificacion
                WHERE berlinale = categoria and rownum = 1;
            ELSE
                IF festival = 'cannes' THEN
                    SELECT unificado INTO categoria FROM categorias_unificacion
                    WHERE cannes = categoria and rownum = 1;
                ELSE
                    IF festival = 'cesars' THEN
                        SELECT unificado INTO categoria FROM categorias_unificacion
                        WHERE cesars = categoria and rownum = 1;
                    ELSE
                        IF festival = 'mtv' THEN
                            SELECT unificado INTO categoria FROM categorias_unificacion
                            WHERE mtv = categoria and rownum = 1;
                        END IF;
                    END IF;
                END IF;
            END IF;
        END IF;
    END IF;

end;
```

- Procedimiento PROC_EDICION: realiza la comprobación en la relación año-edición. Únicamente es llamado desde la tabla que trata los datos del festival Cesars, puesto que en el mismo es el único en el que se han detectado errores en dicha relación. Para realizar la comprobación guarda en primer lugar los valores del registro a tratar en una tabla de apoyo, tabla AÑOEDICION, para posteriormente averiguar el valor más repetido, dando como bueno dicho valor. El código del procedimiento es el siguiente:

```
create or replace procedure proc_edicion
(edi IN VARCHAR2, year IN OUT VARCHAR2, ediciones IN OUT
NUMBER)
is
begin
IF year IS NOT NULL THEN
    INSERT INTO añoedicion(año, edicion) values(year, edi);
    SELECT count(*) INTO ediciones FROM añoedicion WHERE edicion =
edi;
    IF ediciones > 0 THEN
        SELECT año, count(*) INTO year, ediciones FROM añoedicion
WHERE edicion = edi AND rownum = 1 GROUP BY año ORDER BY 2
desc;
    END IF;
END IF;
end;
```

- Procedimiento PROC_NOMINADO: realiza acciones sobre el campo 'nominees', asignándole el valor del título original de la película, en caso de que el mismo esté vacío. El código del procedimiento es el siguiente:

```
create or replace procedure proc_nominado
(titoriginal IN OUT VARCHAR2, nominado IN OUT VARCHAR2)
is
begin
IF titoriginal IS NOT NULL THEN
    IF ((nominado IS NULL) OR (nominado = '')) THEN
        nominado := titoriginal;
    END IF;
END IF;
end;
```

- Procedimiento PROC_PAIS: realiza acciones sobre el valor del campo 'year', asignándole el valor encontrado en la tabla de unificación de países para el valor pasado como parámetro. En caso de tratarse del festival MTV, en el que no se indica ningún país, se asigna directamente USA, tal y como ha indicado la ACA en un inicio. El código del procedimiento es el siguiente:

```
create or replace procedure proc_pais
(festival IN VARCHAR2, pais IN OUT VARCHAR2)
is
begin
IF festival = 'academy' THEN
    SELECT unificado INTO pais FROM paises_traduccion WHERE
academy = pais;
ELSE
    IF festival = 'berlinale' THEN
        SELECT unificado INTO pais FROM paises_traduccion WHERE
berlinale = pais;
    ELSE
        IF festival = 'cannes' THEN
            SELECT unificado INTO pais FROM paises_traduccion WHERE
cannes = pais;
        ELSE
            IF festival = 'cesars' THEN
                SELECT unificado INTO pais FROM paises_traduccion WHERE
cesars = pais;
            ELSE
                IF festival = 'mtv' THEN
                    pais := 'USA';
                END IF;
            END IF;
        END IF;
    END IF;
END IF;
end;
```

- Procedimiento PROC_TITFESTIVAL: realiza acciones sobre el campo correspondiente al título utilizado en el festival, asignándole el valor del título original de la película, en caso de que el primero esté vacío. El código del procedimiento es el siguiente:

```
create or replace procedure proc_titfestival
(original IN OUT VARCHAR2, festival IN OUT VARCHAR2)
is
begin
IF original IS NOT NULL THEN
    IF ((festival IS NULL) OR (festival = '')) THEN
        festival := original;
    END IF;
END IF;
end;
```

- Procedimiento PROC_TITORIGINAL: realiza acciones sobre el campo correspondiente al título original, asignándole el valor del título original de la película, en caso de que el primero esté vacío. El código del procedimiento es el siguiente:

```
create or replace procedure proc_titoriginal
(original IN OUT VARCHAR2, festival IN OUT VARCHAR2)
is
begin
IF festival IS NOT NULL THEN
  IF ((original IS NULL) OR (original = '')) THEN
    original := festival;
  END IF;
END IF;
end;
```

- Procedimiento PROC_WINNER: realiza acciones sobre el campo que indica si la nominación ha sido premiada, eliminando cualquier codificación e insertando un claro valor 'SI' en el caso de haber sido galardonada y 'NO' en caso contrario. El código del procedimiento es el siguiente:

```
create or replace procedure proc_winner
(premio IN OUT VARCHAR2)
is
begin
  IF premio = 'X' THEN
    premio := 'SI';
  ELSE
    premio := 'NO';
  END IF;
end;
```

Tras la fase de transformación, en la que se han tratado los datos en las tablas temporales, se realiza la última fase, la de carga de datos al Almacén de Datos. Para la misma, se ha utilizado la misma técnica que para la anterior, es decir hacer uso de disparadores tras la inserción de los registros ya tratados y realizar la llamada al procedimiento, en este caso el mismo, que realiza la carga de los datos desde las tablas temporales a las tablas propias del Almacén de Datos, logrando poblar al mismo con los valores definitivos que permitan su posterior explotación. Los disparadores y procedimientos que se utilizan en esta fase son los siguientes:

- Disparador AFTER sobre tabla TMP_ACADEMY:

```
create or replace trigger after_academy
AFTER
insert or update on "TMP_ACADEMY"
for each row

declare
num NUMBER := 0;
dest VARCHAR2(255);
idPe NUMBER := 0;
idNo NUMBER := 0;
idFe NUMBER := 0;
idCa NUMBER := 0;
idAñ NUMBER := 0;
idPa NUMBER := 0;

begin
proc_carga(num, :new.original_title, :new.english_title,
:new.nominees, :new.award, :new.category, dest, :new.year,
:new.country, idPe, idNo, idFe, idCa, idAñ, idPa, :new.winner);
end;
```

- Disparador AFTER sobre tabla TMP_BERLINALE:

```
create or replace trigger after_berlinale
AFTER
insert or update on "TMP_BERLINALE"
for each row

declare
num NUMBER := 0;
dest VARCHAR2(255);
idPe NUMBER := 0;
idNo NUMBER := 0;
idFe NUMBER := 0;
idCa NUMBER := 0;
idAñ NUMBER := 0;
idPa NUMBER := 0;

begin
proc_carga(num, :new.film_title, :new.localized_title, :new.nominees,
:new.award, :new.category, dest, :new.year, :new.country, idPe, idNo,
idFe, idCa, idAñ, idPa, :new.winner);
end;
```

- Disparador AFTER sobre tabla TMP_CANNES:

```
create or replace trigger after_cannes
AFTER
insert or update on "TMP_CANNES"
for each row

declare
num NUMBER := 0;
dest VARCHAR2(255);
idPe NUMBER := 0;
idNo NUMBER := 0;
idFe NUMBER := 0;
idCa NUMBER := 0;
idAñ NUMBER := 0;
idPa NUMBER := 0;

begin
proc_carga(num, :new.film_title, :new.localized_title, :new.nominees,
:new.award, :new.category, dest, :new.year, :new.country, idPe, idNo,
idFe, idCa, idAñ, idPa, :new.winner);
end;
```

- Disparador AFTER sobre tabla TMP_CESARS:

```
create or replace trigger after_cesars
AFTER
insert or update on "TMP_CESARS"
for each row

declare
edi NUMBER := 0;
num NUMBER := 0;
dest VARCHAR2(255);
idPe NUMBER := 0;
idNo NUMBER := 0;
idFe NUMBER := 0;
idCa NUMBER := 0;
idAñ NUMBER := 0;
idPa NUMBER := 0;

begin
proc_carga(num, :new.film_title, :new.localized_title, :new.nominees,
:new.award, :new.category, dest, :new.year, :new.country, idPe, idNo,
idFe, idCa, idAñ, idPa, :new.winner);
end;
```

- Disparador AFTER sobre tabla TMP_MTV:

```
create or replace trigger after_mtv
AFTER
insert or update on "TMP_MTV"
for each row

declare
num NUMBER := 0;
dest VARCHAR2(255);
idPe NUMBER := 0;
idNo NUMBER := 0;
idFe NUMBER := 0;
idCa NUMBER := 0;
idAñ NUMBER := 0;
idPa NUMBER := 0;

begin
proc_carga(num, :new.original_title, :new.localized_title,
:new.nominees, :new.award, :new.category, dest, :new.year,
:new.country, idPe, idNo, idFe, idCa, idAñ, idPa, :new.winner);
end;
```

- Procedimiento PROC_CARGA: realiza la carga de datos en las tablas propias del Almacén de Datos con los parámetros que el disparador que lo ha llamado le ha pasado. El proceso se inicia en las tablas correspondientes a las dimensiones, comprobando antes de hacer la inserción que no exista un registro con los mismos valores. Una vez pobladas las dimensiones, se insertan los datos correspondientes al registro (seleccionando el identificador del valor de cada campo) en la tabla de hechos. El código del procedimiento de carga es el siguiente:

```
create or replace procedure proc_carga
(num OUT NUMBER,
titOr IN VARCHAR2,
titFe IN VARCHAR2,
nomNo IN VARCHAR2,
nomFe IN VARCHAR2,
nomCa IN VARCHAR2,
dest OUT VARCHAR2,
numAñ IN NUMBER,
nomPa IN VARCHAR2,
idPe OUT NUMBER,
idNo OUT NUMBER,
idFe OUT NUMBER,
idCa OUT NUMBER,
idAñ OUT NUMBER,
idPa OUT NUMBER,
prem IN VARCHAR2
```

```
)  
is  
begin  
  
SELECT count(*) INTO num FROM peliculas WHERE titOriginal = titOr;  
IF num = 0 THEN  
    INSERT INTO peliculas(titOriginal, titFestival) values(titOr, titFe);  
END IF;  
  
SELECT count(*) INTO num FROM nominados WHERE nomNominado =  
nomNo;  
IF num = 0 THEN  
    INSERT INTO nominados(nomNominado) values(nomNo);  
END IF;  
  
SELECT count(*) INTO num FROM festivales WHERE nomFestival =  
nomFe;  
IF num = 0 THEN  
    INSERT INTO festivales(nomFestival) values(nomFe);  
END IF;  
  
SELECT count(*) INTO num FROM categorias WHERE nomCategoria =  
nomCa;  
IF num = 0 THEN  
    SELECT destino INTO dest FROM categorias_unificacion WHERE  
unificado = nomCa and rownum = 1;  
    INSERT INTO categorias(nomCategoria, destino) values(nomCa, dest);  
END IF;  
  
SELECT count(*) INTO num FROM año WHERE año = numAñ;  
IF num = 0 THEN  
    INSERT INTO año(año) values(numAñ);  
END IF;  
  
SELECT count(*) INTO num FROM paises WHERE nomPais = nomPa;  
IF num = 0 THEN  
    INSERT INTO paises(nomPais) values(nomPa);  
END IF;  
  
SELECT idPeli INTO idPe FROM Peliculas WHERE titOriginal = titOr;  
SELECT idNominado INTO idNo FROM Nominados WHERE  
nomNominado = nomNo;  
SELECT idFestival INTO idFe FROM Festivales WHERE nomFestival =  
nomFe;  
SELECT idCategoria INTO idCa FROM Categorias WHERE nomCategoria  
= nomCa;  
SELECT Año INTO idAñ FROM Año WHERE año = numAñ;  
SELECT idPais INTO idPa FROM Paises WHERE nomPais = nomPa;
```

```
IF idPe > 0 AND idNo > 0 AND idFe > 0 AND idCa > 0 AND idAñ > 0 AND  
idPa > 0 THEN  
    INSERT INTO Premios(idPeli, idNominado, idFestival, idCategoria,  
idAño, idPais, premiado) values(idPe, idNo, idFe, idCa, idAñ, idPa,  
prem);  
END IF;  
end;
```

Con ello se completan todos los elementos que intervienen en el proceso de ETL, por lo que una vez realizadas las fases indicadas, sería posible lanzar dicho proceso. Para ello se utilizará la herramienta SQL Loader. La misma realizará la inserción de los datos, ya preparados en los ficheros de origen CSV, en las tablas temporales correspondientes. Dicha inserción provocará que se disparen tanto los disparadores BEFORE, antes de la inserción, como los AFTER, tras la misma. Cada disparador llamará a los procedimientos oportunos para que, de forma automática, con la ejecución de la herramienta SQL Loader se logre poblar el Almacén de Datos.

Para lanzar el proceso, la herramienta comentada precisa de unos ficheros de control, archivos CTL, que le indiquen una serie de parámetros que la misma precisa, como son si existe línea de título (mediante la instrucción "OPTIONS(SKIP=1)"), el juego de caracteres que se utiliza (por medio de la instrucción "CHARACTERSET 'WE8MSWIN1252' "), el archivo en el que se encuentran los datos (mediante la instrucción "INFILE 'nombreDelFichero.csv' "), si se realizan tareas de carga manteniendo los datos de la tabla de destino o bien se eliminan los datos antes de cargarlos (mediante APPEND o REPLACE), la tabla en la que se realizará dicha inserción (mediante "INTO TABLE nombretabla"), el carácter que indica el final de un campo (mediante "FIELDS TERMINATED BY ';' ") y el orden en que aparecen los datos en el fichero de carga entre paréntesis y separados por comas. El contenido de los ficheros de control que se utilizan para indicar a SQL Loader las tareas que debe realizar son los siguientes:

- Fichero ACADEMY.CTL:

```
OPTIONS (SKIP=1)  
LOAD DATA  
CHARACTERSET 'WE8MSWIN1252'  
INFILE 'Academy_Awards_2006.csv'  
REPLACE  
INTO TABLE tmp_academy  
FIELDS TERMINATED BY ';' (  
    original_title,  
    english_title,  
    sort_title,  
    year,  
    country,  
    award,  
    category,
```

```
winner,  
nominees,  
academy_awards,  
item  
)
```

- Fichero BERLINALE.CTL:

```
OPTIONS (SKIP=1)  
LOAD DATA  
CHARACTERSET 'WE8MSWIN1252'  
INFILE 'berlinale_2004.csv'  
REPLACE  
INTO TABLE tmp_berlinale  
FIELDS TERMINATED BY ';'   
(  
film_title,  
localized_title,  
sort_title,  
year,  
country,  
award,  
category,  
winner,  
nominees,  
berlinale,  
item  
)
```

- Fichero CANNES.CTL:

```
OPTIONS (SKIP=1)  
LOAD DATA  
CHARACTERSET 'WE8MSWIN1252'  
INFILE 'Cannes_Awards_2005.csv'  
REPLACE  
INTO TABLE tmp_cannes  
FIELDS TERMINATED BY ';'   
(  
film_title,  
localized_title,  
sort_title,  
year,  
country,  
award,  
category,  
winner,  
nominees,  
cannes_film_festival,  
item  
)
```

- Fichero CESARS.CTL:

```
OPTIONS (SKIP=1)
LOAD DATA
CHARACTERSET 'WE8MSWIN1252'
INFILE 'Cesar_Awards_2005.csv'
REPLACE
INTO TABLE tmp_cesars
FIELDS TERMINATED BY ';'
(
film_title,
localized_title,
sort_title,
year,
country,
award,
category,
winner,
nominees,
cesars,
item
)
```

- Fichero MTV.CTL:

```
OPTIONS (SKIP=1)
LOAD DATA
CHARACTERSET 'WE8MSWIN1252'
INFILE 'MTV_Movie_Awards_2006.csv'
REPLACE
INTO TABLE tmp_mtv
FIELDS TERMINATED BY ';'
(
original_title,
localized_title,
sort_title,
year,
country,
award,
category,
winner,
nominees,
mtv_movie_awards,
item
)
```

- Fichero CATEGORIAS.CTL:

```
OPTIONS (SKIP=1)
LOAD DATA
CHARACTERSET 'WE8MSWIN1252'
INFILE 'categorias_unificacion.csv'
APPEND
INTO TABLE categorias_unificacion
FIELDS TERMINATED BY ';'
(
id,
academy,
berlinale,
cannes,
cesars,
mtv,
unificado,
destino
)
```

- Fichero PAISES.CTL:

```
OPTIONS (SKIP=1)
LOAD DATA
CHARACTERSET 'WE8MSWIN1252'
INFILE 'Paises_Traduccion.csv'
APPEND
INTO TABLE paises_traduccion
FIELDS TERMINATED BY ';'
(
id,
academy,
berlinale,
cannes,
cesars,
unificado
)
```

Con los ficheros de control descritos, únicamente es necesario ejecutar una sentencia que lance el proceso para cada uno de los archivos y todos los pasos descritos anteriormente se realizarán de forma automática. Las sentencias a ejecutar, en el directorio donde se encuentren los ficheros de control y de datos, son las siguientes:

- Sentencia PAISES:

```
Sqllldr userid=PR/pr control=paises.ctl
```

- Sentencia CATEGORIAS:

```
Sqllldr userid=PR/pr control=categorias.ctl
```


- Sentencia ACADEMY:

 Sqlldr userid=PR/pr control=academy.ctl
- Sentencia BERLINALE:

 Sqlldr userid=PR/pr control=berlinale.ctl
- Sentencia CANNES:

 Sqlldr userid=PR/pr control=cannes.ctl
- Sentencia CESARS:

 Sqlldr userid=PR/pr control=cesars.ctl
- Sentencia MTV:

 Sqlldr userid=PR/pr control=mtv.ctl

Los resultados de la ejecución de cada una de las sentencias indicadas, pueden observarse en el apartado correspondiente a Capturas de pantalla.

Informes

Tras poblar el Almacén de Datos, es posible realizar su explotación por medio de la herramienta Oracle Discoverer. Los pasos y las decisiones tomadas en cada uno de los informes solicitados por la ACA son los siguientes:

- Número de premios conseguidos por un actor/actriz por año y festival
 - o Elementos seleccionados:
 - Año.Año
 - Festivales.Nomfestival
 - Nominados.Nomnominado
 - Premios.Premiado(count)
 - o Diseño de matriz:
 - Filas: Año.Año y Nominados.Nomnominado
 - Columnas: Premios.Premiado(count) y Festivales.Nomfestival
 - o Condiciones:
 - Premios.Premiado = 'SI'
 - Categorías.Destino = 'Actor' OR Categorías.Destino = 'Actress'
- Número de premios conseguidos por un actor/actriz por año, festival y categoría
 - o Elementos seleccionados:
 - Año.Año

- Categorías.Nomcategoría
 - Festivales.Nomfestival
 - Nominados.Nomnominado
 - Premios.Premiado(count)
 - Diseño de matriz:
 - Filas: Año.Año y Nominados.Nomnominado
 - Columnas: Premios.Premiado(count), Festivales.Nomfestival y Categorías.Nomcategoría
 - Condiciones:
 - Premios.Premiado = 'SI'
 - Categorías.Destino = 'Actor' OR Categorías.Destino = 'Actress'
- Número de premios conseguidos por un director por año, festival y película
 - Elementos seleccionados:
 - Año.Año
 - Festivales.Nomfestival
 - Nominados.Nomnominado
 - Películas.Titoriginal
 - Premios.Premiado(count)
 - Diseño de matriz:
 - Filas: Año.Año, Nominados.Nomnominado y Películas.Titoriginal
 - Columnas: Premios.Premiado(count), Festivales.Nomfestival
 - Condiciones:
 - Premios.Premiado = 'SI'
 - Categorías.Destino = 'Director'
- Número de premios conseguidos por año, festival, categoría y película
 - Elementos seleccionados:
 - Año.Año
 - Categorías.Nomcategoría
 - Festivales.Nomfestival
 - Películas.Titoriginal
 - Premios.Premiado(count)
 - Diseño de matriz:
 - Filas: Año.Año, Festivales.Nomfestival y Películas.Titoriginal
 - Columnas: Premios.Premiado(count), Categorías.Nomcategoría
 - Condiciones:
 - Premios.Premiado = 'SI'
- Dada una fecha, para cada festival, mostrar el número de ediciones realizadas del festival, número total de actores que han sido nominados, número total de actrices que han sido nominadas y número total de películas nominadas
 - Elementos seleccionados:

- Festivales.Nomfestival
 - Diseño de matriz:
 - Filas: Festivales.Nomfestival
 - Condiciones:
 - Premios.Idaño <= "Año"
 - Cálculos:
 - Actores Nominados distintos: COUNT_DISTINCT(CASE WHEN Destino = 'Actor' THEN Idnominado END)
 - Años distintos: COUNT_DISTINCT("Idaño")
 - Actrices nominadas distintas: COUNT_DISTINCT(CASE WHEN Destino = 'Actress' THEN Idnominado END)
 - Películas nominadas distintas: COUNT_DISTINCT(CASE WHEN Destino = 'Film' THEN Idnominado END)
 - Parámetros:
 - Año
- Para cada año, número de premios ganados por cada película y festival
- Elementos seleccionados:
 - Año.Año
 - Festivales.Nomfestival
 - Películas.Titoriginal
 - Premios.Premiado(count)
 - Diseño de matriz:
 - Filas: Año.Año y Películas.Titoriginal
 - Columnas: Premios.Premiado(count) y Festivales.Nomfestival
 - Condiciones:
 - Premios.Premiado = 'SI'

Asimismo, los pasos y decisiones tomadas en cada uno de las consultas realizadas de forma adicional son los siguientes:

- Para cada año, indicar las películas con mayor éxito y mayor fracaso, según unos criterios que definan cada concepto:
 - Método de valoración: ratio entre nominaciones y premios
 - Elementos seleccionados:
 - Películas.Titoriginal
 - Premios.Idaño
 - Premios.Premiado(count)
 - Diseño de matriz:
 - Filas: Premios.Idaño y Películas.Titoriginal
 - Columnas: Premios.Premiado(count)
 - Condiciones:

- $\text{Ratio} = \text{MAX}(\text{Ratio}) \text{ OVER}(\text{PARTITION BY "Idaño" }) \text{ OR } \text{Ratio} = \text{MIN}(\text{Ratio}) \text{ OVER}(\text{PARTITION BY "Idaño" })$
 - Cálculos:
 - Premios conseguidos: $\text{COUNT}(\text{CASE WHEN Premiado} = \text{'SI'} \text{ THEN Premiado ELSE NULL END})$
 - Ratio: Premios conseguidos/Premiado COUNT
 - Orden:
 - Ratio
- Número de premios por categoría ganados por el país más galardonado, dejando de lado a EE.UU.
 - Parte1 (obtener el país más galardonado):
 - Elementos seleccionados:
 - Paises.Nompais
 - Premios.Premiado(count)
 - Diseño de matriz:
 - Filas: Paises.NomPais
 - Premios.Premiado(count)
 - Condiciones:
 - Premios.Premiado = 'SI'
 - Paises.Nompais <> 'USA'
 - Orden = 1
 - Cálculos:
 - Orden: $\text{ROW_NUMBER}() \text{ OVER}(\text{ORDER BY Premiado COUNT DESC })$
 - Parte2 (informe solicitado):
 - Elementos seleccionados:
 - Categorías.Nomcategoria
 - Paises.Nompais
 - Premios.Premiado(count)
 - Diseño de matriz:
 - Filas: Paises.Nompais y Categorías.Nomcategoria
 - Columnas: Premios.Premiado(count)
 - Condiciones:
 - Premios.Premiado = 'SI'
 - Paises.Nompais = Parte1 (subconsulta)
- Analizar una posible relación entre ganar un premio en un festival y ganar el mismo, o equivalente, en otro festival

Para este análisis en lugar de obtener un listado a través de Oracle Discoverer, se ha utilizado el entorno web de Oracle para obtener, en primer lugar un listado en el que aparezca, para cada categoría y nominado, el número de premios obtenidos por festival y la suma de los

mismos. Dicho listado se ha exportado a CSV desde el propio entorno web y se ha tratado con Microsoft Excel, añadiendo una nueva columna que cuente el número de festivales en los que no se ha obtenido premio alguno y utilizando las funciones “CONTAR.SI” y “CONTAR.SI.CONJUNTO” para obtener la siguiente información, la cual permite llegar a la conclusión que no existe una clara relación entre ganar un premio en un festival y ganar el mismo en un festival distinto:

ANALISIS DATOS			
	Número	% sobre total	% sobre premiados
Con 5 ceros	1761	51,21%	-
Con 4 ceros	1659	48,24%	98,87%
Con 3 ceros	17	0,49%	1,01%
Con 2 ceros	2	0,06%	0,12%
Con 1 cero	0	0,00%	0,00%
Sin ceros	0	0,00%	0,00%
Total nominados	3439	100,00%	-
Total con premio	1678	-	100,00%

La sentencia SQL ejecutada para obtener el listado previo exportado a CSV es la siguiente:

```
SELECT
ca.nomCategoria,
no.nomNominado,
count(case when fe.nomFestival = 'Academy' then premiado else null
end) as Academy,
count(case when fe.nomFestival = 'Berlinale' then premiado else null
end) as Berlinale,
count(case when fe.nomFestival = 'Cannes Festival' then premiado
else null end) as Cannes,
count(case when fe.nomFestival = 'César' then premiado else null
end) as Césars,
count(case when fe.nomFestival = 'MTV Movie Awards' then
premiado else null end) as Mtv,
(count(case when fe.nomFestival = 'Academy' then premiado else
null end) +
count(case when fe.nomFestival = 'Berlinale' then premiado else null
end) +
count(case when fe.nomFestival = 'Cannes Festival' then premiado
else null end) +
count(case when fe.nomFestival = 'César' then premiado else null
end) +
count(case when fe.nomFestival = 'MTV Movie Awards' then
premiado else null end)) as Suma
FROM categorias ca, nominados no, premios pr, festivales fe
WHERE ca.idCategoria = pr.idCategoria
AND no.idNominado = pr.idNominado
```

AND fe.idFestival = pr.idFestival
 AND premiado = 'SI'
 GROUP BY ca.nomCategoria, no.nomNominado
 ORDER BY 8 desc;

Capturas de pantalla referente a este análisis pueden observarse más adelante en este mismo documento.

Capturas de pantalla

Planificación

A continuación se muestra el detalle de las tareas descritas en el apartado Planificación del proyecto, así como los diagramas de Gantt correspondientes a las mismas:








	Nombre de tarea	Duración	Comienzo	Fin	Predecesoras
1	 Inicio del proyecto	3 días	mié 02/03/11	sáb 05/03/11	
2	Descarga del material de la asignatura, lectura del plan docente, presentación e	1 día	mié 02/03/11	jue 03/03/11	
3	Preparación hardware y software inicial	1 día	vie 04/03/11	sáb 05/03/11	2
4	Búsqueda preliminar de información web	1 día	sáb 05/03/11	sáb 05/03/11	3
5	 PAC1	7 días	sáb 05/03/11	mié 16/03/11	1
6	 Elaboración Boceto	5 días	sáb 05/03/11	lun 14/03/11	
7	Descarga y lectura del enunciado del TFC	1 día	sáb 05/03/11	sáb 05/03/11	
8	Toma de contacto y comprensión datos proporcionados (ficheros Excel)	1 día	sáb 05/03/11	sáb 05/03/11	
9	Descarga y lectura del modelo ejemplo	1 día	sáb 05/03/11	sáb 05/03/11	
10	Elaboración borrador documento "Análisis de datos"	1 día	dom 06/03/11	mar 08/03/11	7,8,9
11	Elaboración borrador documento "Documento de especificaciones"	1 día	mar 08/03/11	jue 10/03/11	10
12	Entrega borradores "Análisis de datos" y "Documento de especificaciones"	0 días	jue 10/03/11	jue 10/03/11	10;11
13	Elaboración borrador documento "Planificación del proyecto"	2 días	jue 10/03/11	lun 14/03/11	11
14	Entrega borrador "Planificación del proyecto"	0 días	lun 14/03/11	lun 14/03/11	13
15	 Elaboración Entrega	2 días	lun 14/03/11	mié 16/03/11	6
16	Lectura comentarios sobre borradores entregados	1 día	lun 14/03/11	mar 15/03/11	
17	Inicio lectura material extraordinario suministrado (MMMD)	2 días	lun 14/03/11	mié 16/03/11	
18	Elaboración documento "Análisis de datos"	1 día	lun 14/03/11	mar 15/03/11	
19	Elaboración documento "Documento de especificaciones"	1 día	lun 14/03/11	mar 15/03/11	
20	Elaboración documento "Planificación del proyecto"	2 días	lun 14/03/11	mié 16/03/11	
21	Entrega documentos PAC1	0 días	mié 16/03/11	mié 16/03/11	18;19;20
22	 PAC2	24 días	vie 18/03/11	mar 19/04/11	5
23	 Elaboración Boceto	19 días	vie 18/03/11	mar 12/04/11	
24	Descarga y lectura del enunciado de la PAC2	1 día	vie 18/03/11	sáb 19/03/11	
25	Instalación y puesta en marcha máquina virtual proporcionada	1 día	vie 18/03/11	sáb 19/03/11	
26	Continuación lectura material extraordinario suministrado (MMMD)	4 días	vie 18/03/11	mar 22/03/11	
27	Consulta bibliografía sobre "Análisis Multidimensional"	4 días	vie 18/03/11	mar 22/03/11	
28	Inicio del proceso de limpieza y depuración de los datos iniciales	4 días	vie 18/03/11	mar 22/03/11	
29	Elaboración borrador documento "Análisis de requisitos"	8 días	sáb 19/03/11	mar 29/03/11	24;25
30	Entrega borrador "Análisis de requisitos"	0 días	mar 29/03/11	mar 29/03/11	29
31	Elaboración borrador documento "Diseño del Data Warehouse"	10 días	mié 30/03/11	mar 12/04/11	30
32	Entrega borrador "Diseño del Data Warehouse"	0 días	mar 12/04/11	mar 12/04/11	31
33	 Elaboración Entrega	5 días	mié 13/04/11	mar 19/04/11	23
34	Lectura comentarios sobre borradores entregados	1 día	mié 13/04/11	jue 14/04/11	
35	Elaboración documento "Análisis de requisitos"	4 días	vie 15/04/11	mar 19/04/11	34
36	Elaboración documento "Diseño del Data Warehouse"	4 días	vie 15/04/11	mar 19/04/11	34
37	Entrega documentos PAC2	0 días	mar 19/04/11	mar 19/04/11	35;36

Fig.3: Planificación del proyecto (parte1)

		Nombre de tarea	Duración	Comienzo	Fin	Predecesoras
38		PAC3	21 días	vie 22/04/11	mar 24/05/11	22
39		Elaboración Boceto	15 días	vie 22/04/11	sáb 14/05/11	
40		Descarga y lectura del enunciado de la PAC3	1 día	vie 22/04/11	sáb 23/04/11	
41		Finalización proceso de limpieza y depuración de los datos iniciales	3 días	vie 22/04/11	sáb 23/04/11	28
42		Creación de tablas, procedimientos y disparadores	3 días	sáb 23/04/11	mar 26/04/11	40
43		Carga de datos	1 día	mié 27/04/11	jue 28/04/11	42
44		Instalación herramienta de explotación de datos	1 día	lun 02/05/11	mar 03/05/11	43
45		Construcción de los informes detallados en el enunciado del TFC	4 días	mié 04/05/11	sáb 07/05/11	44
46		Análisis de la información	4 días	mié 04/05/11	sáb 07/05/11	44
47		Elaboración boceto documento "Implementación del Data Warehouse"	5 días	lun 09/05/11	sáb 14/05/11	45;46
48		Entrega borrador "Implementación del Data Warehouse"	0 días	sáb 14/05/11	sáb 14/05/11	47
49		Elaboración Entrega	6 días	lun 16/05/11	mar 24/05/11	39
50		Lectura comentarios sobre borrador entregado	1 día	lun 16/05/11	mar 17/05/11	
51		Pruebas sobre informes creados	2 días	lun 16/05/11	jue 19/05/11	45
52		Resolución incidencias en informes iniciales	2 días	vie 20/05/11	sáb 21/05/11	51
53		Elaboración documento "Implementación del Data Warehouse"	5 días	mié 18/05/11	mar 24/05/11	50
54		Entrega documento PAC3	0 días	mar 24/05/11	mar 24/05/11	53
55		Entrega Final	16 días	vie 27/05/11	dom 12/06/11	38
56		Elaboración Boceto	6 días	vie 27/05/11	sáb 04/06/11	
57		Elaboración boceto documento final	6 días	vie 27/05/11	sáb 04/06/11	54
58		Entrega borrador documento final	0 días	sáb 04/06/11	sáb 04/06/11	57
59		Elaboración Entrega	10 días	sáb 04/06/11	dom 12/06/11	56
60		Lectura comentarios sobre borrador entregado	1 día	sáb 04/06/11	sáb 04/06/11	
61		Elaboración documento final	6 días	sáb 04/06/11	sáb 11/06/11	60
62		Elaboración presentación	3 días	sáb 11/06/11	dom 12/06/11	61
63		Entrega TFC (documento final y presentación)	0 días	dom 12/06/11	dom 12/06/11	61;62
64		Tribunal	3 días	mié 15/06/11	sáb 18/06/11	55
65		Recepción y resolución cuestiones propuestas por Tribunal	3 días	mié 15/06/11	sáb 18/06/11	

Fig.4: Planificación del proyecto (parte2)

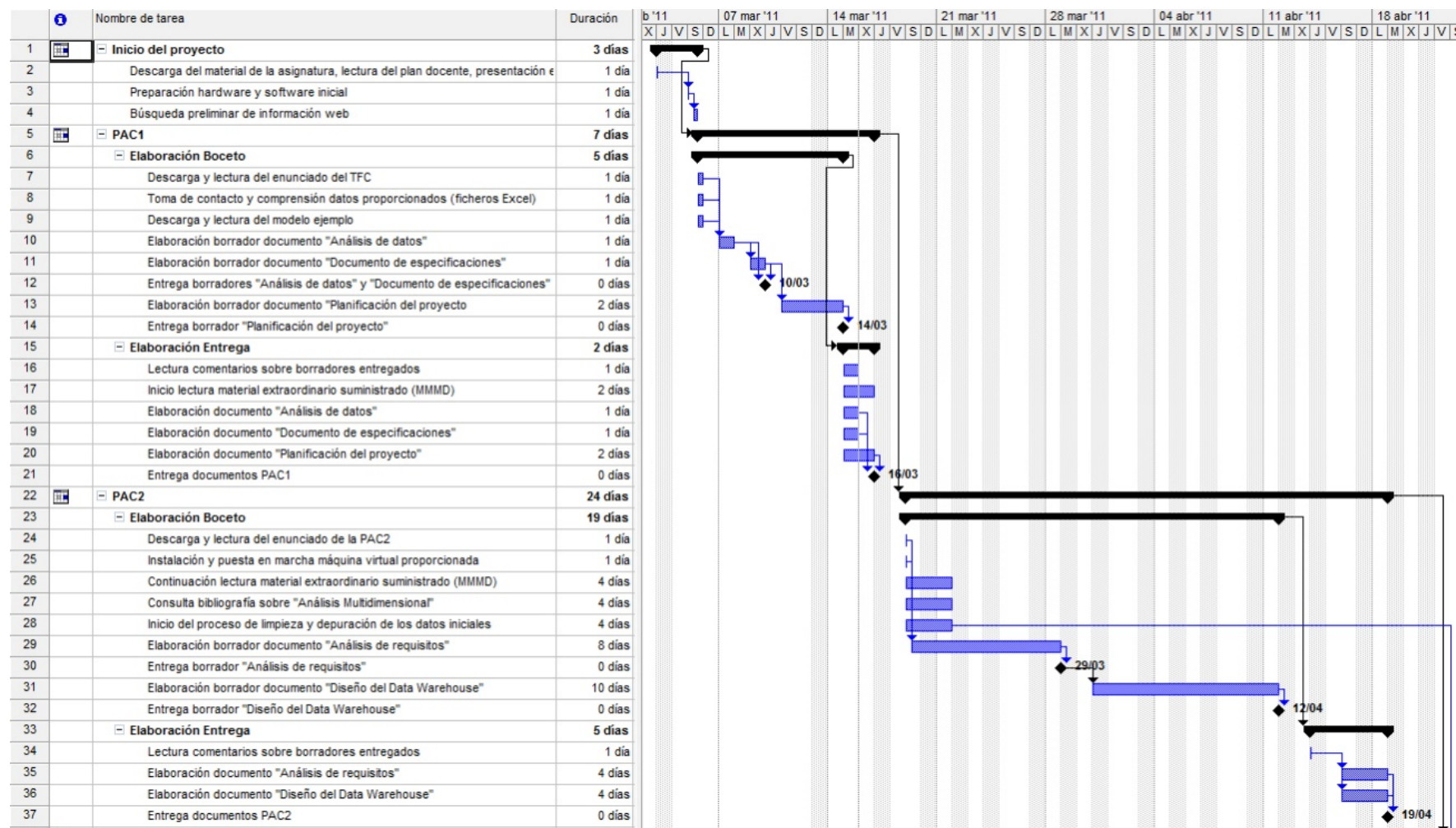


Fig.5: Diagrama de Gantt (parte1)

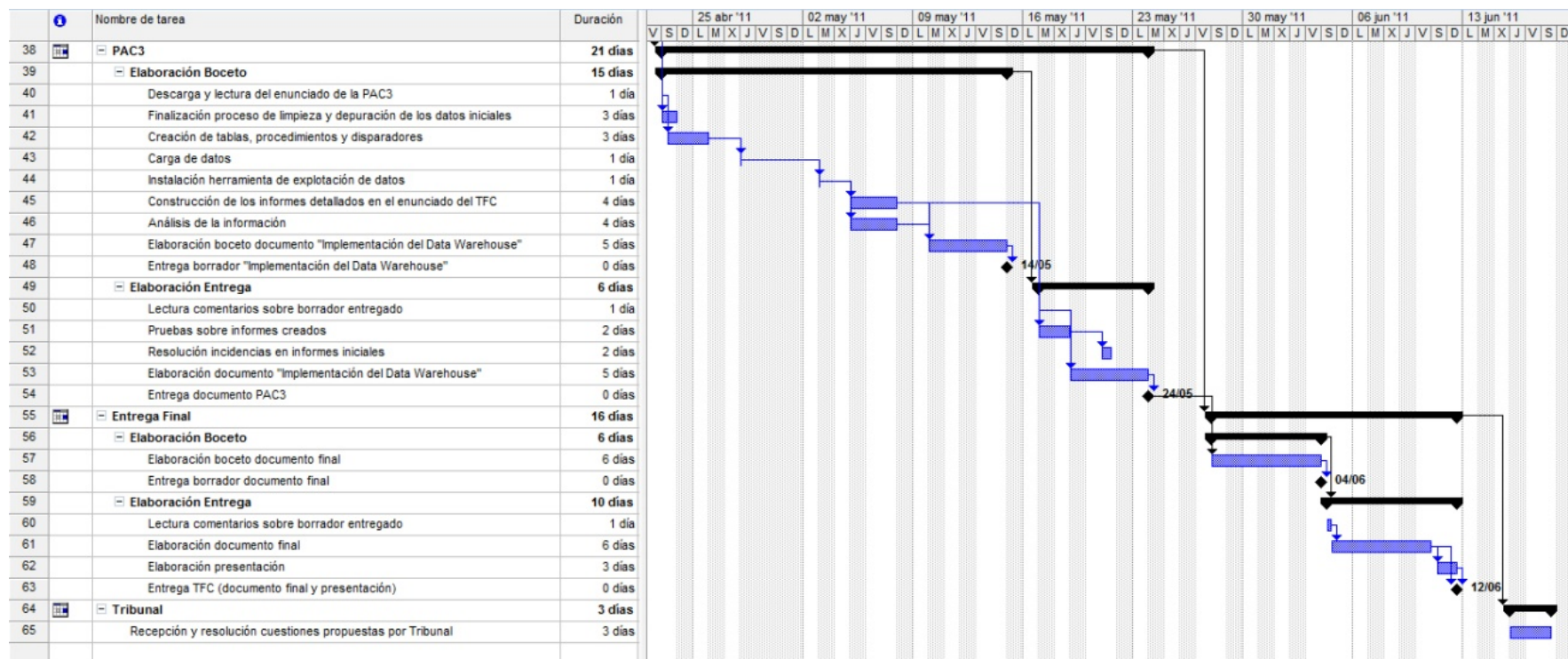
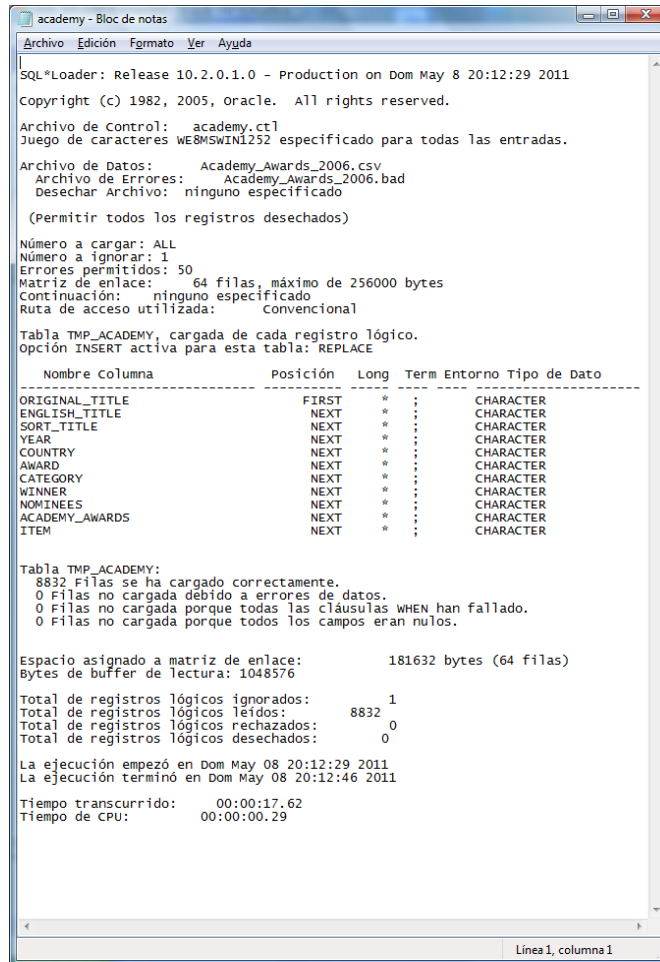


Fig.6: Diagrama de Gantt (parte2)

SQL Loader

Los resultados de las ejecuciones de la herramienta SQL Loader referentes a las tablas de los festivales son los siguientes:

Log ACADEMY:



```
academy - Bloc de notas
Archivo  Edición  Formato  Ver  Ayuda

SQL*Loader: Release 10.2.0.1.0 - Production on Dom May 8 20:12:29 2011
Copyright (c) 1982, 2005, oracle. All rights reserved.

Archivo de Control: academy.ct1
Juego de caracteres WE8MSWIN1252 especificado para todas las entradas.

Archivo de Datos: Academy_Awards_2006.csv
Archivo de Errores: Academy_Awards_2006.bad
Desechar Archivo: ninguno especificado

(Permitir todos los registros desechados)

Número a cargar: ALL
Número a ignorar: 1
Errores permitidos: 50
Matriz de enlace: 64 filas, máximo de 256000 bytes
Continuación: ninguno especificado
Ruta de acceso utilizada: Convencional

Tabla TMP_ACADEMY, cargada de cada registro lógico.
opción INSERT activa para esta tabla: REPLACE

  Nombre Columna      Posición  Long  Term Entorno Tipo de Dato
-----
ORIGINAL_TITLE        FIRST    *    ;    CHARACTER
ENGLISH_TITLE         NEXT    *    ;    CHARACTER
SORT_TITLE            NEXT    *    ;    CHARACTER
YEAR                  NEXT    *    ;    CHARACTER
COUNTRY               NEXT    *    ;    CHARACTER
AWARD                 NEXT    *    ;    CHARACTER
CATEGORY              NEXT    *    ;    CHARACTER
WINNER                NEXT    *    ;    CHARACTER
NOMINEES              NEXT    *    ;    CHARACTER
ACADEMY_AWARDS        NEXT    *    ;    CHARACTER
ITEM                  NEXT    *    ;    CHARACTER

Tabla TMP_ACADEMY:
8832 Filas se ha cargado correctamente.
0 Filas no cargada debido a errores de datos.
0 Filas no cargada porque todas las cláusulas WHEN han fallado.
0 Filas no cargada porque todos los campos eran nulos.

Espacio asignado a matriz de enlace: 181632 bytes (64 filas)
Bytes de buffer de lectura: 1048576

Total de registros lógicos ignorados: 1
Total de registros lógicos leídos: 8832
Total de registros lógicos rechazados: 0
Total de registros lógicos desechados: 0

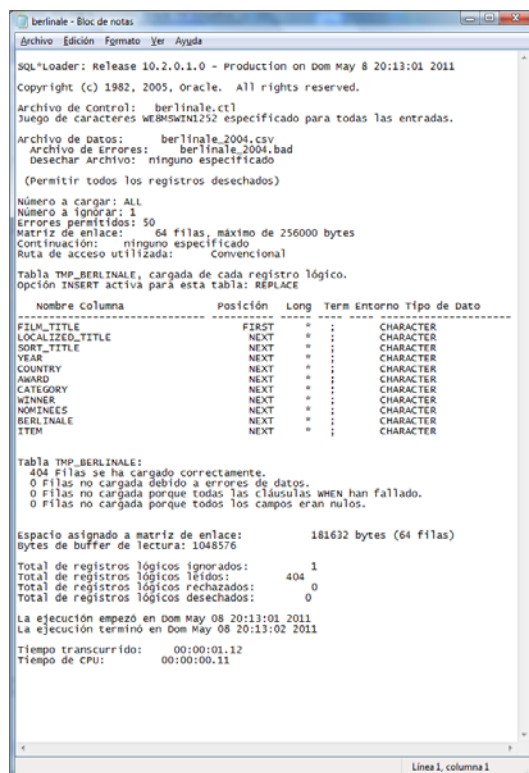
La ejecución empezó en Dom May 08 20:12:29 2011
La ejecución terminó en Dom May 08 20:12:46 2011

Tiempo transcurrido: 00:00:17.62
Tiempo de CPU: 00:00:00.29

Línea 1, columna 1
```

Fig.7: Resultado carga fichero ACADEMY

Log BERLINALE:



```
berlinale - Bloc de notas
Archivo  Edición  Formato  Ver  Ayuda

SQL*Loader: Release 10.2.0.1.0 - Production on Dom May 8 20:13:01 2011
Copyright (c) 1982, 2005, Oracle. All rights reserved.

Archivo de Control:  berlinale.ctl
Juego de caracteres WE8MSWIN1252 especificado para todas las entradas.

Archivo de datos:      berlinale_2004.csv
Archivo de Errores:    berlinale_2004.bad
Desechar Archivo:     ninguno especificado

(Permittir todos los registros desechados)

Número a cargar: ALL
Número a ignorar: 1
Errores permitidos: 50
Matriz de enlace:     64 filas, máximo de 256000 bytes
Continuación:         ninguno especificado
Ruta de acceso utilizada: Convencional

Tabla TMP_BERLINALE, cargada de cada registro lógico.
Opción INSERT activa para esta tabla: REPLACE

-----
Nombre columna      Posición  Long  Term Entorno Tipo de dato
-----
FILM_TITLE          FIRST     *    :    CHARACTER
LOCALIZED_TITLE     NEXT     *    :    CHARACTER
SORT_TITLE          NEXT     *    :    CHARACTER
YEAR                NEXT     *    :    CHARACTER
COUNTRY             NEXT     *    :    CHARACTER
AWARD               NEXT     *    :    CHARACTER
CATEGORY            NEXT     *    :    CHARACTER
WINNER              NEXT     *    :    CHARACTER
NOMINEES            NEXT     *    :    CHARACTER
BERLINALE           NEXT     *    :    CHARACTER
ITEM                NEXT     *    :    CHARACTER

Tabla TMP_BERLINALE:
  404 Filas se ha cargado correctamente.
  0 Filas no cargada debido a errores de datos.
  0 Filas no cargada porque todas las cláusulas WHEN han fallado.
  0 Filas no cargada porque todos los campos eran nulos.

Espacio asignado a matriz de enlace:      181632 bytes (64 filas)
Bytes de buffer de lectura: 1048576

Total de registros lógicos ignorados:      1
Total de registros lógicos leídos:         404
Total de registros lógicos rechazados:      0
Total de registros lógicos desechados:      0

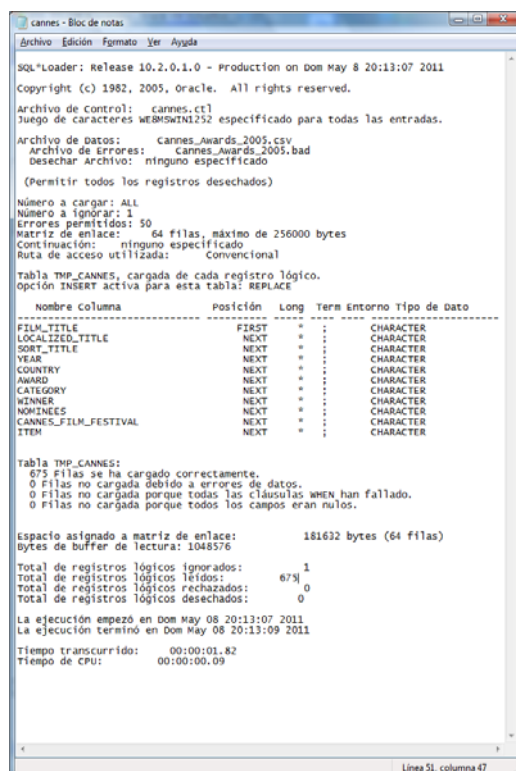
La ejecución empezó en Dom May 08 20:13:01 2011
La ejecución terminó en Dom May 08 20:13:02 2011

Tiempo transcurrido: 00:00:01.12
Tiempo de CPU:      00:00:00.11

Línea 1, columna 1
```

Fig.8: Resultado carga fichero BERLINALE

Log CANNES:



```
cannes - Bloc de notas
Archivo  Edición  Formato  Ver  Ayuda

SQL*Loader: Release 10.2.0.1.0 - Production on Dom May 8 20:13:07 2011
Copyright (c) 1982, 2005, Oracle. All rights reserved.

Archivo de Control:  cannes.ctl
Juego de caracteres WE8MSWIN1252 especificado para todas las entradas.

Archivo de datos:      cannes_awards_2005.csv
Archivo de Errores:    cannes_awards_2005.bad
Desechar Archivo:     ninguno especificado

(Permittir todos los registros desechados)

Número a cargar: ALL
Número a ignorar: 1
Errores permitidos: 50
Matriz de enlace:     64 filas, máximo de 256000 bytes
Continuación:         ninguno especificado
Ruta de acceso utilizada: Convencional

Tabla TMP_CANNES, cargada de cada registro lógico.
Opción INSERT activa para esta tabla: REPLACE

-----
Nombre columna      Posición  Long  Term Entorno Tipo de dato
-----
FILM_TITLE          FIRST     *    :    CHARACTER
LOCALIZED_TITLE     NEXT     *    :    CHARACTER
SORT_TITLE          NEXT     *    :    CHARACTER
YEAR                NEXT     *    :    CHARACTER
COUNTRY             NEXT     *    :    CHARACTER
AWARD               NEXT     *    :    CHARACTER
CATEGORY            NEXT     *    :    CHARACTER
WINNER              NEXT     *    :    CHARACTER
NOMINEES            NEXT     *    :    CHARACTER
CANNES_FILM_FESTIVAL NEXT     *    :    CHARACTER
ITEM                NEXT     *    :    CHARACTER

Tabla TMP_CANNES:
  675 Filas se ha cargado correctamente.
  0 Filas no cargada debido a errores de datos.
  0 Filas no cargada porque todas las cláusulas WHEN han fallado.
  0 Filas no cargada porque todos los campos eran nulos.

Espacio asignado a matriz de enlace:      181632 bytes (64 filas)
Bytes de buffer de lectura: 1048576

Total de registros lógicos ignorados:      1
Total de registros lógicos leídos:         675
Total de registros lógicos rechazados:      0
Total de registros lógicos desechados:      0

La ejecución empezó en Dom May 08 20:13:07 2011
La ejecución terminó en Dom May 08 20:13:09 2011

Tiempo transcurrido: 00:00:01.82
Tiempo de CPU:      00:00:00.09

Línea 51, columna 47
```

Fig.9: Resultado carga fichero CANNES

Log CESARS:

```
cesars - Bloc de notas
Archivo Edición Formato Ver Ayuda

SQL*Loader: Release 10.2.0.1.0 - Production on Dom May 8 20:13:13 2011
Copyright (c) 1982, 2005, Oracle. All rights reserved.

Archivo de Control: cesars.ctl
Juego de caracteres WE8MSWIN1252 especificado para todas las entradas.

Archivo de Datos: cesar_Awards_2005.csv
Archivo de Errores: cesar_Awards_2005.bad
Desechar Archivo: ninguno especificado
(Permittir todos los registros desechados)

Número a cargar: ALL
Número a ignorar: 1
Errores permitidos: 50
Matriz de enlace: 64 filas, máximo de 256000 bytes
Continuación: ninguno especificado
Ruta de acceso utilizada: Convencional

Tabla TMP_CESARS, cargada de cada registro lógico.
Opción INSERT activa para esta tabla: REPLACE

-----
Nombre columna      Posición  Long  Term Entorno Tipo de dato
-----
FILM_TITLE           FIRST     *    ;    CHARACTER
LOCALIZED_TITLE      NEXT     *    ;    CHARACTER
SORT_TITLE           NEXT     *    ;    CHARACTER
YEAR                 NEXT     *    ;    CHARACTER
COUNTRY              NEXT     *    ;    CHARACTER
AWARD                NEXT     *    ;    CHARACTER
CATEGORY             NEXT     *    ;    CHARACTER
WINNER               NEXT     *    ;    CHARACTER
NOMINEES             NEXT     *    ;    CHARACTER
CESARS               NEXT     *    ;    CHARACTER
ITEM                 NEXT     *    ;    CHARACTER

Tabla TMP_CESARS:
2457 Filas se ha cargado correctamente.
0 Filas no cargada debido a errores de datos.
0 Filas no cargada porque todas las cláusulas WHEN han fallado.
0 Filas no cargada porque todos los campos eran nulos.

Espacio asignado a matriz de enlace: 181632 bytes (64 filas)
Bytes de buffer de lectura: 1048576

Total de registros lógicos ignorados: 1
Total de registros lógicos leídos: 2457
Total de registros lógicos rechazados: 0
Total de registros lógicos desechados: 0

La ejecución empezó en Dom May 08 20:13:13 2011
La ejecución terminó en Dom May 08 20:13:22 2011

Tiempo transcurrido: 00:00:09.14
Tiempo de CPU: 00:00:00.08

-----
Línea 1, columna 1
```

Fig.10: Resultado carga fichero CESARS

Log MTV:

```
mtv - Bloc de notas
Archivo Edición Formato Ver Ayuda

SQL*Loader: Release 10.2.0.1.0 - Production on Dom May 8 20:13:32 2011
Copyright (c) 1982, 2005, Oracle. All rights reserved.

Archivo de Control: mtv.ctl
Juego de caracteres WE8MSWIN1252 especificado para todas las entradas.

Archivo de Datos: MTV_Movie_Awards_2006.csv
Archivo de Errores: MTV_Movie_Awards_2006.bad
Desechar Archivo: ninguno especificado
(Permittir todos los registros desechados)

Número a cargar: ALL
Número a ignorar: 1
Errores permitidos: 50
Matriz de enlace: 64 filas, máximo de 256000 bytes
Continuación: ninguno especificado
Ruta de acceso utilizada: Convencional

Tabla TMP_MTV, cargada de cada registro lógico.
Opción INSERT activa para esta tabla: REPLACE

-----
Nombre columna      Posición  Long  Term Entorno Tipo de dato
-----
ORIGINAL_TITLE       FIRST     *    ;    CHARACTER
LOCALIZED_TITLE      NEXT     *    ;    CHARACTER
SORT_TITLE           NEXT     *    ;    CHARACTER
YEAR                 NEXT     *    ;    CHARACTER
COUNTRY              NEXT     *    ;    CHARACTER
AWARD                NEXT     *    ;    CHARACTER
CATEGORY             NEXT     *    ;    CHARACTER
WINNER               NEXT     *    ;    CHARACTER
NOMINEES             NEXT     *    ;    CHARACTER
MTV_MOVIE_AWARDS     NEXT     *    ;    CHARACTER
ITEM                 NEXT     *    ;    CHARACTER

Registro 617: Rechazado - Error en tabla TMP_MTV.
ORA-00001: restricción única (PR_PK_PR_PREMIOS) violada
ORA-06512: en "PR_PROC_CARGA", línea 61
ORA-06512: en "PR_AFTER_MTV", línea 12
ORA-04088: error durante la ejecución del disparador 'PR_AFTER_MTV'

Tabla TMP_MTV:
910 Filas se ha cargado correctamente.
1 fila no cargada debido a errores de datos.
0 Filas no cargada porque todas las cláusulas WHEN han fallado.
0 Filas no cargada porque todos los campos eran nulos.

Espacio asignado a matriz de enlace: 181632 bytes (64 filas)
Bytes de buffer de lectura: 1048576

Total de registros lógicos ignorados: 1
Total de registros lógicos leídos: 911
Total de registros lógicos rechazados: 1
Total de registros lógicos desechados: 0

La ejecución empezó en Dom May 08 20:13:32 2011
La ejecución terminó en Dom May 08 20:13:34 2011

Tiempo transcurrido: 00:00:02.34
Tiempo de CPU: 00:00:00.12

-----
Línea 1, columna 1
```

Fig.11: Resultado carga fichero MTV

Informes solicitados por la ACA

Número de premios conseguidos por un actor/actriz por año y festival

		Número de premios conseguidos				
Elementos de Página:						
		Nº premios				
		Academy Award	Berlinale	Cannes Festival	César	MTV Movie Awards
Adrien Brody	2002	1			1	2
Agnès Jaoui	1997	1			1	1
Al Pacino	1992	1				1
Alain Delon	1984	1			1	1

Fig.12: Resultado primer informe en primer formato

Elementos de Página:		Nº premios	
Nomnominado: Alicia Silverstone ▼		MTV Movie Awards	Total premios
1994	1	1	
1996	2	2	

Fig.13: Resultado primer informe en segundo formato

Número de premios conseguidos por un actor/actriz por año, festival y categoría

		Número de premios conseguidos por un actor/a						
Elementos de Página:								
		Nº premios						
		Academy Award				Berlinale		
		Best Actor	Best Actress	Best Supporting Actor	Best Supporting Actress	Best Actor	Best Actress	Best Actor
Adrien Brody	2002	1						
Agnès Jaoui	1997	1						
Al Pacino	1992	1						
Alain Delon	1984							
Alan Arkin	2006			1				
Albert Finney	1984					1		
Alberto Sordi	1972					1		
Alec Guinness	1957	1						
Alfredo Landa, Francisco Rabal	1984							1

Fig.14: Resultado segundo informe en primer formato

Elementos de Página:		Nº premios			
Nomnominado: Denzel Washington ▼		Academy Award		Berlinale	Movie Awards
		Best Actor	Best Supporting Actor	Best Actor	Best Actor
1989			1		
1992				1	
1993					1
1999				1	
2001	1				

Fig.15: Resultado segundo informe en segundo formato

Número de premios conseguidos por un director por año, festival y película

Número de premios conseguidos por un director por año, festival y película				
Elementos de Página:		Nº premios		
		Academy Award	Berlinale	Cannes Festival
Abdel Kechiche				1
2003				1
L'esquive				1
Alain Kurosawa		1		
1959		1		
Kakushi toride no san-akunin		1		
Alan Cavalier				1
1905				1
Thérèse				1
Alan Corneau				1
1991				1
Tous les matins du monde				1
Alan Parris				2
1977				1
Providence				1
1993				1
Smoking/No Smoking				1

Fig.16: Resultado tercer informe en primer formato

Elementos de Página:	
Nomnominado:	John Ford ▼
Nº premios	
Academy Award	
1935	1
The Informer	1
1940	1
The Grapes of Wrath	1
1941	1
How Green Was My Valley	1
1952	1
The Quiet Man	1

Fig.17: Resultado tercer informe en segundo formato

Número de premios conseguidos por año, festival, categoría y película

Número de premios conseguidos por año, festival, categoría y película						
Elementos de Página:						
		Best Costume Design, Black-and-White	Best Costume Design, Color	Best Dance Direction	Best Dance Sequence	Best Dance Sequence Performance
Romeo i Zhuletta						
1956						
Academy Award		1	1			
Anastasia		1	1			
Around the World in 80 Days						
Crashing the Water Barrier						
Grand						1
La Strada						
Le Ballon rouge						
Le Monde du silence						
Lust for Life						
Mister Magoo's Puddle Jumper						
[no specific film title]						
Somebody Up There Likes Me						
The Bespoke Overcoat						
The Brave One						
The King and I			1			
The Man who Knew Too Much						
The Solid Gold Cadillac		1				
The Ten Commandments						
The True Story of the Civil War						
Written on the Wind						
Berlinale						1
Autumn Leaves						1
Donatella						
Invitation To The Dance						
La sorcière						
Richard III.						
Trapeze						

Fig.18: Resultado cuarto informe

Dada una fecha indicada por parámetro, número de ed				
Elementos de Página: <input type="text"/>				
	Años distintos	Actores Nominados distintos	Actrices nominadas distintas	Películas nominadas distintas
Academy Award	68	369	353	3.487
Berlinale	44	36	38	194
Cannes Festival	48	45	52	320
César	21	139	137	717
MTV Movie Awards	4	23	36	130

Para cada año, número de premios que ganó					Elementos de Página: Año: 1928 ▼		
Elementos de Página:		Nº premios				Nº premios	
		Academy Award	Berlinale	Cannes Festival		Academy Award	Total premios
1928		10				1	10
	[no specific film title]	1				1	1
	Street Angel	1				1	1
	Sunrise	3				3	3
	Tempest	1				1	1
	The Circus	1				1	1
	The Dove	1				1	1
	The Jazz Singer	1				1	1
	The Last Command	1				1	1
	The Way of All Flesh	1				1	1
	Two Arabian Knights	1				1	1
	Underworld	1				1	1
	Wings	2				2	2
	7th Heaven	3				3	3
1929		7				1	7
	Cocquette	1				1	1
	In Old Arizona	1				1	1
	The Bridge of San Luis Rey	1				1	1
	The Broadway Melody	1				1	1
	The Divine Lady	1				1	1
	The Patriot	1				1	1
	White Shadows in the South Seas	1				1	1
1930		6				1	6
	All Quiet on the Western Front	2				2	2

Página 54 de 59

Para cada año, indicar las películas con mayor éxito y mayor fracaso, según unos criterios que definan cada concepto

	Total nominaciones	Premios conseguidos	Ratio
1928	41	18,00	0,44
Street Angel	1	1,00	1,00
Tempest	1	1,00	1,00
The Circus	1	1,00	1,00
The Dove	1	1,00	1,00
The Way of All Flesh	1	1,00	1,00
Two Arabian Knights	1	1,00	1,00
Underworld	1	1,00	1,00
Wings	2	2,00	1,00
Sunrise	4	3,00	0,75
7th Heaven	5	3,00	0,60
The Last Command	2	1,00	0,50
[no specific film title]	3	1,00	0,33
The Jazz Singer	3	1,00	0,33
A Ship Comes In	1	0,00	0,00
Chang: A Drama of the Wilderness	1	0,00	0,00
Glorious Betsy	1	0,00	0,00
Sadie Thompson	2	0,00	0,00
Sorrell and Son	1	0,00	0,00
Speedy	1	0,00	0,00
The Crowd	2	0,00	0,00
The Devil Dancer	1	0,00	0,00
The Magic Flame	1	0,00	0,00
The Moose	1	0,00	0,00
The Patent Leather Kid	1	0,00	0,00
The Private Life of Helen of Troy	1	0,00	0,00
The Racket	1	0,00	0,00
1929	46	7,00	0,15
Coquette	1	1,00	1,00

Fig.22: Resultado primer informe adicional

Número de premios por categoría ganados por el país más galardonado, dejando de lado a EE.UU.

Elementos de Página:		
	Nº Premios	OrdenTotalPorPremios
France	699	1,00

Fig.23: Resultado segundo informe adicional en primer formato

Elementos de Página:	
	Nº Premios
France	699
Adventure and Detective Film	1
AGICOA-Preis Der Blaue Engel für den besten europäischen Film	2
Artistic Contribution	1
Best Actor	42
Best Actress	42
Best Breakthrough Female	24
Best Breakthrough Male	24
Best Director	37
Best Documentary	1
Best Documentary, Feature	5
Best Dressed	22
Best Film	37
Best Film Editing	31
Best Foreign Language Film	9
Best Music, Musical Picture	2
Best Short Film	1

Fig.24: Resultado segundo informe adicional en segundo formato

Analizar una posible relación entre ganar un premio en un festival y ganar el mismo, o equivalente, en otro festival

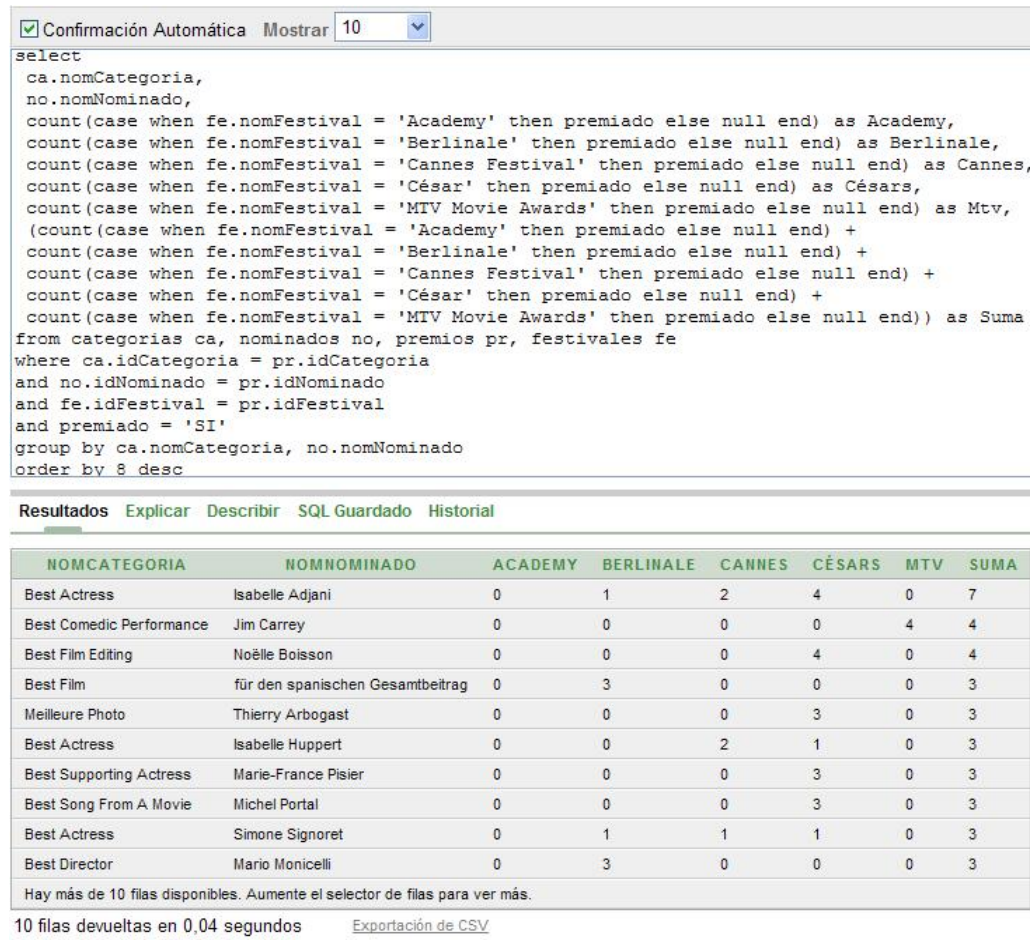


Fig.25: Consulta y datos obtenidos para tercer informe adicional

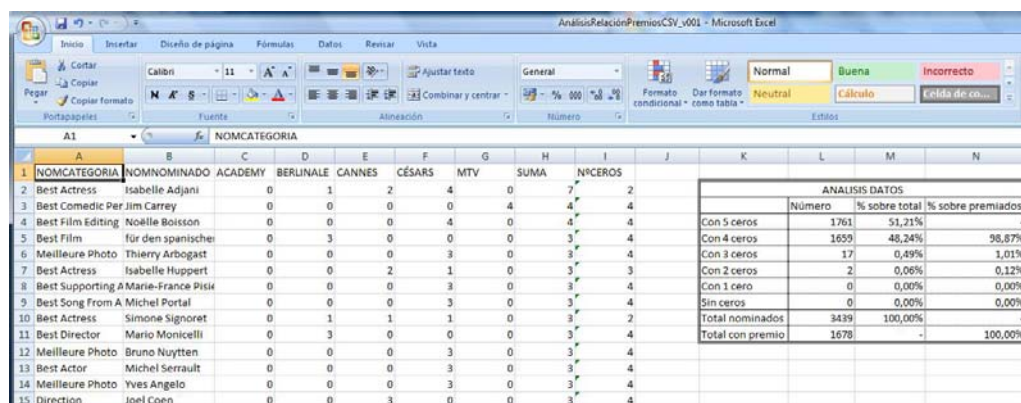


Fig.26: Análisis para tercer informe adicional

Conclusiones

La asignación de cualquier proyecto a un profesional siempre pone de manifiesto la confianza que en el mismo se deposita, con la certeza que el resultado obtenido será el esperado.

El propio profesional se enfrenta en ese instante al reto, tanto personal como profesional, de retornar la confianza puesta en él con la entrega de un producto que satisfaga los requisitos que el proyecto presentaba en un inicio e, incluso, que ofrezca nuevas expectativas de cara a la explotación del mismo en un futuro.

En el caso específico del TFC debo, como estudiante de la Ingeniería Técnica en Informática de Sistemas, ser capaz de poner en práctica los conocimientos adquiridos durante mi formación y ofrecer un producto que cumpla con los requisitos que en el enunciado del TFC se exponen. He de reconocer que personalmente pienso que dicho objetivo ha sido logrado, el cual se ha plasmado en el presente documento, en el que se puede observar la aplicación práctica de gran parte de los conocimientos adquiridos durante mi estancia en la UOC.

A su vez, de forma adicional, se ha logrado adquirir nuevos conocimientos en el ámbito de las Bases de Datos que se desconocían antes del inicio del proyecto. Los mismos hacen referencia a los almacenes de datos y cubos multidimensionales principalmente y han sido puestos en práctica durante la elaboración del presente proyecto, como si de un caso real se tratara. Precisamente dotándolo de dicha característica, se ha incluido una serie de consejos que se creen oportunos comentar cuyo objetivo es optimizar los resultados de cara a la explotación del proyecto en un futuro.

Líneas de evolución futura

Todo proyecto informático consta de diferentes fases. Tras la consecución del producto inicial, a partir de los requisitos expuestos por el cliente, el proyecto no finaliza sino que debe continuar con las fases de mantenimiento y evolución. Para esta segunda etapa deben estudiarse posibles nuevas características o funcionalidades a incluir al producto inicial obtenido, para las que se deben tener en cuenta todos los acontecimientos sucedidos durante la elaboración del proyecto, así como nuevos requisitos surgidos durante la misma que, o bien ofrezcan nuevas funcionalidades o bien faciliten la explotación del proyecto optimizando tareas.

Siguiendo las directrices comentadas, a continuación se exponen una serie de consejos que se cree conveniente estudiarlos, a fin de mejorar el producto inicial en diferentes aspectos:

- La aportación inicial de los datos en ficheros Excel precisa de una depuración previa a la carga de los mismos. Este proceso podría optimizarse con la introducción de los datos, a cargar posteriormente en el Almacén de Datos, en tablas relacionales confeccionadas a tal efecto, preferiblemente en el mismo entorno de Oracle para mantener homogeneidad. Dicha estructura facilitaría el traslado de información de la

BBDD relacional al Almacén de Datos, favoreciendo la integración y administración de los datos, así como la optimización de los procesos internos que se realizan sobre dicha información.

- Estudiar posibles adiciones o cambios en campos de las tablas actuales que pudieran resultar interesantes, así como el formato/significado de los actuales. Como iniciales se propone el estudio de la inclusión de posibles jerarquías en datos susceptibles de ello, como los países, agrupándolos en continentes, o los años, agrupándolos en lustros y/o décadas. Los mismos pueden ofrecer mayores posibilidades de análisis en una futura versión del producto inicial.

Los puntos que se acaban de exponer deben ser tratados con la ACA, a fin de estudiar la viabilidad de los mismos y deben tomarse como consejos para la optimización de tareas y dotación de mayor importancia a la información guardada en el Almacén de Datos.

Glosario

Gb: Abreviatura de Gigabyte. Unidad de medida de almacenamiento

Tb: Abreviatura de Terabyte. Unidad de medida de almacenamiento que equivale a 1024Gb

VirtualBox: Software de virtualización creado por Oracle

Diagrama de Gantt: gráfico de visualización temporal de asignación de tareas

ETL: Abreviatura de Extract, Transform and Load. Proceso de extracción, transformación y carga de datos de una fuente a un origen establecido.

Disparador: procedimiento que se ejecuta cuando se cumple una condición sobre una tabla

Procedimiento: programa almacenado en una base de datos

Fichero CSV: fichero de texto plano con registros en su interior en el que se utiliza un indicador como separador de campos

Bibliografía

Material Web:

- http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos
- http://www.sinnexus.com/business_intelligence/datawarehouse.aspx
- <http://www.monografias.com/trabajos17/data-warehouse/data-warehouse.shtml>
- <http://www.sqlmax.com/dataw1.asp>
- http://www.programacion.com/articulo/data_warehousing_201
- <http://www.oracle.com/lad/technologies/datawarehousing/index.html>
- <http://www.dwinfocenter.org/>

- <http://es.wikipedia.org/wiki/OLAP>
- <http://es.wikipedia.org/wiki/ROLAP>
- [http://es.wikipedia.org/wiki/Top-down y Bottom-up](http://es.wikipedia.org/wiki/Top-down_y_Bottom-up)
- [http://es.wikipedia.org/wiki/Extract, transform and load](http://es.wikipedia.org/wiki/Extract,_transform_and_load)
- <http://es.wikipedia.org/wiki/Metadato>
- [http://es.wikipedia.org/wiki/Modelo de datos](http://es.wikipedia.org/wiki/Modelo_de_datos)
- [http://es.wikipedia.org/wiki/Tabla de hechos](http://es.wikipedia.org/wiki/Tabla_de_hechos)
- [http://es.wikipedia.org/wiki/Tabla de dimensi%C3%B3n](http://es.wikipedia.org/wiki/Tabla_de_dimensi%C3%B3n)
- http://www.epidataconsulting.com/tikiwiki/tiki-read_article.php?articleId=64