Construcción y explotación de un almacén de datos de planificación hidrológica

Lucía Varo Ruiz
ETIG

José Ángel Martín Carballo
10 de Enero del 2011

RESUMEN

Este documento describe el plan de trabajo, el análisis y la implementación del trabajo de fin de carrera de los estudios de Ingeniería Técnica de Informática de Gestión, en concreto del área de almacén de datos.

El objetivo de este trabajo es familiarizarse con la arquitectura de un DW: el diseño de un DW, tratamiento de datos, así como conocer las herramientas de que dispone el mercado para trabajar con estos modelos y realizar los análisis dirigidos a usuarios finales.

El proyecto constará de 3 apartados bien diferenciados:

- Diseño y generación de la base de datos
- Procesos ETL
- Generación de informes

El trabajo de fin de carrera se ha desarrollado con la base de datos

- ORACLE 10g Release 2 (10.2) Express Edition

trabajando con las siguientes herramientas:

- "*Oracle Database XE"* → herramienta utilizada para administrar la base de datos, crear tablas, vistas y otros objetos de la base de datos,.. etc.
- "*Oracle BI Discoverer Administrator"*
- "*Oracle BI Discoverer Desktop"* → herramienta utilizada para la generación de informes

todo esto bajo la maquina virtual "*Oracle VM VirtualBox*" .

Para realizar los documentos entregados, se han utilizado:

- Microsoft Office
- Microsoft Project 2003
- Visio
- Argo
- DiaPortable
- Adobe
- Camtasia Studio 7

INDICE DE CONTENIDOS

1.- MEMORIA 5

1.1.- CAPITULO 1 – INTRODUCCIÓN –..... 5
1.1.1. Justificación del TFC. 5
1.1.2. Objetivos TFC. 5
1.1.3. Enfoque y método seguido. 6
1.1.4. Planificación del proyecto. 7
1.1.5. Productos obtenidos. 9
1.1.6. Breve descripción capítulos siguientes. 10
1.2.- CAPÍTULO 2 – ANÁLISIS – 11
1.2.1. Diagramas de casos de usos 11
1.2.2. Diagrama modelo conceptual 13
1.3.- CAPÍTULO 3 – DISEÑO –..... 15
1.3.1. Diagrama de la arquitectura del software 15
1.3.2. Diagrama de la arquitectura del hardware 16
1.3.3. Diseño de la base de datos y diagrama del modelo físico 18
1.3.4. Diseño y descripción de los informes 21
1.4.- CAPÍTULO 4 – CAPTURAS DE PANTALLA – 32
1.4.1. Creación usuario 32
1.4.2. Creación modelo con Discoverer Administrator 33
1.4.3. Creación informes con Discoverer Desktop 40
1.5.- CAPÍTULO 5 – CONCLUSIONES – 42
1.6.- CAPÍTULO 6 – LÍNEAS DE EVOLUCIÓN –..... 43
1.6.1. Nuevas tendencias en el mundo DW. El Datawarehouse 2.0. 43

2.- GLOSARIO 46

3.- BIBLIOGRAFÍA 49

4.- ANEXOS 50

4.1.- CREACIÓN TABLAS AUXILIARES 50
4.2.- CREACIÓN TABLAS BBDD..... 51
4.3.- CARGA BASE DE DATOS (PROCESOS ETL) 52
4.3.1. Proceso ETL 52
4.3.2. Tratamiento hojas Excel 53
4.3.3. Problemas iniciales 54
4.3.4. Carga tablas finales 57
4.3.5. Automatización carga..... 59
4.3.6. Automatización recarga 59

INDICE DE FIGURAS

Figura 1: Implementación DHW 6
Figura 2: Diagrama de Gannt 8
Figura 3: Caso de uso “Usuario Final” 11
Figura 4: Caso de uso “Usuario Administrador” 12
Figura 5: Caso de uso “Usuario Analista” 12
Figura 6: Hecho VOLUMEN 14
Figura 7: Hecho VARIACION_SEMANAL 14
Figura 8: Arquitectura del software 15
Figura 9: Arquitectura del hardware 17
Figura 10: Modelo de relaciones de la BBDD 20
Figura 11: Informe1 (Hm^3) 21
Figura 12: Informe1 (Ml) 22
Figura 13: Literal mes 22
Figura 14: Informe1_2 (Hm^3) 23
Figura 15: Informe2 (Ml) 23
Figura 16: Informe3 (Ml) 24
Figura 17: Informe4 (Ml) 24
Figura 18: Informe5 (Hm^3) 25
Figura 19: Gráfico informe5 (Ml) 25
Figura 20: Informe6 (Ml) 26
Figura 21: Informe7 (Hm^3) 26
Figura 22: Informe8_1 (Hm^3) 27
Figura 23: Informe8_2 (Hm^3) 27
Figura 24: Informe8_2 (Ml) 28
Figura 25: Informe9 30
Figura 26: Gráfico Informe9 31
Figura 27: Ejecución script desde ORACLE DEE 32
Figura 28: Conexión Discoverer 33
Figura 29: Creación EUL 34
Figura 30: Parámetros EUL 34
Figura 31: Seleccionar usuario 35
Figura 32: Aviso de creación correcta del EUL 35
Figura 33: Inicio de creación de una nueva área de negocio 36
Figura 34: Creación de una nueva área de negocio (Cont.) 36
Figura 35: Selección de usuario para el área de negocio. 37
Figura 36: Selección de tablas para el área de negocio 37
Figura 37: Selección de parámetros para el área de negocio 38
Figura 38: Selección de nombre para el área de negocio 38
Figura 39: Visualización de las tablas del área de negocio creada. 39
Figura 40: Conexión Discoverer Desktop 40
Figura 41: Creación nuevo libro de trabajo 40
Figura 42: Creación nuevo libro de trabajo (Cont.) 41
Figura 43: CIF(Corporate Information Factory) 43
Figura 44: Corporate Information Factory 45
Figura 45: Ejecución script creación tablas desde MS-DOS 51
Figura 46: Proceso ETL seguido 52
Figura 47: Ejecución script verificación datos desde MS-DOS 56
Figura 48: Scripts en la MV 59

1.- Memoria

1.1.- Capitulo 1 – Introducción –

1.1.1. Justificación del TFC.

El punto de partida es la petición de generar un nuevo DM para facilitar el trabajo a la CHNE.

La Confederación Hidrográfica Nordeste (CHNE), es la entidad encargada de la gestión del agua e infraestructuras hidráulicas del nordeste de la península ibérica. Realizan un seguimiento periódico del estado de los diferentes embalses que están bajo su responsabilidad, comprobar tendencias y predecir periodos de sequía con el fin que las autoridades competentes tomen las decisiones hidrológicas correctas.

Esta información es suministrada de forma manual, recogidas en Hojas Excel, con el gasto de recursos humanos innecesarios para su explotación.

Nos piden que con la información obtenida de los documentos Excel, se cree un almacén de datos como repositorio de información. Se deberá de diseñar y crear un DW y desarrollar un entorno de explotación estadística que incluya una serie de informes predefinidos para facilitarles su uso.

Importante es el hecho que los volúmenes de agua se han de poder ver tanto en Hm^3 (hectómetros cúbicos) como en Ml (millones de litros), a elección de los técnicos usuarios del Data Mart.

Inicialmente nos proporcionaran una serie de hojas Excel, por año/mes, con el fin de cargar la información en el nuevo DM. Sin embargo, debemos de tener en cuenta que este DM se deberá en un futuro ampliar con la información obtenida mes a mes.

1.1.2. Objetivos TFC.

La finalidad fundamental de este trabajo es el familiarizarse con la estructura de un almacén de datos desarrollando un proyecto como caso práctico, desde la planificación inicial a su desarrollo e implantación final del producto.

El diseño deberá de estar orientado a un almacén de datos físico ROLAP, creando las dimensiones, atributos y hechos necesarios.

A su vez se obtendrán los siguientes conocimientos:

- Profundizar en el conocimiento de las bases de datos y la gestión de la información
- Transformar una base de datos relacional a multidimensional, familiarizándose con la estructura de un Data Warehouse (DW)
- Obtener experiencia en el diseño, construcción y explotación de un almacén de datos
- Analizar y comparar herramientas de mercado que trabajan con modelos físicos de un DW
- Realizar análisis estadísticas dirigidas a los usuarios finales del DW

1.1.3. Enfoque y método seguido.

Si se plantea la metodología propuesta por SAS Institute ("Rapid Warehousing Methodology"), se destaca que las fases de implementación de un DWH son iterativas, basadas en el desarrollo incremental del proyecto, en 5 etapas:

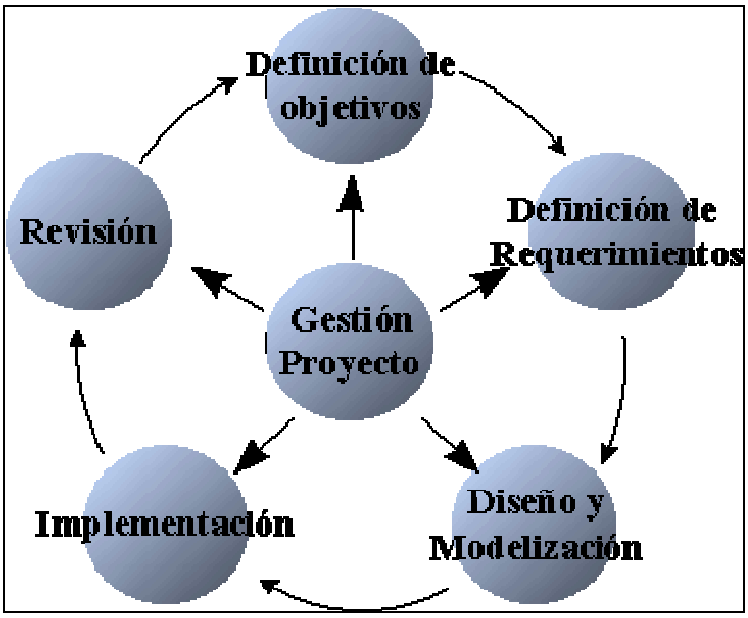


Figura 1: Implementación DHW

- Definición de los objetivos
- Definición de los requerimientos de información
- Diseño y modelización → En esta fase se identificarán las fuentes de los datos (sistema operacional, fuentes externas,..) y las transformaciones necesarias para, a partir de dichas fuentes, obtener el modelo lógico de datos del Data Warehouse. Este modelo estará formado por entidades y relaciones que permitirán resolver las necesidades de negocio de la organización.

  El modelo lógico se traducirá posteriormente en el modelo físico de datos que se almacenará en el Data Warehouse y que definirá la arquitectura de almacenamiento del Data Warehouse adaptándose al tipo de explotación que se realice del mismo.

- Implementación → Extracción, carga y explotación del DWH
- Revisión → La construcción del Data Warehouse no finaliza con la implantación del mismo, sino que es una tarea iterativa en la que se trata de incrementar su alcance aprendiendo de las experiencias anteriores.

  Después de implantarse, debería realizarse una revisión del Data Warehouse planteando preguntas que permitan, después de los seis o nueve meses posteriores a su puesta en marcha, definir cuáles serían los aspectos a mejorar o potenciar en función de la utilización que se haga del nuevo sistema.

Se ha seguido esta metodología para con el fin de llevar a cabo este trabajo.

1.1.4. Planificación del proyecto.

Se muestra a continuación la planificación detallada del proyecto:

Nombre tarea	Inicio	Final
Vacaciones	**.....**	**26/09/10**
T1: Lecturas iniciales	**27/09/10**	**29/09/10**
T2: Instalación software	**30/09/10**	**03/10/10**
PAC1	**27/09/10**	**05/10/10**
Borrador	27/09/10	03/10/10
Corrección	04/10/10	04/10/10
Revisión/entrega	05/10/10	05/10/10
T3: Diseño DW	**06/10/10**	**22/10/10**
Diseño conceptual	06/10/10	16/10/10
Diseño físico	17/10/10	22/10/10
T4: Estudio extracción datos	**23/10/10**	**08/11/10**
Estudio extracción datos	23/10/10	29/10/10
Pseudocódigo extracción datos	30/10/10	08/11/10
PAC2	**06/10/10**	**08/11/10**
Borrador	06/10/10	01/11/10
Corrección	02/11/10	03/11/10
Revisión	04/11/10	06/11/10
Entrega	07/11/10	08/11/10
T5: Generación y carga BD	**09/11/10**	**24/11/10**
Generación BD	09/11/10	14/11/10
Cargas BD	15/11/10	24/11/10
T6: Estudio y generación de informes	**25/11/10**	**20/12/10**
Estudio herramienta explotación BD	25/11/10	28/11/10
Generación informes	29/11/10	08/12/10
Análisis de la información	09/12/10	20/12/10
PAC3	**09/11/10**	**20/12/10**
Borrador	09/11/10	13/12/10
Corrección	14/12/10	15/12/10
Revisión	16/12/10	18/12/10
Entrega	19/12/10	20/12/10
Presentación final	**21/12/10**	**10/01/11**
Borrador	21/12/10	03/01/11
Corrección	04/01/11	05/01/11
Revisión	06/01/11	08/01/11
Entrega	09/01/11	10/01/11
T7: Evaluación tribunal	**11/01/11**	**27/01/11**
Evaluación tribunal	11/01/11	23/01/11
Debate virtual	24/01/11	27/01/11

Así como el diagrama de Gannt, con las tareas definidas en el apartado anterior:

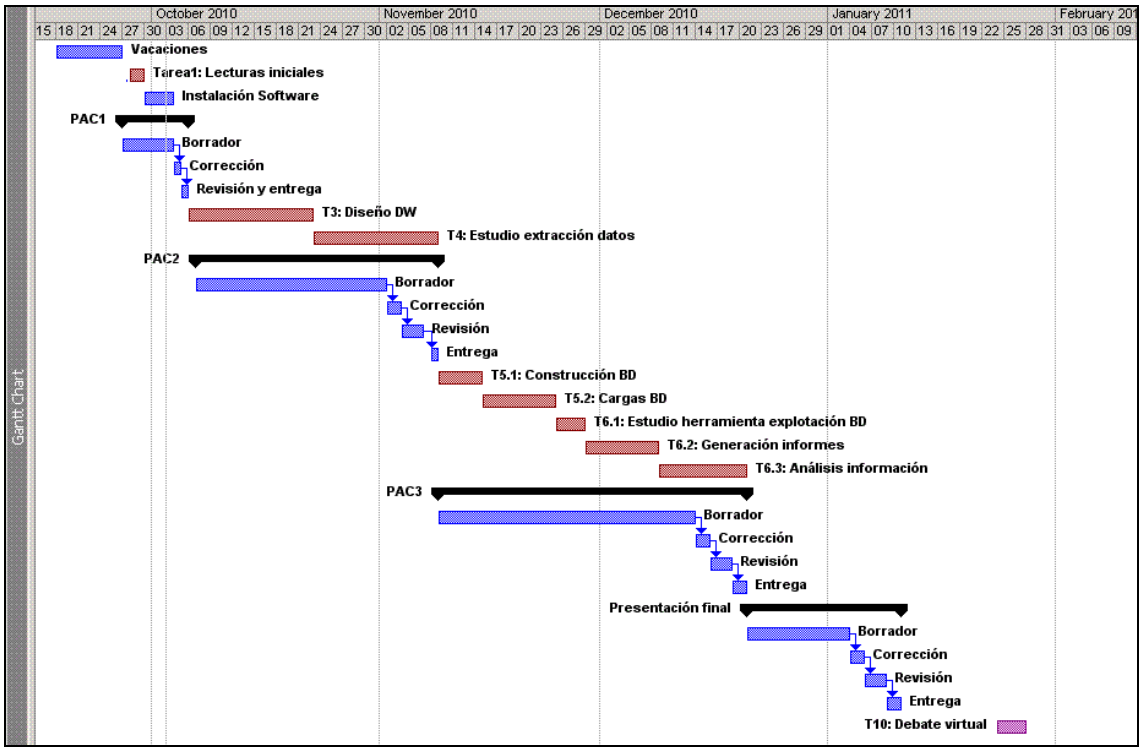


Figura 2: Diagrama de Gannt

1.1.5. Productos obtenidos.

A continuación expondremos los productos obtenidos en este trabajo de fin de carrera, clasificados por tipo:

Nombre	**Tipo**	**Descripción**
Informe1	Libro de trabajo Discoverer Desktop	Informe que muestra la capacidad total versus la capacidad actual por río y año/mes, en Hm^3 y Ml.
Informe1_2	Libro de trabajo Discoverer Desktop	Informe que muestra la capacidad total versus la capacidad actual por río y año/mes, en Hm^3 y Ml (versión matriz).
Informe2	Libro de trabajo Discoverer Desktop	Informe que muestra la capacidad total versus la capacidad actual por embalse y año/mes, en Hm^3 y Ml.
Informe2_2	Libro de trabajo Discoverer Desktop	Informe que muestra la capacidad total versus la capacidad actual por embalse y año/mes, en Hm^3 y Ml (versión matriz).
Informe3	Libro de trabajo Discoverer Desktop	Informe que muestra la capacidad total versus la capacidad actual por comunidad y año/mes, en Hm^3 y Ml.
Informe3_2	Libro de trabajo Discoverer Desktop	Informe que muestra la capacidad total versus la capacidad actual por comunidad y año/mes, en Hm^3 y Ml (versión matriz).
Informe4	Libro de trabajo Discoverer Desktop	Informe que muestra la comparativa anual de la evolución por comunidad de la capacidad agregada de todos sus embalses, en Hm^3 y Ml.
Informe5	Libro de trabajo Discoverer Desktop	Informe que muestra la comparativa de la capacidad media agregada de los embalses de una comunidad entre un año y el anterior, en Hm^3 y Ml.
Informe6	Libro de trabajo Discoverer Desktop	Informe que muestra la aportación (entrada/salida) por río, embalse y año/mes, en Hm^3 y Ml.
Informe7	Libro de trabajo Discoverer Desktop	Informe que muestra el agua diaria en reserva disponible(volumen/día) a toda la confederación hidrográfica por año/mes, en Hm^3 y Ml.
Informe8_1	Libro de trabajo Discoverer Desktop	Informe que muestra el gráfico de evolución del volumen de agua nieve equivalente por año y embalse, en Hm^3 y Ml.
Informe8_2	Libro de trabajo Discoverer Desktop	Informe que muestra los posibles casos de desbordamiento de embalses, en Hm^3 y Ml.
Informe9	Libro de trabajo Discoverer Desktop	Informe que muestra como se puede predecir a partir del Data Mart desarrollos de periodos de sequía, en Hm^3 y Ml.
01_CREATE_Tablas.bat	Fichero .bat	Fichero .bat para creación BBDD
02_CARGA_Tablas.bat	Fichero .bat	Fichero .bat para carga BBDD
03_RECARGA_Tablas.bat	Fichero .bat	Fichero .bat para recarga BBDD
lvaro_memoria	Adobe Reader	Memoria del TFC
lvaro_producte_mv.zip	zip de la MV	Máquina virtual
lvaro_memoria	Fichero .flv	Video de presentación del TFC

1.1.6. Breve descripción capítulos siguientes.

A continuación se definirán brevemente los siguientes capítulos de la memoria.

1.1.6.1. Análisis

Se analizará las necesidades de la empresa, mostrando los diagramas de uso necesarios que detallaran las necesidades del proyecto. En función de lo explicado se expondrá la solución a nivel del modelo conceptual y se justificará su elección.

1.1.6.2. Diseño

Se mostrará los diagramas de arquitectura de software y de hardware, así como el diseño de la base de datos, el diagrama del modelo físico y los informes creados.

1.1.6.3. Capturas de pantalla

Se ha visto oportuno el mostrar las capturas de pantalla, tanto de la creación del usuario de trabajo, de la creación del modelo con Discoverer Administrator, como de la creación de informes con Discoverer Desktop, con el fin que sirva de guía de usuario.

1.1.6.4. Conclusiones

Conclusiones a las que se ha llegado tras la ejecución del TFC.

1.1.6.5. Líneas de evolución futura

Desde hace un par de años Bill Inmon ha hecho referencia a la nueva generación de Data Warehouse nombrada Data Warehouse 2.0. Dicha generación consiste en la evolución del modelo tradicional del DW a un modelo más adaptable y eficiente. En este punto haremos unas breves referencias basadas la visión de Bill Inmon acerca del DW 2.0.

1.1.6.6. Glosario

Glosario de términos.

1.1.6.7. Bibliografía

Bibliografía utilizada en el trabajo de fin de carrera.

1.1.6.8. Anexos

En el anexo de esta memoria hemos añadido cuestiones más técnicas: la creación de las tablas auxiliares, la creación de las tablas finales, así como los procesos ETL generados.

1.2.- Capítulo 2 – Análisis –

1.2.1. Diagramas de casos de usos

En el Lenguaje de Modelado Unificado (UML), un diagrama de casos de uso es una especie de diagrama de comportamiento. El Lenguaje de Modelado Unificado define una notación gráfica para representar casos de uso llamada modelo de casos de uso.

En nuestro entorno de DW, mostrará la relación entre los actores y los casos de uso del sistema. Tendremos los siguientes actores:

- Usuario final
- Usuario administrador
- Usuario analista

El usuario final es el que realiza la importante función de análisis de la información. Es él el que conoce si le es necesario nuevos informes. Por esta razón, será el que realice las peticiones de nuevos informes:

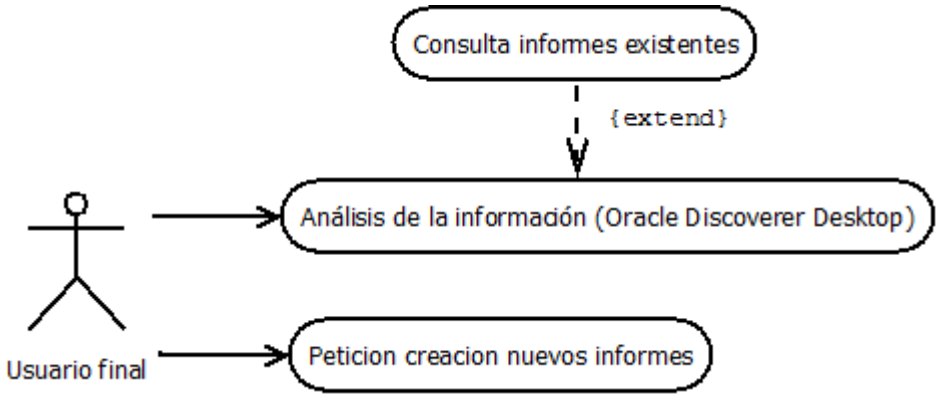


Figura 3: Caso de uso “Usuario Final”

El usuario administrador tiene un rol muy importante. Administra la base de datos, cargándola con la información necesaria para la toma de decisiones de la empresa, gestiona los usuarios y realiza las copias de seguridad:

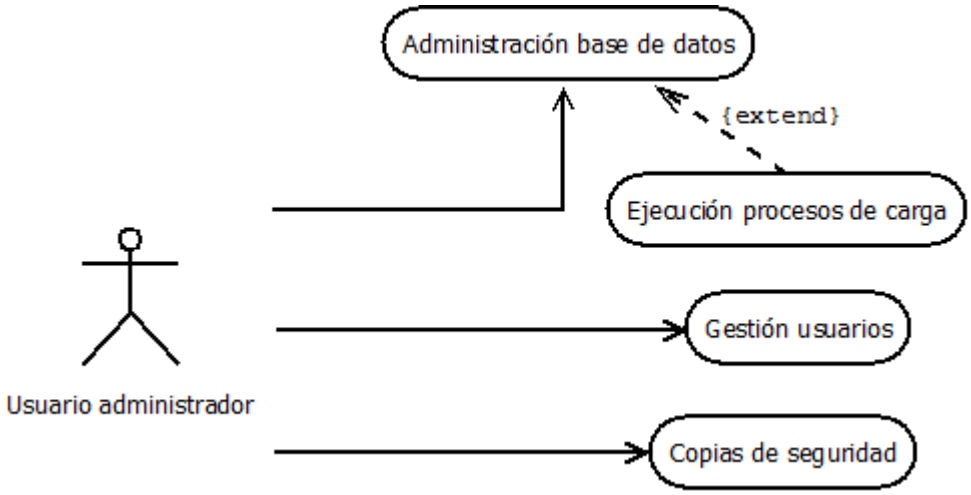


Figura 4: Caso de uso “Usuario Administrador”

El usuario analista es el que recibe la petición del usuario final de generar nuevas áreas de trabajo y/o nuevos informes y ejecuta dicha petición:

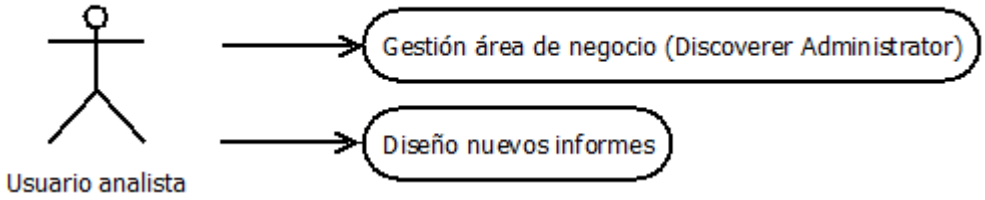


Figura 5: Caso de uso “Usuario Analista”

1.2.2. Diagrama modelo conceptual

El modelo conceptual no es más que un esquema de los datos y de las asociaciones o interrelaciones que se han obtenido a partir del conocimiento de los requerimientos y especificaciones de lo que se quiere obtener al final.

En el diseño conceptual no hemos de considerar si unos datos derivadas se almacenan físicamente o simplemente se calculan cuando hace falta. Esto será una decisión en el diseño físico. En este momento solo hace falta dejar constancia de todo lo que se considere especialmente relevante.

2.1.1.1. Modelo conceptual en una herramienta OLAP

Toda herramienta OLAP ha de ser multidimensional. La multidimensionalidad consiste simplemente a concebir los datos que vamos a analizar en términos de hechos y dimensiones de análisis, de manera que las podamos situar en un espacio n-dimensional.

El modelo multidimensional en el que nos vamos a mover, consta de dos tipos de datos: los *hechos* que queremos analizar y las *dimensiones* que utilizamos para analizarlos.

Un hecho no es más que un objeto de análisis.

Una dimensión representa un punto de vista que utilizaremos en el análisis de datos.

Dentro de una dimensión podemos distinguir grupos de instancias según su tamaño (granularidad).

2.1.1.2. Dimensiones

Las dimensiones encontradas son las siguientes:

- Dimensión **COMUNIDAD**
- Dimensión **EMBALSE**
- Dimensión **RIO**
- Dimensión **TIEMPO**
- Dimensión **ANY_MES**

2.1.1.2. Hechos

En el análisis de la base de datos de este nuevo DWH, el esquema en estrella es el más adecuado: es un modelo de datos que tiene una tabla de hechos (que contiene los datos para el análisis), rodeada de tablas de dimensiones.

- Hecho **VOLUMEN**

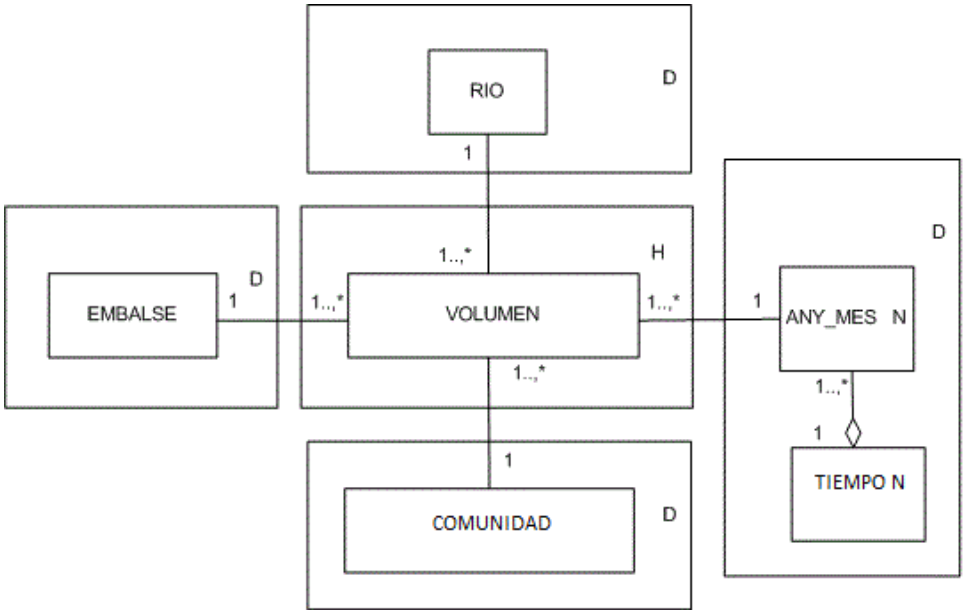


Figura 6: Hecho VOLUMEN

- Hecho **VARIACION _SEMANAL**

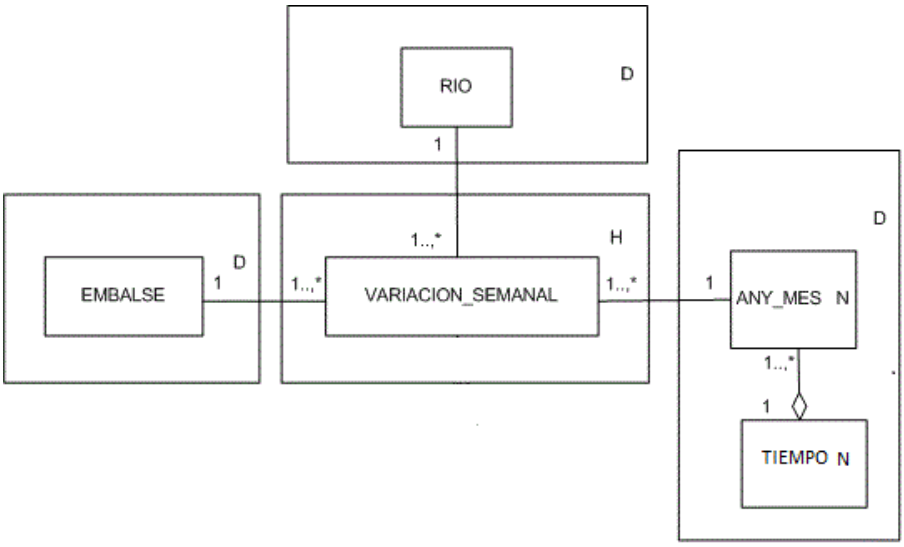


Figura 7: Hecho VARIACION_SEMANAL

1.3.- Capítulo 3 – Diseño –

1.3.1. Diagrama de la arquitectura del software

El trabajo de fin de carrera se ha desarrollado con la base de datos

- ORACLE 10g Release 2 (10.2) Express Edition

trabajando con las siguientes herramientas:

- "*Oracle Database XE"* → herramienta utilizada para administrar la base de datos, crear tablas, vistas y otros objetos de la base de datos,.. etc.
- "*Oracle BI Discoverer Administrator"*
- "*Oracle BI Discoverer Desktop"* → herramienta utilizada para la generación de informes

todo esto bajo la maquina virtual "*Oracle VM VirtualBox*".

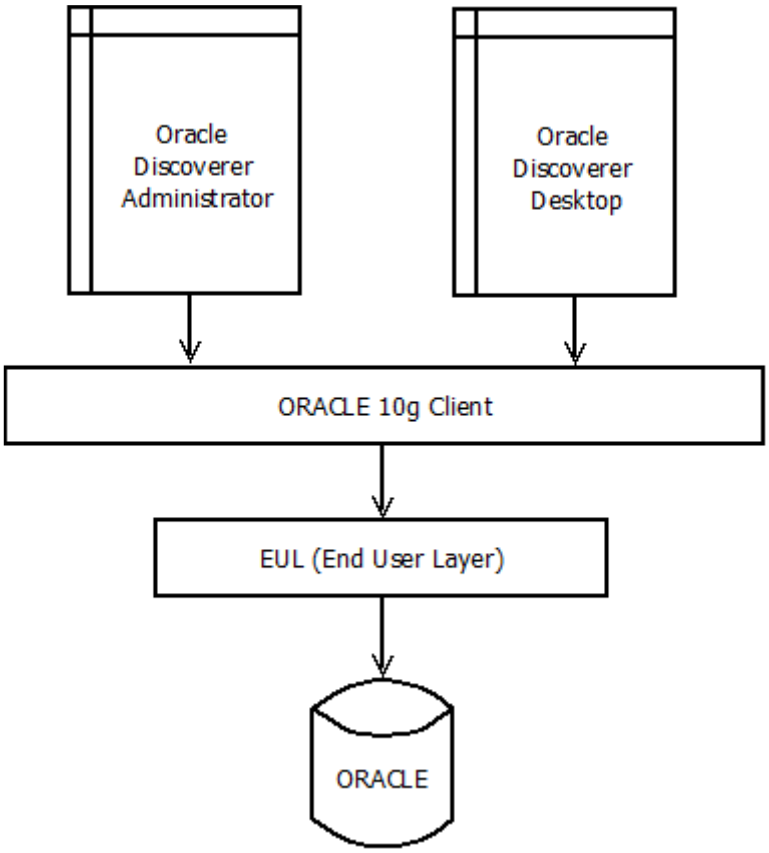


Figura 8: Arquitectura del software

1.3.2. Diagrama de la arquitectura del hardware

Un componente fundamental a la hora de poder contar con un Data Warehouse que responda a las necesidades analíticas avanzadas de los usuarios, es el poder contar con una infraestructura Hardware que la soporte. En este sentido son críticas, a la hora de evaluar uno u otro hardware, dos características principales:

- Por un lado, a este tipo de sistemas suelen acceder pocos usuarios con unas necesidades muy grandes de información, a diferencia de los sistemas operacionales, con muchos usuarios y necesidades puntuales de información. Debido a la flexibilidad requerida a la hora de hacer consultas complejas e imprevistas, y al gran tamaño de información manejada, son necesarias unas altas prestaciones de la máquina.

- Por otro lado, debido a que estos sistemas suelen comenzar con una funcionalidad limitada, que se va expandiendo con el tiempo (situación por cierto aconsejada), es necesario que los sistemas sean escalables para dar soporte a las necesidades crecientes de equipamiento. En este sentido, será conveniente el optar por una arquitectura abierta, que nos permita aprovechar lo mejor de cada fabricante.

En el mercado se han desarrollado tecnologías basadas en tecnología de procesamiento paralelo, dan el soporte necesario a las necesidades de altas prestaciones y escalabilidad de los Data Warehouse. Estas tecnologías son de dos tipos:

- SMP (Symmetric multiprocessing, o Multiprocesadores Simétricos): Los sistemas tienen múltiples procesadores que comparten un único bus y una gran memoria, repartiéndose los procesos que genera el sistema, siendo el sistema operativo el que gestiona esta distribución de tareas. Estos sistemas se conocen como arquitecturas de “casi todo compartido”. El aspecto más crítico de este tipo de sistemas es el grado de rendimiento relativo respecto al número de procesadores presentes, debido a su creciente no lineal.
- MPP (Massively parallel processing, o Multiprocesadores Masivamente Paralelos): Es una tecnología que compite contra la SMP, en la que los sistemas suelen ser casi independientes comunicados por intercambiadores de alta velocidad que permiten gestionarlos como un único sistema. Se conocen por ello como arquitecturas de “nada compartido”. Su escalabilidad es mayor que la de los SMP.

Según Meta Group, las tendencias de mercado indican que las arquitecturas SMP aportan normalmente suficientes características de escalabilidad, con una mayor oferta y un menor riesgo tecnológico. Sin embargo, cuando las condiciones de escalabilidad sean extremas, se puede plantear la opción MPP. No obstante, se están produciendo avances significativos en arquitecturas SMP, que han logrado máquinas con un crecimiento lineal de rendimiento hasta un número de 64 procesadores.

Se asume que en el CHNE tienen una red de servidores conectados a una red de estaciones de trabajo, de la siguiente manera:

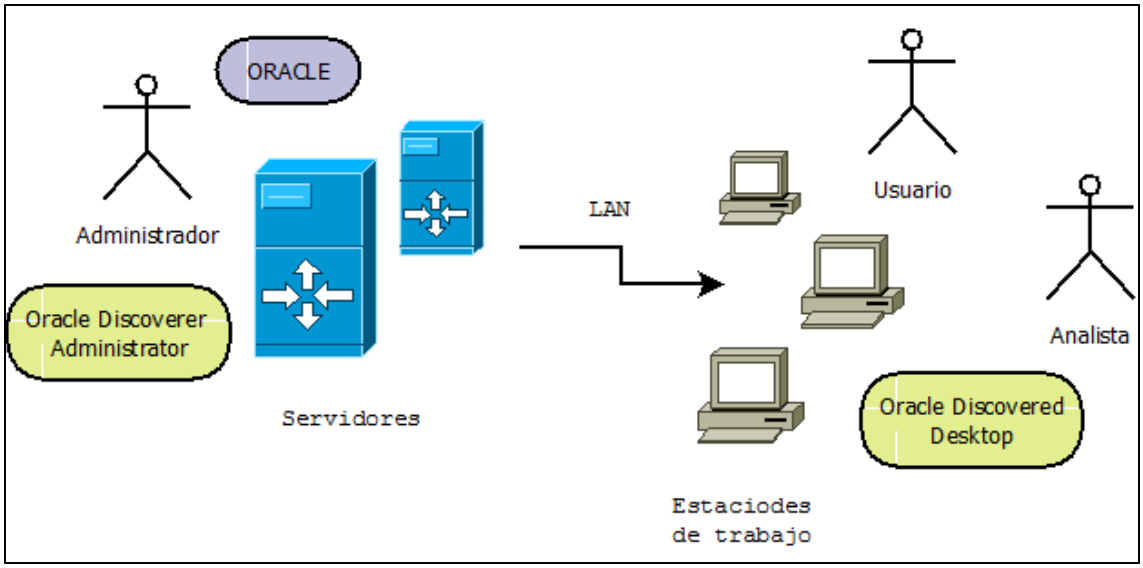


Figura 9: Arquitectura del hardware

Se requiere en principio de un servidor para el almacenamiento y manejo de la base de datos corporativa; este servidor se recomienda que sea altamente escalable, pues como ya hemos comentado, el proyecto de construcción del DWH presenta redimensionamiento a medida que se avanza en la implementación. La capacidad inicial de almacenamiento estará determinada por los requerimientos de información histórica presentados por la empresa y por la perspectiva de crecimiento que se tenga.

Dependiendo del diseño del sistema, puede ser necesario contar con un segundo servidor para las herramientas de consulta de datos. Este equipo debe tener el sistema operativo recomendado por el proveedor de la herramienta a utilizar, siendo el más usado alguna versión de Windows.

Las estaciones de trabajo de cada usuario deberán cumplir con las características recomendadas por el proveedor de la herramienta de consulta seleccionada.

Así mismo, el usuario Analista deberá tener acceso desde su estación de trabajo a la herramienta Discoverer Administrator, con el fin de poder generar nuevas áreas de trabajo cuando sea preciso.

1.3.3. Diseño de la base de datos y diagrama del modelo físico

1.3.3.1. Dimensiones

Las dimensiones son las perspectivas de análisis de las variables. Son catálogos de información complementaria necesaria para la presentación de los datos a los usuarios, como por ejemplo: descripciones, nombres, zonas, rangos de tiempo, etc. Es decir, la información general complementaria a cada uno de los registros de la tabla de Hechos.

Además, las tablas de Hechos estarán ligadas por claves foráneas a las tablas de Dimensión.

Los atributos de las dimensiones detectadas son las siguientes:

- *COMUNIDAD*

COMUNIDAD		
	TIPO	**DESCRICION**
ID_COMUNIDAD	NUM(3)	Identificador de comunidad
Nombre	CHAR(20)	Nombre comunidad

- *EMBALSE*

EMBALSE		
	TIPO	**DESCRICION**
ID_EMBALSE	NUM(3)	Identificador de embalse
Nombre	CHAR(30)	Nombre embalse

- *RIO*

RIO		
	TIPO	**DESCRICION**
ID_RIO	NUM(3)	Identificador de río
Nombre	CHAR(30)	Nombre río

- *TIEMPO*

TIEMPO		
	TIPO	**DESCRICION**
ID_ANY	NUM(3)	Identificador de año
Nany	NUM(4)	Año (AAAA)

- *ANY_MES*:

ANY_MES		
	TIPO	**DESCRICION**
ID_ANYMES	NUM(3)	Identificador de año/mes
ID_ANY	NUM(3)	Identificador de año
Mes	NUM(2)	Mes (MM)

1.3.3.2. Indicadores

Los indicadores en un DWH son los datos que van a ser analizados. Forman parte de las tablas de Hechos.

Las tablas de Hechos junto con sus atributos son los siguientes:

- *VOLUMEN:*

VOLUMEN		
	TIPO	**DESCRICION**
ID_VOLUMEN	NUM(10)	Identificador de volumen
ID_ANYMES	NUM(3)	Identificador de año/mes
ID_EMBALSE	NUM(3)	Identificador de embalse
ID_COMUNIDAD	NUM(3)	Identificador de comunidad
ID_RIO	NUM(3)	Identificador de río
Capacidad_total_embalse	NUM(5)	Capacidad total embalse (Hm^3)
Capacidad_agua_embalsada	NUM(5)	Capacidad agua embalsada en este año/mes (Hm^3)
Capacidad_agua_embalsada_ant	NUM(5)	Capacidad agua embalsada un año/mes anterior (Hm^3)
Agua_diaria_disponible	NUM(4,1)	Agua diaria disponible
Agua_nieve_equivalente	NUM(4,1)	Agua nieve equivalente

- *VARIACION_SEMANAL*:

VARIACION_SEMANAL		
	TIPO	**DESCRICION**
ID_VARSEM	NUM(10)	Identificador de variación semanal por mes/año
ID_ANYMES	NUM(3)	Identificador de año/mes
ID_EMBALSE	NUM(3)	Identificador de embalse
ID_RIO	NUM(3)	Identificador de río
Entrada	NUM(4,1)	Entrada (Hm^3)
Salida	NUM(4,1)	Salida (Hm^3)
Incremento	NUM(3)	Incremento, diferencia entre la salida y la entrada(Hm^3)

1.3.3.3. Diagrama

Finalmente, las tablas del nuevo DWH a desarrollar se relacionaran de la siguiente manera:

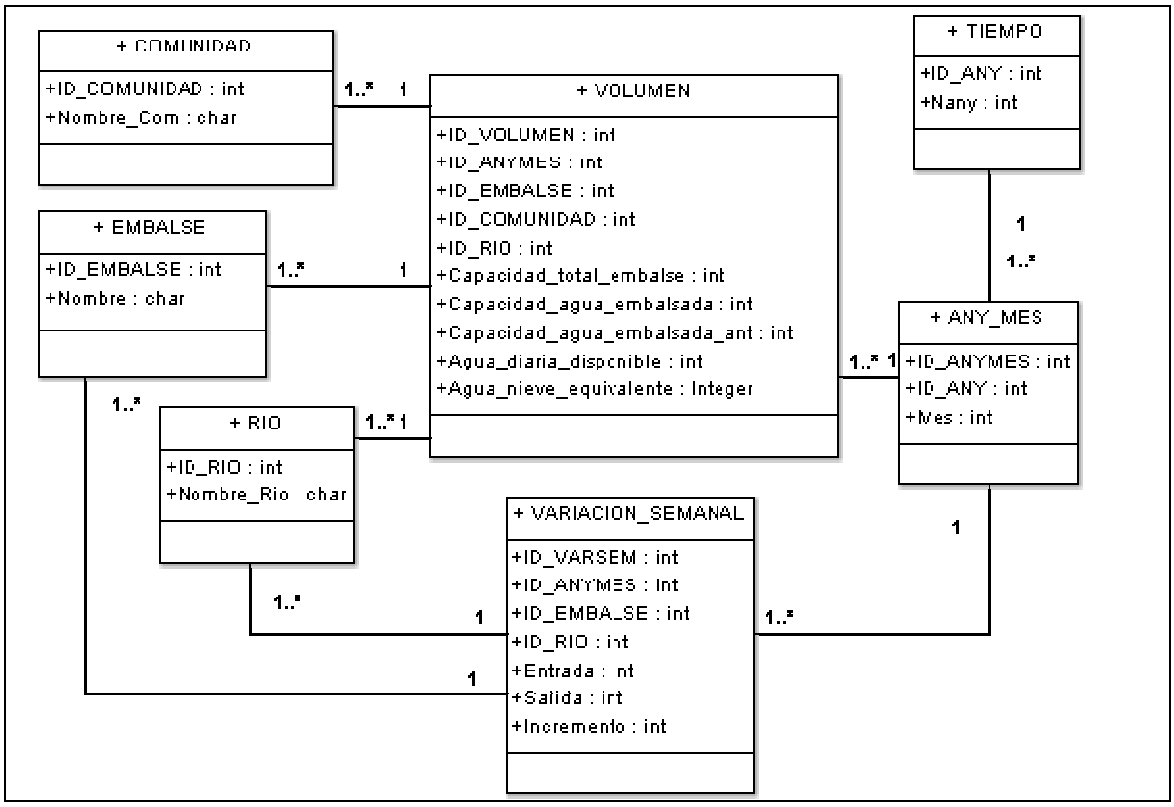


Figura 10: Modelo de relaciones de la BBDD

1.3.4. Diseño y descripción de los informes

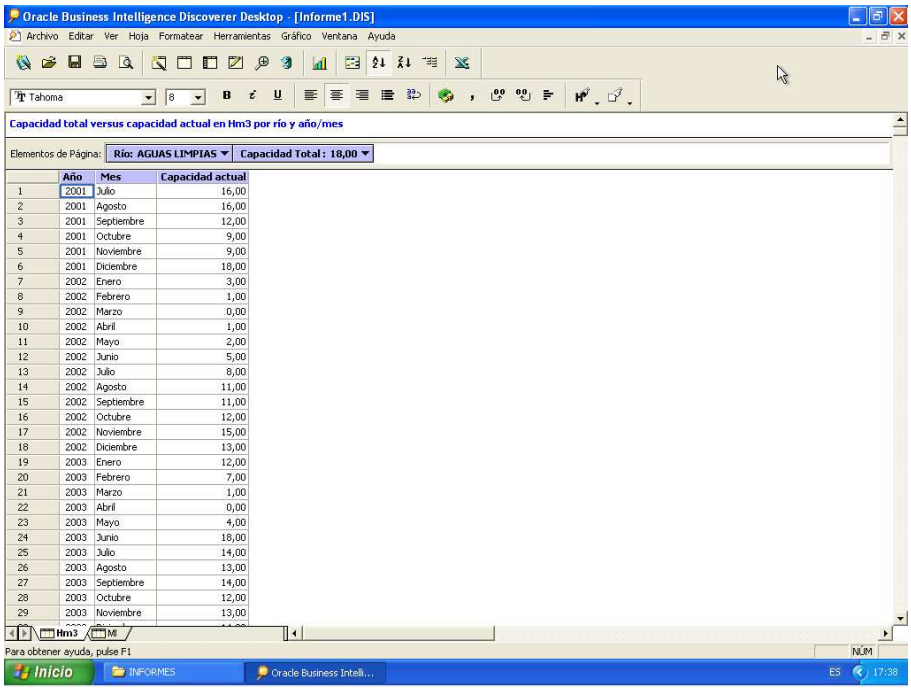
Los informes se han dejado en el disco UOCPRACT, en la carpeta:

- E:/TFC/INFORMES

A destacar que en las especificaciones, el cliente ha pedido obtener la información tanto en Hm^3 como en Ml. Para ello, cada informe tendrá dos pestañas, una con la información en Hm^3 y otra en Ml. Quedará sólo el seleccionar una u otra para obtener la información adecuada.

A continuación se muestran lo obtenido.

1. Capacidad total versus capacidad actual por río y año/mes



	Año	Mes	Capacidad actual
1	2001	Julio	16,00
2	2001	Agosto	16,00
3	2001	Septiembre	12,00
4	2001	Octubre	9,00
5	2001	Noviembre	9,00
6	2001	Diciembre	18,00
7	2002	Enero	3,00
8	2002	Febrero	1,00
9	2002	Marzo	0,00
10	2002	Abril	1,00
11	2002	Mayo	2,00
12	2002	Junio	5,00
13	2002	Julio	8,00
14	2002	Agosto	11,00
15	2002	Septiembre	11,00
16	2002	Octubre	12,00
17	2002	Noviembre	15,00
18	2002	Diciembre	13,00
19	2003	Enero	12,00
20	2003	Febrero	7,00
21	2003	Marzo	1,00
22	2003	Abril	0,00
23	2003	Mayo	4,00
24	2003	Junio	18,00
25	2003	Julio	14,00
26	2003	Agosto	13,00
27	2003	Septiembre	14,00
28	2003	Octubre	12,00
29	2003	Noviembre	13,00

Figura 11: Informe1 (Hm^3)

Si seleccionamos la pestaña de MI, obtendremos lo siguiente:

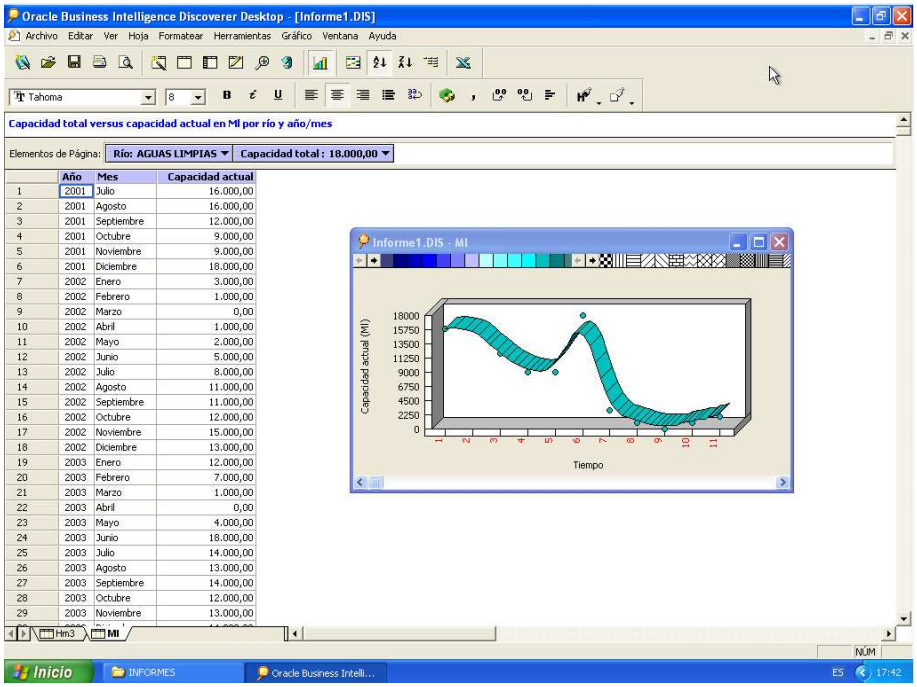


Figura 12: Informe1 (MI)

Se puede observar que los meses salen con el nombre del mes, no con el número de mes que es lo que se ha guardado en las tablas. Para obtener esto se ha realizado lo siguiente:

- Dentro del Discoverer Administrator, en primer lugar, se ha realizado una función de traducción del campo “Mes” de la tabla **ANY_MES**, dentro del área de negocio creada anteriormente.
- A continuación, recogiendo la tabla **ANY_MES**, se ha seleccionado el insertar clase de elemento, con ordenación alternativa por el campo “Mes”, con el fin que se ordene por número de mes y no por orden alfabético en los informes.

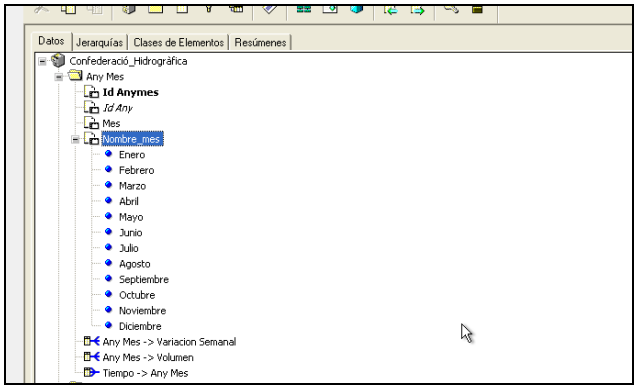
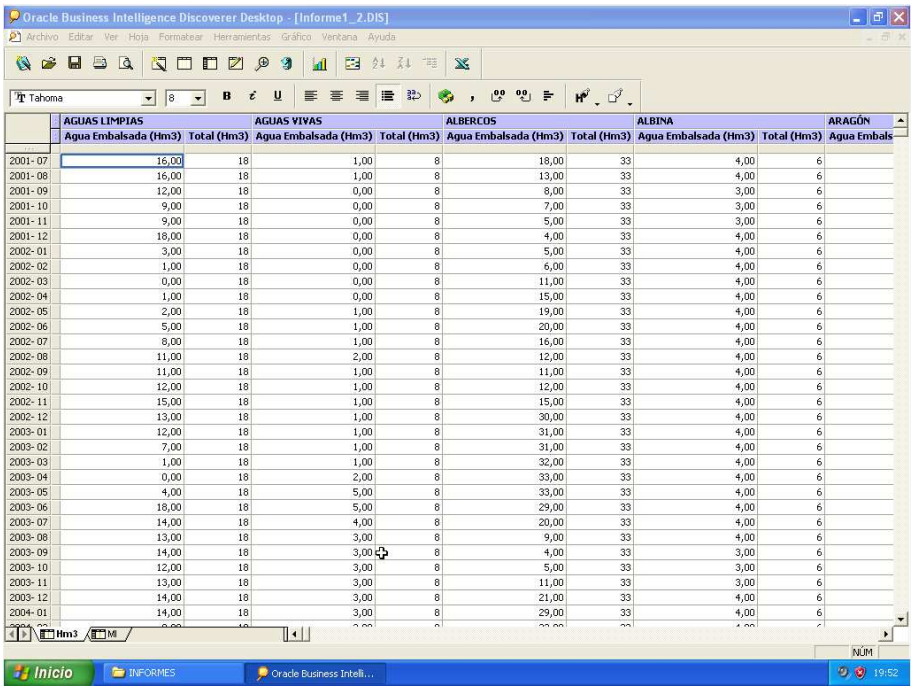


Figura 13: Literal mes

Seleccionando por “Nombre_mes” en los informes, obtendremos los literales del mes en cuestión.

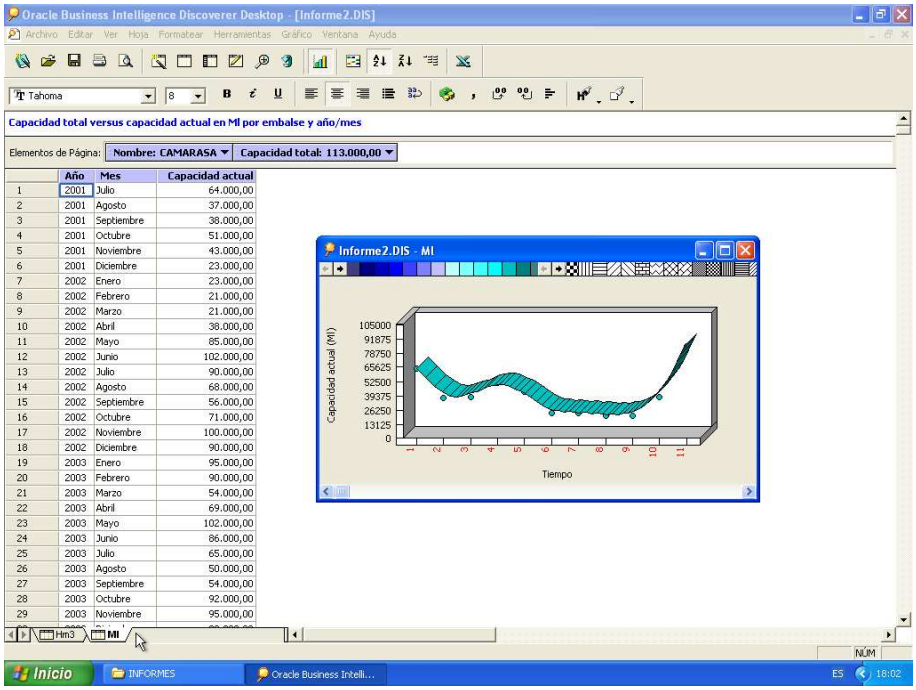
También se ha hecho otra versión de este informe, con una matriz, quedando de la siguiente manera:



	AGUAS LIMPIAS		AGUAS VIVAS		ALBERCOS		ALBINA		ARAGÓN
	Agua Embalsada (Hm3)	Total (Hm3)	Agua Embalsada (Hm3)	Total (Hm3)	Agua Embalsada (Hm3)	Total (Hm3)	Agua Embalsada (Hm3)	Total (Hm3)	Agua Embals
2001- 07	16,00	18	1,00	8	18,00	33	4,00	6	
2001- 08	16,00	18	1,00	8	13,00	33	4,00	6	
2001- 09	12,00	18	0,00	8	8,00	33	3,00	6	
2001- 10	9,00	18	0,00	8	7,00	33	3,00	6	
2001- 11	9,00	18	0,00	8	5,00	33	3,00	6	
2001- 12	18,00	18	0,00	8	4,00	33	4,00	6	
2002- 01	3,00	18	0,00	8	5,00	33	4,00	6	
2002- 02	1,00	18	0,00	8	6,00	33	4,00	6	
2002- 03	0,00	18	0,00	8	11,00	33	4,00	6	
2002- 04	1,00	18	0,00	8	15,00	33	4,00	6	
2002- 05	2,00	18	1,00	8	19,00	33	4,00	6	
2002- 06	5,00	18	1,00	8	20,00	33	4,00	6	
2002- 07	8,00	18	1,00	8	16,00	33	4,00	6	
2002- 08	11,00	18	2,00	8	12,00	33	4,00	6	
2002- 09	11,00	18	1,00	8	11,00	33	4,00	6	
2002- 10	12,00	18	1,00	8	12,00	33	4,00	6	
2002- 11	15,00	18	1,00	8	15,00	33	4,00	6	
2002- 12	13,00	18	1,00	8	30,00	33	4,00	6	
2003- 01	12,00	18	1,00	8	31,00	33	4,00	6	
2003- 02	7,00	18	1,00	8	31,00	33	4,00	6	
2003- 03	1,00	18	1,00	8	32,00	33	4,00	6	
2003- 04	0,00	18	2,00	8	33,00	33	4,00	6	
2003- 05	4,00	18	5,00	8	33,00	33	4,00	6	
2003- 06	18,00	18	5,00	8	29,00	33	4,00	6	
2003- 07	14,00	18	4,00	8	20,00	33	4,00	6	
2003- 08	13,00	18	3,00	8	9,00	33	4,00	6	
2003- 09	14,00	18	3,00	8	4,00	33	3,00	6	
2003- 10	12,00	18	3,00	8	5,00	33	3,00	6	
2003- 11	13,00	18	3,00	8	11,00	33	3,00	6	
2003- 12	14,00	18	3,00	8	21,00	33	4,00	6	
2004- 01	14,00	18	3,00	8	29,00	33	4,00	6	

Figura 14: Informe1_2 (Hm^3)

2. Capacidad total versus capacidad actual por embalse y año/mes



	Año	Mes	Capacidad actual
1	2001	Julio	64.000,00
2	2001	Agosto	37.000,00
3	2001	Septiembre	38.000,00
4	2001	Octubre	51.000,00
5	2001	Noviembre	43.000,00
6	2001	Diciembre	23.000,00
7	2002	Enero	23.000,00
8	2002	Febrero	21.000,00
9	2002	Marzo	21.000,00
10	2002	Abril	38.000,00
11	2002	Mayo	85.000,00
12	2002	Junio	102.000,00
13	2002	Julio	90.000,00
14	2002	Agosto	68.000,00
15	2002	Septiembre	56.000,00
16	2002	Octubre	71.000,00
17	2002	Noviembre	100.000,00
18	2002	Diciembre	90.000,00
19	2003	Enero	95.000,00
20	2003	Febrero	90.000,00
21	2003	Marzo	54.000,00
22	2003	Abril	69.000,00
23	2003	Mayo	102.000,00
24	2003	Junio	86.000,00
25	2003	Julio	65.000,00
26	2003	Agosto	50.000,00
27	2003	Septiembre	54.000,00
28	2003	Octubre	92.000,00
29	2003	Noviembre	95.000,00

Figura 15: Informe2 (Ml)

La versión matriz se puede consultar en E:/TFC/INFORMES/Informe2_2.DIS.

3. Capacidad total versus capacidad actual por comunidad y año/mes

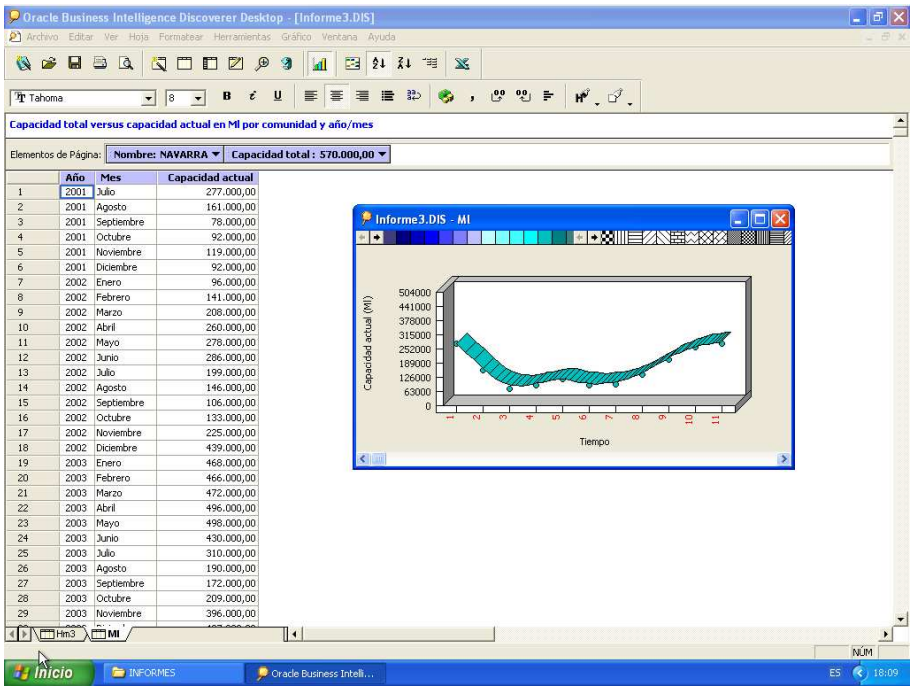


Figura 16: Informe3 (MI)

La versión matriz se puede consultar en E:/TFC/INFORMES/Informe3_2.DIS.

4. Comparativa anual de la evolución por comunidad de la capacidad agregada de todos sus embalses

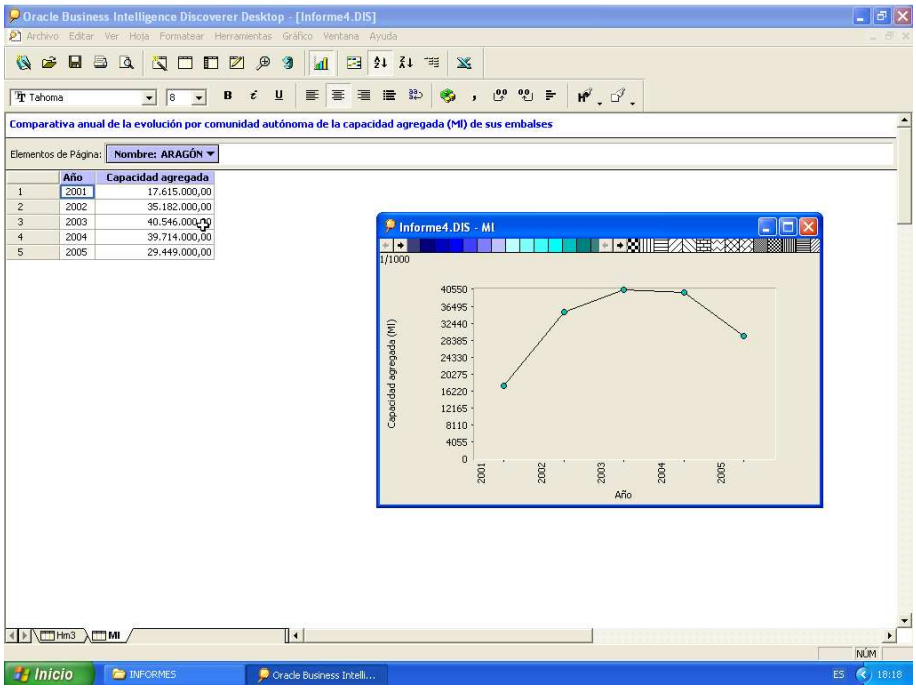
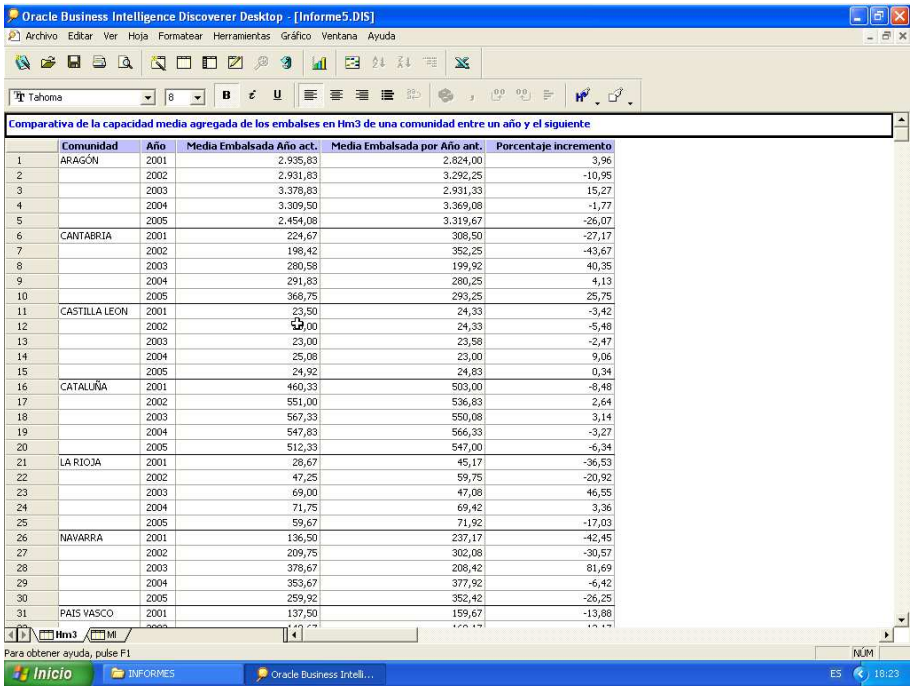


Figura 17: Informe4 (MI)

5. Comparativa de la capacidad media agregada de los embalses de una comunidad entre un año y el año anterior



	Comunidad	Año	Media Embalsada Año act.	Media Embalsada por Año ant.	Porcentaje incremento
1	ARAGÓN	2001	2.935,83	2.824,00	3,96
2		2002	2.931,83	3.292,25	-10,95
3		2003	3.378,83	2.931,33	15,27
4		2004	3.309,50	3.369,08	-1,77
5		2005	2.454,08	3.319,67	-26,07
6	CANTABRIA	2001	224,67	308,50	-27,17
7		2002	198,42	352,25	-43,67
8		2003	280,58	199,92	40,35
9		2004	291,83	280,25	4,13
10		2005	368,75	293,25	25,75
11	CASTILLA LEON	2001	23,50	24,33	-3,42
12		2002	[illegible],00	24,33	-5,48
13		2003	23,00	23,58	-2,47
14		2004	25,08	23,00	9,06
15		2005	24,92	24,83	0,34
16	CATALUÑA	2001	460,33	503,00	-8,48
17		2002	551,00	536,83	2,64
18		2003	567,33	550,08	3,14
19		2004	547,83	566,33	-3,27
20		2005	512,33	547,00	-6,34
21	LA RIOJA	2001	28,67	45,17	-36,53
22		2002	47,25	59,75	-20,92
23		2003	69,00	47,08	46,55
24		2004	71,75	69,42	3,36
25		2005	59,67	71,92	-17,03
26	NAVARRA	2001	136,50	237,17	-42,45
27		2002	209,75	302,08	-30,57
28		2003	378,67	208,42	81,69
29		2004	353,67	377,92	-6,42
30		2005	259,92	352,42	-26,25
31	PAIS VASCO	2001	137,50	159,67	-13,88

Figura 18: Informe5 (Hm^3)

Si queremos ver la comparativa en un gráfico de barras, se obtiene lo siguiente:

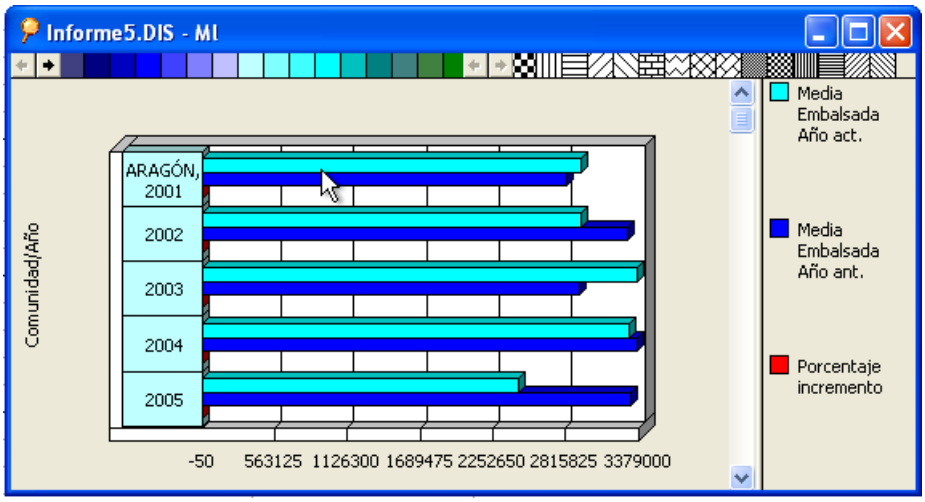
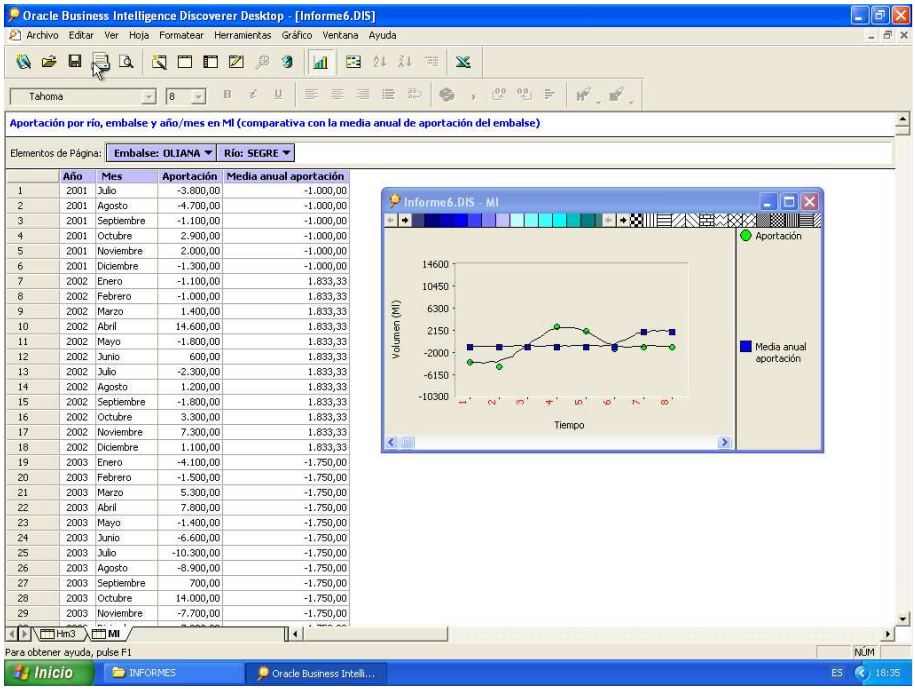


Figura 19: Gráfico informe5 (MI)

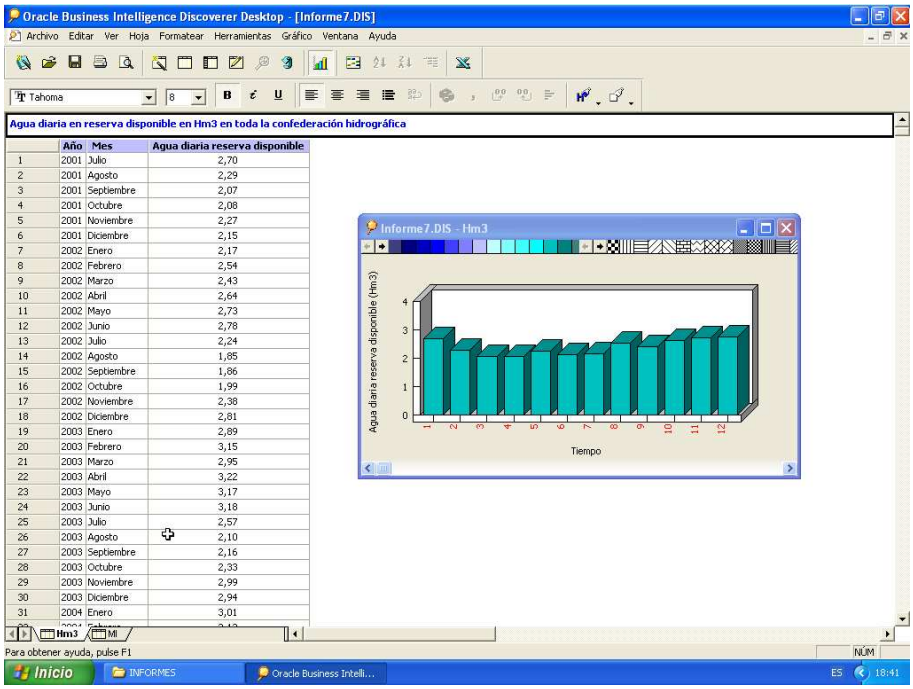
6. Aportación (Entrada/Salida) por río, embalse y año/mes



	Año	Mes	Aportación	Media anual aportación
1	2001	Julio	-3.800,00	-1.000,00
2	2001	Agosto	-4.700,00	-1.000,00
3	2001	Septiembre	-1.100,00	-1.000,00
4	2001	Octubre	2.900,00	-1.000,00
5	2001	Noviembre	2.000,00	-1.000,00
6	2001	Diciembre	-1.300,00	-1.000,00
7	2002	Enero	-1.100,00	1.833,33
8	2002	Febrero	-1.000,00	1.833,33
9	2002	Marzo	1.400,00	1.833,33
10	2002	Abril	14.600,00	1.833,33
11	2002	Mayo	-1.800,00	1.833,33
12	2002	Junio	600,00	1.833,33
13	2002	Julio	-2.300,00	1.833,33
14	2002	Agosto	1.200,00	1.833,33
15	2002	Septiembre	-1.800,00	1.833,33
16	2002	Octubre	3.300,00	1.833,33
17	2002	Noviembre	7.300,00	1.833,33
18	2002	Diciembre	1.100,00	1.833,33
19	2003	Enero	-4.100,00	-1.750,00
20	2003	Febrero	-1.500,00	-1.750,00
21	2003	Marzo	5.300,00	-1.750,00
22	2003	Abril	7.800,00	-1.750,00
23	2003	Mayo	-1.400,00	-1.750,00
24	2003	Junio	-6.600,00	-1.750,00
25	2003	Julio	-10.300,00	-1.750,00
26	2003	Agosto	-8.900,00	-1.750,00
27	2003	Septiembre	700,00	-1.750,00
28	2003	Octubre	14.000,00	-1.750,00
29	2003	Noviembre	-7.700,00	-1.750,00

Figura 20: Informe6 (Ml)

7. Agua diaria en reserva disponible (volumen/día) a toda la confederación hidrográfica por año/mes



	Año	Mes	Agua diaria reserva disponible
1	2001	Julio	2,70
2	2001	Agosto	2,29
3	2001	Septiembre	2,07
4	2001	Octubre	2,08
5	2001	Noviembre	2,27
6	2001	Diciembre	2,15
7	2002	Enero	2,17
8	2002	Febrero	2,54
9	2002	Marzo	2,43
10	2002	Abril	2,64
11	2002	Mayo	2,73
12	2002	Junio	2,78
13	2002	Julio	2,24
14	2002	Agosto	1,85
15	2002	Septiembre	1,86
16	2002	Octubre	1,99
17	2002	Noviembre	2,38
18	2002	Diciembre	2,81
19	2003	Enero	2,89
20	2003	Febrero	3,15
21	2003	Marzo	2,95
22	2003	Abril	3,22
23	2003	Mayo	3,17
24	2003	Junio	3,18
25	2003	Julio	2,57
26	2003	Agosto	2,10
27	2003	Septiembre	2,16
28	2003	Octubre	2,33
29	2003	Noviembre	2,99
30	2003	Diciembre	2,94
31	2004	Enero	3,01

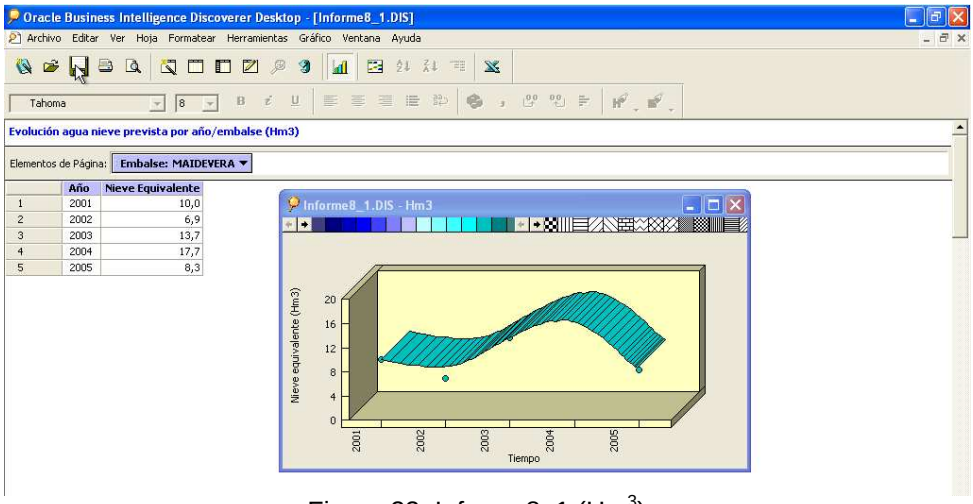
Figura 21: Informe7 (Hm^3)

Informes adicionales

Se nos ha pedido un trabajo adicional frente a los informes anteriormente mostrados.

1. Mostrar gráfico de la evolución del volumen de agua nieve equivalente por año y embalse. Si hay, también mostrar casos de desbordamiento del embalse.

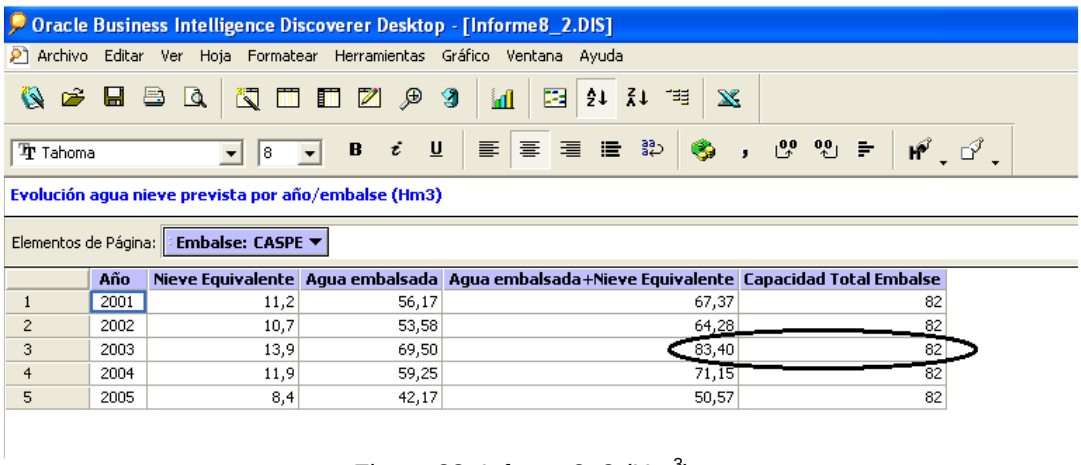
Si queremos mostrar un sencillo gráfico de la evolución de agua nieve equivalente por año y embalse, obtendremos lo siguiente:



	Año	Nieve Equivalente
1	2001	10,0
2	2002	6,9
3	2003	13,7
4	2004	17,7
5	2005	8,3

Figura 22: Informe8_1 (Hm^3)

Sin embargo, si deseamos mostrar posibles casos de desbordamiento de embalses, podemos obtener lo siguiente:

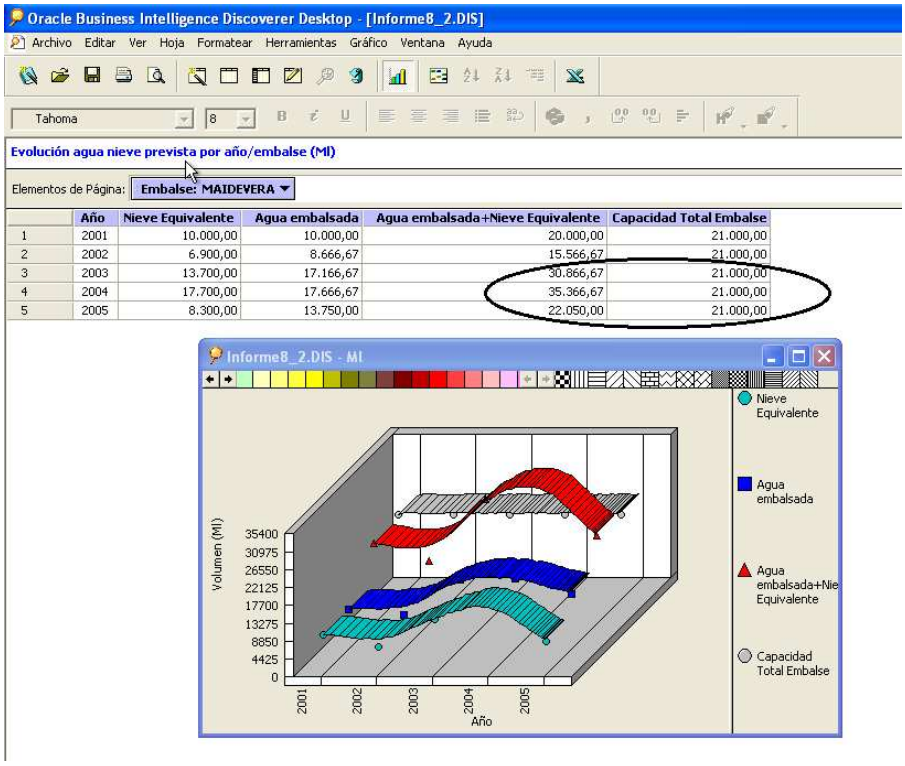


	Año	Nieve Equivalente	Agua embalsada	Agua embalsada+Nieve Equivalente	Capacidad Total Embalse
1	2001	11,2	56,17	67,37	82
2	2002	10,7	53,58	64,28	82
3	2003	13,9	69,50	83,40	82
4	2004	11,9	59,25	71,15	82
5	2005	8,4	42,17	50,57	82

Figura 23: Informe8_2 (Hm^3)

Seleccionando por embalse, se detecta que puede haber un desbordamiento del embalse cuando la suma del agua embalsada y el agua nieve equivalente supera a la capacidad total del embalse. En este caso, con el embalse de CASPE en el año 2003 había un riesgo considerable.

Si queremos tener esta información en gráfico, podemos obtener lo siguiente:



	Año	Nieve Equivalente	Agua embalsada	Agua embalsada+Nieve Equivalente	Capacidad Total Embalse
1	2001	10.000,00	10.000,00	20.000,00	21.000,00
2	2002	6.900,00	8.666,67	15.566,67	21.000,00
3	2003	13.700,00	17.166,67	30.866,67	21.000,00
4	2004	17.700,00	17.666,67	35.366,67	21.000,00
5	2005	8.300,00	13.750,00	22.050,00	21.000,00

Figura 24: Informe8_2 (Ml)

Se observa que el embalse de MAIDEVERA tuvo serio riesgo de desbordamiento en los años 2003, 2004 y 2005.

2. Mostrar como se pueden predecir a partir del Data Mart desarrollos de periodos de sequía.

Los periodos de sequía dependen de una serie de indicadores que afectan a la aportación de agua al embalse. Entre ellos podemos destacar:

- Indicadores pluviométricos
- Indicadores de nieve
- Indicadores piezométricos (acuíferos)
- Indicadores de sequía de aguas superficiales

Incluso investigadores españoles han desarrollado como un nuevo índice de sequía el que considera los efectos de la pluviosidad y el calentamiento térmico, lo que permite cuantificar los efectos del calentamiento global. El indicador, llamado Índice Estandarizado de Precipitación y Evapotranspiración (SPEI, en sus siglas en inglés), cuantifica de forma numérica la intensidad, magnitud y duración de la sequía, además de su extensión superficial, momento de inicio y final.

También vendrá dado por el uso agrícola y por la cantidad de población que deba de abastecer el embalse.

Sin embargo, para facilitar el trabajo, se realizó en su momento una homogenización de los indicadores, mediante la definición del índice de estado.

El índice de estado es un índice mensual que responde a la siguiente expresión:

- Si $V_i \geq V_{med}$

$$Ie = \frac{1}{2}\left[1 + \frac{Vi - Vmed}{V\max - Vmed}\right]$$

- Si $V_i < V_{med}$

$$Ie = \frac{Vi - V\min}{2(Vmed - V\min)}$$

Siendo:

- Vi → el valor de la medida del indicador obtenida en el mes de seguimiento
- Vmed: → el valor medio del indicador en el periodo histórico
- Vmax → el valor máximo del indicador en el periodo histórico
- Vmin → el valor mínimo del indicador en el periodo histórico

El índice será una cifra que oscilará entre 0,5 y 1, con el baremo siguiente:

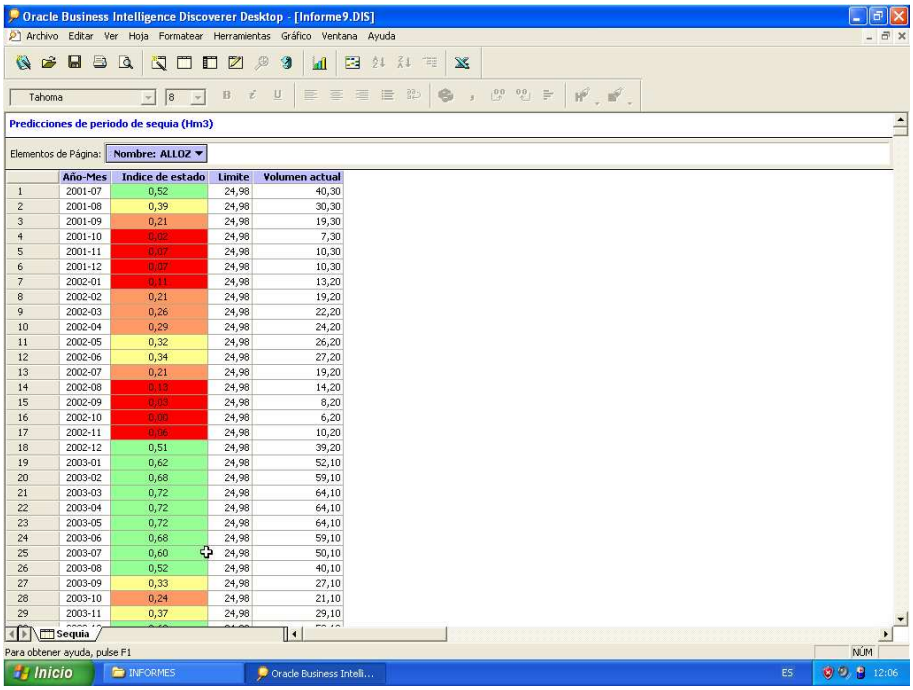
- Ie > 0,5 → Nivel verde (situación estable)
- 0,5 > Ie > 0,3 → Nivel amarillo (situación de prealerta)
- 0,3 > Ie > 0,15 → Nivel naranja (situación de alerta)
- 0,15 > Ie → Nivel rojo (situación de emergencia)

Tomando como indicador el volumen de agua embalsada, calculamos en cada año/mes el índice de estado según las fórmulas descritas.

Según lo obtenido en este índice de estado, marcamos con excepciones los colores según los niveles de situación, con el fin de detectar visualmente estados de alerta.

Además, tomando como punto de alerta cuando el índice esta a un nivel de 0,3, obtenemos el cálculo de la columna "Limite". Esto nos indicará si estamos cerca o no de dicho estado.

El informe obtenido por embalse es el siguiente:



	Año-Mes	Indice de estado	Limite	Volumen actual
1	2001-07	0,52	24,98	40,30
2	2001-08	0,39	24,98	30,30
3	2001-09	0,21	24,98	19,30
4	2001-10	0,02	24,98	7,30
5	2001-11	0,07	24,98	10,30
6	2001-12	0,07	24,98	10,30
7	2002-01	0,11	24,98	13,20
8	2002-02	0,21	24,98	19,20
9	2002-03	0,26	24,98	22,20
10	2002-04	0,29	24,98	24,20
11	2002-05	0,32	24,98	26,20
12	2002-06	0,34	24,98	27,20
13	2002-07	0,21	24,98	19,20
14	2002-08	0,13	24,98	14,20
15	2002-09	0,03	24,98	8,20
16	2002-10	0,00	24,98	6,20
17	2002-11	0,06	24,98	10,20
18	2002-12	0,51	24,98	39,20
19	2003-01	0,62	24,98	52,10
20	2003-02	0,68	24,98	59,10
21	2003-03	0,72	24,98	64,10
22	2003-04	0,72	24,98	64,10
23	2003-05	0,72	24,98	64,10
24	2003-06	0,68	24,98	59,10
25	2003-07	0,60	24,98	50,10
26	2003-08	0,52	24,98	40,10
27	2003-09	0,33	24,98	27,10
28	2003-10	0,24	24,98	21,10
29	2003-11	0,37	24,98	29,10

Figura 25: Informe9

Aquí se puede observar que entramos en estado de alerta (naranja) y de emergencia (rojo) en el 2001 y el 2002 para el embalse de ALLOZ.

A nivel de gráfico, se obtendrá lo siguiente:

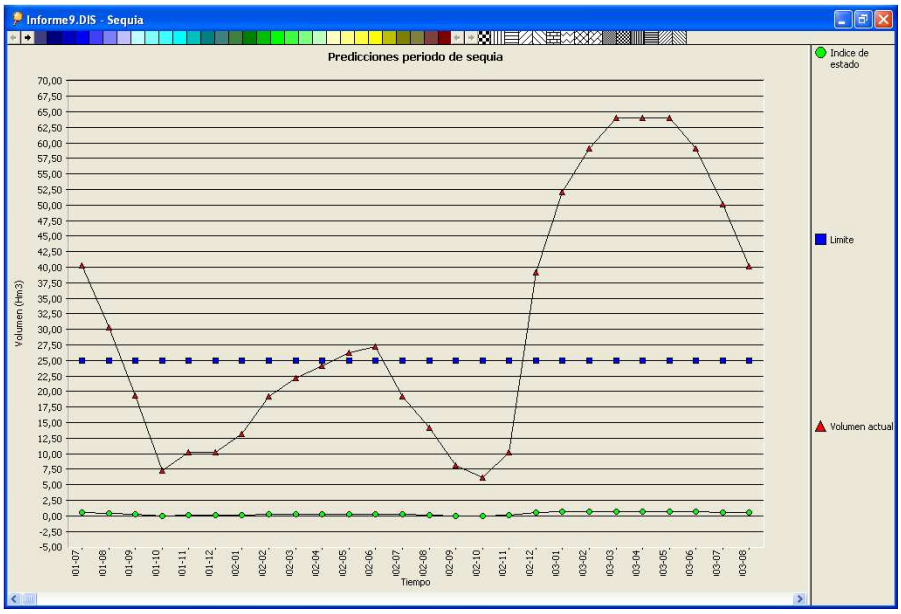


Figura 26: Gráfico Informe9

Cuando el volumen actual esté bajo el límite, estamos en estado de alerta

1.4.- Capítulo 4 – Capturas de pantalla –

En este capitulo, se cree oportuno el mostrar las capturas de pantalla, tanto:

- de la creación del usuario
- de la creación del modelo con Discoverer Administrator
- de la creación de informes con Discoverer Desktop

con el fin que sirva de guía de usuario.

1.4.1. Creación usuario

En primer lugar y con el usuario de sistema SYSTEM/UOC, crearemos un nuevo usuario desde ORACLE Data Base Express Edition, llamado LVARO/UOC.

Será desde este usuario que se crearan las tablas y triggers que se necesitan, así como su carga desde el Sql LOAD.

El script para generar dicho usuario se encuentra en:

- E:\TFC\SQL\01_Create_Usuario.txt

Y la forma de trabajar con él será cargándolo y ejecutándolo posteriormente:

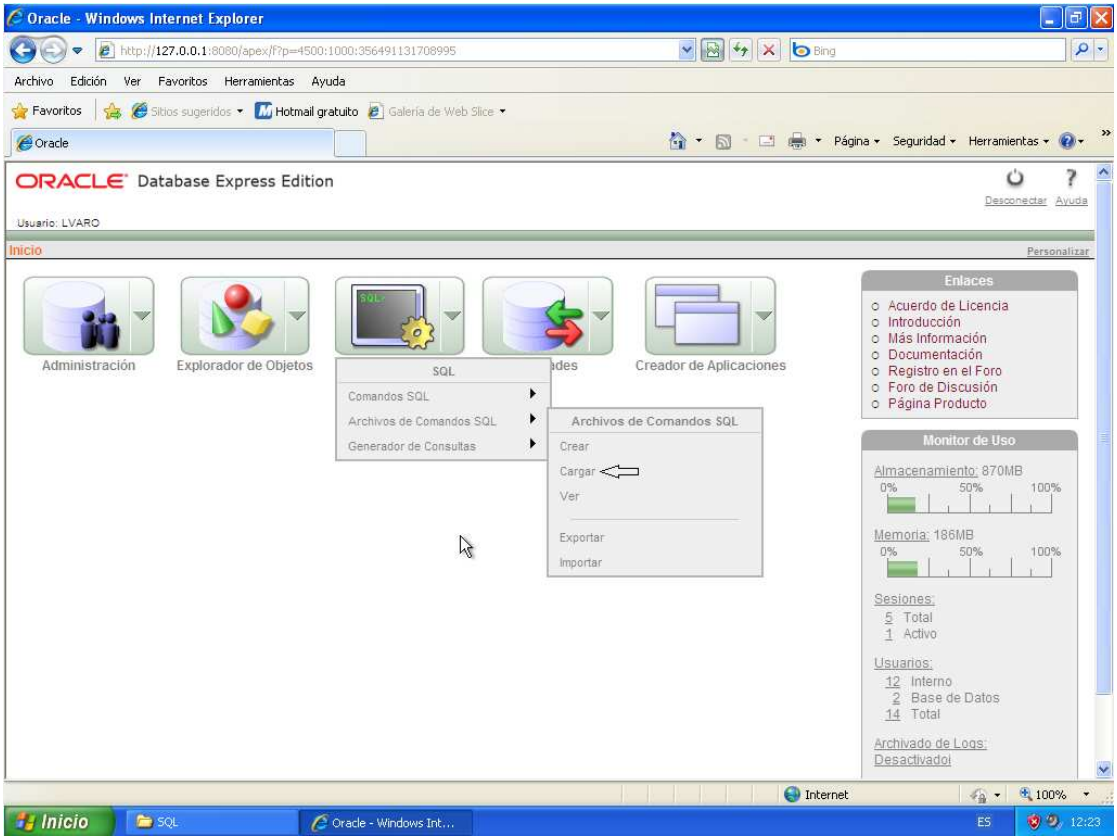


Figura 27: Ejecución script desde ORACLE DEE

1.4.2. Creación modelo con Discoverer Administrator

- Conexión con la Base de datos.

  Nos hemos de conectar con el usuario creado anteriormente (LVARO/UOC). A continuación, nos pide que creemos un EUL (End User Layer) para acceder a Discoverer, ya que dicho usuario aún no tiene ninguno asociado:

Figura 28: Conexión Discoverer

El EUL es una colección de aproximadamente 50 tablas (47 tablas y 3 vistas) en la base de datos. Cuando el administrador de Discoverer crea un EUL para un usuario, estas tablas se crean en el esquema de ese usuario. Estas tablas son usadas para almacenar y mantener los metadatos de las “Business areas” y los otros objetos del EUL. Estas son solamente tablas que pueden ser modificadas a través de Discoverer Administrator. Se puede tener más de un conjunto de tablas del EUL en cualquier base de datos al crear un EUL sobre diferentes esquemas.

Se puede crear un EUL para un usuario existente o un nuevo usuario en la base de datos. El propietario del EUL debe tener los siguientes privilegios en la base de datos:

- Create Session
- Create Table
- Create View
- Create Sequence
- Create Procedure

Además se requiere privilegios adicionales para el manejo de columnas sumarizadas y programaciones batch.

- Creación de EUL

Al seleccionar que se desea crear una nueva EUL, se muestra la siguiente pantalla:

Figura 29: Creación EUL

Se debe de seleccionar un usuario existente en el DBMS. Tomaremos uno ya existente:

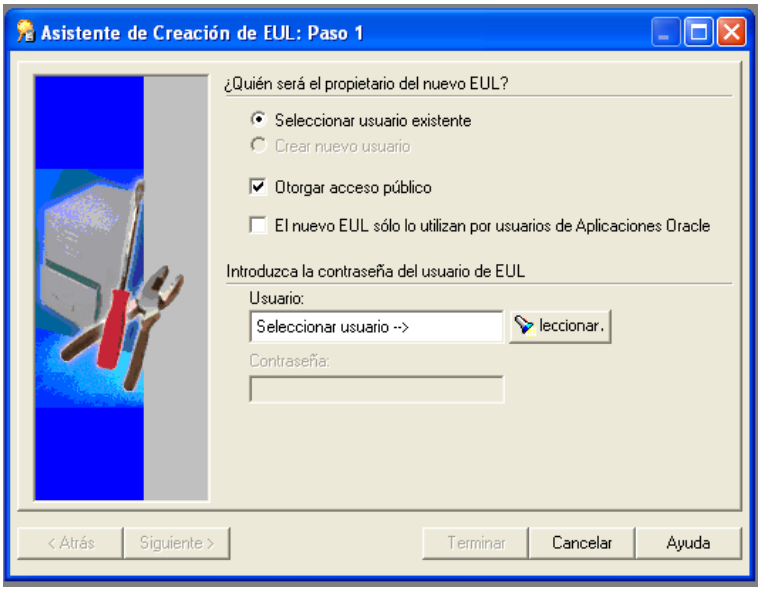


Figura 30: Parámetros EUL

en concreto el usuario LVARO:

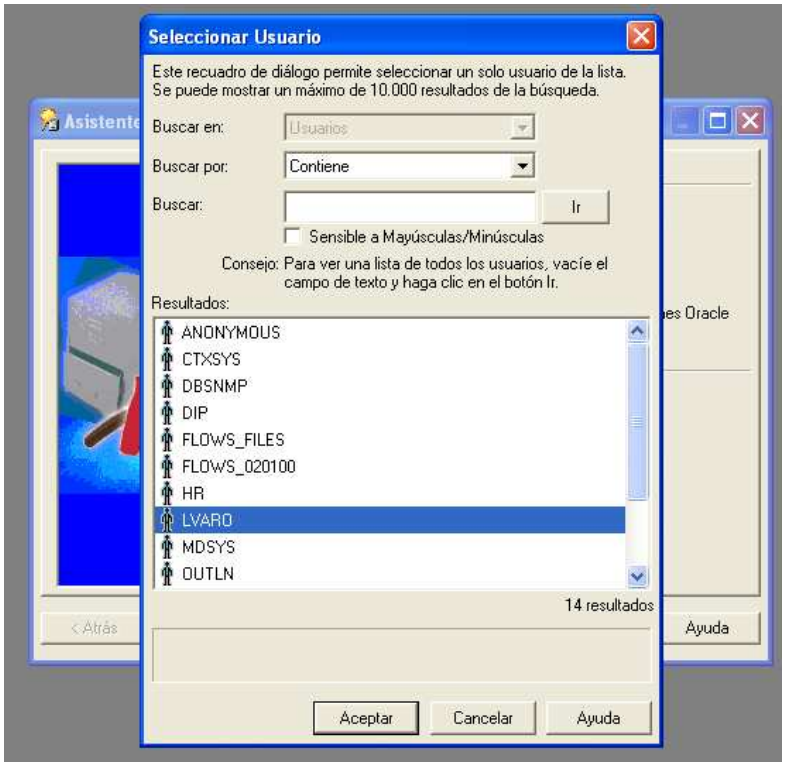


Figura 31: Seleccionar usuario

Una vez ya creados los niveles necesarios para el EUL, finaliza su creación:

Figura 32: Aviso de creación correcta del EUL

- Creación de una nueva área de negocio

A continuación debemos de crear una nueva área de negocio:

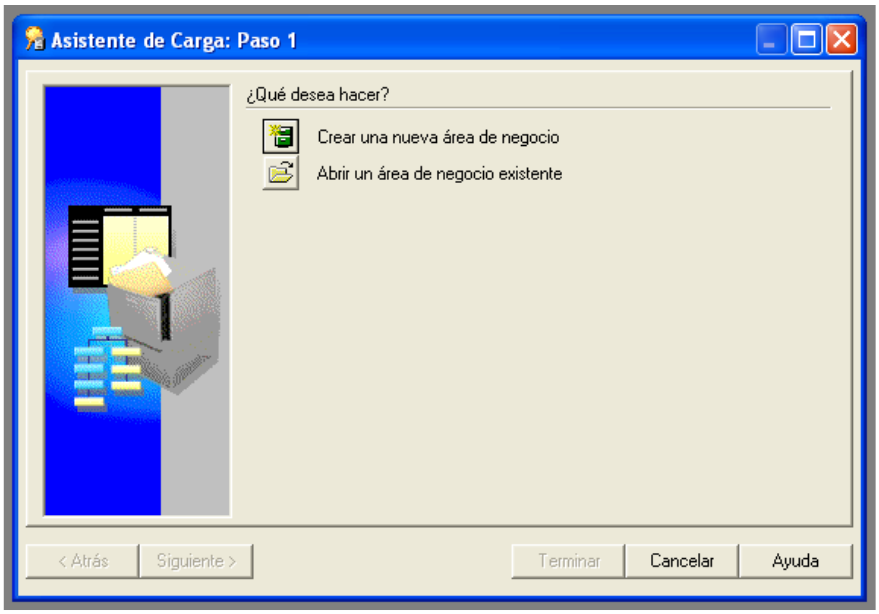


Figura 33: Inicio de creación de una nueva área de negocio

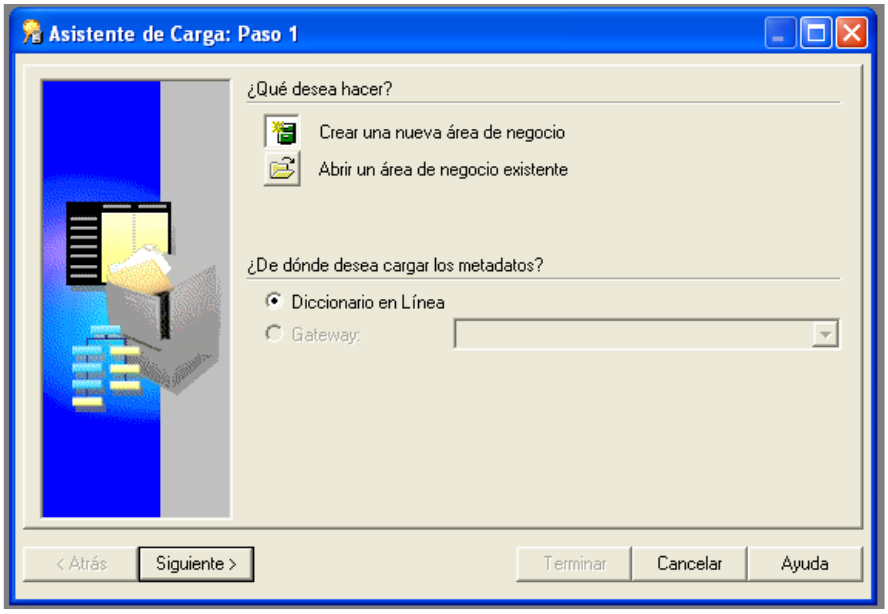


Figura 34: Creación de una nueva área de negocio (Cont.)

Nuevamente seleccionamos el usuario LVARO:

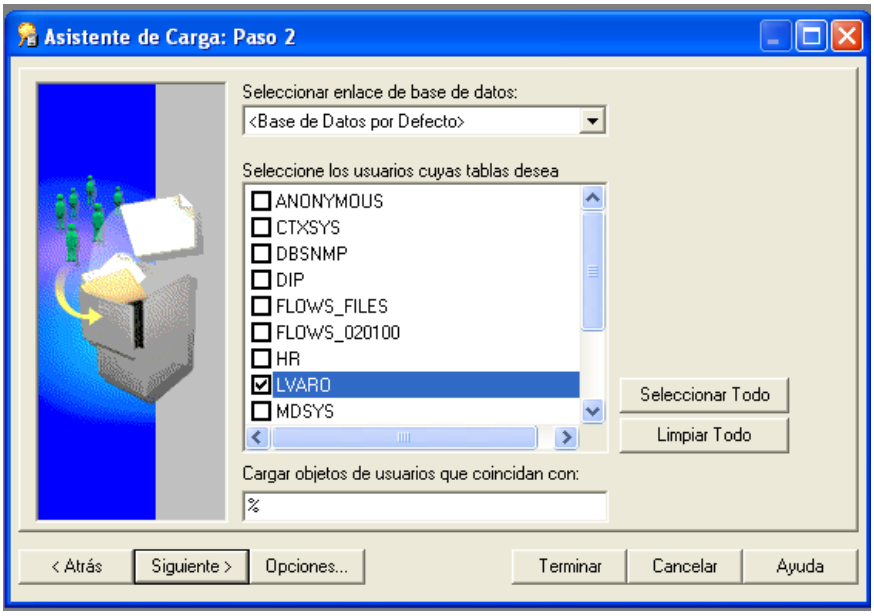


Figura 35: Selección de usuario para el área de negocio.

así como las tablas con las que queremos trabajar en esta nueva área de negocio:

Figura 36: Selección de tablas para el área de negocio

Se parametriza el EUL:

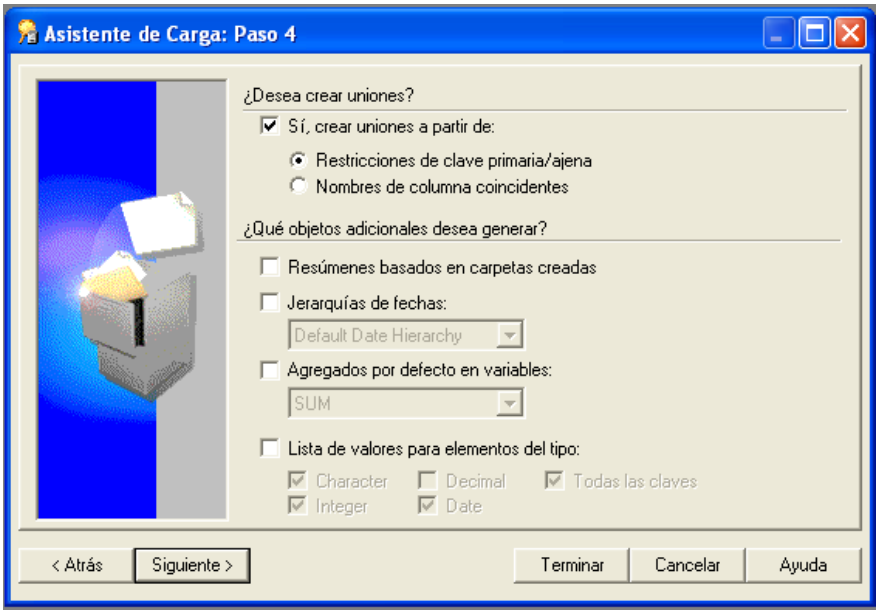


Figura 37: Selección de parámetros para el área de negocio

y se le asigna nombre:

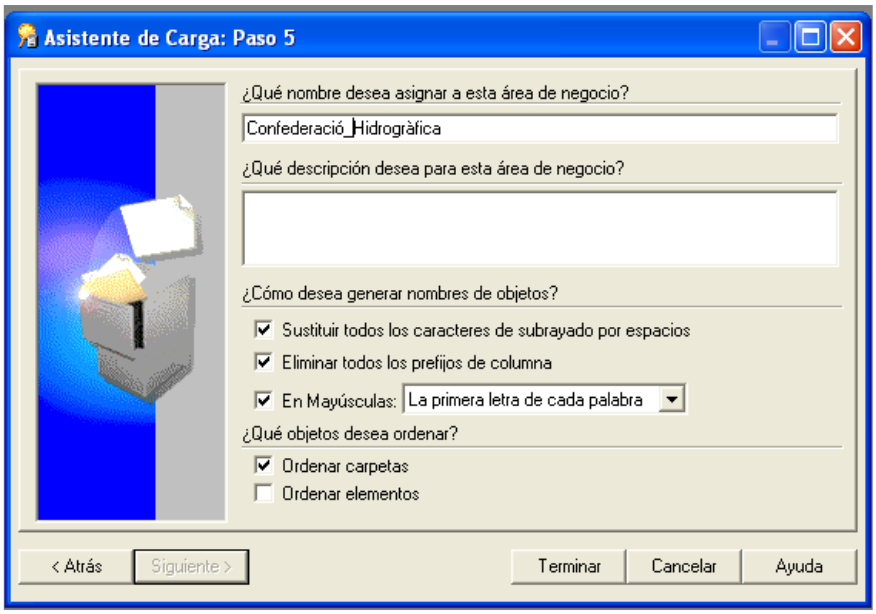


Figura 38: Selección de nombre para el área de negocio

Finalmente podemos visualizar el EUL creado, con las tablas, jerarquías, clases de elementos y los resúmenes correspondientes:

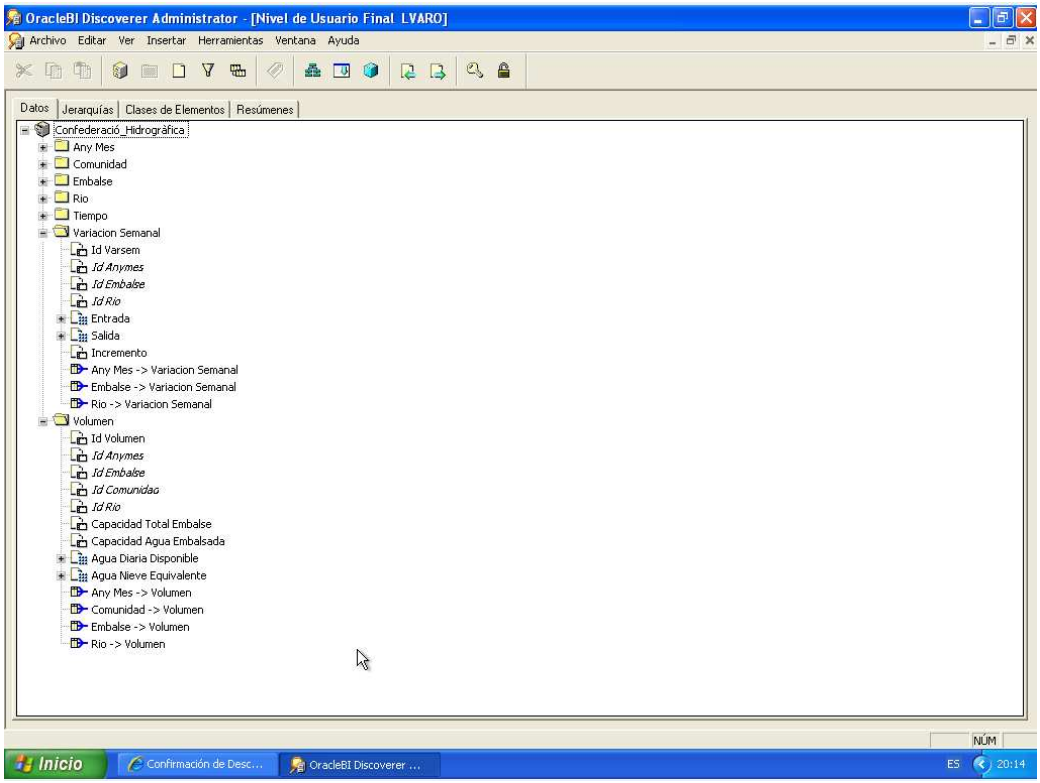


Figura 39: Visualización de las tablas del área de negocio creada.

1.4.3. Creación informes con Discoverer Desktop

Nos conectaremos de nuevo con la clave LVARO/UOC a la base de datos:

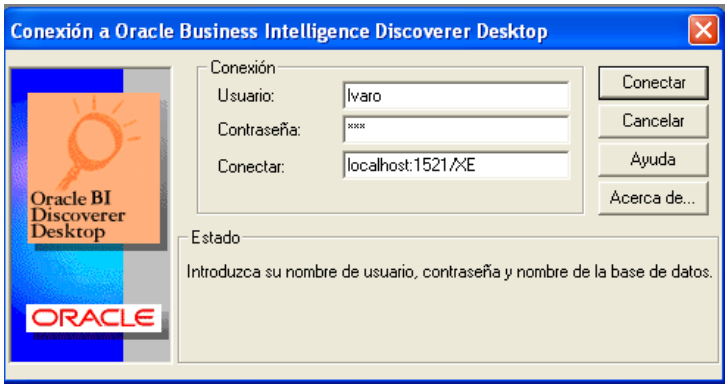


Figura 40: Conexión Discoverer Desktop

A continuación crearemos un libro de trabajo para definir nuestros informes:

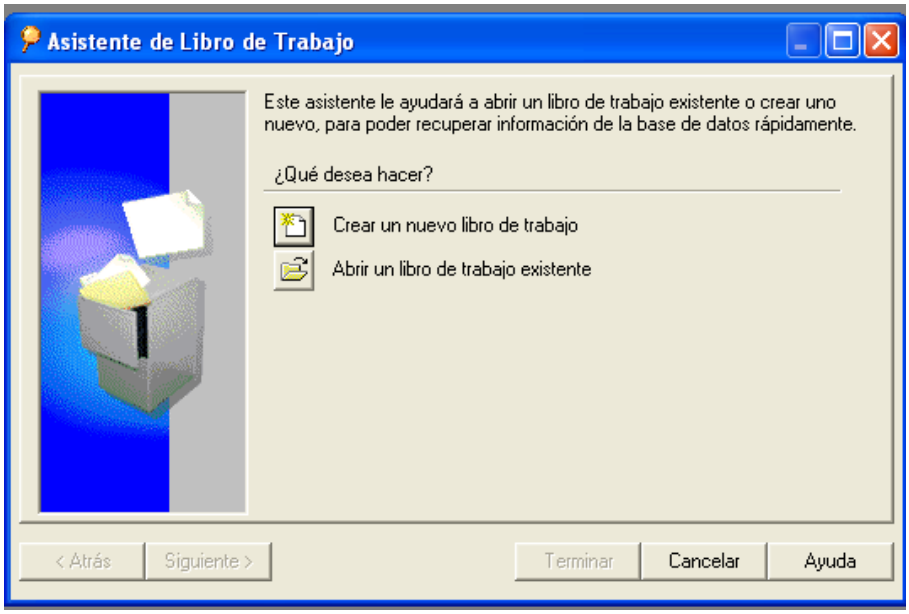


Figura 41: Creación nuevo libro de trabajo

El asistente de trabajo nos guía en la creación del libro de trabajo:

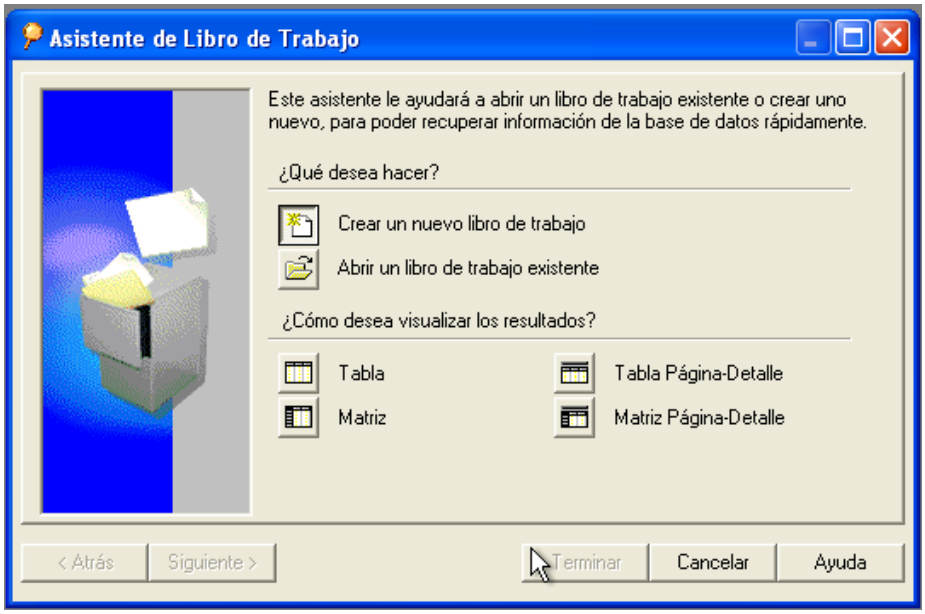


Figura 42: Creación nuevo libro de trabajo (Cont.)

Se pueden crear:

- Tablas, que muestran datos en filas y columnas.
- Matrices, que muestran datos multidimensionales y permite girar las dimensiones vertical y horizontal.
- Tablas Página-Detalle, que permiten mostrar información agrupada según el criterio especificado en el eje de la página.
- Matrices Página-Detalle, que permiten mostrar información agrupada según criterios especificados en el eje de la página.

1.5.- Capítulo 5 – Conclusiones –

Tras finalizar el TFC podemos llegar a la conclusión que:

- Se ha realizado el proceso completo de la creación de un almacén de datos, partiendo del análisis de requerimientos, el diseño del modelo multidimensional, los procesos ETL y, finalmente la implementación del almacén propiamente dicho.

- Una vez creado el almacén de datos, hemos podido realizar todos aquellos informes predeterminados pedidos dando respuesta a las necesidades del cliente.

- Estos informes se pueden visualizar tanto en Hm^3 como en Ml.

- A su vez, hemos podido mostrar fácilmente como se puede detectar casos de desbordamiento del embalse, así como poder predecir periodos de sequía.

- Finalmente podemos afirmar que se han alcanzado los objetivos propuestos.

1.6.- Capítulo 6 – Líneas de evolución –

1.6.1. Nuevas tendencias en el mundo DW. El Datawarehouse 2.0.

Los sistemas DW han evolucionado en los últimos años conforme han surgido nuevas necesidades. Los motivos de esta evolución son varios, y los podemos resumir en:

- Uso de herramientas de análisis que obligaban a estructuras diferentes optimizadas al uso de determinadas tecnologías (por ejemplo el data mining o el uso de herramientas estadísticas).
- Simplificación de la gestión de sistemas DW complejos formados por múltiples datamarts orientados a cada departamento en los que se pierde el concepto de Corporativo (que hace que se pierdan oportunidades).
- De la unión de múltiples aplicaciones pequeñas (Datamarts o Datawarehouse), no surge toda la información corporativa. Sería necesario construir este Centro a partir del cual se van a generar todos los DW necesarios para todos los ámbitos de análisis.
- Proceso Online: los procesos de actualización hacían que hubiera muchos momentos en los que no se podía acceder a los datos. Igualmente, podría haber cierto retardo en la disponibilidad de la información, lo que nos impedía poder hacer análisis inmediatos (análisis más orientados a la operación del negocio).
- Evolución tecnológica en las herramientas ETL, costes de la tecnología (los costes han bajando de tal forma que permiten abordar los proyectos de una forma mas amplia), etc.

Por todo esto surge el concepto de CIF (Corporate Information Factory), que podría incluir todos los elementos que vemos en la siguiente imagen:

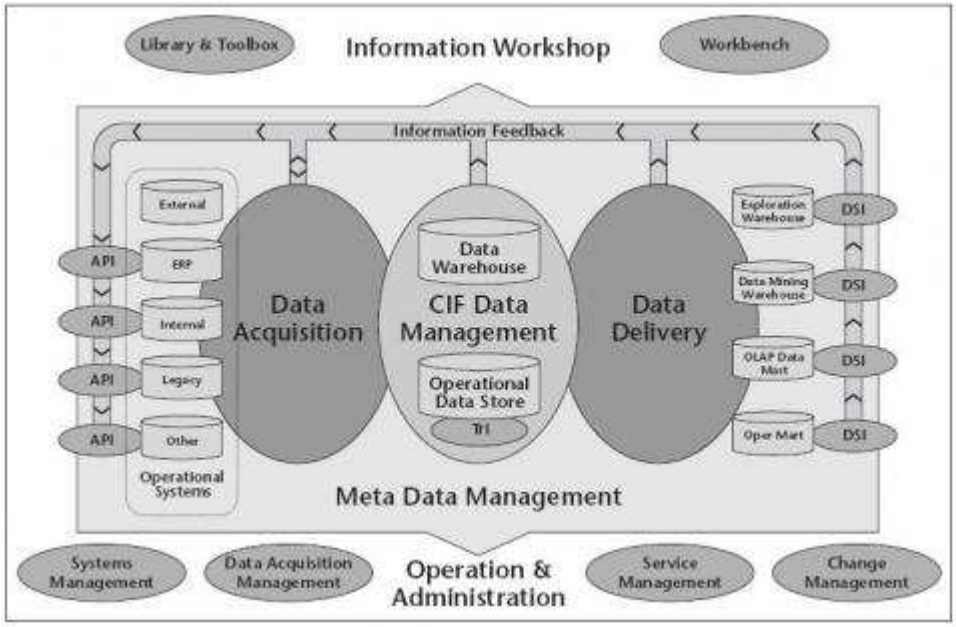


Figura 43: CIF(Corporate Information Factory)

El *Corporate Information Factory (CIF)* es una arquitectura conceptual que describe y categoriza los almacenes de información usados para operar y gestionar con éxito una infraestructura de BI robusta.

El uso de esta arquitectura o de otras más sencillas va a depender del tipo de compañía, los requerimientos de análisis y hasta donde se quiera llegar en el uso del BI. Los elementos que forman el CIF, de forma resumida, son los siguientes:

- *Data Warehouse*: es el almacén de datos, según las definiciones vistas hasta ahora. Pero además, en esta arquitectura, es el punto central de la integración de datos. Centraliza toda la información, nos da una visión en común de la información de toda la organización y proporciona los datos para llenar de contenido el resto almacenes de datos específicos, a través de los procesos de Data delivery (extracción de datos con condiciones de filtrado, sumarización, etc. para otros tipos de análisis).

- *Operational Data Store*: es un almacén de datos, como el DW, pero orientado a las toma de decisiones tácticas. Se alimenta de datos actuales de los sistemas operacionales, nos es un sistema histórico, tiene la información mucho mas en detalle y los tiempos de actualización suelen ser mucho mas rápidos para permitir la toma de decisiones rápidas sobre los datos de operación del negocio. Sería un sistema cercano al tiempo real y suele incluir información sobre clientes, materiales, stocks, ventas, etc.

- *Data Acquisition*: son todas las herramientas y sistemas de gestión que nos permite la extracción, transformación y carga de los datos provenientes de los diferentes sistemas origen (sistemas externos, ERP, sistemas internos, ficheros, etc.), en nuestro Datawarehouse. Serian las herramientas ETL y los sistemas de gestión de la adquisición de datos (Data Acquisition Management).

- *Data Delivery*: son las operaciones de agregación de la información, filtrado por dimensiones específicas o requerimientos de negocio, reformateo o procesamiento de la información para soportar el uso de herramientas de BI específicas, y finalmente, la transmisión de la información a través de la organización (para dar contenido a los Datamarts o Warehouse específicos).

A partir del DW podremos construir subconjuntos de él orientados al uso de técnicas específicas de BI:

- *Exploration Warehouse*: almacén de exploración para utilizar herramientas de tipo estadístico y de exploración.
- *Data Mining Warehouse*: almacén para el uso de técnicas de datamining.
- *Olap Data Mart*: almacén de datos para el uso de análisis multidimensionales (tipo OLAP).
- *Operational Mart*: subconjunto del ODS (operational Data Store), para permitir análisis operacional restringido a un ámbito menor.

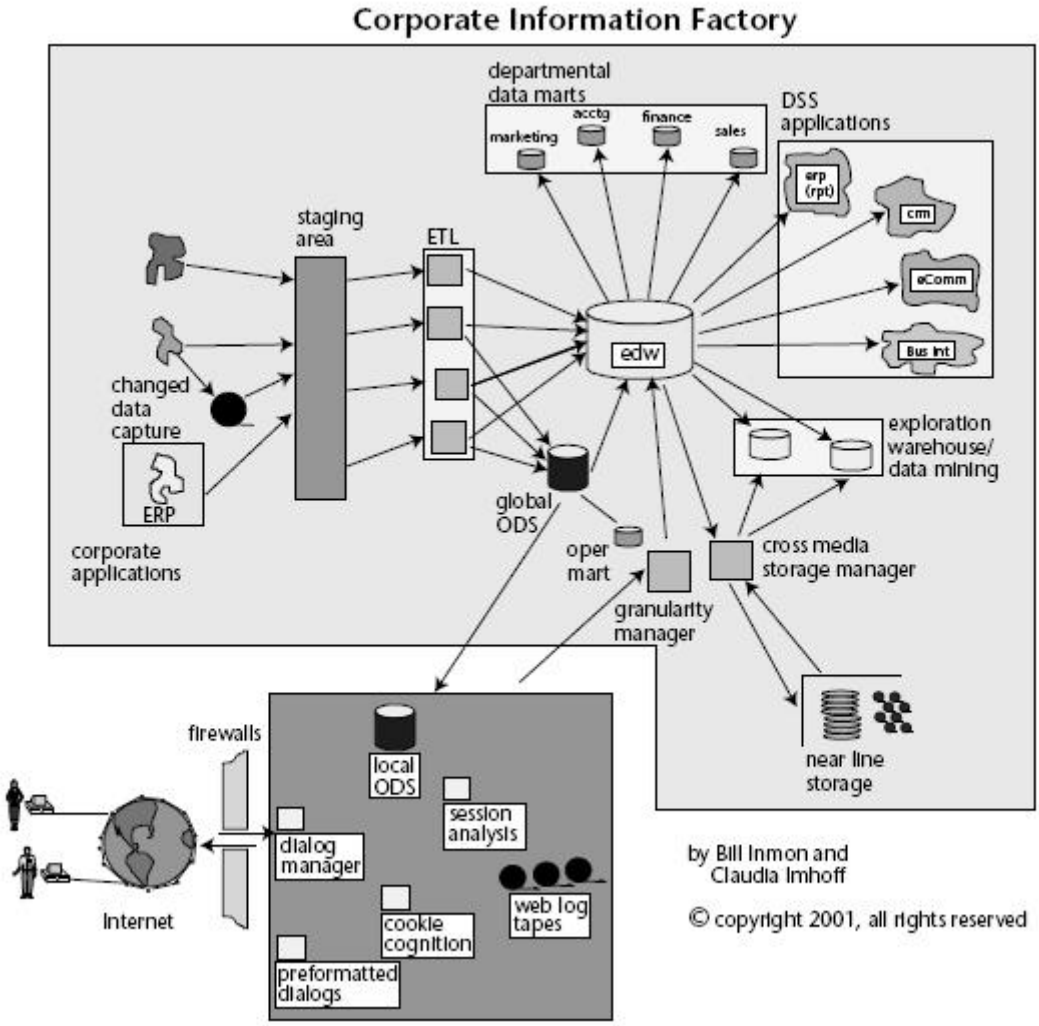


Figura 44: Corporate Information Factory

2.- Glosario

- **Análisis multidimensional**: El análisis multidimensional consiste en analizar los datos que hacen referencia a hechos desde la perspectiva de sus componentes o dimensiones, abarcando también los diferentes niveles que puedan tener las respectivas dimensiones.

- **Base de Datos**: Es un sistema que almacena datos que están relacionados.

- **Corporate Information Factory (CIF)**: es una arquitectura conceptual que describe y categoriza los almacenes de información usados para operar y gestionar con éxito una infraestructura de BI robusta.

- **Cubo OLAP**: Los cubos de información o cubos OLAP funcionan como los cubos de rompecabezas en los juegos, en el juego se trata de armar los colores y en el DWH se trata de organizar los datos por tablas o relaciones; los primeros (el juego) tienen 3 dimensiones, los cubos OLAP tienen un número indefinido de dimensiones, razón por la cual también reciben el nombre de hipercubos. Un cubo OLAP contendrá datos de una determinada variable que se desea analizar, proporcionando una vista lógica de los datos provistos por el sistema de información hacia el DWH, esta vista estará dispuesta según unas dimensiones y podrá contener información calculada. El análisis de los datos está basado en las dimensiones del hipercubo, por lo tanto, se trata de un análisis multidimensional.

- **Data Warehouse**: Es la integración de datos consolidados, almacenados en un dispositivo de memoria no volátil, proveniente de múltiples y posiblemente diferentes fuentes de datos. Con el propósito del análisis y a partir de este tomar decisiones en función de mejorar la gestión del negocio. Contiene un conjunto de cubos de datos que permiten a través de técnicas de OLAP consolidar, ver y resumir los datos acorde a diferentes dimensiones de estos. (Chaudhuri & Dayal, 1997)

- **Data Marts:** Es un subconjunto del Data Warehouse, usado normalmente para el análisis parcial de los datos. Por ejemplo, el Data Mart de los datos del departamento ventas y el Data Mart de Inventarios. El objetivo de subdividir está dado por la complejidad computacional del análisis global de todas las dimensiones del Data Warehouse y por la necesidad de rapidez. (Microsoft Data Warehouse Training, 2000)

- **Data Mining:** Es el descubrimiento de conocimiento oculto en las bases de datos. Relaciones entre estos y tendencias que permiten una toma de decisiones acertada. Incluye Asociación, Caracterización, Clasificación, Análisis de Series Cronológicas, etc. (Chaudhuri & Dayal, 1997).

- **Diagrama de casos:** En el Lenguaje de Modelado Unificado (UML), un diagrama de casos de uso es una especie de diagrama de comportamiento.

- **Dimensiones:** Las dimensiones de un cubo son atributos relativos a las variables, son las perspectivas de análisis de las variables (forman parte de la tabla de dimensiones). Son catálogos de información complementaria necesaria para la presentación de los datos a los usuarios, como por ejemplo: descripciones, nombres, zonas, rangos de tiempo, etc. Es decir, la información general complementaria a cada uno de los registros de la tabla de hechos.

- **Esquema de estrella**: En las bases de datos usadas para data warehousing, un esquema en estrella es un modelo de datos que tiene una tabla de hechos (o tabla fact) que contiene los datos para el análisis, rodeada de las tablas de dimensiones. Este aspecto, de tabla de hechos (o central) más grande rodeada de radios o tablas más pequeñas es lo que asemeja a una estrella, dándole nombre a este tipo de construcciones.

- **Proceso ETL (extracción, transformación y carga):** Los procesos de extracción, transformación y carga (ETL) son importantes ya que son la forma en que los datos se guardan en un almacén de datos (o en cualquier base de datos). Implican la extracción (acción de obtener la información deseada a partir de los datos almacenados en fuentes externas), la transformación (cualquier operación realizada sobre los datos para que puedan ser cargados en el *data warehouse* o se puedan migrar de éste a otra base de datos) y la carga (almacenar los datos en la base de datos final, por ejemplo el almacén de datos objetivo normal)

- **Granularidad :** se refiere a la especificidad a la que se define un nivel de detalle en una tabla, es decir, si hablamos de una jerarquía la granularidad empieza por la parte más alta de la jerarquía, siendo la granularidad mínima, el nivel más bajo. En Data Warehouse, no sólo existe granularidad para las tablas de hechos (Fact's), también existe granularidad en las dimensiones.

- **Indicadores**: También llamadas “indicadores de gestión”, son los datos que están siendo analizados. Forman parte de la tabla de hechos. Más formalmente, las variables representan algún aspecto cuantificable o medible de los objetos o eventos a analizar.

- **Lenguaje de modelado unificado (UML)**: El Lenguaje de Modelado Unificado define una notación gráfica para representar casos de uso llamada modelo de casos de uso.

- **Modelo conceptual:** El modelo conceptual no es más que un esquema de los datos y de las asociaciones o interrelaciones que se han obtenido a partir del conocimiento de los requerimientos y especificaciones de lo que se quiere obtener al final.

- **OLTP (Online Transaction Proccesing)**: Se les llama así a las aplicaciones orientadas principalmente a la inserción, actualización y eliminación de datos, diseñada casi siempre usando el modelo Relacional. Estos sistemas están optimizados para realizar estas operaciones en un tiempo corto. (Microsoft Books Online, 2000)

- **OLAP (Online Analitical Proccesing)**: Son los sistemas que se usan para analizar los datos que las OLTP introducen en la Base de Datos. A diferencia de los primeros estos casi siempre usan el modelo multidimensional para organizar los datos en la Base de Datos ya que brindan mejores resultados a la hora del análisis de estos. (Microsoft Books Online, 2000)

- **ROLAP:** significa **P**rocesamiento **A**nalítico **O**n**L**ine **R**elacional, es decir, se trata de sistemas y herramientas OLAP (*Procesamiento Analítico OnLine*) construidos sobre una base de datos relacional. Es una alternativa a la tecnología MOLAP (**M**ultidimensional OLAP) que se construye sobre bases de datos multidimensionales. Ambos tipos de herramientas, tanto ROLAP como MOLAP, están diseñadas para realizar análisis de datos a través del uso de modelos de datos multidimensionales, aunque en el caso de **ROLAP** estos modelos no se implementan sobre un sistema multidimensional, sino sobre un sistema relacional clásico.

- **Sistemas de soporte a decisiones (DSS)**: es un sistema informático utilizado para servir de apoyo, más que automatizar, el proceso de toma de decisiones. La decisión es una elección entre alternativas basadas en estimaciones de los valores de esas alternativas. El apoyo a una decisión significa ayudar a las personas que trabajan solas o en grupo a reunir inteligencia, generar alternativas y tomar decisiones. Apoyar el proceso de toma de decisión implica el apoyo a la estimación, la evaluación y/o la comparación de alternativas. En la práctica, las referencias a DSS suelen ser referencias a aplicaciones informáticas que realizan una función de apoyo.

- **Tablas de hecho**: es la tabla central de un esquema dimensional (en estrella o en copo de nieve) y contiene los valores de las medidas de negocio. Cada medida se toma mediante la intersección de las dimensiones que la definen, dichas dimensiones estarán reflejadas en sus correspondientes tablas de dimensiones que rodearán la tabla de hechos y estarán relacionadas con ella.

- **Toma de decisiones**: Es el proceso durante el cual se debe de escoger entre dos o más alternativas.

3.- Bibliografía

- Documento "La factoria d' informació corporativa", Alberto Abelló Gamazo y José Samos Jiménez, de la asignatura Magatzem de dades i models multidimensionals de la Universitat Oberta de Catalunya.
- Documento "Disseny multidimensional", Alberto Abelló Gamazo, de la asignatura Magatzem de dades i models multidimensionals de la Universitat Oberta de Catalunya.
- "Building the Data Warehouse", William H.Inmon (Fourth Edition)
- "Building a Data Warehouse, with examples in SQL Server", Vincent Rainardi (2008)
- "The Data Warehouse Toolkit, the complete guide to dimensional modele", Ralph Kimball, Margy Ross (Second Edition)
- "Oracle9*i* Data Warehousing Guide, Release 2 (9.2)", Paul Lane, Valerie Moore (Marzo 2002).
- http://es.wikipedia.org
- http://www.google.es
- http://www.dataprix.com
- http://www.businessintelligence.info/
- http://www.monografias.com
- http://cisnet.baruch.cuny.edu/holowczak/oracle/sqlplus/
- http://www.orafaq.com/wiki/SQL*Loader_FAQ
- http://www.oracleutilities.com/OSUtil/sqlldr.html
- http://www.rkimball.com/
- http://www.monografias.com/trabajos16/warehousing-discoverer/warehousing-discoverer.shtml
- http://download.oracle.com/docs/html/B10272_01/toc.htm
- http://www.embalses.net/
- http://www.chebro.es/

4.- Anexos

En el anexo de esta memoria hemos añadido cuestiones más técnicas:

- la creación de las tablas auxiliares
- la creación de las tablas finales
- los procesos ETL generados

4.1.- Creación tablas auxiliares

Previa a la carga de la base de datos con la que trabajaremos en el Business Object, vamos a trabajar con tablas intermedias auxiliares. Cargaremos inicialmente la información que tenemos en cada fila de los Excel proporcionados.

El diseño físico de las tablas auxiliares con las que trabajaremos serán las siguientes:

- *LOCALIZACION*

LOCALIZACION		
	TIPO	**DESCRICION**
Comunidad	CHAR(50)	Descripción de comunidad
Provincia	CHAR(50)	Descripción de provincia
Embalse	CHAR(50)	Descripción de embalse
Rio	CHAR(50)	Descripción de río

- *DATOS*

DATOS		
	TIPO	**DESCRICION**
Embalse	CHAR(50)	Descripción de embalse
Rio	CHAR(50)	Descripción de río
CapTotal	NUM(9,2)	Capacidad total embalse
HmEmbalsada	NUM(9,2)	Agua embalsada en esta fecha (Hm^3)
TpcEmbalsada	NUM(9,2)	Porcentaje agua embalsada en esta fecha
VsIn	NUM(9,2)	Variación semanal (entrada) en Hm^3
VsOut	NUM(9,2)	Variación semanal (salida) en Hm^3
VsInc	NUM(9,2)	Incremento semanal
Hm1A	NUM(9,2)	Agua embalsada (Hm^3) hace un año
Tpc1A	NUM(9,2)	Porcentaje agua embalsada hace un año
Ano	NUM(4)	Año
Mes	NUM(2)	Mes
Nieve	NUM(1)	Estimación de nieve según año

- *DESCARTES*

  Tabla idéntica a la tabla auxiliar **DATOS** y que se utilizará para guardar los registros descartados a la hora de cargarlas.

Para generarlas, se deberá de ejecutar desde la consola del MS-DOS el siguiente script:

- E:\TFC\SQL\02_Create_Tablas_aux.txt

con el siguiente comando:

- SQLPLUS usuario/pw@XE @fichero.txt

Obteniendo la siguiente estadística de la ejecución del script:

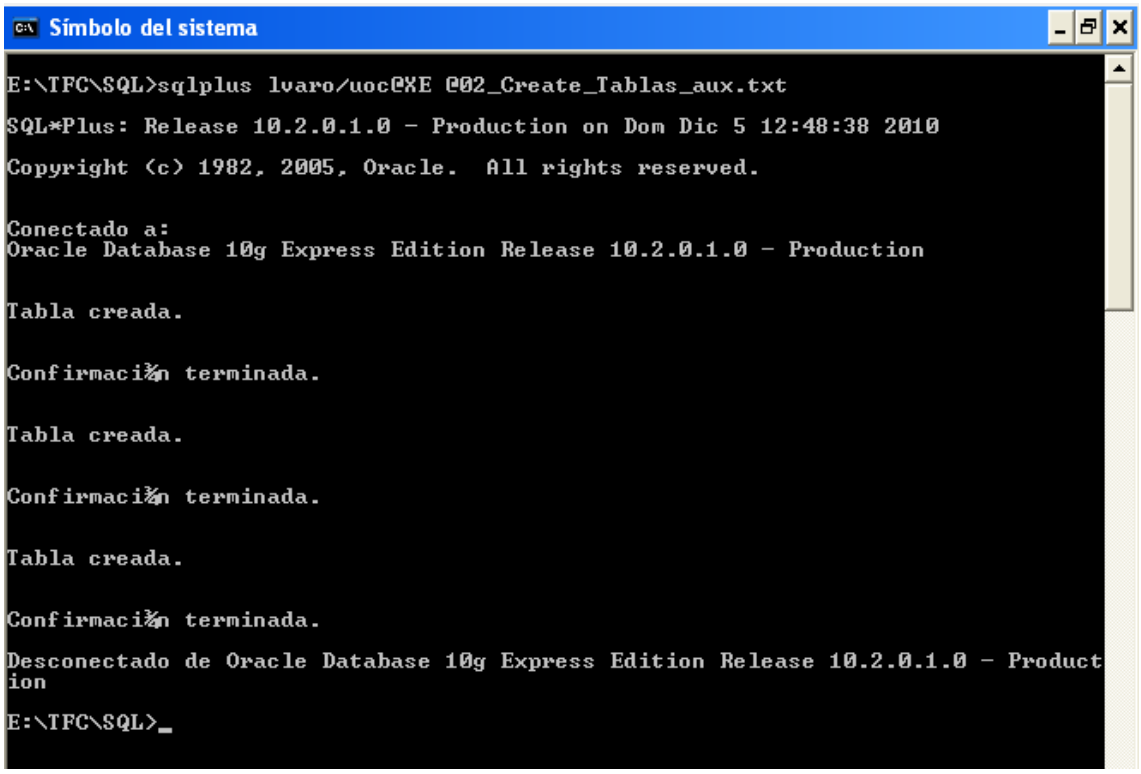


Figura 45: Ejecución script creación tablas desde MS-DOS

4.2.- Creación tablas BBDD

Las tablas finales se podrán generar desde la consola del MS-DOS y con el usuario LVARO/UOC, de la misma manera que se han generado las tablas auxiliares y con el siguiente scripts:

- E:\TFC\SQL\07_Create_Tablas_finales.txt

1.3.3.4. Automatización generación tablas

La automatización de la creación de las tablas, tanto auxiliares como finales, se puede realizar gracias al fichero ".bat" ubicado en:

- E:\TFC\SQL\01_CREATE_Tablas.bat

4.3.- Carga Base de datos (procesos ETL)

4.3.1. Proceso ETL

El proceso ETL seguido se puede resumir en la siguiente figura:

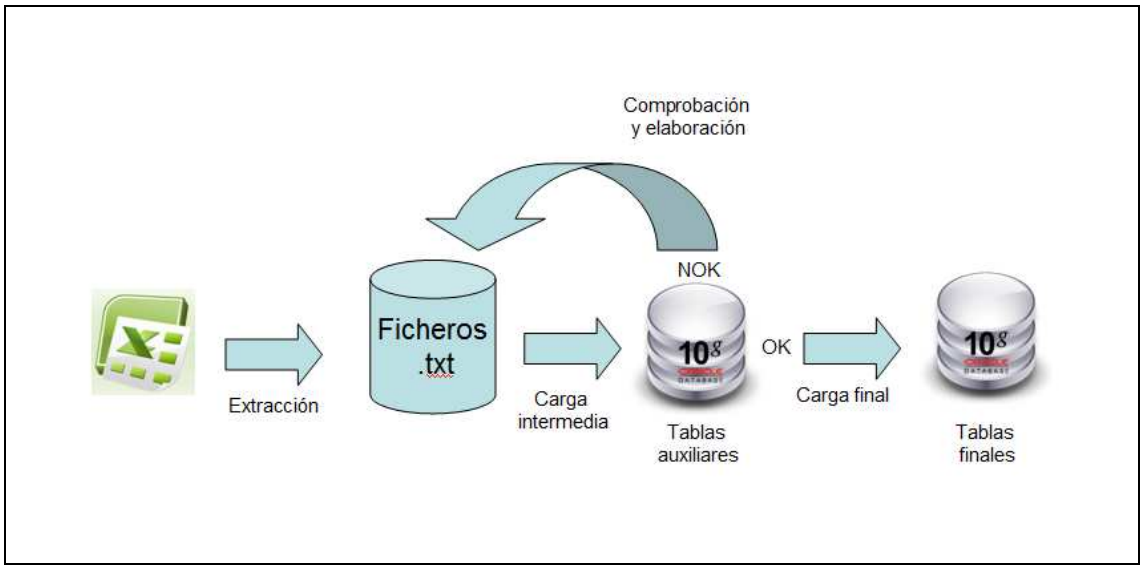


Figura 46: Proceso ETL seguido

- En primer lugar, se ha realizado la extracción de las hojas Excel proporcionadas por el cliente en ficheros .txt
- Con esta información, se han cargado las tablas auxiliares.
- Posteriormente se ha realizado una comprobación y una depuración de los datos. Esto ha hecho necesario de nuevo el cargar las tablas auxiliares y realizar la comprobación de los datos.
- Cuando se ha dado por buena la información cargada en dichas tablas, finalmente se ha realizado la carga de las tablas finales.

En los puntos siguientes explicaremos con más detalle lo realizado.

4.3.2. Tratamiento hojas Excel

Se ha recogido del Excel embalses información para obtener el fichero “embalses.csv”, que nos ayudará a cargar la tabla auxiliar **LOCALIZACION**. Fundamentalmente, se ha tratado para añadir a cada fila la correspondiente comunidad.

A su vez, y gracias a una macro generada en Excel (esta macro se encuentra en E\TFC\CARGA\CHNE\Carga.xls, en la misma carpeta que se encuentran los Excel de datos por año), se crea en el fichero “carga.csv”, que servirá para cargar la tabla auxiliar “DATOS”.

A destacar que en esta macro hemos tenido en cuenta lo siguiente:

- En la columna de “Variación semanal”, hay algunas filas que tanto en los valores de entrada como en los de salida no hay información. De hecho, la entrada está a blancos y la salida está con “---“. Lo que sí está informado es el incremento.

  Asumiremos que el incremento siempre es Entrada-Salida. Por lo tanto, si es negativo:

  - Entrada es igual a 0
  - Salida es igual al incremento

  Y si es positivo:

  - Entrada es igual al incremento
  - Salida es igual a 0

- Se ha añadido una función que obtiene según tabla proporcionada, la estimación de nieve por año y embalse. Este valor nos servirá para calcular el valor de la tupla “Agua_nieve_equivalente “de la tabla VOLUMEN.

NOTA: La ejecución de esta macro se ha realizado fuera de la máquina virtual, al no tener ésta el Office cargado. Por esta razón, internamente tiene la dirección en la que se encontraban las hojas Excel en la máquina de trabajo. Si se desea probar, se deberá de editar y modificar esta dirección.

4.3.3. Problemas iniciales

Una vez ya generadas las tablas auxiliares, desde la consola del MS-DOS, ejecutamos el siguiente comando:

```
sqlldr lvaro/uoc control='e:\tfc\carga\localizacion.ctl' log='e:\tfc\carga\localizacion.log'
```

Este comando carga la tabla auxiliar **LOCALIZACION** desde el fichero “localizacion.ctl” que contiene lo siguiente:

```
LOAD DATA
CHARACTERSET 'WE8ISO8859P1'
INFILE 'e:\tfc\carga\embalses.csv'
INTO TABLE localizacion
REPLACE
FIELDS TERMINATED BY ';'
TRAILING NULLCOLS
( comunidad     "ltrim(rtrim(upper(:comunidad)))",
  provincia     "ltrim(rtrim(upper(:provincia)))",
  embalse       "ltrim(rtrim(upper(:embalse)))",
  rio           "ltrim(rtrim(upper(:rio)))")
```

Según el fichero “localización.log” se cargan 41 tuplas.

A su vez, realizaremos lo mismo con la tabla auxiliar **DATOS**:

```
sqlldr lvaro/uoc control='e:\tfc\carga\datos.ctl' log='e:\tfc\carga\datos.log'
```

Con el fichero “datos.ctl” que contiene lo siguiente:

```
LOAD DATA
CHARACTERSET 'WE8ISO8859P1'
INFILE 'e:\tfc\carga\datos.csv'
INTO TABLE datos
REPLACE
FIELDS TERMINATED BY ';'
TRAILING NULLCOLS
( embalse        "ltrim(rtrim(upper(:embalse)))",
  rio            "ltrim(rtrim(upper(:rio)))",
  CapTotal,
  HmEmbalsada,
  TpcEmbalsada,
  VsIn,
  VsOut,
  VsInc,
  Hm1A,
  Tpc1A,
  Ano,
  Mes,
  Nieve)
```

Según el fichero “datos.log” se cargan 3.078 tuplas.

Sin embargo, si hacemos consultas en las tablas cargadas detectamos que:

- Hay 2052 filas en la tabla “DATOS” con nombres de embalses que no están en la tabla “LOCALIZACION”.
- En concreto en la tabla “DATOS” hay 38 nombres de embalses que no están en la tabla “LOCALIZACION”
- Hay 1458 filas en la tabla “DATOS” con nombres de ríos que no están en la tabla “LOCALIZACION”
- En concreto en la tabla “DATOS” hay 21 nombres de ríos que no están en la tabla “LOCALIZACION”
- Finalmente, en la tabla “DATOS” hay 2106 filas cuyo río o embalse no está en la tabla “LOCALIZACION”

En el fichero ubicado en 'E:\TFC\SQL\04_Verificar_Datos_aux.txt' se encuentran las querys realizadas para obtener esta información.

Esto nos hace detectar que hay embalses/ríos que están en el Excel de Embalses pero no están en los Excel de Totales y viceversa.

Por esta causa, nos vemos obligados a hacer un estudio más exhaustivo de los ríos, embalses y comunidades de dichos embalses que tenemos en los Excel proporcionados.

Tras el estudio y depuración de los datos, se ha generado un fichero con la información obtenida, el “embalses_total.csv”, que será con el que se cargue la tabla auxiliar **LOCALIZACION**. Dicho fichero se encuentra en 'e:\tfc\carga\embalses_total.csv'.

Con este fichero volveremos a cargar las tablas auxiliares **LOCALIZACION** y **DATOS**, siguiendo los pasos explicados anteriormente.

Para verificar la correcta carga de los datos de las tablas auxiliares, utilizaremos la tabla de DESCARTES.

Se realizará su carga con la siguiente query (E:\TFC\SQL\05_Descartar_datos_aux.txt):

```
INSERT INTO descartes
SELECT *
FROM datos
WHERE rio NOT IN (SELECT rio FROM localizacion) or
     embalse NOT IN (SELECT embalse FROM localizacion);
COMMIT;
```

y obtenemos lo siguiente:

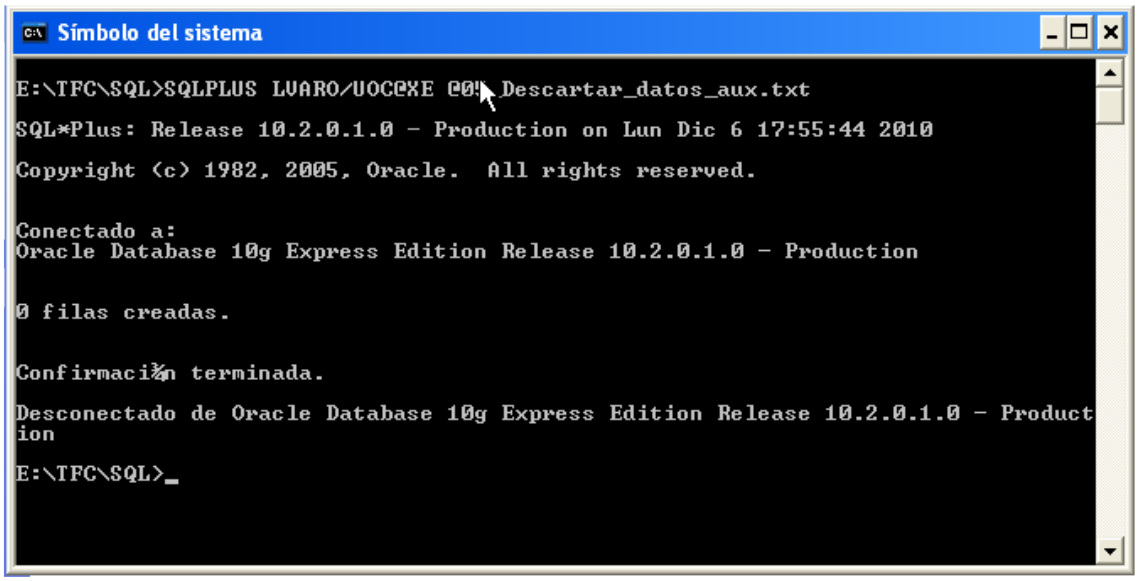


Figura 47: Ejecución script verificación datos desde MS-DOS

Se comprueba que no se ha descartado ningún registro.

4.3.4. Carga tablas finales

- Las tablas de dimensiones se cargaran con las siguientes sentencias:

```
INSERT INTO Comunidad (nombre)
SELECT DISTINCT comunidad
FROM localizacion;
COMMIT;

INSERT INTO Embalse (nombre)
SELECT DISTINCT embalse
FROM localizacion;
COMMIT;

INSERT INTO Rio (nombre)
SELECT DISTINCT rio
FROM localizacion;
COMMIT;

INSERT INTO Tiempo (nany)
SELECT DISTINCT ano
FROM datos
ORDER BY ano;
COMMIT;

INSERT INTO Any_mes (id_any, mes)
SELECT t.id_any, d.mes
FROM tiempo t, datos d
WHERE t.nany = d.ano
GROUP BY t.id_any, d.mes
ORDER BY t.id_any, d.mes;
COMMIT;
```

- La tabla de hecho **VARIACION_SEMANAL** se cargará con la siguiente sentencia:

```
INSERT INTO variacion_semanal (id_anymes, Id_Embalse, Id_Rio, Entrada, Salida, Incremento)
SELECT
  (SELECT a.id_anymes
   FROM any_mes a, tiempo t
   WHERE t.nany = d.ano and
              t.id_any = a.id_any and
             d.mes = a.mes),
  (SELECT e.Id_Embalse
   FROM embalse e
   WHERE d.embalse = e.nombre),
  (SELECT r.Id_Rio
   FROM rio r
   WHERE d.rio = r.nombre),
              d.VsIn, d.VsOut, d.VsInc
FROM datos d
WHERE rio IN (SELECT nombre FROM rio) and
            embalse IN (SELECT nombre FROM embalse);
COMMIT;
```

- La carga de la tabla de hecho **VOLUMEN** tiene unas características que se mencionan a continuación:

  - Para obtener la información de carga de la fila “Agua_diaria_disponible”, se ha tomado la capacidad de agua embalsada en esta fecha (año/mes) y se ha dividido por los días del mes (obtenido por sentencias SQL).
  - Para obtener la información de la carga de la fila “Agua_nieve_equivalente”, se ha seguido la siguiente fórmula:

$$Hm3_agua_nieve_equivalente_anual_embalse = estimacion_nieve_embalse \frac{capacidad_actual_embalse_{mes}}{5}$$

La estimación de nieve por embalse la hemos cargado en la tabla auxiliar **DATOS**. Teniendo la información de la capacidad actual del embalse, se puede calcular sin problemas.

Con lo que finalmente, el script de carga de la tabla **VOLUMENES** será el siguiente:

```
INSERT INTO volumen (id_anymes, Id_Embalse, Id_Comunidad, Id_Rio,
                     Capacidad_total_embalse, Capacidad_agua_embalsada,
                     Capacidad_agua_embalsada_ant,
                      Agua_diaria_disponible, Agua_nieve_equivalente )
SELECT
  (SELECT a.id_anymes
   FROM any_mes a, tiempo t
   WHERE t.nany = d.ano and
            t.id_any = a.id_any and
           d.mes = a.mes),
  (Select e.Id_Embalse
   FROM embalse e
   WHERE d.embalse = e.nombre),
  (Select c.Id_Comunidad
   FROM comunidad c, localizacion l
   WHERE d.embalse = l.embalse  and
            l.comunidad = c.nombre),
  (Select r.Id_Rio
   FROM rio r
   WHERE d.rio = r.nombre),
  d.CapTotal,
  d.HmEmbalsada,
  d.Hm1A,
  d.HmEmbalsada/to_char(last_day(to_date(11*1000000+mes*10000+ano,'ddmmyyyy')),'dd'),
  (SELECT avg(da.HmEmbalsada)*da.nieve/5
   FROM datos da
   WHERE da.ano = d.ano and
            da.embalse = d.embalse
   group by da.embalse, da.ano, da.nieve)
FROM datos d
WHERE rio IN (SELECT nombre FROM rio) and
          embalse IN (SELECT nombre FROM embalse);
COMMIT;
```

4.3.5. Automatización carga

La automatización de la carga final de las tablas, se puede realizar gracias a l fichero ".bat" ubicado en:

- E:\TFC\SQL\02_CARGA_Tablas.bat

4.3.6. Automatización recarga

En este punto explicamos como se debería de actuar a la hora de cargar nueva información.

Previamente a la carga de nuevos datos en el DWH, nos deberán de proporcionar las mismas hojas Excel que nos han suministrado para realizar la primera carga. Serán las siguientes:

1. Excel embalses, para añadir las nuevas combinaciones de comunidad/embalse/río si fuera necesario. Se obtendría un nuevo fichero "embalses.csv".
2. Excel de datos, para que con la macro generada en Excel, podremos obtener un nuevo fichero "datos.csv".

A continuación, bastaría con ejecutar el fichero ".bat" ubicado en:

- E:\TFC\SQL\03_RECARGA_Tablas.bat

A la hora de ejecutar este ".bat" lo que se realiza es:

1. Borrado de las tuplas de las tablas auxiliares
2. Carga de las tablas auxiliares con la nueva información
3. Carga de las tablas finales con la nueva información, con el control interno de claves duplicadas (sentencia SLQ "NOT EXISTS")

Todos los scripts que se han utilizado se pueden encontrar en E:\TFC\SQL\ del disco de la máquina virtual:

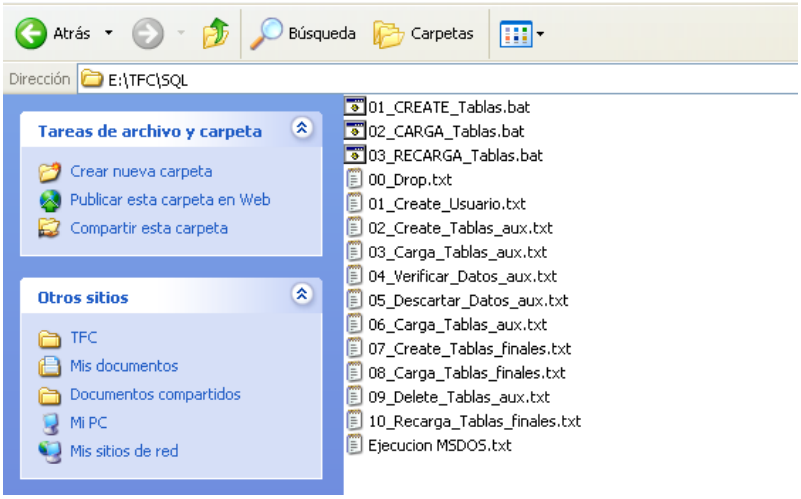


Figura 48: Scripts en la MV