

Construcción y explotación de un almacén de datos de planificación hidrológica

José Manuel Ibáñez Lora
ETIG

José Angel Martín Carballo

10/01/2011

RESUMEN

Este documento es la memoria del trabajo de final de carrera (TFC) del área de almacenes de datos, dentro de los estudios de Ingeniería Técnica en Informática de Gestión de la Universitat Oberta de Catalunya (UOC).

El proyecto a desarrollar en dicho TFC versa sobre el diseño, implementación y explotación de un almacén de datos para la Confederación Hidrográfica del Noreste (CHNE). Partiendo de sus fuentes de datos construiremos un almacén de datos del cual poder obtener un conjunto de informes predefinidos, según las especificaciones del CHNE.

Contenido

1.	Introducción.....	4
1.1.	Justificación del proyecto y contexto en el cual se desarrolla: punto de partida y aportación del TFC.....	4
1.2.	Objetivos.....	4
1.3.	Enfoque y método seguido.....	5
1.4.	Planificación del proyecto.....	5
1.4.1.	Calendario laboral del proyecto.....	5
1.4.2.	Fechas relevantes del TFC.....	5
1.4.3.	Recursos humanos disponibles.....	5
1.4.4.	Plan de contingencia.....	6
1.4.5.	Solapamiento con otras asignaturas.....	6
1.4.5.1.	Avería del punto de trabajo.....	6
1.4.6.	Motivos de salud.....	6
1.4.7.	Cuadro de tareas.....	7
1.4.8.	Diagrama de Gantt.....	8
1.5.	Productos obtenidos.....	9
1.6.	Breve descripción de los otros capítulos de la memoria.....	9
2.	Análisis.....	10
2.1.	Diagramas de casos de usos con una explicación de cada uso.....	10
2.2.	Diagrama del modelo.....	11
2.2.1.	Estudio de las dimensiones del modelo.....	12
2.2.2.	Modelo conceptual.....	14
3.	Diseño.....	15
3.1.	Diagrama de la arquitectura software.....	15
3.1.1.	Requisitos de carga y actualización.....	15
3.1.2.	Diagrama de la arquitectura del software utilizado.....	16
3.2.	Diagrama de la arquitectura hardware.....	17
3.3.	Diseño de la base de datos y diagrama del modelo físico.....	18
3.3.1.	Almacén de datos departamental.....	18
3.3.2.	Componentes de extracción y carga.....	19
3.3.3.	Modelo conceptual de la capa de extracción y transformación.....	19
3.4.	Diseño y descripción de los informes creados.....	20
4.	Capturas de pantalla.....	22
5.	Conclusiones.....	26
6.	Líneas de evolución futuras.....	26
7.	Glosario.....	27
8.	Bibliografía.....	27
9.	Anexos.....	28

1. Introducción.

1.1. Justificación del proyecto y contexto en el cual se desarrolla: punto de partida y aportación del TFC.

La Confederación Hidrográfica Nord-Este (CHNE) es la entidad encargada de la gestión del agua y de las infraestructuras hidráulicas del Nord-Este de la península ibérica. Parte de su actividad consiste en realizar un seguimiento periódico del estado de los diferentes embalses de su responsabilidad, comprobar tendencias y predecir periodos de sequía, de cara a la toma de correctas decisiones hidrológicas por parte de las autoridades competentes. Actualmente este trabajo se realiza mediante datos que se reciben periódicamente en hojas de Excel y de un tratamiento manual realizado por los técnicos del departamento responsable de planificación hidrológico con herramientas ofimáticas. Este trabajo es muy pesado y supone disponer de un elevado número de recursos humanos.

Las aportaciones de este proyecto al CHNE son:

- Automatización de la centralización de los datos. La unificación de las fuentes de datos en un único almacén de datos departamental, facilitaran el mantenimiento y la integridad de los mismos. También permitirán un acceso único y estándar para todos los usuarios.
- Una herramienta para los usuarios finales fácil de utilizar que no requiera de conocimiento informáticos técnicos. Esta facilidad en el uso permitirá a los usuarios ganar en productividad, al dedicar sus esfuerzos y su tiempo a lo que realmente interesa al CHNE: el análisis de los datos.

1.2. Objetivos.

Dada la situación descrita en el apartado anterior, la CHNE ha decidido encargarnos como consultoría externa al organismo, la creación de un almacén de datos que permita automatizar la recogida de la información proveniente de las hojas Excel y explotar esta información con herramienta de Business Intelligent (BI) adecuadas.

Se desea obtener al menos la siguiente información distribuida en dos bloques:

- Bloque primero:
 - Capacidad total Vs capacidad actual en Hm3 por rio y año-mes.
 - Capacidad total Vs capacidad actual en Hm3 por embalse año-mes.
 - Capacidad total Vs capacidad actual en Hm3 por comunidad autónoma año-mes.
 - Comparativa anual de la evolución por comunidad autónoma de la capacidad de agregada de todos los embalses.
 - Comparativa de la capacidad media agregada de todos los embalses de una comunidad autónoma entre un año y el anterior (valores i porcentajes de incremento año anterior-año actual).
 - Aportación (Entrada-Salida) en Hm3 por rio, embalse y año-mes, con comparativa con la media anual de aportación del embalse.
 - Agua diaria en reserva disponible (Hm3/día) en toda la confederación hidrográfica por año-mes.
- Bloque segundo:
 - Gráfico de la evolución de los Hm3 d agua de nieve equivalente prevista por año y embalse.
 - Mostrar cómo podemos predecir, a partir del almacén de datos, periodos de sequía.

1.3. Enfoque y método seguido.

El enfoque y método seguido para la realización de este proyecto se corresponde con el ciclo de vida clásico de un proyecto informático, en el cual se han establecido las siguientes fases:

- Planificación y análisis previo de requisitos.
- Análisis y diseño conceptual y físico.
- Construcción.
- Pruebas.

En nuestro proyecto, estas etapas se corresponden con la elaboración de documentación y la entrega parcial del trabajo realizado. En cada una de estas fases, las conclusiones obtenidas y documentadas han sido el punto de partida de la siguiente fase. Aun siendo un enfoque y método orientado hacia el ciclo de vida en cascada, se ha dado un cariz iterativo e incremental, mediante entregas preliminares al consultor con la intención de recibir las indicaciones de este en cada una de las fases y corregir posibles deficiencias e incorporar las mejoras propuestas.

1.4. Planificación del proyecto.

1.4.1. Calendario laboral del proyecto.

Se establece el siguiente calendario laboral para el desarrollo del proyecto.

- Lunes a viernes de 21:30 a 23:30
- Sábados y domingos de 10:00 a 13:00
- Días festivos:
 - El 25 y 26 de diciembre 2010.
 - 1 y 2 de diciembre de 2010.

1.4.2. Fechas relevantes del TFC.

Se adjunta el calendario a cumplir según requerimiento de la asignatura TFC-Almacenes de Datos de la UOC. La pretensión es entregar las diferentes fases de proyecto con la suficiente antelación a la fecha tope de entrega, a fin de que el consultor pueda hacer las puntualizaciones que encuentre oportunas e incluirlas en el documento final.

Título	Fecha inicio	Fecha entrega
Pac1	23/09/2010	05/10/2010
Pac2	06/10/2010	08/1/2010
Pac3	09/11/2010	20/12/2010
Entrega final	21/12/2010	10/01/2010

1.4.3. Recursos humanos disponibles

Se dispone de los siguientes activos humanos:

- José Manuel Ibáñez Lora: estudiante de Ingeniería Técnica Informática de Gestión.
- José Ángel Martín Carballo: consultor del TFC área almacenes de datos.
- Foro de estudiantes de la asignatura TFC-Almacén de datos.

1.4.4. Plan de contingencia.

1.4.5. Solapamiento con otras asignaturas

Junto con la asignatura TFC-Almacén de datos en la cual se engloba este proyecto de construcción de explotación de almacén de datos para la CHNE, también se cursan las asignaturas de Análisis y Matemática discreta, de las cuales se adjuntan sus calendarios con las fechas más relevantes.

Análisis.			Matemática discreta		
Título	Fecha inicio	Fecha entrega	Título	Fecha inicio	Fecha entrega
Pac1	04/10/2010	10/10/2010	Pac1 – 1ª	24/10/2010	09/11/2010
Pac2	18/10/2010	24/10/2010	Pac1 – 2ª	07/10/2010	09/11/2010
Pac3	01/11/2010	07/11/2010	Pac2 – 1ª	25/11/2010	27/12/2010
Pac4	15/11/2010	21/11/2010	Pac2 – 2ª	07/12/2010	27/12/2010
Pac5	29/11/2010	05/12/2010			
Pac6	13/12/2010	19/12/2010			

En rojo se han marcado las fechas que pueden entrar en conflicto. Se dará prioridad al calendario del TFC, ya que es primordial cumplir con la evaluación continua para la ejecución del proyecto. Por el contrario, las asignaturas de Análisis y Matemática discreta, aunque recomendable, no es obligatorio seguir la evaluación continua.

1.4.5.1. Avería del punto de trabajo.

Para el desarrollo del proyecto se dispone un ordenador portátil de la empresa con la actualmente colaboro. En el supuesto de que dicho ordenador sufriera una avería que no permitiera la recuperación del puesto de trabajo, también se dispone de otro portátil de uso doméstico, y que se podría utilizar para continuar con el proyecto del TFC. Para la recuperación del puesto de trabajo en caso de una avería crítica, se realizan copias de seguridad semanales de todas las carpetas del proyecto en unidad externa, por lo que el daño implicaría la pérdida de los documentos actualizados o creados como mucho en una semana.

En el supuesto de un fallo en la conexión a Internet se dispone de las siguientes opciones:

- Modem USB Movistar datos con conexión a Internet.
- Conexión a Internet en la empresa.

1.4.6. Motivos de salud.

En el caso de que por motivos de salud no fuera posible cumplir con los plazos estipulados según la planificación adjunto, se evaluarán las siguientes posibilidades:

1. Rehacer la planificación con el fin de poder cumplir con la fecha de entrega de las PAC.
2. Solicitar al consultar aplazamiento en la fecha de entrega.
3. Abandonar el proyecto ante la imposibilidad de asumir los compromisos que el proyecto del TFC exige.

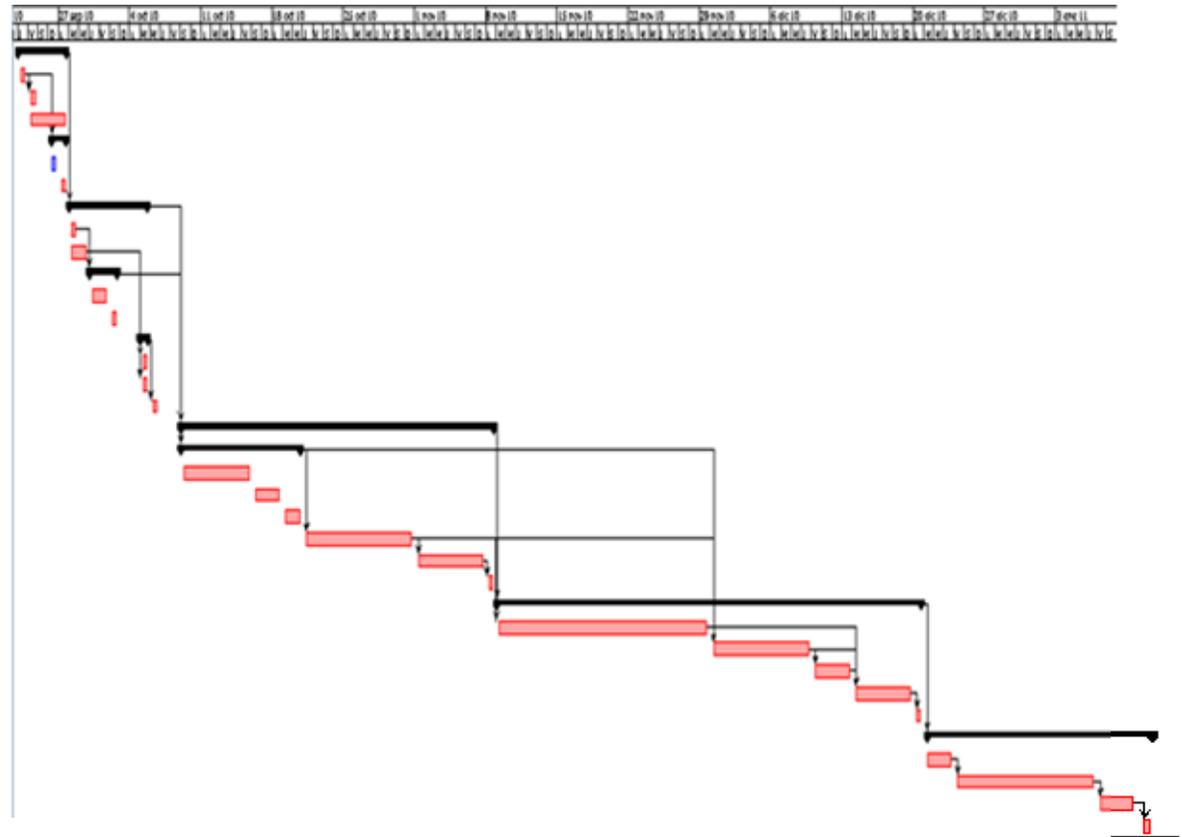
1.4.7. Cuadro de tareas.

Después de una lectura detallada del enunciado del TFC se identifican las siguientes tareas, con la planificación temporal y dependencias que se muestra en el siguiente cuadro:

Tarea	Fechas	Predecesores	Días
1. Tareas previas	23/09/10 a 29/09/10		5
1.1. Descarga del enunciado y de los recursos del aula	23/09/10		
1.2. Primera lectura y comprensión del enunciado	23/09/10	1.1	
1.3. Lectura de los módulos didácticos	24/09/10 a 29/09/10		
1.4. Instalación de las herramientas básicas	26/09/10 a 27/09/10	1.1	
1.4.1. Instalación de la MV	26/09/10	1.1	
1.4.2. Instalación Open Project y Gantt	27/09/10	1.1	
2. Planificación y análisis previo de requisitos	28/09/10 a 05/10/10	1	8
2.1. Lectura detallada del enunciado	28/09/10		
2.2. Planificación del proyecto y diagrama de Gantt	28/09/10 a 29/09/10	2.1	
2.3. Análisis previo de los requisitos	30/09/10 a 04/10/10	2.1	
2.3.1. Dimensiones y atributos	30/09/10 a 01/10/10		
2.3.2. Fuentes de datos	02/10/10 a 04/10/10		
2.4. Redacción documentos finales	05/10/10		
2.4.1. Redacción del documento de planificación	05/10/10	2.2	
2.4.2. Redacción del documento del análisis previo	05/10/10	2.3	
2.5. Entrega Final Pac1	05/10/10	2.4	
3. Análisis de requisitos - Diseño conceptual y físico	09/10/10 a 08/11/10	2	31
3.1. Análisis de requisitos	09/10/10 a 20/10/10	2.3	
3.1.1. Introducción, objetivo, punto de partida y alcance del proyecto	09/10/10 a 15/10/10		
3.1.2. Análisis detallado de las fuentes de datos	16/10/10 a 18/10/10		
3.1.3. Requisitos de almacén de datos	19/10/10 a 20/10/10		
3.2. Diseño conceptual y físico	21/10/10 a 31/10/10	3.1	
3.3. Presentación propuesta Pac2	01/11/10 a 07/11/10	3.2	
3.4. Entrega final Pac2	08/11/10	3.3	
4. Implementación	09/11/10 a 20/12/10	3	42
4.1. Construcción del almacén de datos	09/11/10 a 29/11/10	3.2	
4.2. Implementación de los informes	30/11/10 a 09/12/10	3.1,3.2	
4.3. Análisis de la información obtenida	10/12/10 a 13/12/10	4.2	
4.4. Presentación propuesta Pac3	14/12/10 a 19/12/10	4.1,4.2,4.3	
4.5. Entrega final Pac3	20/12/10	4.4	
5. Entrega Final	21/12/10 a 10/01/11	4	21
5.1. Revisión de las indicaciones del consultor	21/12/10 a 23/12/10		
5.2. Elaboración de la memoria final	24/12/10 a 06/01/11	5.1	
5.3. Preparación de la presentación virtual	07/01/11 a 09/01/11	5.2	
5.4. Entrega Final	10/01/11	5.3	
6. Debate virtual	24/01/11 a 27/01/11	5	4

1.4.8. Diagrama de Gantt.

	Nombre	Duración	Inicio	Terminado	Predecesores
1	<input type="checkbox"/> Tareas previas	5 days	23/09/10 0:00	27/09/10 17:00	
2	Descarga del enunciado y de los recursos del aula	1 day	23/09/10 0:00	23/09/10 17:00	
3	Primera lectura y comprensión del enunciado	1 day	23/09/10 0:00	23/09/10 17:00	2
4	Lectura de los módulos didácticos	4 days	24/09/10 0:00	27/09/10 17:00	
5	<input type="checkbox"/> Instalación de las herramientas básicas	2 days	26/09/10 8:00	27/09/10 17:00	2
6	Instalación de la MV	1 day	26/09/10 8:00	26/09/10 17:00	
7	Instalación OpenProject y Gant	1 day	27/09/10 0:00	27/09/10 17:00	
8	<input checked="" type="checkbox"/> Planificación y análisis previo de requisitos	8 days	28/09/10 0:00	5/10/10 17:00	1
9	Lectura detallada del enunciado	1 day	28/09/10 0:00	28/09/10 17:00	
10	Planificación del proyecto y diagrama de Gant	2 days	28/09/10 0:00	29/09/10 17:00	
11	<input checked="" type="checkbox"/> Análisis previo de los requerimientos	3 days	30/09/10 0:00	2/10/10 17:00	9
12	Dimensiones y atributos	2 days	30/09/10 0:00	1/10/10 17:00	
13	Fuentes de datos	1 day	2/10/10 0:00	2/10/10 17:00	
14	<input checked="" type="checkbox"/> Redacción documentos finales	1 day	5/10/10 0:00	5/10/10 17:00	
15	Redacción de la documento de planificación	1 day	5/10/10 0:00	5/10/10 17:00	10
16	Redacción del documento del análisis previo	1 day	5/10/10 0:00	5/10/10 17:00	11
17	Entrega Final Pac1	1 day	5/10/10 0:00	5/10/10 17:00	14
18	<input checked="" type="checkbox"/> Análisis de requisitos - Diseño conceptual y físico	31 days	9/10/10 0:00	8/11/10 17:00	8
19	<input type="checkbox"/> Análisis de requisitos	12 days	9/10/10 0:00	20/10/10 17:00	11
20	Introducción, objetivo, punto de partida y alcance del proyecto	7 days	9/10/10 0:00	15/10/10 17:00	
21	Análisis detallado de las fuentes de datos	3 days	16/10/10 0:00	18/10/10 17:00	
22	Requerimientos de almacén de datos	2 days	19/10/10 0:00	20/10/10 17:00	
23	Diseño conceptual y físico	11 days	21/10/10 8:00	31/10/10 17:00	19
24	Presentación propuesta Pac2	7 days	1/11/10 8:00	7/11/10 17:00	23
25	Entrega final Pac2	1 day	8/11/10 0:00	8/11/10 17:00	24
26	<input type="checkbox"/> Implementación	42 days	9/11/10 0:00	20/12/10 17:00	18
27	Construcción del almacén de datos	21 days	9/11/10 0:00	29/11/10 17:00	23
28	Construcción de informes	10 days	30/11/10 8:00	9/12/10 17:00	19,23
29	Análisis de la información obtenida	4 days	10/12/10 8:00	13/12/10 17:00	28
30	Presentación propuesta Pac3	6 days	14/12/10 8:00	19/12/10 17:00	27,28,29
31	Entrega final Pac3	1 day	20/12/10 8:00	20/12/10 17:00	30
32	<input type="checkbox"/> Entrega Final	21 days	21/12/10 8:00	10/01/11 17:00	26
33	Revisión de las indicaciones del consultor	3 days	21/12/10 8:00	23/12/10 17:00	
34	Elaboración de la memoria final	14 days	24/12/10 8:00	6/01/11 17:00	33
35	Preparación de la presentación virtual	3 days	7/01/11 8:00	9/01/11 17:00	34
36	Entrega Final	1 day	10/01/11 8:00	10/01/11 17:00	35



1.5. Productos obtenidos.

El proceso de desarrollo del proyecto ha dado como resultado los siguientes productos:

- Plan de trabajo: documento en el que se hace una breve introducción sobre la justificación, punto de partida y objetivos del proyecto, junto con un detallado plan de trabajo de todas las fases.
- Análisis de requisitos y diseño técnico: documento en el que se describe el análisis funcional del proyecto, el análisis de las fuentes de datos suministradas y de las decisiones tomadas a raíz de dicho estudio; junto con el diseño conceptual y técnico del almacén de datos propuesto.
- Implementación: en este producto se incluyen los ficheros correspondientes a las bases de datos construidas, el código necesario para crear dichas bases de datos y al código para tratar y cargar las fuentes de datos. También incluye los informes generados con Oracle Discoverer, a partir de elementos definidos en el almacén departamental y que cumplen con las necesidades del CHNE según sus especificaciones.
- Memoria del proyecto: este documento recopila los anteriores productos descritos, en uno único revisado y actualizado con las mejoras propuestas por el consultor.
- Presentación del proyecto: es una exposición de esta memoria en forma de audiovisual, en la que se recoge una visión general del proyecto y sus puntos más importante.

1.6. Breve descripción de los otros capítulos de la memoria.

Capítulo Análisis.

Este capítulo se divide en dos apartados:

- Casos de uso de la aplicación, en el que se muestran esquemáticamente los diferentes casos de uso del proyecto.
- Diagrama del modelo conceptual, del almacén de datos presentado como solución a las necesidades presentadas por el CHNE y que recoge todas sus especificaciones.

Capítulo Diseño.

Este capítulo recoge los siguientes apartados:

- Diagrama de la arquitectura software, donde se muestra la estructura a nivel de software de todos el proyecto.
- Diagrama de la arquitectura hardware, donde se detalla la estructura a nivel de hardware necesaria para dar soporte físico al proyecto.
- Diseño de la base de datos y diagrama del modelo físico, donde se muestran la estructura del almacén de datos de una forma gráfica utilizando herramientas case para su generación.
- Diseño y descripción de los informes creados, en el que se describe el diseño y los resultados obtenidos con los informes generados con la herramienta Oracle Discoverer.

Capítulo Captura de pantallas.

En este capítulo se muestran los resultados obtenidos con la ejecución de los informes diseñados con Oracle Discoverer Desktop.

Capítulo Conclusiones.

Conclusiones obtenidas con la realización del proyecto.

Capítulo Líneas de evolución futuras.

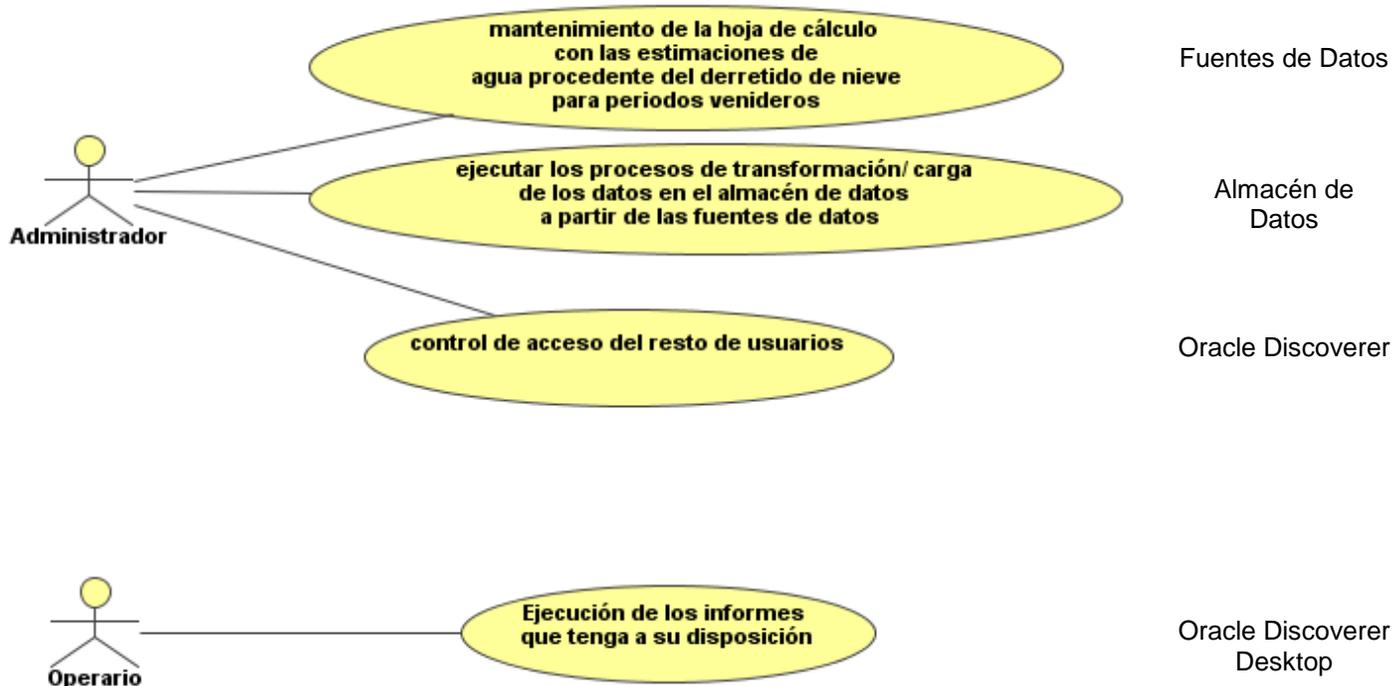
Sugerencias, posibles acciones a tomar para trabajos futuros después de la toma de contacto con el entorno actual ofrecido por el proyecto.

2. Análisis.

2.1. Diagramas de casos de usos con una explicación de cada uso.

Se identifican dos perfiles de usuario para este almacén de datos:

- **Administrador:**
 - Será el encargado del mantenimiento de la hoja de cálculo con las estimaciones de agua procedente del derretido de nieve para periodos venideros.
 - Será quien ejecute los procesos de transformación y carga de los datos en el almacén de datos a partir de las fuentes de datos suministras por la CHNE.
 - Se encargará del control de acceso del resto de usuarios, por mediación de un sistema de altas, bajas y modificaciones de usuarios en el sistema.
- **Operario:**
 - Ejecutará los informes que tenga a su disposición según su nivel de acceso.



2.2. Diagrama del modelo.

Como resultado del análisis de los requisitos establecidos por el CHNE para el almacén de datos y de las fuentes de datos suministradas, se han detectado las siguientes dimensiones:

- **Comunidad autónoma:** define las comunidades autónomas del CHNE.
- **Embalse:** conjunto de embalses.
- **Río:** Ríos con origen o que desembocan en los embalses.
- **Temporal:** permite establecer los periodos de año, mes y día para las consultas.
- **Medida:** permite establecer en Hm3 o en MI los volúmenes de agua.

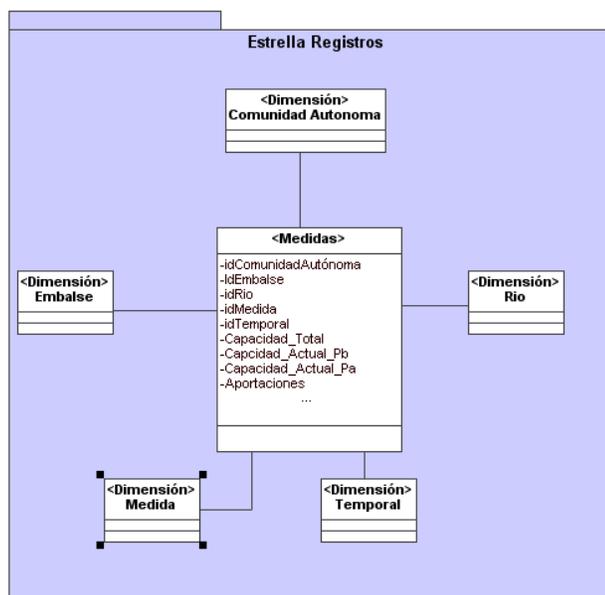
En cuanto los hechos definidos y sus medidas detectadas, estas son:

- **Hecho Registros:**
 - **Capacidad total:** capacidad total del embalse.
 - **Capacidad actual Pb:** capacidad utilizada periodo básico o actual.
 - **Capacidad actual Pa:** capacidad utilizada periodo anterior.
 - **Aportaciones:** aportación del río al embalse.
- **Hecho Nieve:**
 - **Estimación nieve:** estimación de agua por fundido de nieve.
 - **Capacidad total:** capacidad total del embalse.
 - **Capacidad actual:** capacidad utilizada periodo.
 - **Capacidad media:** capacidad media en el periodo.

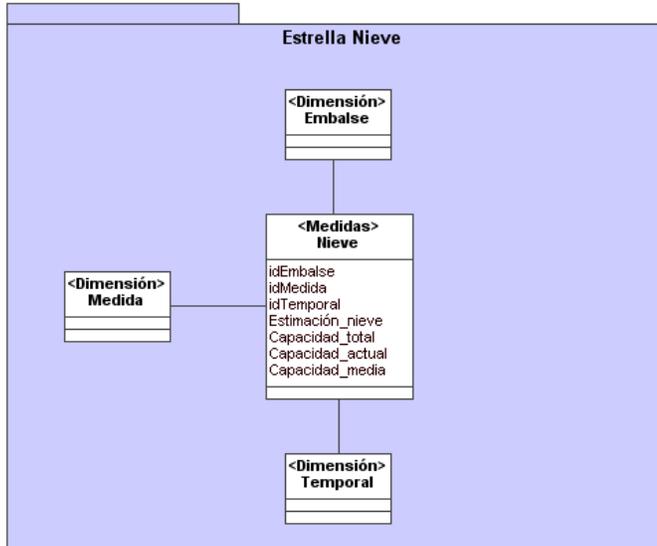
Con el hecho “Registros” podemos obtener los informes solicitados en el primer bloque de los requisitos establecidos para el proyecto, mientras que con el hecho “Nieve” obtendremos exclusivamente los informes que directamente se relacionan con la cantidad de agua procedente del fundido de nieve y de las estimaciones que se tienen sobre este aspecto y que forman parte del segundo bloque de requisitos. Se han establecido dos hechos diferentes por que las estimaciones de agua procedente del fundido de la nieve son a nivel de embalse, mientras que el resto de información se debe detallar a nivel comunidad autónoma, embalse y río. El hecho de no disponer de las previsiones a nivel de río, impide establecer el diseño del almacén de datos con un único hecho que pudiera satisfacer todos los requisitos.

Se establecen las siguientes relaciones entre las dimensiones y medidas, obteniendo los siguientes hechos para el almacén de datos:

Modelo para el hecho “Registros”.



Modelo para el hecho "Nieve".



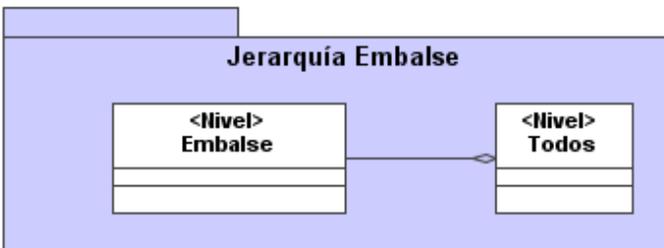
2.2.1. Estudio de las dimensiones del modelo.

Dimensión Comunidad Autónoma.



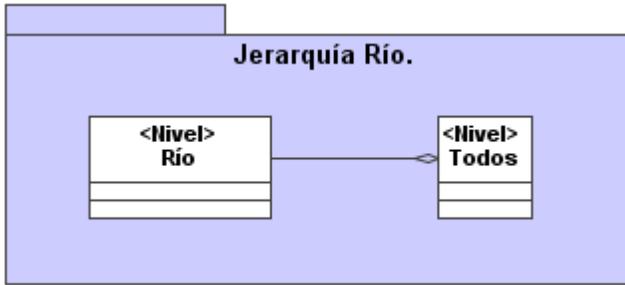
La dimensión Comunidad Autónoma contendrá la información concerniente a todas las comunidades autonómicas que configuran la CHNE. Esta dimensión contiene un único nivel, que contendrá el nombre de la comunidad autónoma perfectamente identificado.

Dimensión Embalse.



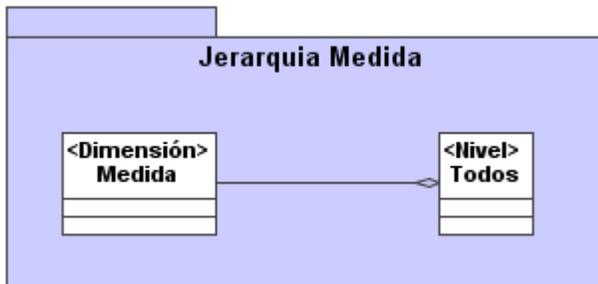
La dimensión embalse contiene la información concerniente a los diferentes embalses que forma parte de la red del CHNE. Esta dimensión al igual que la anterior tiene un único nivel con el nombre del embalse perfectamente identificado.

Dimensión Río.



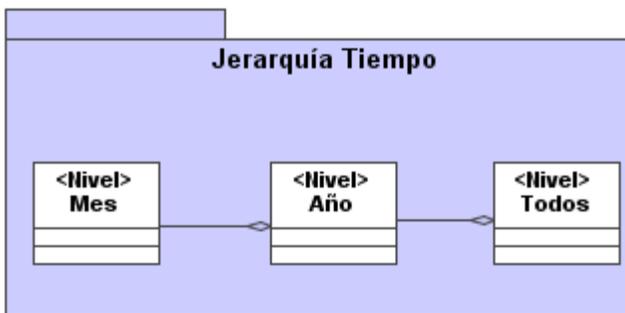
La dimensión Río almacena la información referente a los ríos del CHNE. Esta dimensión al igual que las anteriores, tiene un único nivel con el nombre del río perfectamente identificado.

Dimensión Medida



La dimensión Medida almacena las dos medidas en las que se pueden expresar las unidades y que son hectómetros cúbicos (Hm³) y millones de litros (MI). Es una medida con un único nivel.

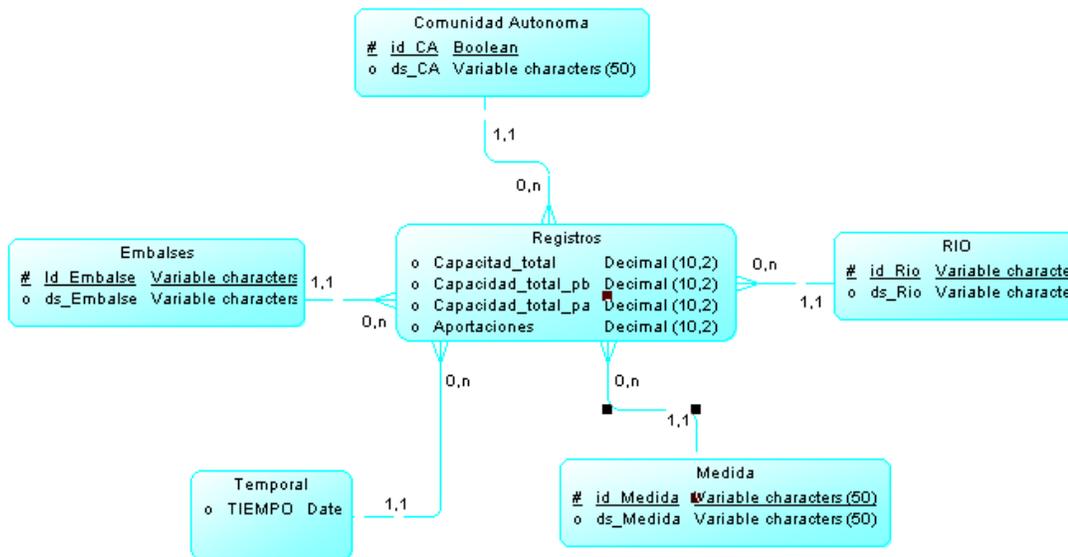
Dimensión Temporal.



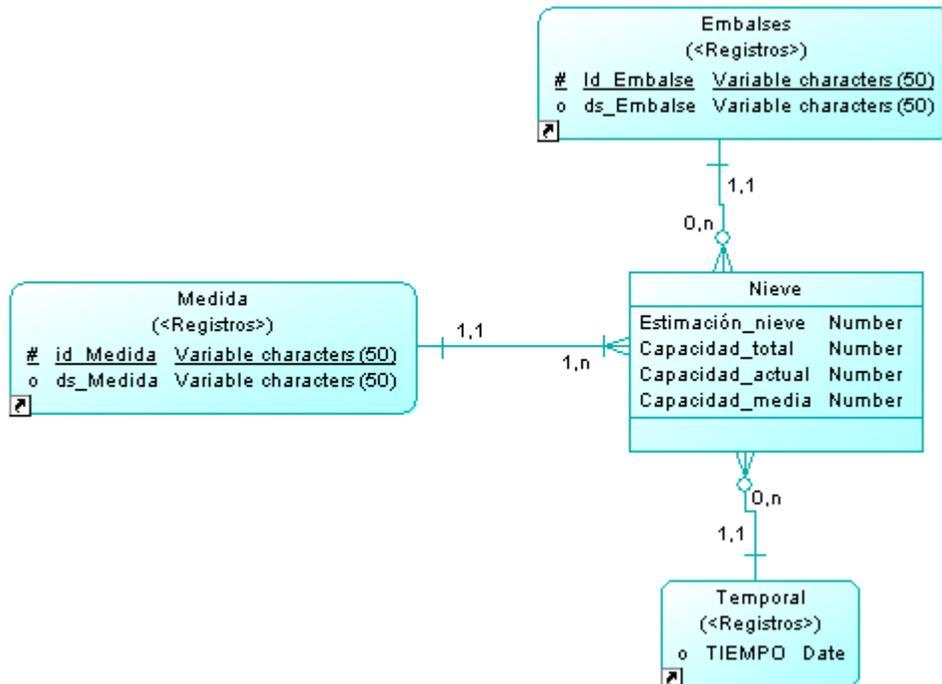
La dimensión Tiempo contiene la información necesaria para poder situar históricamente todos los datos del almacén de datos. Tal y como se ha recogido en el análisis de los requisitos, la dimensión tiempo incluirá dos categorías: mes y año.

2.2.2. Modelo conceptual.

Para el hecho "Registros" obtenemos el siguiente modelo conceptual.



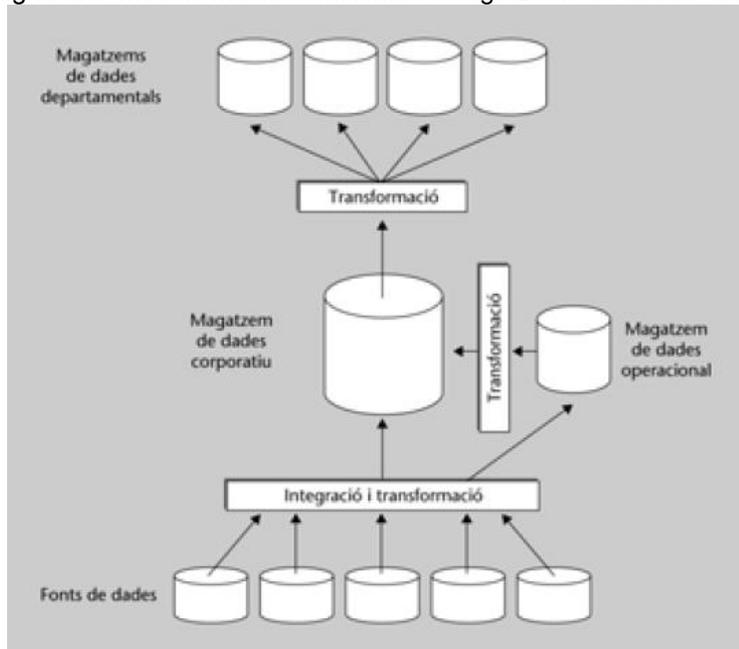
Para el hecho "Nieve" tenemos el siguiente modelo conceptual.



3. Diseño.

3.1. Diagrama de la arquitectura software.

Tras el análisis de los requisitos del proyecto, entendemos que no es necesario crear un almacén de datos o "Factoría de información corporativa", según el esquema siguiente extraído de los materiales oficiales de la asignatura "Models Multidimensionals i magatzems de dades" de la Universitat Oberta de Catalunya.



Sobre la estructura mencionada se todas las siguientes decisiones:

- No será necesario crear un almacén de datos corporativo, dado que no se necesita toda la historia de todos los datos del CHNE integrados en una única base de datos.
- Se creará un almacén de datos operacional en el cual se cargaran todos los datos de las fuentes de datos por completo, por lo que no será necesario un proceso de actualización ya que la carga de datos será siempre completa. Este almacén de datos operacional es el que tras un proceso de transformación dará lugar al almacén de datos departamental.
- Si se construirá un almacén de datos departamental en el cual se almacenarán los datos resultantes de la integración y transformación del almacén de datos operacional, poniéndose a disposición de los usuarios por mediación de los informes predefinidos.

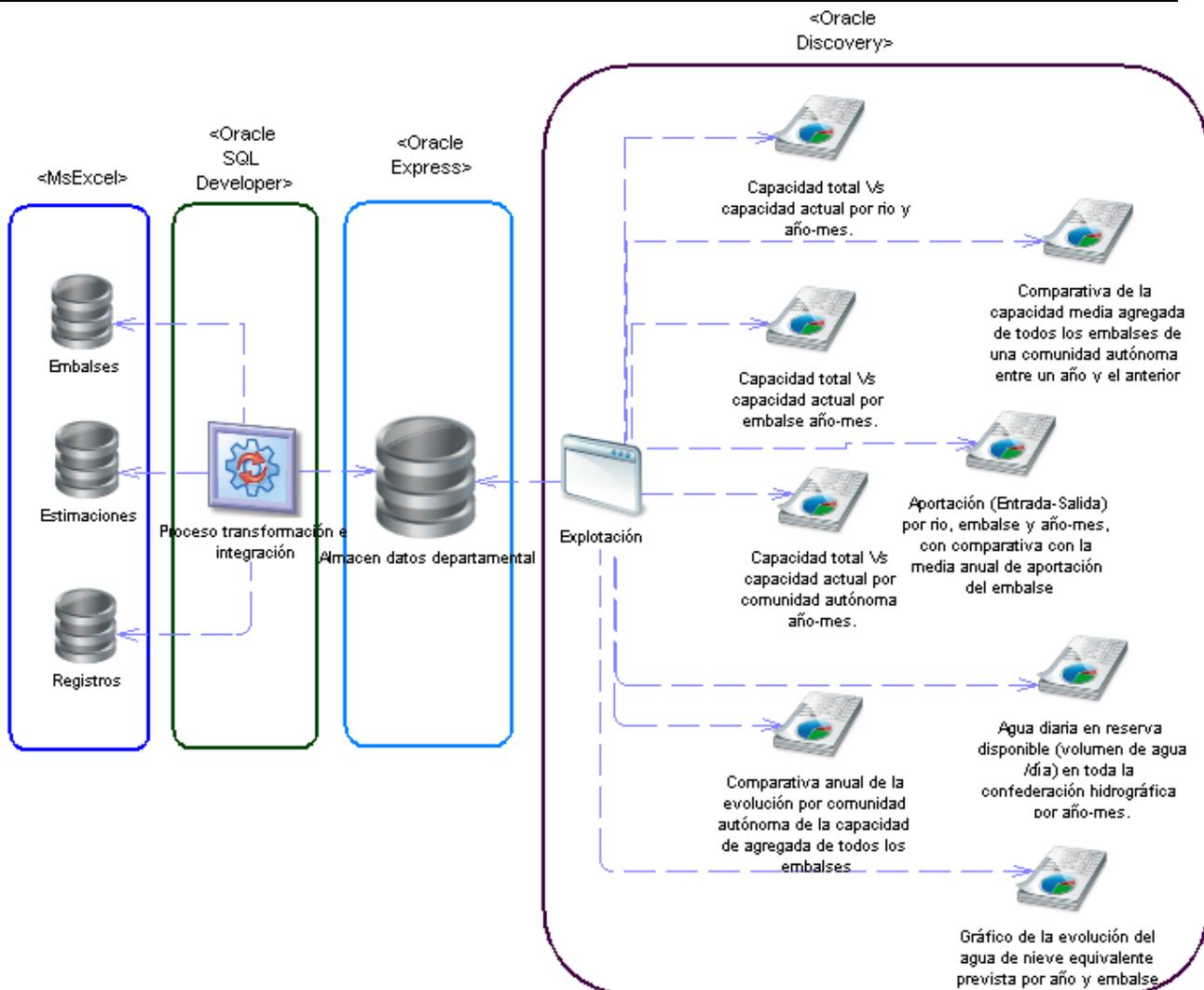
3.1.1. Requisitos de carga y actualización.

Tras el análisis de las fuentes de datos se aprecia la necesidad de establecer controles sobre el formato de los datos de las fuentes de datos suministradas. Hay deficiencias que se deberán subsanar antes de iniciar el proceso de transformación e integración en el almacén de datos, como:

- Normalización de la tabla embalses a un formato adecuado para el proceso de transformación e integración en el almacén de datos.
- Creación de la tabla de estimaciones de agua procedente del fundido de nieve y su posterior mantenimiento.

En cuanto al resto de datos suministrados, se establecerán controles que verifiquen la validez tanto del formato de cada dato como de sus valores.

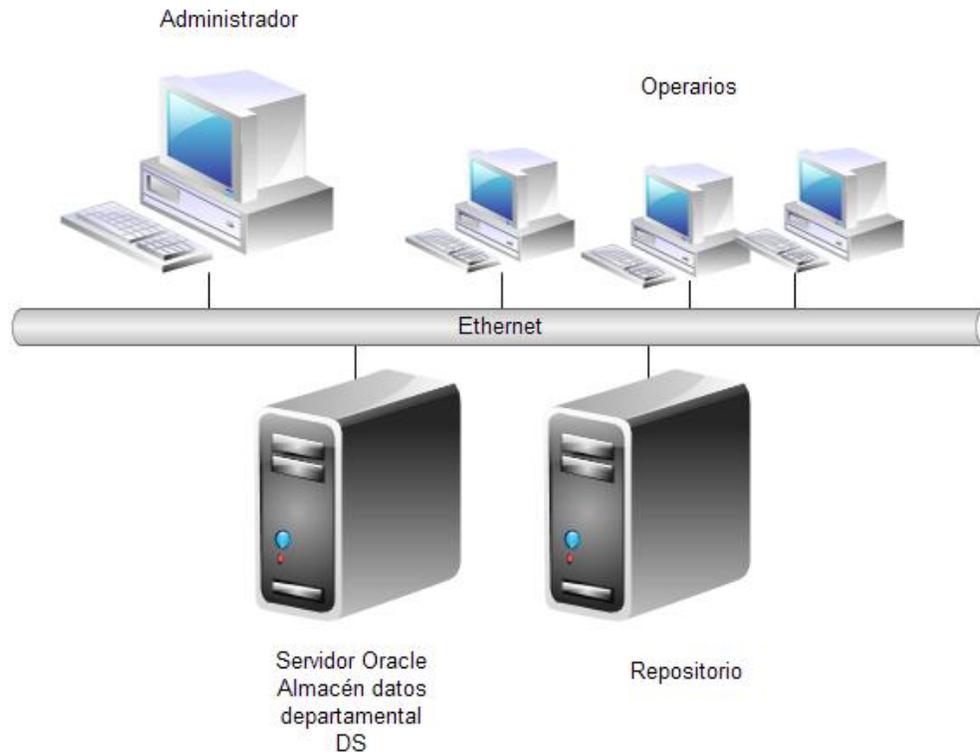
3.1.2. Diagrama de la arquitectura del software utilizado.



- **Microsoft Excel:** se utiliza en las fuentes de datos. El usuario Administrador que se ha definido en apartados anterior, deberá de disponer de licencia para usar esta aplicación, ya que deberá realizar las correcciones necesarias que aseguren la integridad de los datos de las fuentes suministras, a la vez que deberá normalizar y mantener la tabla embalses e introducir la tabla estimaciones.
- **Oracle Sql Developer:** esta herramienta permitirá la implementación del componte de transformación e integración de los datos en el almacén de datos. Puesto que esta parte es responsabilidad nuestra, la licencia de uso deberá ir a nuestro cargo.
- **Oracle Express:** esta versión de Oracle gratuita, a pesar de sus limitaciones se muestra como idónea para albergar nuestro almacén de datos.
- **Oracle Discoverer:** esta herramienta permite la administración de los usuarios y sus permisos y la generación de los informes solicitados. Será necesario la adquisición de tantas licencias como usuarios deban explotar el almacén de datos, más la licencia del usuario administrador por sus tareas de mantenimiento de usuarios.

3.2. Diagrama de la arquitectura hardware.

Aunque en los requisitos del proyecto no se hace una mención concreta sobre las necesidades de la arquitectura de hardware que se dispone o de la que se podría disponer, podemos suponer una estructura mínima que asegure el correcto funcionamiento del proyecto en su fase de explotación. Dicha estructura se muestra en el siguiente gráfico:

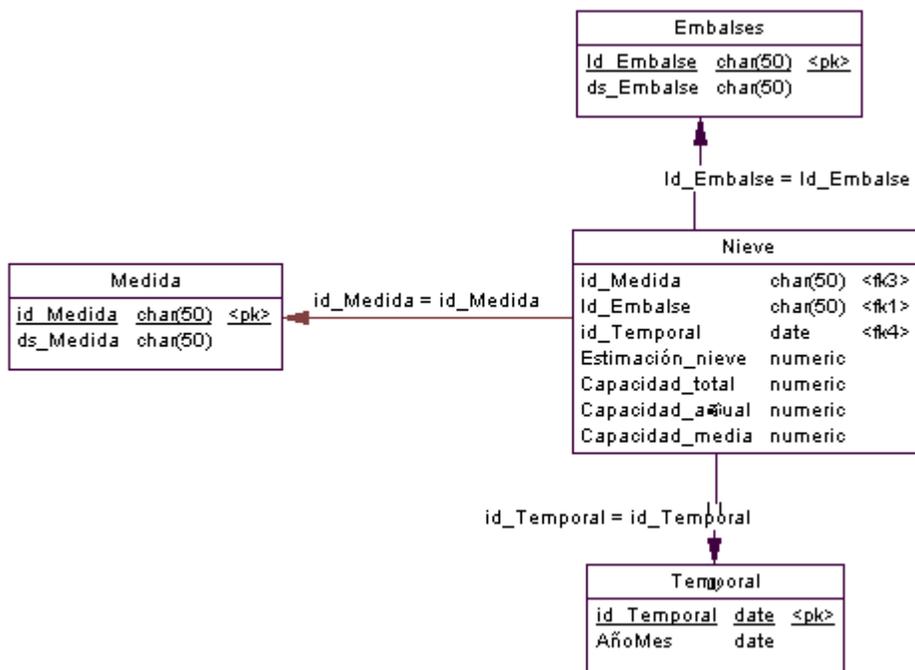
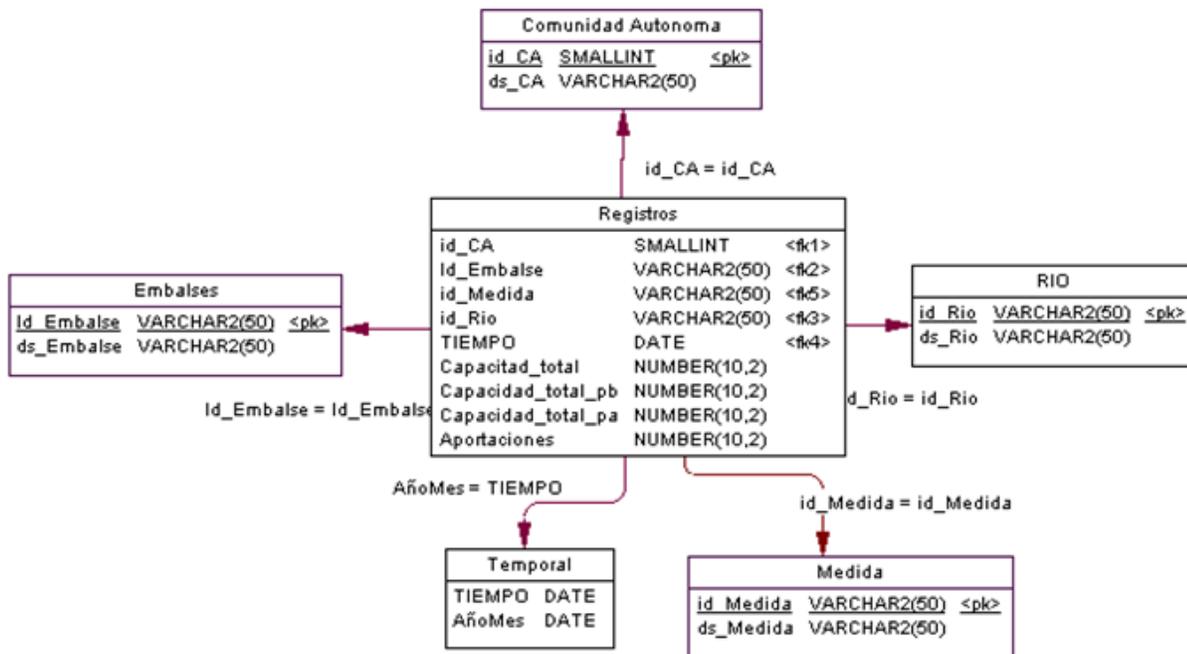


En la estructura se pueden observar los siguientes elementos:

- Servidor Oracle DS: se trata de un equipo informático que dará soporte tanto al almacén de datos departamental, como contendrá y ejecutará las herramientas de transformación e integración de los datos procedentes de las fuentes de datos en el almacén de datos departamental.
- Repositorio: puesto que las fuentes de datos se generan en diferentes equipos y después hay que reunirlos para procesarlos, ponemos a disposición del CHNE de un repositorio común donde se irán depositando los datos.
- Administrador: es el equipo del usuario administrador del sistema.
- Operarios: usuarios que explotarán el almacén de datos por mediación de los informes predefinidos.
- Ethernet: es el soporte físico que une los diferentes elementos y permite la comunicación entre ellos.

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

3.3.1. Almacén de datos departamental.



3.3.2. Componentes de extracción y carga.

La carga de datos no podrá ser un proceso totalmente automático, puesto que dos de los ficheros de las fuentes de datos requerirán de un mantenimiento manual por parte del usuario administrador.

Por este motivo definiremos un proceso previo de acondicionamiento de los datos que tendrá las siguientes tareas:

- Normalización de la tabla “embalses.xls”. Para esta tarea se creará un macro en hoja Excel auxiliar, que será la encargada de reajustar la hoja al siguiente formato:
 - Columna1: Nombre de la comunidad autónoma.
 - Columna2: Nombre del embalse.
 - Columna3: Nombre del rio.

Como resultado de la ejecución de la macro, obtendremos una nueva hoja Excel en la que cada fila contendrá todos los valores de las columnas anteriormente reseñadas, perdiéndose el dato de la provincia de la comunidad autónoma que no es necesario para nuestro almacén de datos. El fichero Excel original se renombrará con otro nombre a fin de poder comprobar el correcto funcionamiento del proceso de normalización.

- Creación y mantenimiento de las estimaciones de nieve. Esta tarea es totalmente manual y requiere que el usuario administrador cree inicialmente y después mantenga los datos de la tabla Excel “estimaciones.xls”.

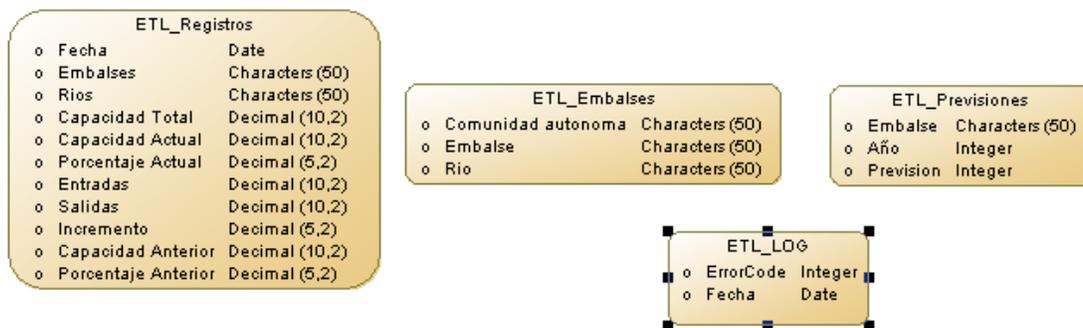
Una vez completada esta primera fase de acondicionamiento de los datos, se podrá iniciar el proceso de transformación e integración en el almacén de datos.

Para el proceso de extracción y transformación de los datos de las fuentes de datos será necesario disponer de un almacenamiento temporal. Dichas tablas no se consideran parte del almacén de datos departamental, y solo contendrá información durante el proceso de carga y transformación.

Puesto que los datos serán actualizados incrementalmente, y dado el formato propenso a errores de codificación, cada carga borrará los datos anteriores existentes en el almacén de datos, a fin de corregir posibles errores estadísticos.

3.3.3. Modelo conceptual de la capa de extracción y transformación.

A continuación se detalla las tablas que forman parte de la estructura temporal que se menciona:



- **ETL_Registros:** tabla donde se almacenará una copia de los datos contenidos en las tablas Excel que contienen las lecturas realizadas por el CHNE de los niveles de los diferentes embalses que configuran su red.
- **ETL_Embalses:** en esta tabla se almacenará una copia de la tabla “embalses.xls”, tabla resultante del proceso previo ya descrito, y que por mediación de una macro se normaliza la tabla original.
- **ETL_Estimaciones:** en esta tabla se copiarán los datos de la tabla Excel Estimaciones.
- **ETL_LOG:** tabla que permitirá registrar las diferentes incidencias que puedan producirse tanto en el proceso de carga como en el de transformación.

3.4. Diseño y descripción de los informes creados.

A continuación se describen los informes creados según las especificaciones del CHNE para el proyecto. De cada informe se indica el formato utilizado, si tiene elementos de página que se han utilizado, las columnas incluidas y una descripción de su funcionalidad.

- *Bloque primero:*
 - **Capacidad total Vs capacidad actual en Hm3 por río y año-mes.**
Formato: tabla.
Elemento de página: Medida (Hm3 / MI).
Columnas: río, capacidad total, año, mes, capacidad actual, porcentaje de uso.
Funcionalidad: aporta información en Hm3 o MI de la situación actual de cada río por año y mes respecto a sus capacidades totales.
 - **Capacidad total Vs capacidad actual en Hm3 por embalse año-mes.**
Formato: tabla.
Elemento de página: Medida (Hm3 / MI).
Columnas: embalse, capacidad total, año, mes, capacidad actual, porcentaje de uso.
Funcionalidad: este informe aporta la situación actual en Hm3 o MI de los embalses por año y mes respecto a sus capacidades totales.
 - **Capacidad total Vs capacidad actual en Hm3 por comunidad autónoma año-mes.**
Formato: tabla.
Elemento de página: Medida (Hm3 / MI).
Columnas: comunidad autónoma, capacidad total, año, mes, capacidad actual, porcentaje de uso.
Funcionalidad: muestra información en Hm3 o MI de la capacidad actual de cada comunidad autónoma por año y mes respecto a sus capacidades totales.
 - **Comparativa anual de la evolución por comunidad autónoma de la capacidad de agregada de todos los embalses.**
Formato: matriz.
Elemento de página: Medida (Hm3 / MI).
Columnas: capacidad agregada anual.
Filas: comunidades autónomas.
Funcionalidad: muestra un resumen anual de las capacidades agregadas de cada comunidad autónoma.
 - **Comparativa de la capacidad media agregada de todos los embalses de una comunidad autónoma entre un año y el anterior (valores i porcentajes de incremento año anterior-año actual).**
Formato: tabla.
Elemento de página: Medida (Hm3 / MI), comunidad autónoma y año.
Columnas: embalse, capacidad media, capacidad media año anterior, incremento y porcentaje de incremento.
Funcionalidad: muestra una comparativa de las capacidades medias agregadas de los embalses de una comunidad autónoma por año sobre al año anterior.

- **Aportación (Entrada-Salida) en Hm3 por río, embalse y año-mes, con comparativa con la media anual de aportación del embalse.**
 Formato: tabla.
 Elemento de página: Medida (Hm3 / MI).
 Columnas: río, embalse, media anual, año mes y aportaciones.
 Funcionalidad: muestra información sobre la media anual de aportaciones por embalse y las aportaciones por año y mes.

- **Agua diaria en reserva disponible (Hm3/día) en toda la confederación hidrográfica por año-mes.**
 Formato: tabla.
 Elemento de página: Medida (Hm3 / MI).
 Columnas: año, mes y agua disponible por día.
 Funcionalidad: muestra información a nivel de año y mes de la disponibilidad de agua diaria (disponibilidad mensual / 30 días).

- *Bloque segundo:*
 - **Gráfico de la evolución de los Hm3 d agua de nieve equivalente prevista por año y embalse.**
 Formato: tabla.
 Elemento de página: Medida (Hm3 / MI) y año.
 Columnas: embalse, capacidad uso, capacidad total y desbordamiento (Si / No).
 Funcionalidad: este informe muestra si existe posible desbordamiento del embalse según la previsión de agua de nieve estimada y la capacidad total de cada embalse por año.

 - **Mostrar cómo podemos predecir, a partir del almacén de datos, periodos de sequía.**
 Formato: tabla.
 Elemento de página: Medida (Hm3 / MI) y año.
 Columnas: embalse, estimación agua por nieve, capacidad actual, capacidad media, situación.
 Funcionalidad: muestra que embalses se encuentran con una capacidad baja o en riesgo de sequía. El campo situación tiene dos posibles valores:
 - Bueno: cuando la suma de la estimación de agua por nieve y la capacidad actual son mayores o iguales a un 75% de la capacidad media anual.
 - Sequia: cuando la suma de la estimación de agua por nieve y la capacidad actual son menores que el 75% de la capacidad media anual.

4. Capturas de pantalla.

- *Bloque primero:*
 - **Capacidad total Vs capacidad actual en Hm3 por rio y año-mes.**

Elementos de Página: Medida: HM3 ▼						
	Río	Capacidad Total	▶ Año	▶ Mes	Capacidad actual	% Uso
▶ 1	AGUAS VIVAS	8	2001	7	1	12,50
▶ 2				8	1	12,50
▶ 3				9	0	0,00
▶ 4				10	0	0,00
▶ 5				11	0	0,00
▶ 6				12	0	0,00
▶ 7		8	2002	1	0	0,00
▶ 8				2	0	0,00
▶ 9				3	0	0,00
▶ 10				4	0	0,00
▶ 11				5	1	12,50
▶ 12				6	1	12,50
▶ 13				7	1	12,50
▶ 14				8	2	25,00
▶ 15				9	1	12,50
▶ 16				10	1	12,50
▶ 17				11	1	12,50
▶ 18				12	1	12,50
▶ 19		8	2003	1	1	12,50
▶ 20				2	1	12,50
▶ 21				3	1	12,50
▶ 22				4	2	25,00

- **Capacidad total Vs capacidad actual en Hm3 por embalse año-mes.**

Elementos de Página: Medida: HM3 ▼						
	Embalse	Capacidad Total	▶ Año	▶ Mes	Capacidad actual	% Uso
▶ 1	ALLOZ	65,00	2001	7	37	56,92
▶ 2				8	27	41,54
▶ 3				9	16	24,62
▶ 4				10	4	6,15
▶ 5				11	7	10,77
▶ 6				12	7	10,77
▶ 7		65,00	2002	1	10	15,38
▶ 8				2	16	24,62
▶ 9				3	19	29,23
▶ 10				4	21	32,31
▶ 11				5	23	35,38
▶ 12				6	24	36,92
▶ 13				7	16	24,62
▶ 14				8	11	16,92
▶ 15				9	5	7,69
▶ 16				10	3	4,62
▶ 17				11	7	10,77
▶ 18				12	36	55,38
▶ 19		65,00	2003	1	44	67,69
▶ 20				2	51	78,46
▶ 21				3	56	86,15
▶ 22				4	56	86,15

- **Capacidad total Vs capacidad actual en Hm3 por comunidad autónoma año-mes.**

Elementos de Página: Medida: HM3 ▼

	Comunidad Autónoma	Capacidad total	Año	Mes	Capacidad actual	Porcentaje de Uso
▶ 1	ARAGÓN	794	2001	7	572	72,04
▶ 2				8	448	56,42
▶ 3				9	310	39,04
▶ 4				10	381	47,98
▶ 5				11	413	52,02
▶ 6				12	399	50,25
▶ 7		794	2002	1	386	48,61
▶ 8				2	393	49,50
▶ 9				3	439	55,29
▶ 10				4	487	61,34
▶ 11				5	559	70,40
▶ 12				6	549	69,14
▶ 13				7	366	46,10
▶ 14				8	254	31,99
▶ 15				9	289	36,40
▶ 16				10	366	46,10
▶ 17				11	405	51,01
▶ 18				12	549	69,14
▶ 19		794	2003	1	608	76,57
▶ 20				2	580	73,05
▶ 21				3	608	76,57
▶ 22				4	693	87,28
...				-

- **Comparativa anual de la evolución por comunidad autónoma de la capacidad de agregada de todos los embalses.**

Elementos de Página: Medida: HM3 ▼

	Capacidad agregada				
	▶ 2001	▶ 2002	▶ 2003	▶ 2004	▶ 2005
ARAGÓN	2523	5042	6814	6395	3272
CANTABRIA	1348	2381	3367	3502	4425
CATALUÑA	297	872	902	915	760
LA RIOJA	117	395	569	578	481
NAVARRA	811	2410	4442	4130	3003

- **Comparativa de la capacidad media agregada de todos los embalses de una comunidad autónoma entre un año y el anterior (valores i porcentajes de incremento año anterior-año actual).**

Elementos de Página: Medida: HM3 ▼ Comunidad Autónoma: ARAGÓN ▼ Año: 2001 ▼

	Embalse	Capacidad media	Capacidad media años anterior	Incremento	Incremento %
1	BÚBAL	21,67	27,83	-6,17	-28,46
2	CALANDA	27,17	17,83	9,33	34,36
3	CASPE	56,17	30,83	25,33	45,10
4	CUEVA FORADAC	2,17	3,17	-1,00	-46,15
5	LANUZA	10,83	12,67	-1,83	-16,92
6	MAIDEVERA	10,00	13,83	-3,83	-38,33
7	MEDIANO	234,33	187,67	46,67	19,91
8	MONEVA	0,33	1,00	-0,67	-200,00
9	PENA	14,00	6,50	7,50	53,57
10	SANTOLEA	35,17	32,33	2,83	8,06
11	VADIELLO	8,67	10,00	-1,33	-15,38

- **Aportación (Entrada-Salida) en Hm3 por río, embalse y año-mes, con comparativa con la media anual de aportación del embalse.**

Elementos de Página: Medida: HM3 ▼

	Río	Embalse	Media anual embalse	▶ Año	▶ Mes	Aportaciones
▶ 1	AGUAS VIVAS	MONEVA	0,00	2001	7	,00
▶ 2					8	,00
▶ 3					9	,00
▶ 4					10	,00
▶ 5					11	,00
▶ 6					12	,00
▶ 7			0,08	2002	1	,00
▶ 8					2	,00
▶ 9					3	,00
▶ 10					4	,00
▶ 11					5	,00
▶ 12					6	,00
▶ 13					7	,00
▶ 14					8	,90
▶ 15					9	,00
▶ 16					10	,00
▶ 17					11	,00
▶ 18					12	,00
▶ 19			-0,05	2003	1	,00
▶ 20					2	,00
▶ 21					3	,00
▶ 22					4	,10
▶ 23					5	,00

- **Agua diaria en reserva disponible (Hm3/día) en toda la confederación hidrográfica por año-mes.**

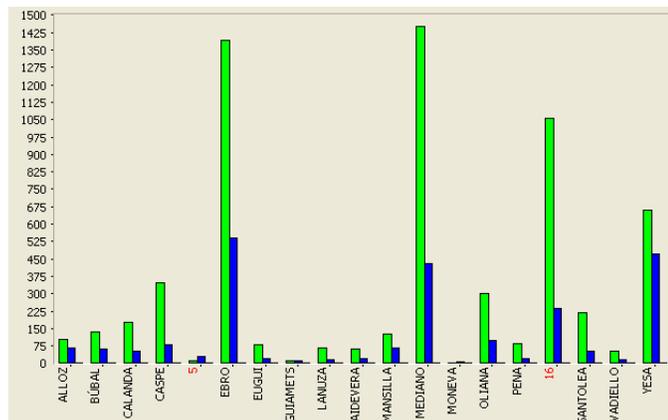
Elementos de Página: Medida: HM3 ▼

	▶ Año	▶ Mes	Agua diaria disponible
▶ 1	2001	7	43,83
▶ 2		8	31,53
▶ 3		9	20,93
▶ 4		10	23,13
▶ 5		11	25,70
▶ 6		12	24,73
▶ 7	2002	1	25,00
▶ 8		2	27,20
▶ 9		3	32,00
▶ 10		4	37,77
▶ 11		5	41,60
▶ 12		6	41,67
▶ 13		7	29,37
▶ 14		8	20,00
▶ 15		9	19,17
▶ 16		10	23,23
▶ 17		11	28,70
▶ 18		12	44,30
▶ 19	2003	1	48,20
▶ 20		2	49,77
▶ 21		3	53,07
▶ 22		4	58,13
▶ 23		5	50,00

- **Bloque segundo:**
 - **Gráfico de la evolución de los Hm3 d agua de nieve equivalente prevista por año y embalse.**

Elementos de Página: Medida: HM3 Año: 2001

	Embalse	Capacidad Uso	Capacidad Total	Desbordado
1	ALLOZ	101,27	65,00	SI
2	BÚBAL	134,33	64,00	SI
3	CALANDA	179,30	54,00	SI
4	CASPE	348,23	82,00	SI
5	CUEVA FORADADA	13,43	29,00	NO
6	EBRO	1392,93	540,00	SI
7	EUGUI	78,53	21,00	SI
8	GUIAMETS	13,43	10,00	SI
9	LANUZA	67,17	17,00	SI
10	MAIDEVERA	62,00	21,00	SI
11	MANSILLA	128,70	68,00	SI
12	MEDIANO	1452,87	430,00	SI
13	MONEVA	2,07	8,00	NO
14	OLIANA	302,93	101,00	SI
15	PENA	86,80	19,00	SI
16	SANTA ANA	1054,00	236,00	SI
17	SANTOLEA	218,03	54,00	SI
18	VADIELLO	53,73	16,00	SI
19	YESA	658,23	470,00	SI



- **Mostrar cómo podemos predecir, a partir del almacén de datos, periodos de sequía.**

Elementos de Página: Medida: HM3 Año: 2005

	Embalse	Estación agua nieve	Capacidad Actual	Capacidad Media	Situación
1	ALLOZ	7,834	470,00	39,17	Bueno
2	BÚBAL	4,85	291,00	24,25	Bueno
3	CALANDA	10,35	207,00	17,25	Bueno
4	CASPE	8,434	506,00	42,17	Bueno
5	CUEVA FORADADA	3,434	206,00	17,17	Bueno
6	EBRO	73,75	4425,00	368,75	Bueno
7	EUGUI	3,184	191,00	15,92	Bueno
8	GUIAMETS	1,066	64,00	5,33	Bueno
9	LANUZA	2,234	134,00	11,17	Bueno
10	MAIDEVERA	2,75	165,00	13,75	Bueno
11	MANSILLA	16,032	481,00	40,08	Bueno
12	MEDIANO	20,2	1212,00	101,00	Bueno
13	MONEVA	0,784	47,00	3,92	Bueno
14	OLIANA	11,6	696,00	58,00	Bueno
15	PENA	2,65	159,00	13,25	Bueno
16	SANTA ANA	75,852	1517,00	126,42	Bueno
17	SANTOLEA	5,284	317,00	26,42	Bueno
18	VADIELLO	0,466	28,00	2,33	Bueno
19	YESA	39,034	2342,00	195,17	Bueno

5. Conclusiones

El presente proyecto me ha permitido tomar contacto con una problemática, la de las confederaciones hidrográficas de la cual no tenía experiencia alguna.

También me ha permitido llevar a cabo un proyecto completo desde su planificación hasta su implementación, pasando por sus fases de análisis y diseño; en una materia (la de los almacenes de datos y su explotación para obtener información) que siempre he considerado muy importante sobre todo para el mundo empresarial.

Creo que en general es un proyecto que me ha permitido aproximarme de una forma muy real a lo que sería un proyecto real, en un entorno empresarial cualquiera.

Este proyecto también me ha permitido utilizar los conocimientos que he ido adquiriendo a lo largo de mis estudios en la EGTS, y dar solución a los diferentes retos que se han ido planteando en este trabajo.

Desde un punto más técnico, este proyecto también me ha permitido tener una toma de contacto con herramientas como el Oracle Discoverer, de la cual no tenía experiencia alguna.

6. Líneas de evolución futuras.

En cuanto a las líneas de evolución futuras, siempre con la intención de mejorar el proceso de obtención y procesamiento de los datos, podríamos plantearnos las siguientes:

- Sería recomendable que las lecturas que se introducen manualmente en hojas de cálculo se pudieran introducir en el sistema informático de una forma totalmente automática, por mediación de sistemas electrónicos que realicen las lecturas y las registren directamente en la base de datos.
- Desarrollar una aplicación que permitiera la recogida de las lecturas, sin tener que utilizar hojas de cálculo Excel. Dicha aplicación debería ser accesible desde Internet, para facilitar la interconexión de todos los sistemas de lectura con el sistema informático. Esta aplicación registraría la información directamente en la base de datos operacional, por lo que ya no sería necesario utilizar Excel para el procesamiento de los datos.
- Sustituir el Oracle Discoverer Desktop, por un entorno Web, el cual nos va a permitir mostrar los datos en un entorno más heterogéneo de sistemas.
- Crear un portal Web para publicitar la información y que la comunidad científica y toda persona interesada pueda tener acceso.

7. Glosario

Trabajo final de carrera: Se considerará Trabajo/Proyecto Fin de Carrera un documento destinado a presentar los resultados de una labor de experimentación o actividad de innovación y desarrollo tecnológico en el ámbito de la ingeniería informática. Su realización es un requisito imprescindible para la obtención del título de Ingeniero Técnico en Informática de Gestión.

Siglas: TFC

Confederación Hidrográfica del Noreste: El pasado 12 de marzo de 2008, fue publicado en el Boletín Oficial del Estado el Real Decreto 266/2008, de 22 de febrero, por el cual la **Confederación Hidrográfica del Norte** era segregada en **las Confederaciones Hidrográficas del Cantábrico** y del **Miño-Sil**. Ambas nuevas Confederaciones sucedían, a título universal, a la CH Norte en cuestión de bienes, materias y derechos de ésta, en relación a sus respectivos ámbitos territoriales.

Siglas: CHNE.

Almacén de datos: el almacén de datos es una colección de datos orientados al tema, integradas, no volátiles e historizadas, organizadas para dar soporte a procesos de ayuda a la decisión.

Almacén de datos corporativo: conjunto de datos que guarda integradas todos los datos históricos de la empresa.

Almacén de datos departamental: conjunto de datos que resuelven las necesidades de un cierto departamento o conjunto de usuarios.

Almacén de dato operacional: conjunto de datos integrados y orientados al tema, pero sin datos históricos. Se suele utilizar como paso intermedio en la construcción del almacén de datos corporativo.

8. Bibliografía.

Libros de texto.

- *The Data Warehouse Toolkit. (2nd Edition). Ralph Kimball and Margy Ross.*
- *Building the Data Warehouse (3rd Edition). W. H. Inmon.*
- *The Data Warehouse ETL Toolkit. Joe Caserta and Ralph Kimball.*

Otros materiales.

- Unidades didácticas “Models Multidimensionals i Magatzems de Dades” (UOC). Àngels Rius Gavídia i Montse Serra Vizern.
- PAC de la asignatura “Models Multidimensional i Magatzems de Dades” (UOC). José Manuel Ibáñez Lora.

9. Anexos

IniDS.bat

Descripción: script creación de la base de datos departamental.

Contenido.

```
SQLPLUS TFC/TFC @CREAR_DS
```

CrearDS.sql

Descripción: mandatos sql para la creación de la base de datos departamental.

Contenido.

```
/*=====*/
/* DBMS name:   ORACLE Version 10g           */
/* Created on:  07/12/2010 16:07:08         */
/*=====*/

alter table REGISTROS
  drop constraint FK_REGISTRO_REFERENCE_COMUNIDA;
alter table REGISTROS
  drop constraint FK_REGISTRO_REFERENCE_EMBALSES;
alter table REGISTROS
  drop constraint FK_REGISTRO_REFERENCE_RIO;
alter table REGISTROS
  drop constraint FK_REGISTRO_REFERENCE_TEMPORAL;
alter table REGISTROS
  drop constraint FK_REGISTRO_REFERENCE_MEDIDA;
alter table NIEVE
  drop constraint FK_NIEVE_REFERENCE_MEDIDA;
alter table NIEVE
  drop constraint FK_NIEVE_REFERENCE_EMBALSE;

drop table COMUNIDAD_AUTONOMA cascade constraints;
drop table EMBALSES cascade constraints;
drop table REGISTROS cascade constraints;
drop table NIEVE cascade constraints;
drop table RIO cascade constraints;
drop table TEMPORAL cascade constraints;
drop table MEDIDA cascade constraints;

/*=====*/
/* Tabla: COMUNIDAD_AUTONOMA                 */
/*=====*/
create table COMUNIDAD_AUTONOMA
(
  ID_CA VARCHAR(50) not null,
  constraint PK_COMUNIDAD_AUTONOMA primary key (ID_CA)
);

/*=====*/
/* Tabla: EMBALSES                           */
/*=====*/
create table EMBALSES
(
  ID_EMBALSE VARCHAR2(50) not null,
  constraint PK_EMBALSES primary key (ID_EMBALSE)
);
```

```

/*=====*/
/* Tabla: RIO                                     */
/*=====*/
create table RIO
(
  ID_RIO      VARCHAR2(50)          not null,
  constraint PK_RIO primary key (ID_RIO)
);

```

```

/*=====*/
/* Table: TEMPORAL                               */
/*=====*/
create table TEMPORAL(
  ID_TEMPORAL NUMBER(6,0) NOT NULL,
  ID_ANO      NUMBER(4,0),
  ID_MES      NUMBER(4,0),
  constraint PK_TEMPORAL primary key (ID_TEMPORAL)
);

```

```

/*=====*/
/* Table: MEDIDA                                 */
/*=====*/
create table MEDIDA(
  ID_MEDIDA  VARCHAR2(50) NOT NULL,
  constraint PK_MEDIDA primary key (ID_MEDIDA)
);

```

```

/*=====*/
/* Tabla: REGISTROS                             */
/*=====*/
create table REGISTROS
(
  ID_CA      VARCHAR2(50),
  ID_EMBALSE VARCHAR2(50),
  ID_RIO     VARCHAR2(50),
  ID_TEMPORAL NUMBER(6,0),
  ID_MEDIDA  VARCHAR2(50),
  ANO        NUMBER(4,0),
  CAPACIDAD_TOTAL  NUMBER(10,2),
  CAPACIDAD_TOTAL_PB  NUMBER(10,2),
  CAPACIDAD_TOTAL_PA  NUMBER(10,2),
  APORTACIONES     NUMBER(10,2)
);

```

```

/*=====*/
/* Tabla: NIEVE                                 */
/*=====*/
create table NIEVE
(
  ID_EMBALSE  VARCHAR2(50),
  ANO         NUMBER(4,0),
  ID_MEDIDA   VARCHAR2(50),
  ESTIMACION  NUMBER(10,2),
  CAPACIDAD_TOTAL  NUMBER(10,2),
  CAPACIDAD_ACTUAL  NUMBER(10,2),
  CAPACIDAD_MEDIA  NUMBER(10,2)
);

```

```
alter table REGISTROS
add constraint FK_REGISTRO_REFERENCE_COMUNIDA foreign key (ID_CA)
references COMUNIDAD_AUTONOMA (ID_CA);
```

```
alter table REGISTROS
add constraint FK_REGISTRO_REFERENCE_EMBALSES foreign key (ID_EMBALSE)
references EMBALSES (ID_EMBALSE);
```

```
alter table REGISTROS
add constraint FK_REGISTRO_REFERENCE_RIO foreign key (ID_RIO)
references RIO (ID_RIO);
```

```
alter table REGISTROS
add constraint FK_REGISTRO_REFERENCE_TEMPORAL foreign key (ID_TEMPORAL)
references TEMPORAL (ID_TEMPORAL);
```

```
alter table REGISTROS
add constraint FK_REGISTRO_REFERENCE_MEDIDA foreign key (ID_MEDIDA)
references MEDIDA (ID_MEDIDA);
```

```
alter table NIEVE
add constraint FK_NIEVE_REFERENCE_EMBALSE foreign key (ID_EMBALSE)
references EMBALSES (ID_EMBALSE);
```

```
alter table NIEVE
add constraint FK_NIEVE_REFERENCE_MEDIDA foreign key (ID_MEDIDA)
references MEDIDA (ID_MEDIDA);
```

```
INSERT INTO EMBALSES SELECT UNIQUE EMBALSE FROM ETL_EMBALSES;
INSERT INTO RIO SELECT UNIQUE RIO FROM ETL_EMBALSES;
INSERT INTO COMUNIDAD_AUTONOMA SELECT UNIQUE
ETL_EMBALSES.COMUNIDAD_AUTONOMA FROM ETL_EMBALSES;
INSERT INTO TEMPORAL SELECT UNIQUE (ANO*100)+MES, ANO,MES FROM ETL_REGISTROS;
INSERT INTO MEDIDA VALUES ('HM3');
INSERT INTO MEDIDA VALUES ('ML');
```

```
INSERT INTO REGISTROS
SELECT etl_embalses.comunidad_autonoma,etl_registros.embalse,
etl_registros.rio,(etl_registros.ano*100)+etl_registros.mes,
'HM3',etl_registros.ano,
etl_registros.capacidad_total,etl_registros.capacidad_actual,
etl_registros.capacidad_anterior,etl_registros.entradas - etl_registros.salidas
FROM etl_embalses,etl_registros
WHERE etl_embalses.embalse=etl_registros.embalse AND
etl_embalses.rio=etl_registros.rio;
```

```
INSERT INTO REGISTROS
SELECT etl_embalses.comunidad_autonoma,etl_registros.embalse,
etl_registros.rio,(etl_registros.ano*100)+etl_registros.mes,
'ML',etl_registros.ano,
etl_registros.capacidad_total*1000,etl_registros.capacidad_actual * 1000,
etl_registros.capacidad_anterior*1000,(etl_registros.entradas - etl_registros.salidas) * 1000
FROM etl_embalses,etl_registros
WHERE etl_embalses.embalse=etl_registros.embalse AND
etl_embalses.rio=etl_registros.rio;
```

INSERT INTO NIEVE

```
SELECT unique etl_estimacion.embalse,etl_estimacion.ano,'HM3',etl_estimacion.prevision,
(select sum (distinct(etl_registros.capacidad_total)) from etl_registros
where etl_registros.embalse=etl_estimacion.embalse and
etl_registros.ano=etl_estimacion.ano
group by etl_registros.embalse),
(select SUM(etl_registros.capacidad_actual) from etl_registros
where etl_registros.embalse=etl_estimacion.embalse and
etl_registros.ano=etl_estimacion.ano
group by etl_registros.embalse,etl_registros.ano ),
(select avg(etl_registros.capacidad_actual) from etl_registros
where etl_registros.embalse=etl_estimacion.embalse and
etl_registros.ano=etl_estimacion.ano
group by etl_registros.embalse,etl_registros.ano )
FROM etl_estimacion,etl_registros where etl_estimacion.embalse=etl_registros.embalse;
```

INSERT INTO NIEVE

```
SELECT unique etl_embalses.embalse,etl_registros.ano,'HM3',
(SELECT etl_estimacion.prevision from etl_estimacion where
etl_estimacion.embalse='RESTO' and
etl_estimacion.ano=etl_registros.ano),
(select sum (distinct(etl_registros.capacidad_total)) from etl_registros
where etl_registros.embalse=etl_embalses.embalse and
etl_registros.ano=etl_estimacion.ano
group by etl_registros.embalse),
(select SUM(etl_registros.capacidad_actual) from etl_registros
where etl_registros.embalse=etl_embalses.embalse and
etl_registros.ano=etl_estimacion.ano
group by etl_registros.embalse,etl_registros.ano ),
(select avg(etl_registros.capacidad_actual) from etl_registros
where etl_registros.embalse=etl_embalses.embalse and
etl_registros.ano=etl_estimacion.ano
group by etl_registros.embalse,etl_registros.ano )
FROM
etl_registros,etl_embalses,etl_estimacion
WHERE
etl_embalses.embalse=etl_registros.embalse and
etl_estimacion.embalse='RESTO' and etl_estimacion.ano=etl_registros.ano and
NOT exists (SELECT * from etl_estimacion
where etl_estimacion.embalse=etl_registros.embalse);
```

INSERT INTO NIEVE

```
SELECT unique etl_estimacion.embalse,etl_estimacion.ano,'ML',etl_estimacion.prevision*1000,
(select sum (distinct(etl_registros.capacidad_total*1000)) from etl_registros
where etl_registros.embalse=etl_estimacion.embalse and
etl_registros.ano=etl_estimacion.ano
group by etl_registros.embalse),
(select SUM(etl_registros.capacidad_actual*1000) from etl_registros
where etl_registros.embalse=etl_estimacion.embalse and
etl_registros.ano=etl_estimacion.ano
group by etl_registros.embalse,etl_registros.ano ),
(select avg(etl_registros.capacidad_actual*1000) from etl_registros
where etl_registros.embalse=etl_estimacion.embalse and
etl_registros.ano=etl_estimacion.ano
group by etl_registros.embalse,etl_registros.ano )
FROM etl_estimacion,etl_registros where etl_estimacion.embalse=etl_registros.embalse;
```

INSERT INTO NIEVE

```
SELECT unique etl_embalses.embalse,etl_registros.ano,'ML',
  (SELECT etl_estimacion.prevision * 1000 from etl_estimacion where
    etl_estimacion.embalse='RESTO' and
    etl_estimacion.ano=etl_registros.ano),
  (select sum (distinct(etl_registros.capacidad_total * 1000)) from etl_registros
    where etl_registros.embalse=etl_embalses.embalse and
    etl_registros.ano=etl_estimacion.ano
    group by etl_registros.embalse),
  (select SUM(etl_registros.capacidad_actual * 1000) from etl_registros
    where etl_registros.embalse=etl_embalses.embalse and
    etl_registros.ano=etl_estimacion.ano
    group by etl_registros.embalse,etl_registros.ano ),
  (select avg(etl_registros.capacidad_actual * 1000) from etl_registros
    where etl_registros.embalse=etl_embalses.embalse and
    etl_registros.ano=etl_estimacion.ano
    group by etl_registros.embalse,etl_registros.ano )
FROM
  etl_registros,etl_embalses,etl_estimacion
WhERE
  etl_embalses.embalse=etl_registros.embalse and
  etl_estimacion.embalse='RESTO' and etl_estimacion.ano=etl_registros.ano and
  NOT exists (SELECT * from etl_estimacion
    where etl_estimacion.embalse=etl_registros.embalse);
```

create or replace function "MEDIANUALAPORTACIONES"

```
(p_medida in VARCHAR2,
 p_embalse in VARCHAR2,
 p_anomes in NUMBER)
return number
is
  ano number;
  media number;
Begin
  ano := p_anomes / 100;
  select avg(aportaciones) into media from registros
    where id_medida = p_medida and
    id_embalse=p_embalse and
    id_temporal/100=ano
    group by id_embalse,id_temporal/100;
  return media;
end;
```

IniTemp.bat

Descripción: script para la creación y carga de almacén de datos operacional.

Contenido:

```
DEL *.BAD
```

```
DEL *.LOG
```

```
SQLPLUS TFC/TFC @CREAR_ETL
```

```
SQLldr USERID=TFC/TFC CONTROL=ETL_EMBALSES.CTL SKIP=1
```

```
SQLldr USERID=TFC/TFC CONTROL=ETL_ESTIMACION.CTL SKIP=1
```

```
SQLldr USERID=TFC/TFC CONTROL=ETL_REGISTROS.CTL
```

CrearETL.sql Mandatos en sql para la creación de la base de datos operacional.

Contenido:

```
DROP TABLE "ETL_EMBALSES";
CREATE TABLE "ETL_EMBALSES"
(
  "COMUNIDAD_AUTONOMA" VARCHAR2(50),
  "EMBALSE" VARCHAR2(50),
  "RIO" VARCHAR2(50)
);
```

```
DROP TABLE "ETL_ESTIMACION";
CREATE TABLE "ETL_ESTIMACION"
(
  "EMBALSE" VARCHAR2(50),
  "ANO" NUMBER(4,0),
  "PREVISION" NUMBER(4,0)
);
```

```
DROP TABLE "ETL_REGISTROS";
CREATE TABLE "ETL_REGISTROS"
(
  "DIA" NUMBER(2,0),
  "MES" NUMBER(2,0),
  "ANO" NUMBER(4,0),
  "EMBALSE" VARCHAR2(50),
  "RIO" VARCHAR2(50),
  "CAPACIDAD_TOTAL" NUMBER(10,2),
  "CAPACIDAD_ACTUAL" NUMBER(10,2),
  "PORCENTAJE_ACTUAL" NUMBER(5,2),
  "ENTRADAS" NUMBER(10,2),
  "SALIDAS" NUMBER(10,2),
  "INCREMENTO" NUMBER(5,2),
  "CAPACIDAD_ANTERIOR" NUMBER(10,2),
  "PORCENTAJE_ANTERIOR" NUMBER(5,2)
);
```

QUIT;

ETL_EMBALSES.ctf

Descripción: fichero de control para la carga de la tabla ETL_EMBALSES por mediación sqldr

Contenido:

```
LOAD DATA
INFILE '..\CHND\EMBALSES.CSV'
INTO TABLE ETL_EMBALSES
APPEND
FIELDS TERMINATED BY ';' OPTIONALLY ENCLOSED ""
TRAILING NULLCOLS
(COMUNIDAD_AUTONOMA,EMBALSE,RIO)
BEGINDATA
```

ETL_ESTIMACION.ctf

Descripción: fichero de control para la carga de la tabla ETL_ESTIMACION por mediación sqldr.

Contenido:

```
LOAD DATA
INFILE '..\CHND\ESTIMACION.CSV'
INTO TABLE ETL_ESTIMACION
APPEND
FIELDS TERMINATED BY ';' OPTIONALLY ENCLOSED ""
TRAILING NULLCOLS
(EMBALSE,ANO,PREVISION)
BEGINDATA,
```

ETL_REGISTROS.ctf

Descripción: fichero de control para la carga de la tabla ETL_REGISTROS por mediación sqllldr.

Contenido:

```
LOAD DATA
INFILE '..\CHND\REGISTROS.CSV'
INTO TABLE ETL_REGISTROS
APPEND
FIELDS TERMINATED BY ';' OPTIONALLY ENCLOSED ""
TRAILING NULLCOLS
(DIA,MES,ANO,EMBALSE,RIO,CAPACIDAD_TOTAL,CAPACIDAD_ACTUAL,PORCENTAJE_ACTUAL,EN
TRADAS,
SALIDAS,INCREMENTO,CAPACIDAD_ANTERIOR,PORCENTAJE_ANTERIOR)
BEGINDATA,
```

ETL_PRE.xls

Descripción: macro Excel que integra todas las hojas Excel con las lecturas en una única hoja para poder integrada directamente en la base de datos operacional.

Contenido:

```
Private Sub botonBuscar_Click()
Agin1 = Application.GetOpenFilename("Fuentes de datos, *.xls")
```

```
If Agin1 <> False Then
  For i = Len(Agin1) To 0 Step -1
    If Mid(Agin1, i, 1) = "\" Then
      Exit For
    End If
  Next
  If i = 0 Then
    TextBox1.Value = ThisWorkbook.Path + "\"
  Else
    TextBox1.Value = Left(Agin1, i)
  End If
Else
  TextBox1.Value = ThisWorkbook.Path + "\"
End If
```

```
End Sub
```

```
Private Sub CommandButton1_Click()
Dim wPrimera, MyFile As String
Dim FilalInicial, FilaFinal, Rango, Fila As Range
Application.DisplayAlerts = False
```

'Procesamos los registros para generar un fichero csv con todos los datos

```
Workbooks.Open TextBox1.Value + "registros.xls"
```

```
MyFile = Dir(TextBox1.Value & "e*.xls")
```

```
wPrimera = "SI"
```

```
Do While MyFile <> ""
```

```
  If MyFile <> "ETL_PRE.xls" And _
    MyFile <> "embalses.xls" And _
    MyFile <> "embalses_old.xls" And _
    MyFile <> "estimacion.xls" Then
    Workbooks.Open TextBox1.Value + MyFile
    Windows(MyFile).Activate
    FilaFinal = Range("A65536").End(xlUp).Address
    FilalInicial = Range(FilaFinal).End(xlUp).Address
    FilaFinal = Replace(FilaFinal, "A", "L")
```

```

Set Rango = Range(FilaInicial, FilaFinal)
For Each Fila In Rango.Rows
    Fila.Cells(1, 1).Value = UCase(Fila.Cells(1, 1).Value)
    Fila.Cells(1, 2).Value = UCase(Fila.Cells(1, 2).Value)
    For i = 3 To 13 Step 1
        If Not IsNumeric(Fila.Cells(1, i).Value) Then
            Fila.Cells(1, i).Value = 0
        End If
    Next
Next
Range("J:J").Delete
Range("H:H").Delete
Range("A:A").Select
Selection.Insert Shift:=xlToRight
Selection.Insert Shift:=xlToRight
Selection.Insert Shift:=xlToRight

FilaFinal = Replace(FilaFinal, "L", "M")
Set Rango = Range(FilaInicial, FilaFinal)
For Each Fila In Rango.Rows
    Fila.Cells(1, 1).Value = Mid(MyFile, 8, 2)
    Fila.Cells(1, 2).Value = Mid(MyFile, 6, 2)
    Fila.Cells(1, 3).Value = Mid(MyFile, 2, 4)
Next
Rango.Select
Selection.Copy
Windows("registros.xls").Activate
Sheets("Hoja1").Activate
If wPrimera = "SI" Then
    Range("A1").Select
    wPrimera = "NO"
Else
    Range("A1").End(xlDown).Offset(1, 0).Select
End If
ActiveSheet.Paste
Workbooks(MyFile).Close SaveChanges:=False
End If
MyFile = Dir()
Loop

```

```

Windows("registros.xls").Activate
ActiveWorkbook.SaveAs Filename:=TextBox1.Value + "registros.csv", _
    FileFormat:=xlCSV, _
    CreateBackup:=False

```

```

'Procesamos el fichero embalses.xls para genera su csv
Workbooks.Open TextBox1.Value + "embalses.xls"
FilaFinal = Range("A65536").End(xlUp).Address
FilaInicial = Range(FilaFinal).End(xlUp).Address
FilaFinal = Replace(FilaFinal, "A", "C")
Set Rango = Range(FilaInicial, FilaFinal)
For Each Fila In Rango.Rows
    Fila.Cells(1, 1).Value = UCase(Fila.Cells(1, 1).Value)
    Fila.Cells(1, 2).Value = UCase(Fila.Cells(1, 2).Value)
    Fila.Cells(1, 3).Value = UCase(Fila.Cells(1, 3).Value)
Next

```

```
ActiveWorkbook.SaveAs Filename:=TextBox1.Value + "embalses.csv", _  
    FileFormat:=xlCSV, _  
    CreateBackup:=False
```

```
'Procesamos el fichero estimacion.xls para genera su csv  
Workbooks.Open TextBox1.Value + "estimacion.xls"  
FilaFinal = Range("A65536").End(xlUp).Address  
FilaInicial = Range(FilaFinal).End(xlUp).Address  
FilaFinal = Replace(FilaFinal, "A", "C")  
Set Rango = Range(FilaInicial, FilaFinal)  
For Each Fila In Rango.Rows  
    Fila.Cells(1, 1).Value = UCase(Fila.Cells(1, 1).Value)  
    If Not IsNumeric(Fila.Cells(1, 2).Value) Then  
        Fila.Cells(1, 2).Value = 0  
    End If  
    If Not IsNumeric(Fila.Cells(1, 3).Value) Then  
        Fila.Cells(1, 3).Value = 0  
    End If  
Next
```

```
ActiveWorkbook.SaveAs Filename:=TextBox1.Value + "estimacion.csv", _  
    FileFormat:=xlCSV, _  
    CreateBackup:=False
```

```
Application.DisplayAlerts = True
```

```
End Sub
```

```
Private Sub UserForm_Initialize()  
    ETL_Pre.TextBox1.Value = ThisWorkbook.Path + "\"  
End Sub
```