

# TFC – Almacenes de datos

## Memoria

12/12/2010

**CHNE – CONFEDERACIÓN HIDROGRÁFICA NORTE-ESTE**

Diego Arándiga Cánovas – Ingeniería Técnica de Informática de Gestión

Consultor

Pere Juanola Juanola

## Índice

1.	Introducción .....	4
2.	Objetivos.....	4
3.	Requerimientos del cliente (CHNE) .....	4
3.1	Datos de origen .....	5
3.2	Análisis de los datos de origen .....	6
3.3	Valoración de los datos de origen .....	8
3.4	Informes a realizar .....	11
3.5	Elementos para el análisis multidimensional .....	12
4.	Planificación.....	12
4.1	Fases del proyecto .....	13
4.2	Actividades .....	13
4.3	Tareas .....	14
4.4	Programa de trabajo.....	15
4.4.1	Calendario.....	16
4.4.2	Diagrama de Gannt.....	18
5.	Incidencias y riesgos .....	21
6.	Diseño técnico .....	22
6.1	Diseño del modelo dimensional conceptual .....	22
6.2	Diseño del modelo lógico .....	22
6.3	Diseño del modelo físico .....	23
6.4	Diseño de los procedimientos de extracción de datos.....	25
7.	Proceso de carga.....	38
7.1	Carga de los embalses .....	39
7.2	Carga de los datos de situaciones de los embalses y ríos .....	39
8.	Acceso a la base de datos .....	40

Memoria	TFC – Almacenes de datos	CHNE
9.	Acceso a la herramienta de Bussines Intelligence.....	41
9.1	Acceso a Oracle Discoverer Administrator.....	41
9.2	Acceso a Oracle Discoverer Desktop.....	43
9.3	Tabla de usuarios y contraseñas.....	44
10.	Captura de pantalla de los informes.....	45
10.1	Capacidad total vs capacidad actual por río y año-mes.....	46
10.2	Capacidad total vs capacidad actual para embalse y año-mes.....	47
10.3	Capacidad total vs capacidad actual por comunidad autónoma y año-mes.....	47
10.4	Comparativa anual de la evolución por comunidad autónoma de la capacidad agregada de todos sus embalses.....	48
10.5	Comparación de la capacidad media agregada de todos los embalses de una comunidad autónoma entre un año y el año anterior.....	48
10.6	Aportación (Entrada-Salida) por río, para embalse y año-mes; comparativa con la media anual de aportación del embalse.....	49
10.7	Entrada vs Salida media agregada anual por comunidad autónoma.....	49
10.8	Aportación (Entrada-Salida) por río, embalse y mes; comparativa con la media anual de aportación del embalse.....	50
10.9	Agua diaria en reserva disponible en toda la confederación hidrográfica por año-mes.....	50
10.10	Gráfico de la evolución de los Hm3 de agua de nieve equivalente prevista por año y embalse	51
11.	Predicción de periodos de sequía a partir del data mart desarrollado.....	51
12.	Conclusiones.....	55
13.	Hardware y software utilizado.....	55
13.1	Hardware.....	55
13.2	Software.....	55
14.	Bibliografía.....	56
14.1	Libros de texto.....	56
14.2	Manuales.....	56
14.3	Online.....	57

## 1. Introducción

A continuación se expone la memoria del trabajo realizado para la Confederación Hidrográfica Norte-Este, en adelante (CHNE), entidad encargada de la gestión del agua e infraestructuras hidráulicas del nordeste de la península ibérica. Su actividad consiste en realizar un seguimiento periódico del estado de los embalses que están bajo su responsabilidad, marcar tendencias y predecir periodos de sequía para que las autoridades competentes puedan tomar decisiones hidrológicas correctas en el futuro.

La CHNE, planteó la necesidad de creación de un almacén de datos que permita automatizar la información que los técnicos del departamento responsable de planificación hidrológica, reciben en hojas de cálculo, y explotar esta información, utilizando herramientas de Business Intelligence adecuadas. Actualmente, este trabajo es realizado con la ayuda de aplicaciones ofimáticas, lo cual resulta pesado y requiere disponer de un elevado número de recursos humanos destinado a dicha actividad.

## 2. Objetivos

El objetivo principal del proyecto encargado por la CHNE es la creación de un almacén de datos corporativo, a partir de una base de datos transaccional, para la automatización y explotación de la información de planificación hidrológica, indispensable para la toma de decisiones de la confederación. Esto permitirá a la CHNE mejorar la eficiencia en sus procesos, reduciendo considerablemente los costes en recursos humanos (que no en personal), y por lo tanto económicos, que asume en la actualidad.

A nivel corporativo, los almacenes de datos y las herramientas OLAP permiten la explotación de los datos almacenados en la organización para ofrecer soporte en la toma de decisiones. Un almacén de datos es un repositorio diseñado con esta finalidad y por tanto está orientado a recopilar, resumir y tratar eficientemente un gran volumen de datos presente en cualquier sistema de información, de manera que facilite el análisis de la información desde distintas dimensiones y de esta forma permita detectar tendencias antes que los competidores, consiguiendo así obtener ventaja competitiva respecto a ellos.

Por lo tanto, se han tenido que analizar las técnicas existentes para proyectar la base de datos de un almacén de datos. El diseño está orientado a un almacén de datos físico ROLAP. Se han considerado factores tales como: desnormalización de tablas, inclusión de información agregada, historificación de la información, etc.

## 3. Requerimientos del cliente (CHNE)

La información que se muestra a continuación, es el conjunto de requerimientos definidos por la CHNE, para mejorar la eficiencia de sus procesos de planificación hidrológica, informatizándolos sustancialmente.

### 3.1 Datos de origen

Los datos proporcionados, se encuentran almacenados en hojas de cálculo (una por mes y año), con información cuantitativa del agua de los embalses y ríos que gestiona la CHNE. Las columnas de estas hojas de cálculo son:

- Embalse.
- Río.
- Capacidad total del embalse en Hm<sup>3</sup>.
- Capacidad actual del embalse (en el mes dado) en Hm<sup>3</sup>.
- Porcentaje actual sobre posible.
- Entrada en Hm<sup>3</sup> que ha recibido el embalse.
- Salida en Hm<sup>3</sup> del embalse.
- Incremento en Hm<sup>3</sup> del embalse (respecto la última medida).
- Capacidad del embalse el año anterior en Hm<sup>3</sup>.
- Porcentaje de la capacidad del embalse el año anterior.

Existe también otra hoja de cálculo (embalses.xls) con la relación de los embalses y ríos, y la provincia y la comunidad autónoma a la que pertenecen, así como una tabla de estimación de precipitaciones en forma de nieve que se muestra a continuación:

Embalse	Estimación nieve 2001	Estimación nieve 2002	Estimación nieve 2003	Estimación nieve 2004	Estimación nieve 2005
Maidereva	5	4	4	5	3
Santa Ana	4	4	4	3	3
Búbal	4	3	3	2	3
Valdabra	4	5	5	2	2
Calandra	3	3	3	3	3
Mansilla	3	2	1	2	2
Vicarías	4	5	5	5	5
Oliana	2	1	1	1	1
RESTO	1	1	1	1	1

Por último, la CHNE, nos ha proporcionado una fórmula para calcular la cantidad de agua en Hm<sup>3</sup> proveniente de las precipitaciones en forma de nieve de un embalse en un año:

$$\text{Hm}^3_{\text{agua_nieve_equivalente_anual_embalse}} = \text{estimación_nieve_embalse} \cdot \frac{\text{capacidad\_actual\_embalse}_{\text{mes}}}{5}$$

(Hm<sup>3</sup> de agua equivalente por año es igual a la media de la capacidad agregada anual del embalse multiplicado por la estimación de nieve del año considerado y dividido entre 5).

### ***3.2 Análisis de los datos de origen***

Los datos con la información analítica de los embalses y ríos que gestiona la CHNE, procedentes de las 54 hojas de cálculo, están contenidos en unos 3078 registros, 57 registros por hoja. Cada uno de los contiene los siguientes campos:

- Embalse: Cadena de caracteres que contiene el nombre del embalse. Hay 41 embalses diferentes, con una longitud máxima de 26 caracteres. No existen registros con valores nulos en este campo.
- Río: Cadena de caracteres que contiene el nombre del río. Hay 32 ríos diferentes, con una longitud máxima de 22 caracteres. Existe la posibilidad de que un río tenga más de un embalse. No existen registros con valores nulos en este campo.

#### ***AGUA EMBALSADA:***

- Capacidad total: Valor numérico que contiene la capacidad total del embalse en Hm<sup>3</sup>. Su dominio es el subconjunto de los números naturales dentro del rango [0-1534]. No existen registros con valores nulos en este campo.
- En esta fecha: Valor numérico que contiene la capacidad actual del embalse en Hm<sup>3</sup>. Su dominio es el subconjunto de los números naturales dentro del rango [0-1534]. No existen registros con valores nulos en este campo.
- %: Valor numérico calculado que contiene el porcentaje que supone la capacidad actual sobre el total posible. Su dominio es el subconjunto de los números naturales dentro del rango [0-100]. No existen registros con valores nulos en este campo.

#### ***VARIACIÓN SEMANAL:***

- Entrada: Valor numérico que contiene el volumen de agua en Hm<sup>3</sup> aportado por el río al embalse en un periodo de tiempo (semana, mes). Su dominio es el subconjunto de los números fraccionarios dentro del rango [0-500[, con una precisión de un decimal. Existen registros con valores nulos en este campo.
- Salida: Valor numérico que contiene el volumen de agua en Hm<sup>3</sup> desembalsada en un periodo de tiempo (semana, mes). Su dominio es el subconjunto de los números fraccionarios dentro del rango [0-500[, con una precisión de un decimal. Existen registros con valores nulos en este campo. (en realidad, existen registros con el valor '-'; que será interpretado como nulo).
- Incremento: Valor numérico que contiene el volumen de agua en Hm<sup>3</sup> que aumenta o disminuye el embalse con respecto a la capacidad actual tras sumarle las entradas y

restarle las salidas. Su dominio es el subconjunto de los números enteros dentro del rango  $]-130-95[$ . No existen registros con valores nulos en este campo.

*HACE 1 AÑO:*

- $\text{Hm}^3$ : Valor numérico que contiene la capacidad del embalse el año anterior en  $\text{Hm}^3$ . Su dominio es el subconjunto de los números naturales dentro del rango  $[0-1534]$ . No existen registros con valores nulos en este campo.
- %: Valor numérico calculado que contiene el porcentaje que supone la capacidad del embalse el año anterior sobre el total posible. Su dominio es el subconjunto de los números naturales dentro del rango  $[0-100]$ . No existen registros con valores nulos en este campo.

Además, al final de cada hoja de cálculo, existen unos resúmenes totales que son valores calculados a partir de los registros de dicha hoja. Estos resúmenes son los siguientes:

*AGUA EMBALSADA:*

- Capacidad total de todos los embalses: Valor numérico calculado que contiene el resultado de la suma de las capacidades totales de los embalses en  $\text{Hm}^3$ . Su dominio es el subconjunto de los números naturales dentro del rango  $[0-6504]$ .
- Capacidad actual de todos los embalses: Valor numérico calculado que contiene el resultado de la suma de las capacidades actuales de los embalses en  $\text{Hm}^3$ . Su dominio es el subconjunto de los números naturales dentro del rango  $[0-6504]$ .
- % del total: Valor numérico calculado que contiene el porcentaje total actual sobre el total posible. Su dominio es el subconjunto de los números naturales dentro del rango  $[0-100]$ .

*VARIACIÓN SEMANAL:*

- Incremento total: Valor numérico calculado que contiene el volumen de agua en  $\text{Hm}^3$  resultado de sumar todos los incrementos de todos los embalses. Su dominio es el subconjunto de los números enteros dentro del rango  $]-210-250[$ .
- Incremento porcentual: Valor numérico calculado que contiene el porcentaje que supone el incremento total sobre la capacidad total de todos los embalses. Su dominio es el subconjunto de los números fraccionarios dentro del rango  $[-100-100]$ , con una precisión de un decimal.

*HACE 1 AÑO:*

- Capacidad de todos los embalses el año anterior: Valor numérico calculado que contiene el resultado de la suma de las capacidades de todos los embalses el año

anterior en  $\text{Hm}^3$ . Su dominio es el subconjunto de los números naturales dentro del rango [0-6504].

- % del total el año anterior: Valor numérico calculado que contiene el porcentaje total el año anterior sobre el total posible. Su dominio es el subconjunto de los números naturales dentro del rango [0-100].

Por último, la hoja de cálculo “embalses.xls”, que contiene información geográfica de los embalses y ríos, se compone de los siguientes campos:

- C.A.: Cadena de caracteres que contiene el nombre de la comunidad autónoma donde se encuentra el embalse. En el caso de comunidades autónomas con más de una provincia, este campo contiene también el nombre de la provincia. Hay 6 comunidades autónomas diferentes, con una longitud máxima de 13 caracteres y 9 provincias distintas con un máximo de 9 caracteres. No existen registros con valores nulos en este campo.
- Embalse: Cadena de caracteres que contiene el nombre del embalse. Hay 41 embalses diferentes, con una longitud máxima de 26 caracteres. No existen registros con valores nulos en este campo.
- Río: Cadena de caracteres que contiene el nombre del río. Hay 32 ríos diferentes, con una longitud máxima de 22 caracteres. No existen registros con valores nulos en este campo.

Por lo tanto, se puede observar que la información de este fichero (“embalses.xls”) se encuentra jerarquizada en 4 niveles:

- Primer nivel: La comunidad autónoma. Con 6 comunidades autónomas diferentes: ARAGÓN, NAVARRA, LA RIOJA, CANTABRIA, CASTILLA Y LEÓN Y CATALUÑA.
- Segundo nivel: La provincia. En el caso de comunidades autónomas uniprovinciales, el nombre de la provincia será el de la comunidad autónoma. Son 9 provincias: Zaragoza, Huesca, Teruel, Navarra, La Rioja, Cantabria, Soria, Tarragona y Lérida.
- Tercer nivel: El río. Con 31 los ríos que gestiona la CHNE.
- Cuarto nivel: El embalse. Con los 41 embalses que gestiona la CHNE.

### ***3.3 Valoración de los datos de origen***

Puesto que los datos de origen provienen de hojas de cálculo cuya información se ha incluido mediante procesos no automatizados, se han detectado posibles errores en los mismos.



Por un lado, existen registros cuyos valores en el campo entradas eran negativos, o embalses cuyas entradas y/o salidas eran muy superiores a su capacidad total (por ejemplo, el embalse Bubál abastecido por el río Gallego, con una capacidad total de 64 Hm<sup>3</sup>, y con registros cuya entrada es de 493,5 Hm<sup>3</sup> y su salida de 493,6 Hm<sup>3</sup>) aunque esto podría explicarse en caso de crecidas del río producido por lluvias torrenciales. Este tipo de valoraciones han quedado aclaradas por parte del cliente.

Tras realizar un análisis más profundo de los datos de origen, procedentes de las hojas de cálculo (este análisis consistió finalmente en cruzar los datos de los embalses y ríos de los ficheros proporcionados por el cliente, con el fin de encontrar incoherencias), se ha detectado también que en los ficheros con la situación de los embalses en un periodo de tiempo existen embalses y ríos que no se están en el fichero “embalses.xls” y viceversa.

Por lo tanto, hubo que considerar también como fuente de datos de los embalses y ríos estos ficheros de situación. Esto quiere decir que existen datos de la situación de embalses a los que inicialmente no se les puede asociar una provincia y comunidad autónoma. Sin embargo, gracias a la información encontrada en la página web que nos proporcionó el cliente, <http://www.chebro.es/contenido.visualizar.do?idContenido=3059&idMenu=3081>, este problema ha quedado resuelto.

Además, se han detectado errores de concordancia y de transcripción en los nombres de los embalses y los ríos entre las hojas de cálculo de situación de los embalses y el fichero “embalses.xls”. Sin embargo, si existe concordancia con respecto a los nombres de los embalses y ríos entre todas las hojas de cálculo de la situación de los embalses y ríos en un periodo de tiempo. Todos los errores de concordancia y transcripción, se han corregido en la fase de extracción, transformación y carga, en adelante ETL, quedando la siguiente relación de ríos y embalses:

EMBALSE	RIO	PROVINCIA	CA	CAPACIDAD
ALBINA	ALBINA	Álava	PAÍS VASCO	6
ALCAÑIZ (ESTANCA)	GUADALOPE	Teruel	ARAGÓN	7
ALLOZ	SALADO	Navarra	NAVARRA	65
ALMOCHUEL	AGUAS VIVAS	Zaragoza	ARAGÓN	
ARDISA	GÁLLEGO	Zaragoza	ARAGÓN	
BACHIMAÑA (LAGO)	CALDARÉS	Huesca	ARAGÓN	7
BARASONA	ÉSERA	Huesca	ARAGÓN	92
BASERCA	NOG. RIBAGORZANA	Lérida	CATALUÑA	22
BRAMATUERO ALTO	CALDARÉS	Huesca	ARAGÓN	5
BÚBAL	GÁLLEGO	Huesca	ARAGÓN	64
CALANDA	GUADALOPE	Teruel	ARAGÓN	54
CAMARASA	NOG. PALLARESA	Lérida	CATALUÑA	113
CANELLES	NOG. RIBAGORZANA	Huesca	ARAGÓN	688
CASPE	GUADALOPE	Zaragoza	ARAGÓN	82
CAVALLERS	NOGUERA DE TORT	Lérida	CATALUÑA	16
CERTESCÁNS	CERTESCÁNS	Lérida	CATALUÑA	16
CIENFUENS	FLUMEN	Huesca	ARAGÓN	
CIURANA	CIURANA	Tarragona	CATALUÑA	12
CUEVA FORADADA	MARTÍN	Teruel	ARAGÓN	29
EBRO	EBRO	Cantabria	CANTABRIA	540

ESCALES	NOG. RIBAGORZANA	Huesca	ARAGÓN	152
ESCARRA	ESCARRA	Huesca	ARAGÓN	5
EUGUI	ARGA	Navarra	NAVARRA	21
GALLIPUÉN	GUADALOPILLO	Teruel	ARAGÓN	
GONZALEZ LACASA	ALBERCOS	La Rioja	LA RIOJA	33
GRADO (EL)	CINCA	Huesca	ARAGÓN	399
GUIAMETS	ASMAT	Tarragona	CATALUÑA	10
IP	BARRANCO DE IP	Lérida	CATALUÑA	5
IRABIA	IRATI	Navarra	NAVARRA	14
LANUZA	GÁLLEGO	Huesca	ARAGÓN	17
LLAUSSET	LLAUSSET	Huesca	ARAGÓN	17
MAIDEVERA	ARANDA	Zaragoza	ARAGÓN	21
MAIRAGA	OLORIZ	Navarra	NAVARRA	
MANSILLA	NAJERILLA	La Rioja	LA RIOJA	68
MAR (LAGO)	FLAMISELL	Lérida	CATALUÑA	14
MEDIANO	CINCA	Huesca	ARAGÓN	430
MEQUINENZA	EBRO	Zaragoza	ARAGÓN	1534
MONEVA	AGUAS VIVAS	Zaragoza	ARAGÓN	8
MONTEAGUDO	NÁJIMA	Soria	CASTILLA LEON	9
NEGRO	PEGUERA	Lérida	CATALUÑA	6
OLIANA	SEGRE	Lérida	CATALUÑA	101
PAJARES	PIQUERAS	La Rioja	LA RIOJA	
PENA	PENA	Teruel	ARAGÓN	19
PEÑA (LA)	GÁLLEGO	Huesca	ARAGÓN	15
PIGNATELLI	EBRO	Navarra	NAVARRA	
PINA	EBRO	Zaragoza	ARAGÓN	
RESPOMUSO	AGUAS LIMPIAS	Huesca	ARAGÓN	18
RIALB	CALCÓN	Huesca	ARAGÓN	
RIALB	SEGRE	Lérida	CATALUÑA	
RIBARROJA	EBRO	Tarragona	CATALUÑA	210
SABURO (LAGO)	FLAMISELL	Lérida	CATALUÑA	11
SAN BARTOLOMÉ	ARBA	Zaragoza	ARAGÓN	6
SANTA ANA	NOG. RIBAGORZANA	Huesca	ARAGÓN	236
SANTOLEA	GUADALOPE	Teruel	ARAGÓN	54
S.MARÍA DE BELSUÉ	FLUMEN	Huesca	ARAGÓN	13
SOBRÓN	EBRO	Álava	PAÍS VASCO	20
SOTONERA	SOTÓN	Huesca	ARAGÓN	189
TERRAETS	NOG. PALLARESA	Lérida	CATALUÑA	23
TORCAS (LAS)	HUERVA	Zaragoza	ARAGÓN	7
TORROLLÓN	GRANDE	Huesca	ARAGÓN	
TORT (LAGO)	FLAMISELL	Lérida	CATALUÑA	7
TRANQUERA (LA)	PIEDRA	Zaragoza	ARAGÓN	79
TREMP O TALÁRN	NOG. PALLARESA	Lérida	CATALUÑA	205
ULLIVARRI	ZADORRA	Álava	PAÍS VASCO	147
URDALUR	ALZANIA	Navarra	NAVARRA	
URDICETO	URDICETO	Huesca	ARAGÓN	5
URRUNAGA	SANTA ENGRACIA	Álava	PAÍS VASCO	72

VADIELLO	GUATIZALEMA	Huesca	ARAGÓN	16
VALDABRA	VALDABRA	Huesca	ARAGÓN	
YESA	ARAGÓN	Navarra	NAVARRA	470

Existen también un par de errores de transcripción en la tabla de estimación de nieve, puesto que en el embalse “Maidereva”, el nombre correcto es “Maidevera”, y el embalse Vicarías, se trata en realidad del embalse “Monteagudo de las Vicarías” del fichero “embalses.xls”. Estos errores, también se han corregido en la fase de ETL.

Se entiende que, es responsabilidad del cliente, la interpretación de la valoración de los datos numéricos, quedando delimitada nuestra responsabilidad por la correcta extracción, tratamiento y carga de los datos en las tablas correspondientes.

### ***3.4 Informes a realizar***

Como resultado del trabajo a desarrollar, la CHNE deseaba poder obtener, al menos la siguiente información:

- Capacidad total vs capacidad actual en Hm<sup>3</sup> por río y año-mes.
- Capacidad total vs capacidad actual en Hm<sup>3</sup> para embalse y año-mes.
- Capacidad total vs capacidad actual en Hm<sup>3</sup> por comunidad autónoma y año-mes.
- Comparativa anual de la evolución por comunidad autónoma de la capacidad agregada de todos sus embalses.
- Comparativa de la capacidad media agregada de todos los embalses de una comunidad autónoma entre un año y el año anterior (valores y porcentaje de incremento año anterior / año).
- Aportación (Entrada-Salida) en Hm<sup>3</sup> por río, para embalse y año-mes; comparativa con la media anual de aportación del embalse.
- Entrada vs salida media agregada anual por comunidad autónoma (en Hm<sup>3</sup>).
- Aportación (Entrada-Salida) en Hm<sup>3</sup> por río, embalse y mes; comparativa con la media anual de aportación del embalse.
- Agua diaria en reserva disponible (Hm<sup>3</sup>/día) en toda la confederación hidrográfica por año-mes.
- Gráfico de la evolución de los Hm<sup>3</sup> de agua de nieve equivalente prevista por año y embalse, que muestre, si los hay, los casos de desbordamiento del embalse.
- Predicción detallada e ilustrada con ejemplos, de períodos de sequía, a partir del data mart desarrollado.

Nota: los volúmenes de agua deben de poderse ver en Hm<sup>3</sup> (hectómetros cúbicos) y en MI (millones de litros), a elección de los técnicos usuarios del data mart.

### ***3.5 Elementos para el análisis multidimensional***

Del estudio inicial de los datos aportados por la CHNE, y del análisis preliminar de los informes solicitados por ésta, se pueden extraer los siguientes elementos:

- Indicadores / hechos / medidas: Representan algún aspecto cuantificable o medible de los objetos o eventos a analizar. En este proyecto, se han identificado los siguientes indicadores: la capacidad actual de un embalse, las entradas, las salidas de agua, y la capacidad del año anterior del embalse. El porcentaje, el incremento, y el porcentaje del año anterior son también indicadores derivados que se calculan a partir de la capacidad total (en el caso de los porcentajes tanto actual como del año anterior) y a partir de las entradas y salidas (en el caso del incremento, que no es más que la resta de la entrada y salida). También, hay que tener en cuenta la estimación de nieve de algunos de los embalses.
- Dimensiones: Son las perspectivas de análisis de los indicadores, catálogos de información complementaria, necesaria para la presentación de los datos a los usuarios. Se han identificado las siguientes: Comunidad autónoma, embalse, río, tiempo.
- Atributos: Para cada dimensión se identifican los siguientes atributos:
  - Comunidad autónoma: Nombre, provincia.
  - Embalse: Nombre, río, provincia, capacidad total.
  - Río: Nombre.
  - Tiempo: Año, mes.
- Jerarquías: Dentro de los atributos de las dimensiones se pueden establecer las siguientes jerarquías:
  - Embalse -> Provincia -> Comunidad autónoma.
  - Embalse -> Río.
  - Tiempo: Mes -> Año.

## **4. Planificación**

Una vez definidos los objetivos del proyecto y establecido el análisis preliminar de requerimientos, se ha planificado el proyecto en función de las fechas establecidas en la agenda de este semestre. Se han definido las fases, actividades y tareas que se han realizado en los periodos de tiempo que se han marcado previamente, con el fin de completar todos los trabajos en un tiempo adecuado a la planificación.

#### ***4.1 Fases del proyecto***

Las fases en las que se dividió el proyecto son las siguientes:

- **Análisis preliminar de requerimientos:** En esta fase, se analiza sin entrar en detalle las fuentes de datos operacionales proporcionados por la CHNE, que sirven para cargar cada uno de los elementos, así como los elementos identificados (dimensiones, atributos, indicadores, etc.), disponibles para los usuarios.
- **Análisis de requerimientos y diseño técnico:** En esta fase, se analizan con detalle los requerimientos obtenidos en la fase anterior, y se diseña el modelo dimensional conceptual y físico y los procedimientos de extracción de datos a alto nivel (procesos, pseudocódigo, etc.) que da soporte a las necesidades de los usuarios.
- **Implementación:** En esta fase, se prepara el soporte físico tanto hardware como software (Sistema Gestor de Bases de Datos SGBD), necesario para el almacén de datos. Se genera el modelo de datos a partir del modelo lógico obtenido del análisis de requerimientos de los usuarios. ETL (Extraction, Transformation, Loading), comprende la extracción de datos, su transformación y depuración para la carga de los mismos. Se diseña y crea el cubo OLAP. Se instala la herramienta de explotación de datos (Oracle Discoverer). Se analiza la información y construyen los informes, y por último, se verifica y ajusta el rendimiento de los procesos ETL, las consultas y los informes.

#### ***4.2 Actividades***

Las principales actividades que componen el proyecto se muestran en la siguiente tabla:

<b>Actividad</b>	<b>Fecha Inicio</b>	<b>Fecha Entrega</b>
0. Inicio del proyecto	21/09/2010	23/09/2010
1. Plan de trabajo y análisis preliminar de requerimientos	23/09/2010	05/10/2010
2. Análisis de requerimientos y diseño técnico	06/10/2010	08/11/2010
3. Implementación	09/11/2010	20/12/2010
4. Entrega Final	21/12/2010	10/01/2011
5. Debate	24/01/2011	27/01/2011

### 4.3 Tareas

A su vez, cada actividad se ha dividido en tareas según la siguiente tabla:

Actividad	Tarea
0	Recopilación de la documentación
0	Estudio de la documentación
0	Instalación del software "Gantt Project"
1	<b>1.1 Análisis preliminar de requerimientos</b>
1	1.1.1 Definición de objetivos
1	1.1.2 Análisis de los datos de origen
1	1.1.3 Valoración preliminar de los datos de origen
1	1.1.4 Análisis preliminar de los elementos identificados
1	<b>1.2 Plan de trabajo</b>
1	1.2.1 Definición de actividades
1	1.2.2 Definición de tareas
1	1.2.3 Programación del trabajo
1	1.2.4 Elaboración del calendario
1	1.2.5 Elaboración del diagrama de Gantt
1	<b>1.3 Revisión de la documentación</b>
1	1.3.1 Revisión del borrador del documento del plan de trabajo y análisis preliminar de requerimientos
1	1.3.2 Entrega del borrador del documento del plan de trabajo y análisis preliminar de requerimientos
1	1.3.3 Revisión del plan de trabajo y análisis preliminar de requerimientos
1	1.3.4 Revisión de la documentación del plan de trabajo y análisis preliminar de requerimientos
1	<b>1.4 Entrega del plan de trabajo y análisis preliminar de requerimientos</b>
2	<b>2.1 Instalación del software recomendado</b>
2	<b>2.2 Nueva documentación</b>
2	2.2.1 Recopilación de nueva documentación
2	2.2.2 Estudio de la nueva documentación
2	<b>2.3 Análisis de requerimientos</b>
2	2.3.1 Análisis detallado de los datos de origen
2	2.3.2 Valoración detallada de los datos de origen
2	2.3.3 Análisis detallado de los elementos identificados
2	2.3.4 Feedback del plan de trabajo y análisis preliminar de requerimientos
2	2.3.5 Revisión del plan de trabajo y análisis preliminar de requerimientos
2	<b>2.4 Diseño técnico</b>
2	2.4.1 Diseño del modelo dimensional conceptual
2	2.4.2 Diseño del modelo dimensional físico
2	2.4.3 Diseño de los procedimientos de extracción de datos de alto nivel (procesos, pseudocódigo, etc)
2	<b>2.5 Revisión de la documentación</b>
2	2.5.1 Revisión del borrador del documento del análisis de requerimientos y diseño técnico
2	2.5.2 Entrega del borrador del documento del análisis de requerimientos y diseño

	técnico
2	2.5.3 Revisión del análisis de requerimientos y diseño técnico
2	2.5.4 Revisión de la documentación del análisis de requerimientos y diseño técnico
2	<b>2.6 Entrega del análisis de requerimientos y diseño técnico</b>
3	<b>3.1 Instalación del software necesario</b>
3	3.1.1 Instalación de Virtual BOX + máquina virtual proporcionada por el cliente con todo el software indispensable para el proyecto instalado
3	<b>3.2 Nueva documentación</b>
3	3.2.1 Recopilación de nueva documentación
3	3.2.2 Estudio de la nueva documentación
3	<b>3.3 Construcción del almacén de datos</b>
3	3.3.1 Programación de los procedimientos necesarios para la carga de los datos (ETL – Extraction, Transformation, Loading)
3	3.3.2 Feedback del análisis de requerimientos y diseño técnico
3	3.3.3 Revisión del plan del análisis de requerimientos y diseño técnico
3	3.3.4 Carga de los datos
3	<b>3.4 Análisis de la información</b>
3	<b>3.5 Construcción de los informes</b>
3	<b>3.6 Revisión de la documentación</b>
3	3.6.1 Revisión del borrador del documento de la implementación
3	3.6.2 Entrega del borrador del documento de la implementación
3	3.6.3 Revisión de la implementación
3	3.6.4 Revisión de la documentación de la implementación
3	<b>3.7 Entrega de la implementación</b>
4	<b>4.1 Elaboración de la memoria</b>
4	<b>4.2 Feedback de la implementación</b>
4	<b>4.3 Elaboración de la presentación</b>
4	<b>4.4 Entrega final (memoria + presentación)</b>
5	<b>5.1 Debate Virtual y defensa del proyecto</b>

#### ***4.4 Programa de trabajo***

Una vez concretadas las actividades y tareas que componen el proyecto, se ha realizado la programación del trabajo a desarrollar, teniendo en cuenta las condiciones particulares del personal destinado al proyecto (formado solamente por mí), que son:

- Disponibilidad horaria:
  - Lunes a viernes de 17:00 a 21:30
  - Sábados de 10:00 a 13:00 y de 16:00 a 20:00
  - Domingos de 10:00 a 12:00
- Posibilidad de atender de manera esporádica, otros proyectos en el horario citado con anterioridad, debido a la inminente apertura de un nuevo centro sanitario en el que

desempeño mi actividad profesional como técnico especialista en informática, desarrollando aplicaciones web (Hospital Universitario “Los Arcos del Mar Menor” del Servicio Murciano de Salud).

#### 4.4.1 Calendario

Actividad	Tarea	Fecha Inicio	Fecha Fin
0	Recopilación de la documentación	21/09/2010	23/09/2010
0	Estudio de la documentación	21/09/2010	23/09/2010
0	Instalación del software “Gantt Project”	22/09/2010	22/09/2010
1	<b>1.1 Análisis preliminar de requerimientos</b>	23/09/2010	26/09/2010
1	1.1.1 Definición de objetivos	23/09/2010	23/09/2010
1	1.1.2 Análisis de los datos de origen	24/09/2010	24/09/2010
1	1.1.3 Valoración preliminar de los datos de origen	25/09/2010	25/09/2010
1	1.1.4 Análisis preliminar de los elementos identificados	26/09/2010	26/09/2010
1	<b>1.2 Plan de trabajo</b>	27/09/2010	29/09/2010
1	1.2.1 Definición de actividades	27/09/2010	27/09/2010
1	1.2.2 Definición de tareas	27/09/2010	27/09/2010
1	1.2.3 Programación del trabajo	28/09/2010	28/09/2010
1	1.2.4 Elaboración del calendario	28/09/2010	28/09/2010
1	1.2.5 Elaboración del diagrama de Gantt	29/09/2010	29/09/2010
1	<b>1.3 Revisión de la documentación</b>	29/09/2010	04/10/2010
1	1.3.1 Revisión del borrador del documento del plan de trabajo y análisis preliminar de requerimientos	29/09/2010	29/09/2010
1	1.3.2 Entrega del borrador del documento del plan de trabajo y análisis preliminar de requerimientos	30/09/2010	30/09/2010
1	1.3.3 Revisión del plan de trabajo y análisis preliminar de requerimientos	01/10/2010	03/10/2010
1	1.3.4 Revisión de la documentación del plan de trabajo y análisis preliminar de requerimientos	04/10/2010	04/10/2010
1	<b>1.4 Entrega del plan de trabajo y análisis preliminar de requerimientos</b>	<b>05/10/2010</b>	<b>05/10/2010</b>
2	<b>2.1 Instalación del software recomendado</b>	06/10/2010	06/10/2010
2	<b>2.2 Nueva documentación</b>	07/10/2010	10/10/2010
2	2.2.1 Recopilación de nueva documentación	07/10/2010	07/10/2010
2	2.2.2 Estudio de la nueva documentación	08/10/2010	10/10/2010
2	<b>2.3 Análisis de requerimientos</b>	11/10/2010	16/10/2010
2	2.3.1 Análisis detallado de los datos de origen	11/10/2010	12/10/2010
2	2.3.2 Valoración detallada de los datos de origen	12/10/2010	13/10/2010
2	2.3.3 Análisis detallado de los elementos identificados	14/10/2010	15/10/2010
2	2.3.4 Feedback del plan de trabajo y análisis preliminar de requerimientos	<b>15/10/2010</b>	<b>15/10/2010</b>
2	2.3.5 Revisión del plan de trabajo y análisis preliminar de requerimientos	15/10/2010	16/10/2010
2	<b>2.4 Diseño técnico</b>	18/10/2010	28/10/2010

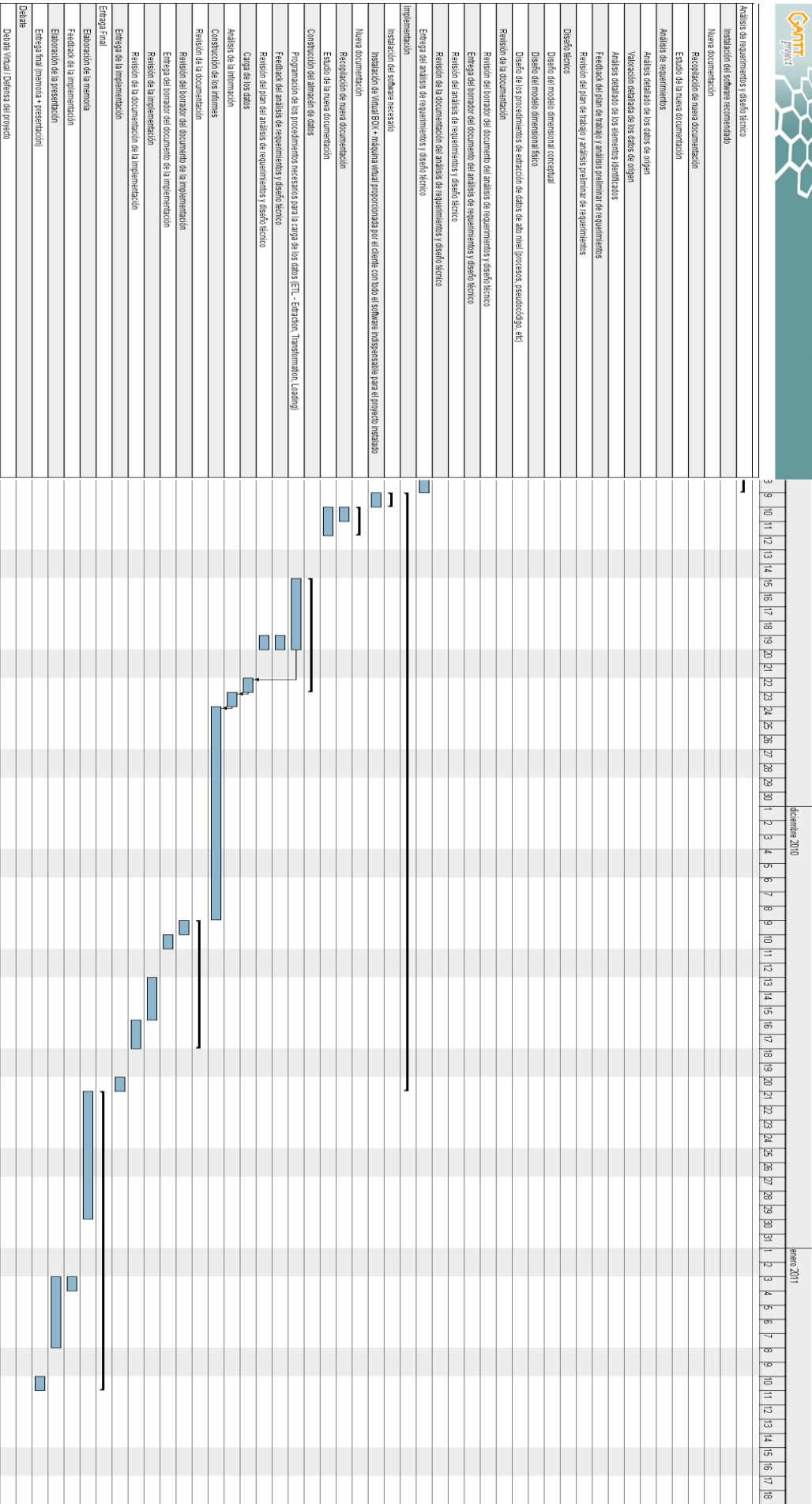


2	2.4.1 Diseño del modelo dimensional conceptual	18/10/2010	20/10/2010
2	2.4.2 Diseño del modelo dimensional físico	20/10/2010	22/10/2010
2	2.4.3 Diseño de los procedimientos de extracción de datos de alto nivel (procesos, pseudocódigo, etc)	23/10/2010	28/10/2010
2	<b>2.5 Revisión de la documentación</b>	29/10/2010	07/11/2010
2	2.5.1 Revisión del borrador del documento del análisis de requerimientos y diseño técnico	29/10/2010	29/10/2010
2	2.5.2 Entrega del borrador del documento del análisis de requerimientos y diseño técnico	30/10/2010	30/10/2010
2	2.5.3 Revisión del análisis de requerimientos y diseño técnico	31/10/2010	04/11/2010
2	2.5.4 Revisión de la documentación del análisis de requerimientos y diseño técnico	05/11/2010	07/11/2010
2	<b>2.6 Entrega del análisis de requerimientos y diseño técnico</b>	<b>08/11/2010</b>	<b>08/11/2010</b>
3	<b>3.1 Instalación del software necesario</b>	09/11/2010	09/11/2010
3	3.1.1 Instalación de Virtual BOX + máquina virtual proporcionada por el cliente con todo el software indispensable para el proyecto instalado	09/11/2010	09/11/2010
3	<b>3.2 Nueva documentación</b>	10/11/2010	12/11/2010
3	3.2.1 Recopilación de nueva documentación	10/10/2010	10/10/2010
3	3.2.2 Estudio de la nueva documentación	10/10/2010	12/10/2010
3	<b>3.3 Construcción del almacén de datos</b>	12/11/2010	13/11/2010
3	3.3.1 Programación de los procedimientos necesarios para la carga de los datos (ETL – Extraction, Transformation, Loading)	13/11/2010	18/11/2010
3	3.3.2 Feedback del análisis de requerimientos y diseño técnico	<b>19/11/2010</b>	<b>19/11/2010</b>
3	3.3.3 Revisión del plan del análisis de requerimientos y diseño técnico	19/11/2010	20/11/2010
3	3.3.4 Carga de los datos	20/11/2010	20/11/2010
3	<b>3.4 Análisis de la información</b>	22/11/2010	22/11/2010
3	<b>3.5 Construcción de los informes</b>	23/11/2010	09/12/2010
3	<b>3.6 Revisión de la documentación</b>	10/12/2010	19/12/2010
3	3.6.1 Revisión del borrador del documento de la implementación	10/12/2010	10/12/2010
3	3.6.2 Entrega del borrador del documento de la implementación	11/12/2010	11/12/2010
3	3.6.3 Revisión de la implementación	13/12/2010	16/12/2010
3	3.6.4 Revisión de la documentación de la implementación	17/12/2010	19/12/2010
3	<b>3.7 Entrega de la implementación</b>	<b>20/12/2010</b>	<b>20/12/2010</b>
4	<b>4.1 Elaboración de la memoria</b>	21/12/2010	30/12/2010
4	<b>4.2 Feedback de la implementación</b>	<b>03/01/2011</b>	<b>03/01/2011</b>
4	<b>4.3 Elaboración de la presentación</b>	03/01/2011	08/01/2011
4	<b>4.4 Entrega final (memoria + presentación)</b>	<b>10/01/2011</b>	<b>10/01/2011</b>
5	<b>5.1 Debate Virtual y defensa del proyecto</b>	<b>24/01/2011</b>	<b>27/01/2011</b>

### **4.4.2 Diagrama de Gantt**

El diagrama de Gantt correspondiente a la programación del proyecto se muestra a continuación. Se ha adjuntado también dicho diagrama en formato xml y mpx (Microsoft Project) a la documentación del proyecto entregada ya al cliente.





## 5. Incidencias y riesgos

Las incidencias y/o riesgos que se han podido producir durante el transcurso del proyecto y sus soluciones previstas para cada caso concreto, que se han tenido en cuenta son las siguientes:

- Incidencias por motivos laborales: Como se menciona en el apartado “4.4 Programa de trabajo”, existe la posibilidad de desatender mínimamente y de manera esporádica este proyecto en el horario de lunes a viernes, citado en dicho punto (4.4), debido a la inminente apertura de un nuevo centro sanitario en el que desempeño mi actividad profesional como técnico especialista en informática, desarrollando aplicaciones web (Hospital Universitario “Los Arcos del Mar Menor” del Servicio Murciano de Salud). **Solución:** Recuperar el tiempo no invertido en el proyecto aumentando el horario de trabajo en caso necesario, o con los días de vacaciones y días de libre disposición de los que dispongo.
- Incidencias por motivos de salud: Ésta es sin duda la contingencia más impredecible a la que se enfrenta cualquier proyecto. **Solución:** Hacer una planificación flexible, llevar al día o incluso adelantada la redacción de algunos capítulos y el desarrollo de ciertas tareas, son medidas de prevención ante este tipo de problemas. Estaba previsto también prescindir de la entrega de borradores del proyecto si se presentaban incidencias de este tipo, lo cual no fue necesario.
- Cursos de formación: Los días 13, 18 y 20 de octubre, de 17 a 19:30 (en horario de trabajo del proyecto), asistí a un curso de formación de “Microstrategy”. **Solución:** Recuperar el tiempo no invertido en el proyecto aumentando el horario de trabajo en caso necesario. Además, en el curso, se vieron temas relacionados con este proyecto como: diseño de proyectos, modelo lógico de datos, hechos, atributos, jerarquías, creación de indicadores, diseño de informes, etc. con lo que sirvió de estudio con vistas a fases posteriores del proyecto. Más información en: <http://www.ffis.es/pec2010/mostrarcursos.php?id=747>
- Incidencias por motivos técnicos: A lo largo del proyecto se pueden producir diversos problemas técnicos con el equipamiento empleado para llevarlo a cabo. **Solución:** Se ha adelantado la fase de instalación del software necesario para desarrollar el proyecto (Instalación de Virtual Box y de la máquina virtual proporcionada por el cliente). Se ha realizado también una copia de seguridad de dicha máquina una vez configurada, para una rápida restauración en caso de que sea necesario. Se dispone también, de hardware alternativo (2 ordenadores de sobremesa y 1 portátil adicional) y suficiente para sustituir el equipo principal en caso de avería y conexión a internet móvil por si falla la conexión a internet principal. Además se programaron copias de seguridad a diario de todo el material modificado, entregable al cliente y relacionado con el proyecto, estando estas copias de seguridad disponibles en varios servidores de correo.

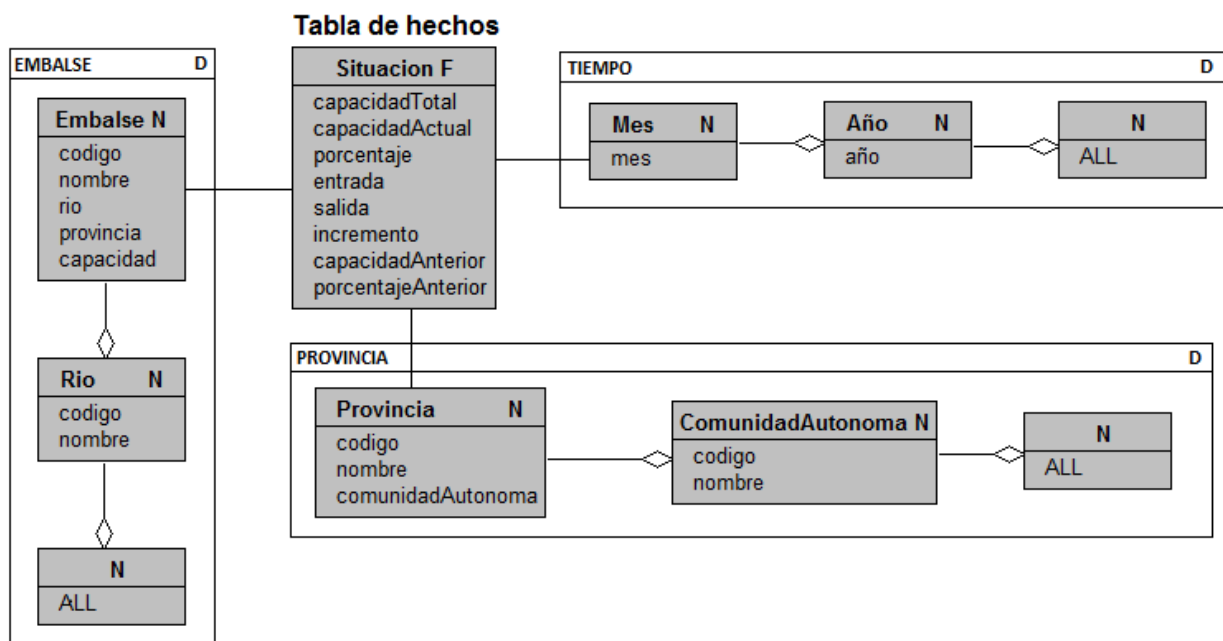
## 6. Diseño técnico

### 6.1 Diseño del modelo dimensional conceptual

Como en cualquier proyecto data warehouse (almacenes de datos en inglés, en adelante DW), éste gira en torno a una tabla central del esquema dimensional, también conocida como tabla de hechos que contiene los valores de las medidas de negocio o indicadores. Cada medida se toma mediante la intersección de las dimensiones que la definen, reflejadas en sus correspondientes tablas de dimensiones que rodean la tabla de hechos y relacionándose con ella.

A la hora de diseñar el modelo conceptual de este proyecto se ha tenido en cuenta que, la capacidad total del embalse, es un atributo propio de éste, puesto que es una característica que aunque podría cambiar en el tiempo (por ejemplo, debido a una obra para ampliar la capacidad del embalse), en este proyecto en concreto, permanece constante en el tiempo según el análisis realizado a los datos de origen. Sin embargo, por motivos de eficiencia, se ha añadido también la capacidad total a la tabla de hechos a pesar de la redundancia. Esta es una característica muy típica de los almacenes de datos (desnormalización de tablas).

A continuación se muestra el modelo conceptual de datos de este proyecto mediante un diagrama en estrella en el que se pueden apreciar también las jerarquías:



### 6.2 Diseño del modelo lógico

Partiendo del diagrama en estrella del apartado anterior, el diseño lógico se obtiene creando una entidad para la tabla de hechos y una por cada una de las dimensiones que componen el modelo. Una vez hecho esto, hay que identificar los identificadores, las restricciones y relaciones que puedan existir entre las diferentes entidades. El modelo lógico obtenido se muestra a continuación:

Entidades:

- **ComunidadAutonoma** (codigo, nombre)
- **Provincia** (codigo, nombre, *comunidadAutonoma*)
- **Rio** (codigo, nombre)
- **Embalse** (codigo, nombre, *rio*, *provincia*, capacidad)
- **Situacion** (fecha, embalse, capacidadTotal, capacidadActual, porcentaje, entrada, salida, incremento, capacidadAnterior, porcentajeAnterior)
- **Nieve** (embalse, anio, estimación, capacidadMedia)

Relaciones:

- Provincia.comunidadAutonoma -> ComunidadAutonoma.codigo.
- Embalse.provincia -> Provincia.codigo
- Embalse.rio -> Rio.codigo
- Situacion.embalse -> Embalse.codigo
- Nieve.embalse -> Embalse.codigo

### 6.3 Diseño del modelo físico

El modelo físico de la base de datos de este proyecto se compone de las siguientes tablas o relaciones:

- Comunidad autónoma: Contiene los datos de las comunidades que intervienen en el proyecto. Su estructura es la siguiente:

#### CA

Campo	Tipo	Nulo	Restricción
<u>codigo</u>	VARCHAR2(3)	NO	PK (Clave principal)
nombre	VARCHAR2(50)	NO	UK (Clave única)

- Provincia: Contiene los datos de las provincias que intervienen en el proyecto. Su estructura es la siguiente:

#### PROVINCIA

Campo	Tipo	Nulo	Restricción
<u>codigo</u>	VARCHAR2(3)	NO	PK
nombre	VARCHAR2(50)	NO	UK
ca	VARCHAR(3)	NO	FK (Clave foránea) -> (CA.codigo)

- Río: Contiene los datos de los ríos que intervienen en el proyecto. Su estructura es la siguiente:

#### RIO

Campo	Tipo	Nulo	Restricción
<u>codigo</u>	VARCHAR2(3)	NO	PK
nombre	VARCHAR2(50)	NO	UK

- Embalse: Contiene los datos de los embalses que intervienen en el proyecto. Su estructura es la siguiente:

**EMBALSE**

Campo	Tipo	Nulo	Restricción
codigo	VARCHAR2(3)	NO	PK
nombre	VARCHAR2(50)	NO	
rio	VARCHAR2(3)	SI	FK -> (RIO.codigo)
provincia	VARCHAR2(3)	SI	FK -> (PROVINCIA.codigo)
capacidad	NUMBER(4)	SI	

- Situación: Tabla de hechos que contiene los datos de la situación de los embalses y ríos así como sus indicadores que intervienen en el proyecto. Su estructura es la siguiente:

**SITUACION**

Campo	Tipo	Nulo	Restricción	
<u>fecha</u>	DATE	NO	PK	FK -> (EMBALSE.codigo)
<u>embalse</u>	VARCHAR2(3)	NO		
capacidadTotal	NUMBER(4)	NO		
capacidadActual	NUMBER(4)	NO		
porcentaje	NUMBER(3)	NO		
entrada	NUMBER(4,1)	SI		
salida	NUMBER(4,1)	SI		
incremento	NUMBER(3)	NO		
capacidadAnterior	NUMBER(4)	NO		
porcentajeAnterior	NUMBER(3)	NO		

- Nieve: Tabla que contiene la estimación de nieve de algunos embalses. Su estructura es la siguiente:

**NIEVE**

Campo	Tipo	Nulo	Restricción	
<u>embalse</u>	VARCHAR2(3)	NO	PK	FK -> (EMBALSE.codigo)
<u>anio</u>	VARCHAR2(4)	NO		
estimacion	NUMBER(1)	NO		
capacidadMedia	NUMBER(6,3)	SI		

También se han creado tablas temporales en las que se cargan directamente los datos de las hojas de cálculo sin transformar, para después insertar los datos resultantes elaborados a partir de los procesos (ETL) en las tablas anteriormente descritas. Estas tablas son:

- Embalse\_tmp: Tabla temporal que contiene el nombre de los ríos, embalses, provincias y comunidades autónomas cargados directamente de la hoja de cálculo “embalses.xls”.

**EMBALSE\_TMP**

Campo	Tipo	Nulo	Restricción
codigo	VARCHAR2(2)	NO	PK



ca	VARCHAR2(50)	SI	
embalse	VARCHAR2(50)	SI	
rio	VARCHAR2(50)	SI	

- Situacion\_tmp: Tabla temporal que contiene los datos de la situación de los embalses y ríos cargados directamente de las distintas hojas de cálculo. Su estructura es la siguiente:

**SITUACION\_TMP**

Campo	Tipo	Nulo	Restricción
fecha	DATE	SI	
embalse	VARCHAR2(50)	SI	
rio	VARCHAR2(50)	SI	
capacidadTotal	VARCHAR2(60)	SI	
capacidadActual	VARCHAR2(50)	SI	
porcentaje	VARCHAR2(50)	SI	
entrada	VARCHAR2(50)	SI	
salida	VARCHAR2(50)	SI	
incremento	VARCHAR2(50)	SI	
capacidadAnterior	VARCHAR2(50)	SI	
porcentajeAnterior	VARCHAR2(50)	SI	

## ***6.4 Diseño de los procedimientos de extracción de datos***

El proceso de ETL (Extracción, transformación y carga), se divide en los siguientes procesos:

- Transformación de las hojas de cálculo (xls) a formato csv. En el caso del fichero “embalses.xls” simplemente se ha abierto con Excel (o con cualquier otro software similar) y guardado como fichero en formato csv. Con los ficheros de la situación de los embalses y ríos, además, se han suprimido primero, de cada hoja, las columnas H y J, para todos y cada uno de los 54 ficheros (xls), tal y como se muestra en la siguiente captura de pantalla:

EMBALSES	RÍOS	AGUA EMBALSADA			VARIACIÓN SEMANAL		
		Capacidad Total Hm³	En esta fecha Hm³	%	Entrada Hm³	Salida Hm³	Incremento Hm³
10 ALBINA	ALBINA	6	4	67	---	---	0
11 ALCAÑIZ (ESTANCA)	GUADALOPE	7	6	86	1	1	0
12 ALLOZ	SALADO	65	37	57	0	2,4	-3
13 BACHIMAÑA (LAGO)	CALDERÉS	7	6	86	---	---	0
14 BARASONA	ÉSERA	92	65	71	9,6	18,7	-9
15 BASERCA	NOG. RIBAGORZANA	22	20	91	3,8	3,1	1
16 BRAMATUERO ALTO	CALDERÉS	5	4	80	---	---	0
17 BUBAL	GÁLLEGO	64	39	61	4,2	7,3	-3
18 CALANDA	GUADALOPE	54	36	67	1,8	3,6	-2
19 CAMARASA	NOG. PALLARESA	113	64	57	6,9	20	-13
20 CANELLES	NOG. RIBAGORZANA	688	560	81	24	24	0
21 CASPE	GUADALOPE	82	63	77	0,7	1,9	-2
22 CAVALLERS	NOGUERA DE TORT	16	13	81	1,5	0,4	1
23 CERTESCÁNS	CERTESCÁNS	16	12	75	---	---	-1
24 CIURANA	CIURANA	12	4	33	0	0	0
25 CUEVA FORADADA	MARTÍN	29	2	7	0,2	0,2	0
26 EBRO	EBRO	540	353	65	0,9	24,8	-24
27 ESCALES	NOG. RIBAGORZANA	152	137	90	2,7	8,3	-6
28 ESCARRA	ESCARRA	5	2	40	---	---	-1
29 EUGUJ	ARGA	21	17	81	0,3	0,8	0
30 GONZALEZ LACASA	ALBERCOS	33	18	55	0	2,2	-3
31 GRADO (EL)	CINCA	399	377	94	34,3	42,8	-9
32 GUIAMETS	ASMAT	10	3	30	0	0	0

- Carga automática del fichero “embales.csv”: Sólo deberá realizarse una vez salvo que cambien los datos de los embales y ríos, o se añadan registros nuevos. Se utilizará para este fin la herramienta SQL\* Loader, y se cargarán los datos en la tabla temporal “CHNE.embalse\_tmp”, después de transformar el fichero a formato CSV.

```
sqlldr chne/uoc@XE control=embalse_tmp.ctl
```

```
OPTIONS (SKIP=1)
LOAD DATA
  INFILE 'csv\embales.csv'
  REPLACE
  INTO TABLE CHNE.embalse_tmp
  FIELDS TERMINATED BY ";" OPTIONALLY ENCLOSED BY '"' TRAILING NULLCOLS
  (ca,
   embalse,
   rio)
```

- Carga automática de los ficheros de la situación de los embales: Deberá realizarse una carga por cada uno de los 54 ficheros con las situaciones de los embales en un periodo de tiempo. Esta carga será individual, y no se procederá a cargar otro fichero sin haber terminado todo el proceso. Se utilizará para este fin, la herramienta SQL\* Loader, y se cargarán los datos en la tabla temporal “CHNE.situacion\_tmp”, después de transformar el fichero a formato CSV. Esta tabla, será borrada antes de cargar el siguiente fichero.

```
sqlldr chne/uoc@XE control=situacion_tmp.ctl
```

```
LOAD DATA
  INFILE 'csv\eYYYYMMDD.csv'
  REPLACE
  INTO TABLE CHNE.situacion_tmp
  FIELDS TERMINATED BY ";" OPTIONALLY ENCLOSED BY '"' TRAILING NULLCOLS
  (embalse,
   rio,
   capacidadtotal,
   capacidadactual,
```

```

porcentaje,
entrada,
salida,
incremento,
capacidadanterior,
porcentajeanterior)

```

Para los siguientes procesos, se ha creado un paquete de base de datos con procedimientos almacenados por cada una de las tablas que interviene en el proceso, con el fin de estructurar mejor todo el procedimiento. Estos procesos (procedimientos almacenados), son llamados por el procedimiento principal **pSituacion.ETL** y se describen a continuación:

- **CHNE.pEmbalse\_tmp.actualiza.** Tras la carga automática del fichero “embalses.csv” y con algunos de los embalses y ríos en la tabla “CHNE.embalse\_tmp”, se procede a preparar los datos de ésta, para luego poder copiarlos a las tablas definitivas de comunidades autónomas, provincias, ríos y embalses. Para ello se empieza actualizando las provincias de los embalses:

```

PROCEDURE actualiza IS
BEGIN
    UPDATE CHNE.embalse_tmp SET ca = 'Zaragoza' WHERE codigo BETWEEN 3 AND 9;
    UPDATE CHNE.embalse_tmp SET ca = 'Huesca' WHERE codigo BETWEEN 11 AND 22;
    UPDATE CHNE.embalse_tmp SET ca = 'Teruel' WHERE codigo BETWEEN 24 AND 28;
    UPDATE CHNE.embalse_tmp SET ca = 'Navarra' WHERE codigo BETWEEN 30 AND 35;
    UPDATE CHNE.embalse_tmp SET ca = 'La Rioja' WHERE codigo BETWEEN 37 AND 39;
    UPDATE CHNE.embalse_tmp SET ca = 'Cantabria' WHERE codigo = 41;
    UPDATE CHNE.embalse_tmp SET ca = 'Lérida' WHERE codigo = 47;

    COMMIT;

END actualiza;

```

Este proceso, en principio sólo debería lanzarse una vez, salvo que se vuelva a cargar el fichero “embalses.csv”. En tal caso, siempre se le asignarán el mismo identificador a los registros puesto que la secuencia que selecciona el identificador de la tabla, está definida como:

```

CREATE SEQUENCE "CHNE"."SEMBALSE_TMP" MINVALUE 1 MAXVALUE 47 INCREMENT BY 1 START
WITH 1 NOCACHE ORDER CYCLE;

```

Sin embargo, para simplificar el proceso ETL, será llamado cada vez que se ejecute el procedimiento principal **pSituacion.ETL**.

- **CHNE.pSituacion\_tmp.actualiza.** Tras la carga automática de cada fichero de situación de los embalses y ríos en la tabla “CHNE.situacion\_tmp”, se procede a extraer la fecha que se encuentra a modo de título, con un formato adecuado, para que pase a formar parte de los registros correspondientes. También se transforman los nulos (null, ‘---’) en ceros, y se borran las filas que no se necesitan:

```

PROCEDURE actualiza IS
    contador NUMBER := 0;
    fecha_ CHNE.situacion_tmp.fecha%TYPE;

```

```

BEGIN

SELECT COUNT(*)
  INTO contador
  FROM CHNE.situacion_tmp;

IF contador > 0 THEN

  SELECT TO_DATE(fecha, 'dd Month YYYY')
    INTO fecha_
    FROM (SELECT REPLACE(SUBSTR(capacidadtotal, 30), 'DE', '') AS fecha
          FROM CHNE.situacion_tmp
          WHERE capacidadtotal LIKE '%SITUACI%');

  UPDATE CHNE.situacion_tmp
    SET fecha = fecha_
  WHERE embalse IS NOT NULL
    OR embalse NOT LIKE '%EMBALSES%';

  UPDATE CHNE.situacion_tmp
    SET entrada = 0
  WHERE entrada IS NULL
    AND (embalse IS NOT NULL
        OR embalse NOT LIKE '%EMBALSES%');

  UPDATE CHNE.situacion_tmp
    SET salida = 0
  WHERE salida LIKE '%---%'
    AND (embalse IS NOT NULL
        OR embalse NOT LIKE '%EMBALSES%');

  DELETE
    FROM CHNE.situacion_tmp
  WHERE embalse IS NULL
    OR embalse LIKE '%EMBALSES%';

  COMMIT;

END IF;

END actualiza;

```

Este proceso, deberá lanzarse cada vez que se cargue un fichero de situación de los embalses y ríos en la tabla “CHNE.situacion\_tmp”. Si no existieran registros en dicha tabla, el procedimiento no hará nada.

- **CHNE.pRio.inserta.** Este procedimiento, inserta los datos de los ríos procedentes de las tablas “CHNE.embalse\_tmp” y “CHNE.situacion\_tmp” **que no se hayan insertado ya**, en la tabla “CHNE.rio”, corrigiendo los errores de concordancia y transcripción detectados entre ambas tablas:

```

PROCEDURE inserta IS

BEGIN

  INSERT
    INTO CHNE.rio(nombre)
  SELECT DISTINCT
    DECODE(rio,
      'CALDERÉS', 'CALDARÉS',
      'ESERA', 'ÉSERA',
      'FLAMISEL', 'FLAMISELL',
      'GUADALOPE (DERIVACIÓN)', 'GUADALOPE',
      'NOGUERA-RIBAGORZANA', 'NOG. RIBAGORZANA',
      rio) AS rio

```

```

FROM (SELECT UPPER(rio) AS rio
      FROM CHNE.embalse_tmp
      WHERE embalse IS NOT NULL
UNION
SELECT rio
      FROM CHNE.situacion_tmp)
WHERE DECODE(rio,
             'CALDERÉS', 'CALDARÉS',
             'ESERA', 'ÉSERA',
             'FLAMISEL', 'FLAMISELL',
             'GUADALOPE (DERIVACIÓN)', 'GUADALOPE',
             'NOGUERA-RIBAGORZANA', 'NOG. RIBAGORZANA',
             rio) NOT IN (SELECT nombre FROM CHNE.rio)

ORDER BY rio;

COMMIT;

END inserta;

```

Este proceso, en principio sólo debería lanzarse una vez, salvo que se vuelva a cargar el fichero “embalses.csv” y contenga nuevos ríos o cambios en éstos. En cualquier caso, si se lanza varias veces, no da error, ya que sólo inserta aquellos registros que no se hayan insertado ya en la tabla “CHNE.rio” que en caso de lanzarse más de una vez, deben ser ninguno. Sin embargo, para simplificar el proceso ETL, será llamado cada vez que se ejecute el procedimiento principal ***pSituacion.ETL***.

- ***CHNE.pCA.inserta***. Este procedimiento, inserta los datos de las comunidades autónomas procedentes de la tabla “CHNE.embalse\_tmp” **que no se hayan insertado ya**, en la tabla “CHNE.ca”, además del País Vasco, que falta como comunidad autónoma según los embalses que aparecen en las hojas de cálculo de situación de los embalses y ríos.

```

PROCEDURE inserta IS
BEGIN
  INSERT
  INTO CHNE.ca(nombre)
  SELECT DISTINCT ca
  FROM CHNE.embalse_tmp
  WHERE embalse IS NULL
  AND rio IS NULL
  AND ca NOT IN (SELECT nombre FROM CHNE.ca)
  UNION
  SELECT DISTINCT 'PAÍS VASCO' AS ca
  FROM CHNE.ca
  WHERE 'PAÍS VASCO' NOT IN (SELECT nombre FROM CHNE.ca);

  COMMIT;

END inserta;

```

Este proceso, en principio sólo debería lanzarse una vez, salvo que se vuelva a cargar el fichero “embalses.csv” y contenga nuevas comunidades autónomas o cambios en éstas. En cualquier caso, si se lanza varias veces, no da error, ya que sólo inserta aquellos registros que no se hayan insertado ya en la tabla “CHNE.ca” que en caso de lanzarse más de una vez, deben ser ninguno. Sin embargo, para simplificar el proceso ETL, será llamado cada vez que se ejecute el procedimiento principal ***pSituacion.ETL***.

- **CHNE.pProvincia.inserta.** Este procedimiento, inserta los datos de las provincias procedentes de las tablas “CHNE.embalse\_tmp” y “CHNE.ca” en la tabla “CHNE.provincia”, además de Álava, que falta como provincia según los embalses que aparecen en las hojas de cálculo de situación de los embalses y ríos.

```

PROCEDURE inserta IS

    contador NUMBER := 0;

BEGIN

    SELECT COUNT(*)
        INTO contador
        FROM CHNE.provincia;

    IF contador <= 0 THEN

        -- Provincias de ARAGÓN.
        INSERT INTO CHNE.provincia
        SELECT (SELECT TO_CHAR(NVL(MAX(TO_NUMBER(codigo)), 0) + 1)
            FROM CHNE.provincia),
            'Zaragoza',
            codigo
        FROM CHNE.ca
        WHERE nombre LIKE '%ARAG%';

        INSERT INTO CHNE.provincia
        SELECT (SELECT TO_CHAR(NVL(MAX(TO_NUMBER(codigo)), 0) + 1)
            FROM CHNE.provincia),
            'Huesca',
            codigo
        FROM CHNE.ca
        WHERE nombre LIKE '%ARAG%';

        INSERT INTO CHNE.provincia
        SELECT (SELECT TO_CHAR(NVL(MAX(TO_NUMBER(codigo)), 0) + 1)
            FROM CHNE.provincia),
            'Teruel',
            codigo
        FROM CHNE.ca
        WHERE nombre LIKE '%ARAG%';

        -- NAVARRA
        INSERT INTO CHNE.provincia
        SELECT (SELECT TO_CHAR(NVL(MAX(TO_NUMBER(codigo)), 0) + 1)
            FROM CHNE.provincia),
            'Navarra',
            codigo
        FROM CHNE.ca
        WHERE nombre LIKE '%NAVARRA%';

        -- LA RIOJA
        INSERT INTO CHNE.provincia
        SELECT (SELECT TO_CHAR(NVL(MAX(TO_NUMBER(codigo)), 0) + 1)
            FROM CHNE.provincia),
            'La Rioja',
            codigo
        FROM CHNE.ca
        WHERE nombre LIKE '%RIOJA%';

        -- CANTABRIA
        INSERT INTO CHNE.provincia
        SELECT (SELECT TO_CHAR(NVL(MAX(TO_NUMBER(codigo)), 0) + 1)
            FROM CHNE.provincia),
            'Cantabria',
            codigo
        FROM CHNE.ca
        WHERE nombre LIKE '%CANTABRIA%';
    
```

```

-- CASTILLA LEÓN
INSERT INTO CHNE.provincia
SELECT (SELECT TO_CHAR(NVL(MAX(TO_NUMBER(codigo)), 0) + 1)
        FROM CHNE.provincia),
        'Soria',
        codigo
FROM CHNE.ca
WHERE nombre LIKE '%CASTILLA LEON%';

-- Provincias de CATALUÑA.
INSERT INTO CHNE.provincia
SELECT (SELECT TO_CHAR(NVL(MAX(TO_NUMBER(codigo)), 0) + 1)
        FROM CHNE.provincia),
        'Tarragona',
        codigo
FROM CHNE.ca
WHERE nombre LIKE '%CATALU%';

INSERT INTO CHNE.provincia
SELECT (SELECT TO_CHAR(NVL(MAX(TO_NUMBER(codigo)), 0) + 1)
        FROM CHNE.provincia),
        'Lérida',
        codigo
FROM CHNE.ca
WHERE nombre LIKE '%CATALU%';

-- Provincias del PAÍS VASCO.
INSERT INTO CHNE.provincia
SELECT (SELECT TO_CHAR(NVL(MAX(TO_NUMBER(codigo)), 0) + 1)
        FROM CHNE.provincia),
        'Álava',
        codigo
FROM CHNE.ca
WHERE nombre LIKE '%VASCO%';

COMMIT;

END IF;

```

Este proceso, en principio sólo debería lanzarse una vez, salvo que se vuelva a cargar el fichero “embales.csv” y contenga nuevas provincias o cambios en éstas. En cualquier caso, si se lanza varias veces, no da error, ya que sólo inserta en caso de que no existan registros en la tabla “CHNE.provincia”. La consulta `SELECT TO_CHAR(NVL(MAX(TO_NUMBER(codigo)), 0) + 1) FROM CHNE.provincia` presente en cada una de las inserciones anteriores, se utiliza para seleccionar el identificador del registro a insertar (similar al NEXTVAL de una secuencia). Sin embargo, para simplificar el proceso ETL, será llamado cada vez que se ejecute el procedimiento principal *pSituacion.ETL*.

- **CHNE.pEmbalse.inserta.** Este procedimiento, inserta los datos de los embalses procedentes de las tablas “CHNE.embalse\_tmp” y “CHNE.situacion\_tmp” **que no se hayan insertado ya**, en la tabla “CHNE.embalse”, corrigiendo los errores de concordancia y transcripción detectados entre ambas tablas:

```

PROCEDURE inserta IS
BEGIN
    INSERT
    INTO CHNE.embalse(nombre, rio, provincia)
    SELECT DISTINCT
        e.embalse, e.rio, e.provincia
    FROM CHNE.embalse_tmp t,
        (SELECT DISTINCT

```

```

        DECODE(e.embalse,
                'EL GRADO I', 'GRADO (EL)',
                'ESTANCA DE ALCAÑIZ', 'ALCAÑIZ (ESTANCA)',
                'GONZÁLEZ LACASA', 'GONZALEZ LACASA',
                'JOAQUÍN COSTA (BARASONA)', 'BARASONA',
                'LA SOTONERA', 'SOTONERA',
                'LAS TORCAS', 'TORCAS (LAS)',
                'MONTEAGUDO DE LAS VICARÍAS', 'MONTEAGUDO',
                'SANTA Mª BELSUÉ', 'S.MARÍA DE BELSUÉ',
                'TRANQUERA', 'TRANQUERA (LA)',
                e.embalse) AS embalse,
        r.codigo AS rio,
        DECODE(e.rio,
                'CALDERÉS', 'CALDARÉS',
                'ESERA', 'ÉSERA',
                'FLAMISEL', 'FLAMISELL',
                'GUADALOPE (DERIVACIÓN)', 'GUADALOPE',
                'NOGUERA-RIBAGORZANA', 'NOG. RIBAGORZANA',
                e.rio) AS nombre,
        e.provincia
FROM CHNE.rio r,
     (SELECT UPPER(e.embalse) AS embalse,
          UPPER(e.rio) AS rio, p.codigo AS provincia
      FROM CHNE.embalse_tmp e,
           CHNE.provincia p
      WHERE e.ca = p.nombre
            AND e.embalse IS NOT NULL
      UNION
      SELECT s.embalse, s.rio, p.codigo AS provincia
      FROM CHNE.situacion_tmp s,
           CHNE.embalse_tmp e,
           CHNE.provincia p
      WHERE DECODE(s.embalse,
                    'GRADO (EL)', 'EL GRADO I',
                    'ALCAÑIZ (ESTANCA)', 'ESTANCA DE ALCAÑIZ',
                    'GONZALEZ LACASA', 'GONZÁLEZ LACASA',
                    'BARASONA', 'JOAQUÍN COSTA (BARASONA)',
                    'SOTONERA', 'LA SOTONERA',
                    'TORCAS (LAS)', 'LAS TORCAS',
                    'MONTEAGUDO', 'MONTEAGUDO DE LAS VICARÍAS',
                    'S.MARÍA DE BELSUÉ', 'SANTA Mª BELSUÉ',
                    'TRANQUERA (LA)', 'TRANQUERA',
                    s.embalse) = UPPER(e.embalse(+))
            AND e.ca = p.nombre(+)) e
WHERE DECODE(e.rio,
             'CALDERÉS', 'CALDARÉS',
             'ESERA', 'ÉSERA',
             'FLAMISEL', 'FLAMISELL',
             'GUADALOPE (DERIVACIÓN)', 'GUADALOPE',
             'NOGUERA-RIBAGORZANA', 'NOG. RIBAGORZANA',
             e.rio) = r.nombre) e
WHERE e.embalse = DECODE(UPPER(t.embalse(+)),
                         'EL GRADO I', 'GRADO (EL)',
                         'ESTANCA DE ALCAÑIZ', 'ALCAÑIZ (ESTANCA)',
                         'GONZÁLEZ LACASA', 'GONZALEZ LACASA',
                         'JOAQUÍN COSTA (BARASONA)', 'BARASONA',
                         'LA SOTONERA', 'SOTONERA',
                         'LAS TORCAS', 'TORCAS (LAS)',
                         'MONTEAGUDO DE LAS VICARÍAS', 'MONTEAGUDO',
                         'SANTA Mª BELSUÉ', 'S.MARÍA DE BELSUÉ',
                         'TRANQUERA', 'TRANQUERA (LA)',
                         UPPER(t.embalse(+)))
      AND e.nombre = UPPER(t.rio(+))
      AND e.embalse NOT IN (SELECT nombre FROM CHNE.embalse)
ORDER BY e.embalse, e.rio;

COMMIT;

END inserta;

```



Este proceso, en principio sólo debería lanzarse una vez, salvo que se vuelva a cargar el fichero “embales.csv” y contenga nuevos embalses o cambios en éstos. En cualquier caso, si se lanza varias veces, no da error, ya que sólo inserta aquellos registros que no se hayan insertado ya en la tabla “CHNE.embalse” que en caso de lanzarse más de una vez, deben ser ninguno. Sin embargo, para simplificar el proceso ETL, será llamado cada vez que se ejecute el procedimiento principal **pSituacion.ETL**.

- **CHNE.pEmbalse.actualiza.** Este procedimiento, actualiza la provincia de los embalses con provincia aún sin definir (NULO) y la capacidad de los embalses insertados en el procedimiento anterior:

```

PROCEDURE actualiza IS

    contador NUMBER := 0;

BEGIN

    SELECT COUNT(*)
    INTO contador
    FROM CHNE.situacion_tmp;

    IF contador > 0 THEN
        -- Embalses de Álava (PAÍS VASCO) sin provincia aún definida.
        UPDATE CHNE.embalse
            SET provincia = (SELECT codigo
                            FROM CHNE.provincia
                            WHERE UPPER(nombre) = 'ÁLAVA')
        WHERE UPPER(nombre) IN ('ALBINA', 'ULLIVARRI', 'URRUNAGA', 'SOBRÓN');

        -- Embalses de Huesca (ARAGÓN) sin provincia aún definida.
        UPDATE CHNE.embalse
            SET provincia = (SELECT codigo
                            FROM CHNE.provincia
                            WHERE UPPER(nombre) = 'HUESCA')
        WHERE UPPER(nombre) IN ('BACHIMAÑA (LAGO)', 'BRAMATUERO ALTO',
                                'CANELLES', 'ESCALES', 'ESCARRA', 'IP',
                                'LLAUSER', 'PEÑA (LA)', 'RESPOMUSO', 'URDICETO');

        -- Embalses de Zaragoza (ARAGÓN) sin provincia aún definida.
        UPDATE CHNE.embalse
            SET provincia = (SELECT codigo
                            FROM CHNE.provincia
                            WHERE UPPER(nombre) = 'ZARAGOZA')
        WHERE UPPER(nombre) IN ('MEQUINENZA', 'SAN BARTOLOMÉ');

        -- Embalses de Lérida (CATALUÑA) sin provincia aún definida.
        UPDATE CHNE.embalse
            SET provincia = (SELECT codigo
                            FROM CHNE.provincia
                            WHERE UPPER(nombre) = 'LÉRIDA')
        WHERE UPPER(nombre) IN ('BASERCA', 'CAMARASA', 'CAVALLERS', 'CERTESCÁNS',
                                'MAR (LAGO)', 'NEGRO', 'IP', 'SABURO (LAGO)',
                                'TERRADETS', 'TREMP O TALÀRN', 'TORT (LAGO)');

        -- Embalses de Tarragona (CATALUÑA) sin provincia aún definida.
        UPDATE CHNE.embalse
            SET provincia = (SELECT codigo
                            FROM CHNE.provincia
                            WHERE UPPER(nombre) = 'TARRAGONA')
        WHERE UPPER(nombre) IN ('CIURANA', 'RIBARROJA');

        -- Embalses de Navarra (NAVARRA) sin provincia aún definida.
        UPDATE CHNE.embalse
            SET provincia = (SELECT codigo
                            FROM CHNE.provincia
    
```

```

        WHERE UPPER(nombre) = 'NAVARRA')
WHERE UPPER(nombre) IN ('IRABIA');

-- Actualización de la capacidad total de los embalses.
UPDATE CHNE.embalse e
  SET capacidad = (SELECT s.capacidadtotal
                   FROM CHNE.situacion_tmp s
                   WHERE e.nombre = s.embalse);

COMMIT;

END IF;

END actualiza;

```

Este proceso, en principio sólo debería lanzarse una vez, puesto que la capacidad total de los embalses permanece constante en todas las hojas de cálculo de situación de los embalses y ríos. Si no existieran registros en la tabla “CHNE.situacion\_tmp”, el procedimiento no hará nada. Sin embargo, para simplificar el proceso ETL será llamado cada vez que se ejecute el procedimiento principal ***pSituacion.ETL***.

- ***CHNE.pNieve.inserta***. Este procedimiento, inserta los datos de las estimaciones de nieve de los embalses para los años proporcionados en el cuadro correspondiente de los requerimientos del cliente:

```

PROCEDURE inserta IS
  contador NUMBER := 0;

BEGIN

  SELECT COUNT(*)
  INTO contador
  FROM CHNE.nieve;

  IF contador <= 0 THEN

    -- Embalse MAIDEVERA.
    INSERT INTO CHNE.nieve(embalse, anio, estimacion)
    SELECT codigo, '2001', 5
      FROM CHNE.embalse
     WHERE nombre = 'MAIDEVERA';

    INSERT INTO CHNE.nieve(embalse, anio, estimacion)
    SELECT codigo, '2002', 4
      FROM CHNE.embalse
     WHERE nombre = 'MAIDEVERA';

    INSERT INTO CHNE.nieve(embalse, anio, estimacion)
    SELECT codigo, '2003', 4
      FROM CHNE.embalse
     WHERE nombre = 'MAIDEVERA';

    INSERT INTO CHNE.nieve(embalse, anio, estimacion)
    SELECT codigo, '2004', 5
      FROM CHNE.embalse
     WHERE nombre = 'MAIDEVERA';

    INSERT INTO CHNE.nieve(embalse, anio, estimacion)
    SELECT codigo, '2005', 3
      FROM CHNE.embalse
     WHERE nombre = 'MAIDEVERA';

    -- Embalse SANTA ANA.
    INSERT INTO CHNE.nieve(embalse, anio, estimacion)

```

```
SELECT codigo, '2001', 4
  FROM CHNE.embalse
 WHERE nombre = 'SANTA ANA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2002', 4
  FROM CHNE.embalse
 WHERE nombre = 'SANTA ANA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2003', 4
  FROM CHNE.embalse
 WHERE nombre = 'SANTA ANA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2004', 3
  FROM CHNE.embalse
 WHERE nombre = 'SANTA ANA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2005', 3
  FROM CHNE.embalse
 WHERE nombre = 'SANTA ANA';

-- Embalse BÚBAL.
INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2001', 4
  FROM CHNE.embalse
 WHERE nombre = 'BÚBAL';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2002', 3
  FROM CHNE.embalse
 WHERE nombre = 'BÚBAL';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2003', 3
  FROM CHNE.embalse
 WHERE nombre = 'BÚBAL';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2004', 2
  FROM CHNE.embalse
 WHERE nombre = 'BÚBAL';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2005', 3
  FROM CHNE.embalse
 WHERE nombre = 'BÚBAL';

-- Embalse VALDABRA.
INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2001', 4
  FROM CHNE.embalse
 WHERE nombre = 'VALDABRA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2002', 5
  FROM CHNE.embalse
 WHERE nombre = 'VALDABRA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2003', 5
  FROM CHNE.embalse
 WHERE nombre = 'VALDABRA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2004', 2
  FROM CHNE.embalse
 WHERE nombre = 'VALDABRA';
```

```
INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2005', 2
  FROM CHNE.embalse
 WHERE nombre = 'VALDABRA';

-- Embalse CALANDA.
INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2001', 3
  FROM CHNE.embalse
 WHERE nombre = 'CALANDA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2002', 3
  FROM CHNE.embalse
 WHERE nombre = 'CALANDA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2003', 3
  FROM CHNE.embalse
 WHERE nombre = 'CALANDA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2004', 3
  FROM CHNE.embalse
 WHERE nombre = 'CALANDA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2005', 3
  FROM CHNE.embalse
 WHERE nombre = 'CALANDA';

-- Embalse MANSILLA.
INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2001', 3
  FROM CHNE.embalse
 WHERE nombre = 'MANSILLA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2002', 2
  FROM CHNE.embalse
 WHERE nombre = 'MANSILLA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2003', 1
  FROM CHNE.embalse
 WHERE nombre = 'MANSILLA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2004', 2
  FROM CHNE.embalse
 WHERE nombre = 'MANSILLA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2005', 2
  FROM CHNE.embalse
 WHERE nombre = 'MANSILLA';

-- Embalse MONTEAGUDO.
INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2001', 4
  FROM CHNE.embalse
 WHERE nombre = 'MONTEAGUDO';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2002', 5
  FROM CHNE.embalse
 WHERE nombre = 'MONTEAGUDO';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2003', 5
```

```

        FROM CHNE.embalse
        WHERE nombre = 'MONTEAGUDO';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2004', 5
        FROM CHNE.embalse
        WHERE nombre = 'MONTEAGUDO';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2005', 5
        FROM CHNE.embalse
        WHERE nombre = 'MONTEAGUDO';

-- Embalse OLIANA.
INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2001', 2
        FROM CHNE.embalse
        WHERE nombre = 'OLIANA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2002', 1
        FROM CHNE.embalse
        WHERE nombre = 'OLIANA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2003', 1
        FROM CHNE.embalse
        WHERE nombre = 'OLIANA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2004', 1
        FROM CHNE.embalse
        WHERE nombre = 'OLIANA';

INSERT INTO CHNE.nieve(embalse, anio, estimacion)
SELECT codigo, '2005', 1
        FROM CHNE.embalse
        WHERE nombre = 'OLIANA';

COMMIT;

END IF;

END inserta;

```

Este proceso, en principio sólo debería lanzarse una vez. Sin embargo, para simplificar el proceso ETL, será llamado cada vez que se ejecute el procedimiento principal ***pSituacion.ETL***.

- ***CHNE.pSituacion.inserta***. Este procedimiento, inserta los registros ya transformados procedentes de la tabla temporal “CHNE.situacion\_tmp” **que no se hayan insertado ya**, en la tabla de hechos definitiva “CHNE.situacion”:

```

PROCEDURE inserta IS
BEGIN
    INSERT
        INTO CHNE.situacion(fecha, embalse, capacidadtotal, capacidadactual,
                            porcentaje, entrada, salida, incremento,
                            capacidadanterior, porcentajeanterior)
    SELECT DISTINCT
        s.fecha, e.codigo, s.capacidadtotal, s.capacidadactual,
        s.porcentaje, s.entrada, s.salida, s.incremento,
        s.capacidadanterior, s.porcentajeanterior
        FROM CHNE.situacion_tmp s, CHNE.embalse e, CHNE.rio r

```

```

WHERE s.embalse = e.nombre
  AND e.rio = r.codigo
  AND r.nombre = DECODE(s.rio, 'CALDERÉS', 'CALDARÉS',
                        'FLAMISEL', 'FLAMISELL',
                        s.rio)
  AND (s.fecha, e.codigo) NOT IN (SELECT fecha, embalse
                                FROM CHNE.situacion);

COMMIT;

END inserta;

```

Este proceso, deberá lanzarse cada vez que se cargue un fichero de situación de los embalses y ríos en la tabla “CHNE.situacion\_tmp”.

- **CHNE.pNieve.actualiza.** Este procedimiento, actualiza la capacidad media de la tabla “CHNE.nieve”, necesaria para mostrar los posibles casos de desbordamiento del embalse:

```

PROCEDURE actualiza IS
BEGIN
  UPDATE CHNE.nieve n
    SET capacidadmedia = (SELECT AVG(s.capacidadactual)
                        FROM CHNE.situacion s
                        WHERE s.embalse = n.embalse
                        AND TO_NUMBER(TO_CHAR(s.fecha,
                                             'YYYY')) = n.anio);

  COMMIT;

END actualiza;

```

Este proceso, en principio sólo debería lanzarse una vez. Sin embargo, para simplificar el proceso ETL, será llamado cada vez que se ejecute el procedimiento principal **pSituacion.ETL**.

- **CHNE.pSituacion\_tmp.borra.** Este procedimiento, borra los registros de la tabla temporal “CHNE.situacion\_tmp”.

```

PROCEDURE borra IS
BEGIN
  DELETE FROM CHNE.situacion_tmp;

  COMMIT;

END borra;

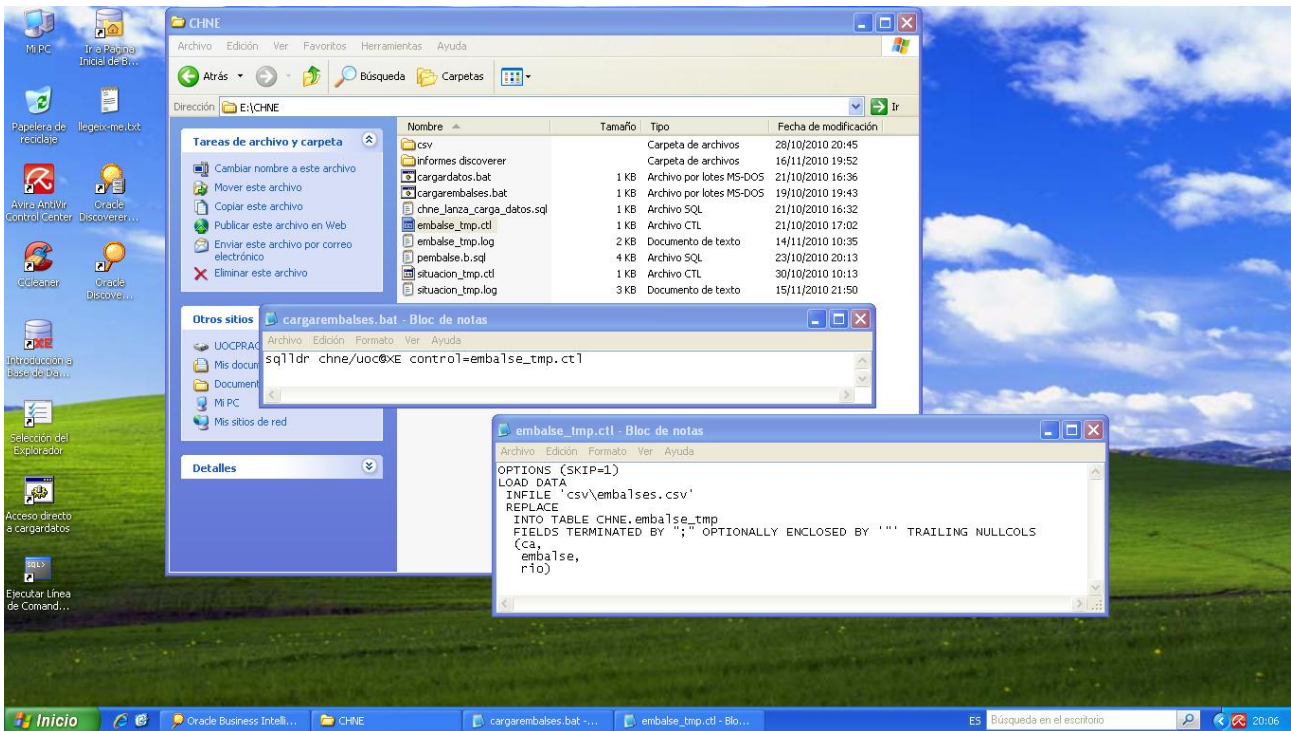
```

Este proceso, deberá lanzarse cada vez que se cargue un fichero de situación de los embalses y ríos en la tabla “CHNE.situacion\_tmp”, una vez finalizada la inserción de los datos transformados, en la tabla definitiva “CHNE.situacion”.

## 7. Proceso de carga

## 7.1 Carga de los embalses

Se ha creado el ejecutable **cargarembalses.bat** en la ruta E:\CHNE de la máquina virtual proporcionada por el cliente para realizar este proceso de manera automática. Tras la ejecución, se genera el fichero “*embalse\_tmp.log*” en el que se indica si la carga de los embalses en la base de datos ha transcurrido correctamente. En caso de que algunos registros no se hayan podido insertar, se creará de manera automática el fichero “*embalse\_tmp.bad*” con los registros fallidos.



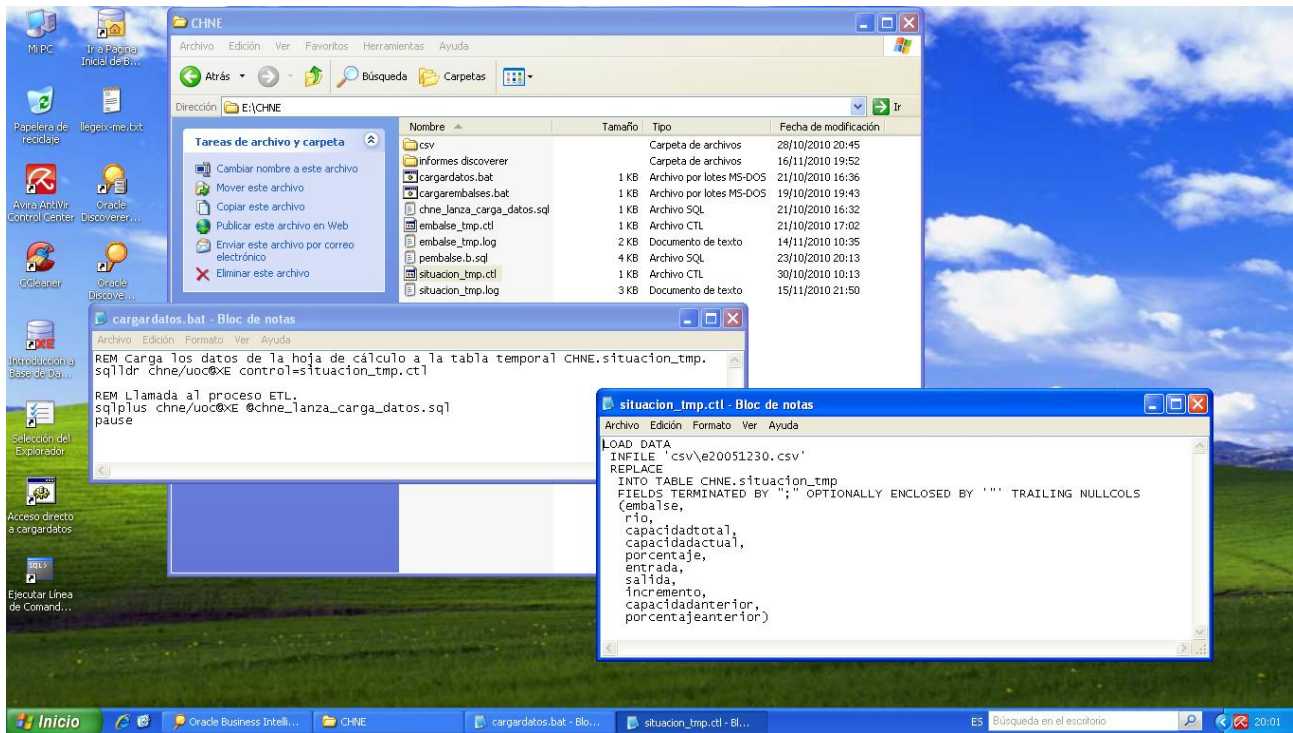
## 7.2 Carga de los datos de situaciones de los embalses y ríos

Se ha creado el ejecutable **cargardatos.bat** en la ruta E:\CHNE de la máquina virtual proporcionada por el cliente para realizar este proceso de manera automática. Tras la ejecución, se genera el fichero “*situacion\_tmp.log*” en el que se indica si la carga de los registros en la base de datos ha transcurrido correctamente. En caso de que algunos registros no se hayan podido insertar, se creará de manera automática el fichero “*situacion\_tmp.bad*” con los registros fallidos.

Este ejecutable, además del proceso de carga de datos de las situaciones de los embalses procedentes cada una de las 54 hojas de cálculo en la tabla “CHNE.situación\_tmp”, lanza todo el proceso de transformación de datos que se comenta más abajo, tal y como indica la línea de código: `sqlplus chne/uoc@XE @chne_lanza_carga_datos.sql`, donde el fichero `chne_lanza_carga_datos.sql` contiene la llamada al procedimiento almacenado que lanza todo el proceso:

```
BEGIN
  CHNE.pSituacion.ETL;
END;
/
QUIT
```

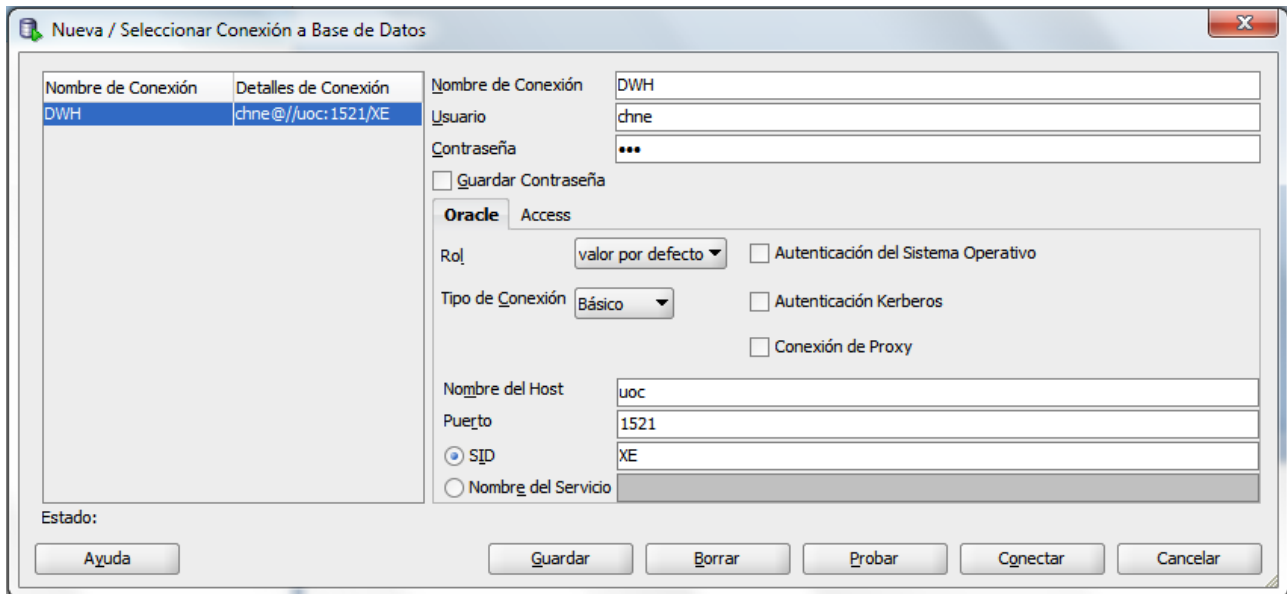
Por lo tanto, cada vez que se quiera cargar una nueva hoja de cálculo (transformada a formato csv) con los datos de las situaciones de los embalses y ríos, se debe cambiar en el archivo `situacion_tmp.ctl`, en la línea `INFILE 'csv\YYYYMMDD.csv'` el nombre del fichero que contiene los datos a cargar. En caso de que se ejecute varias veces **cargardatos.bat** para la misma hoja de cálculo, no se produce error puesto que sólo se insertarán los registros que no estén ya en la base de datos.



## 8. Acceso a la base de datos

Para montar el almacén de datos, se ha creado utilizando el usuario *SYSTEM* y contraseña proporcionada por el cliente (CHNE), el usuario/esquema de base de datos *CHNE* con contraseña "uoc" cuyo tablespace por defecto es el *USERS* y tablespace temporal *TEMP*. A este usuario se le han otorgado los roles "CONNECT" y "RESOURCE". Para acceder a la base de datos de dicho almacén, la cadena de conexión es `chne@//uoc:1521/XE`, tal y como se muestra en la siguiente captura de pantalla:





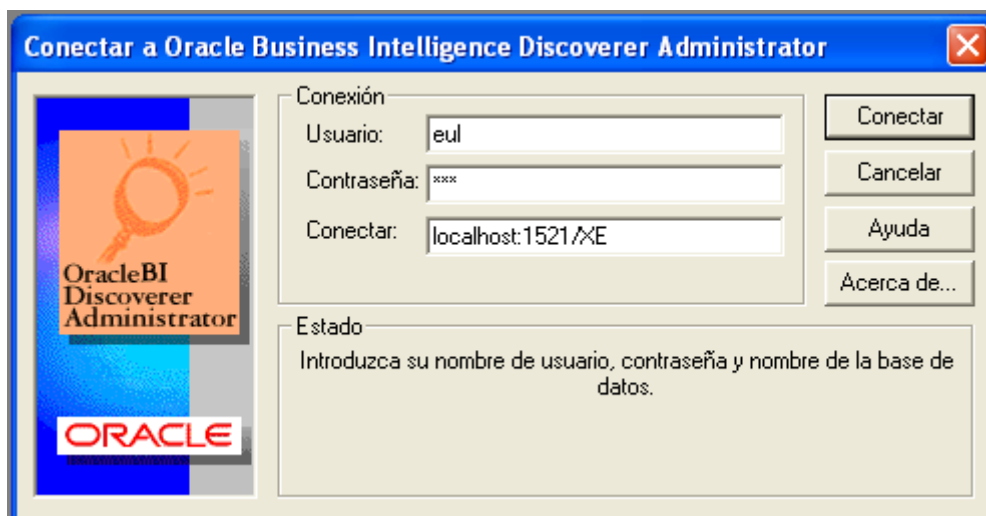
Una vez conectado, ya se puede acceder a todos los objetos que componen el modelo físico del almacén de datos (tablas, índices, secuencias, disparadores, paquetes, procedimientos almacenados, etc.).

## 9. Acceso a la herramienta de Bussines Intelligence

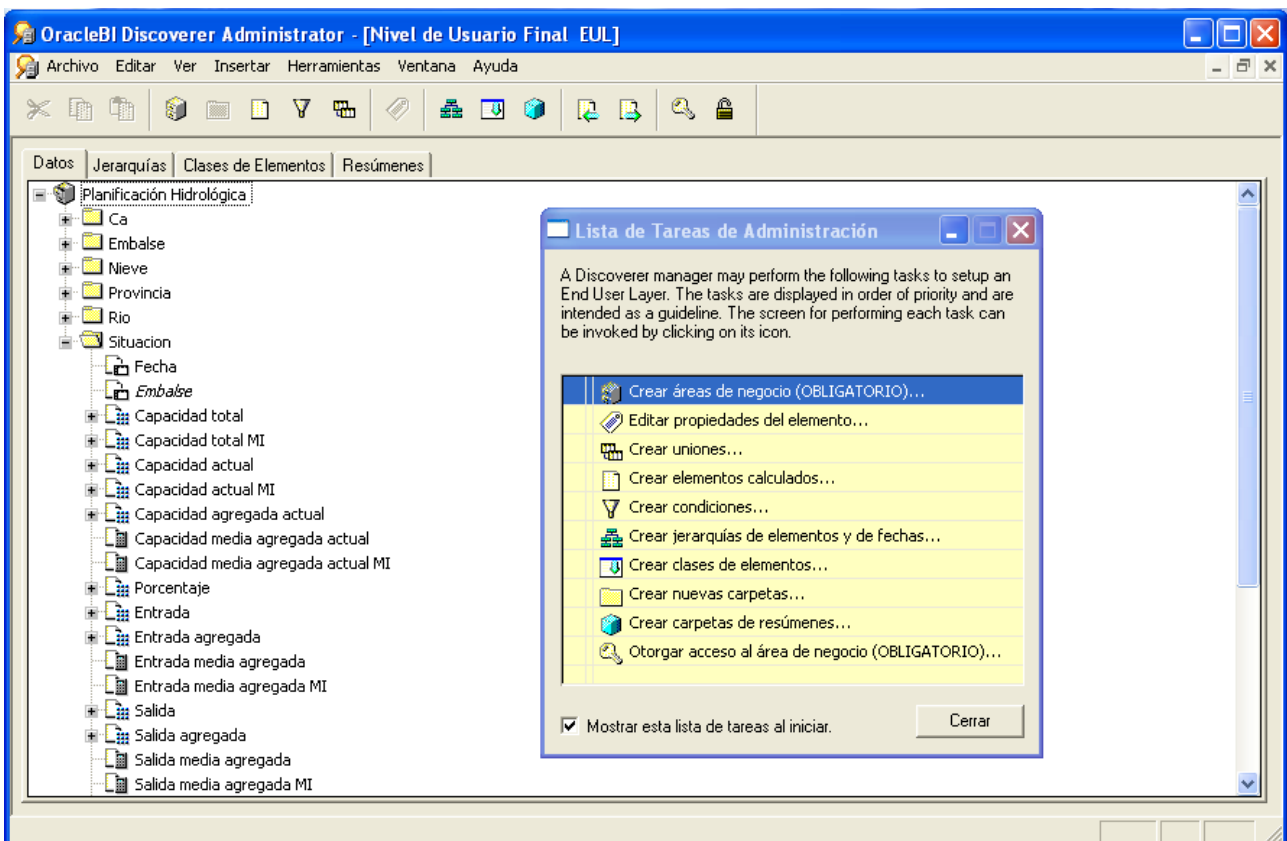
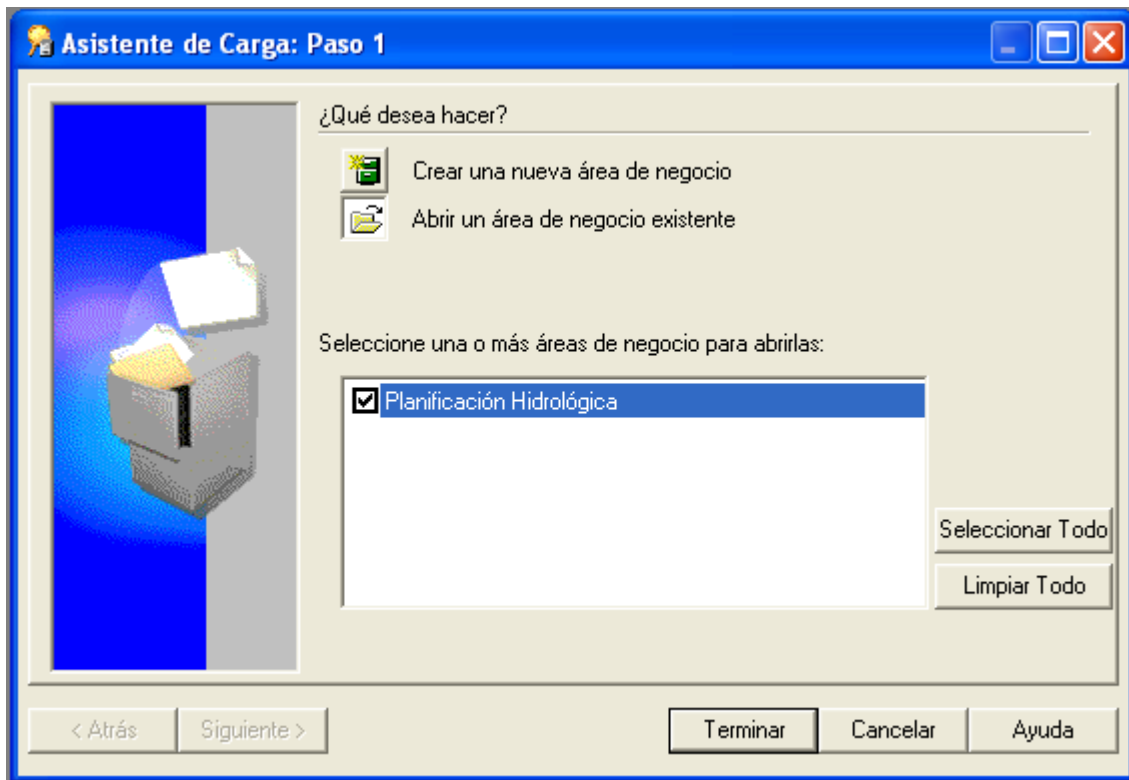
### 9.1 Acceso a Oracle Discoverer Administrator

Para crear la capa final de usuario (EUL), se ha creado el usuario/esquema de bases de datos *EUL* con contraseña “uoc” cuyo tablespace por defecto es el *USERS* y tablespace temporal *TEMP*. A este usuario se le han otorgado los roles “CONNECT” y “RESOURCE”.

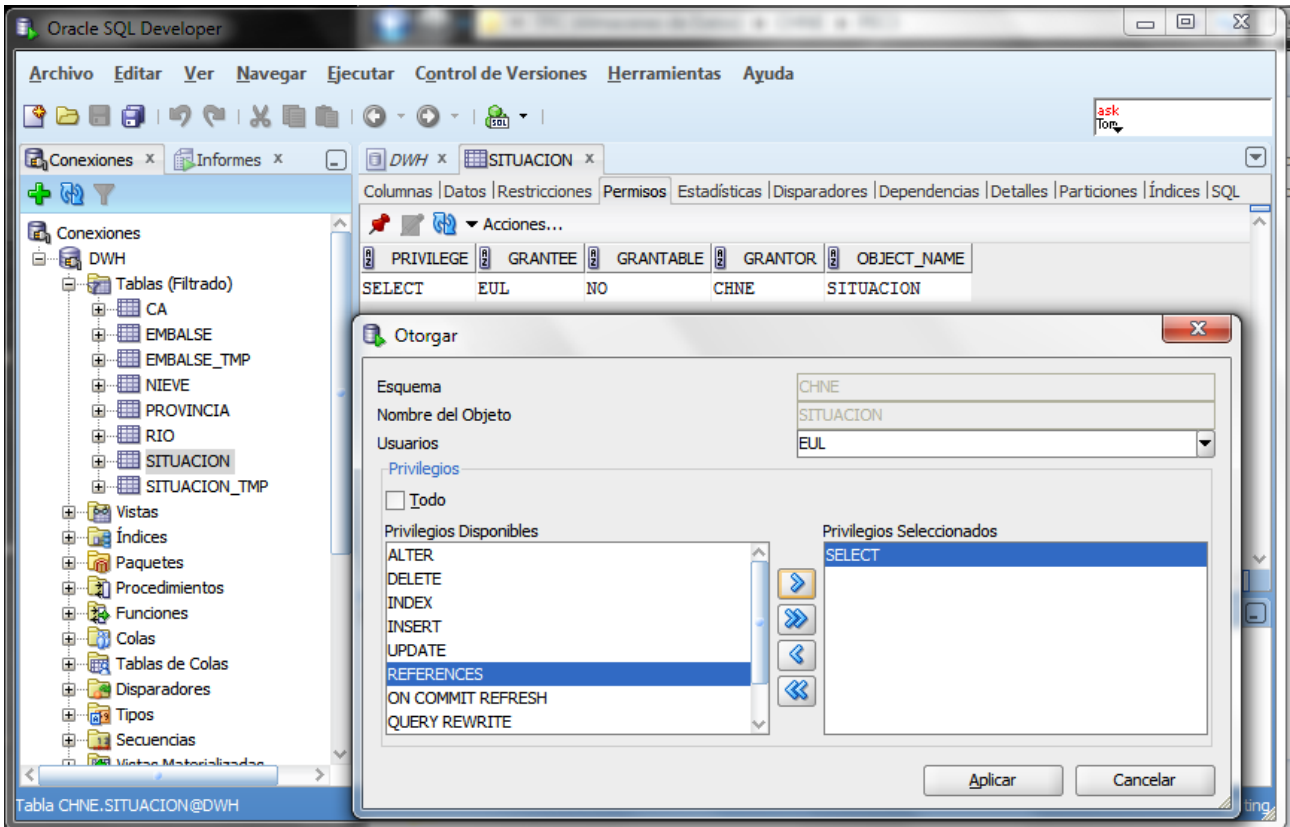
Este usuario se utiliza única y exclusivamente para conectar a Oracle Discoverer Administrator y gestionar las distintas áreas de negocio, o crear nuevas.



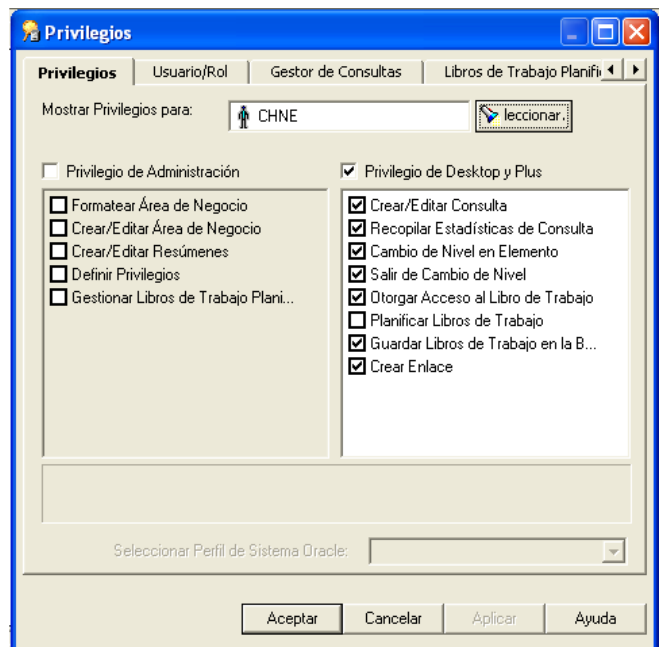
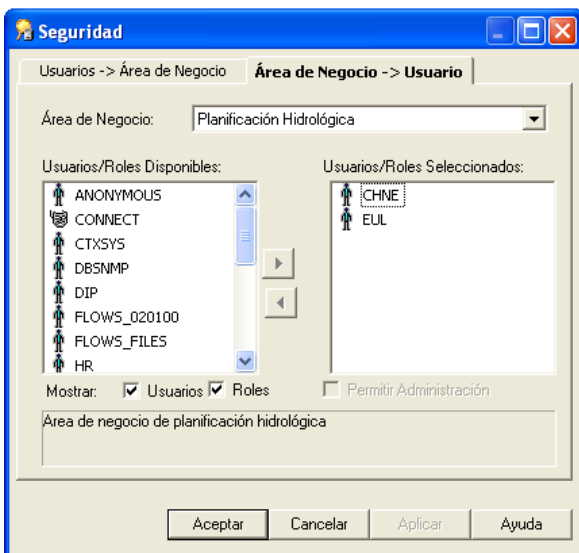
En el caso de este proyecto, se ha creado el área de negocio “Planificación hidrológica” donde se ha definido todo el modelo de explotación del almacén de datos.



Para ello, se ha tenido que conceder al usuario *EUL* el privilegio “SELECT” sobre las tablas CHNE.ca, CHNE.embalse, CHNE.nieve, CHNE.provincia, CHNE.rio y CHNE.situacion, ya que el área de negocio de “Planificación hidrológica” utiliza la información de dichas tablas, tal y como se muestra a continuación:

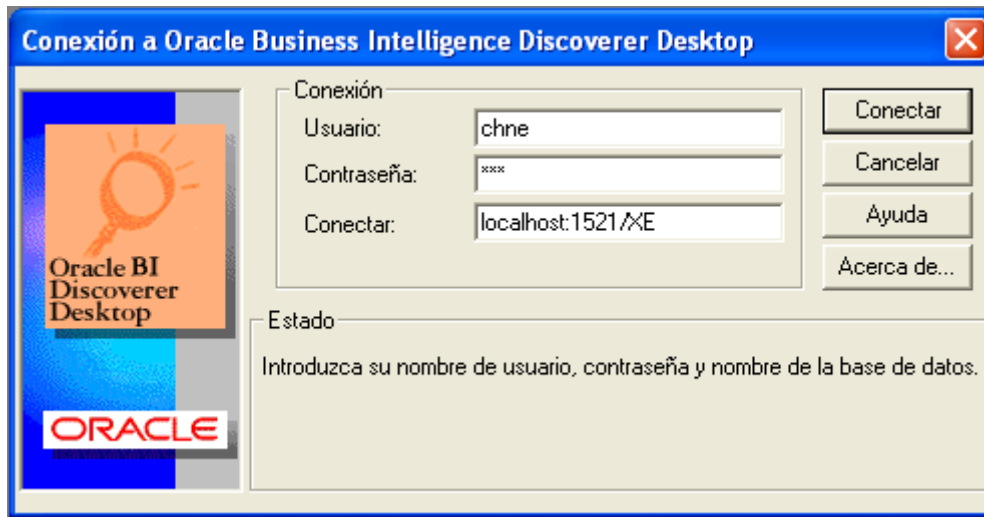


Por último, se ha otorgado permiso de acceso al área de negocio de “Planificación hidrológica” al usuario *CHNE* con los privilegios que se muestran en la captura de pantalla siguiente:

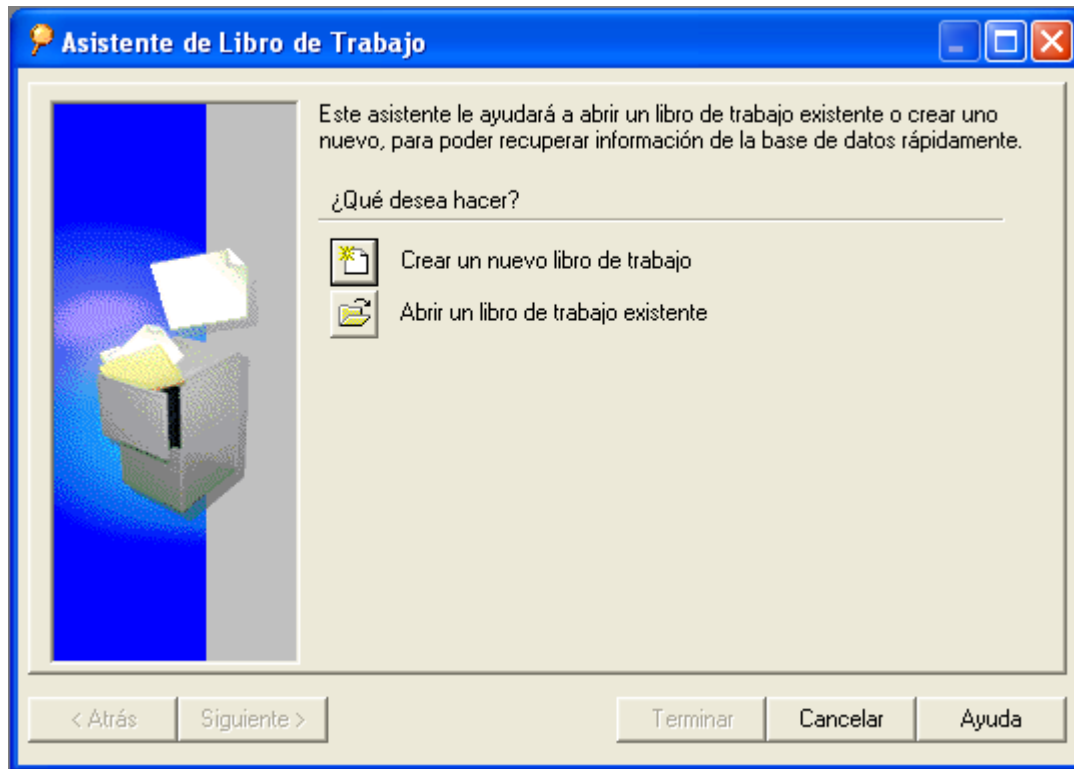


## 9.2 Acceso a Oracle Discoverer Desktop

Para acceder a Oracle Discoverer Desktop, se utilizará el usuario *CHNE*:



Una vez conectado, ya se podrá crear un libro de trabajo nuevo o bien abrir uno ya existente:

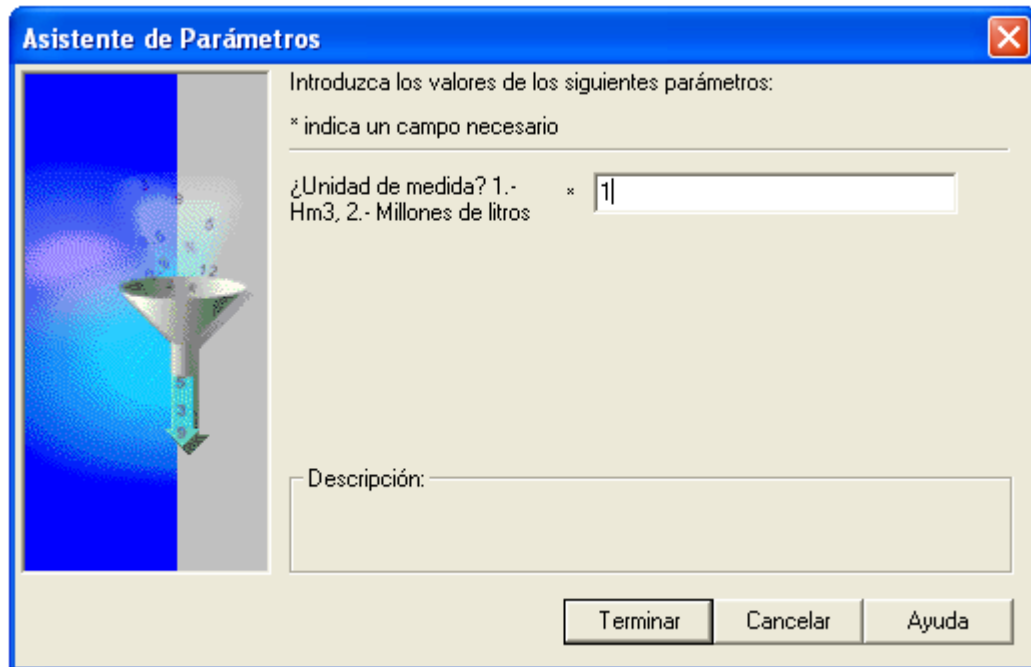



### 9.3 Tabla de usuarios y contraseñas

Usuario	Contraseña	Observaciones
system	uoc	Usuario administrador de la bbdd
eul	uoc	Usuario Oracle Discoverer Administrator
chne	uoc	Usuario bbdd y Oracle Discoverer Desktop

## 10. Captura de pantalla de los informes

Al iniciar cualquiera de los informes del proyecto, se le muestra al usuario la siguiente pantalla para que elija el volumen en el que desea ver los datos del informe:



1.- Hm<sup>3</sup>, 2.- Millones de litros, cualquier otro valor y valor por defecto: 1.- Hm<sup>3</sup>. Si se desea volver a seleccionar el volumen de los datos del informe, habrá que hacer clic en el botón . **En las siguientes capturas de pantallas se mostrarán los informes en millones de litros.**

### 10.1 Capacidad total vs capacidad actual por río y año-mes

Oracle Business Intelligence Discoverer Desktop - [1.- Capacidad total vs capacidad actual en Hm3 por río y año-mes]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 11 B i U

Elementos de Página: Año: 2001 Mes: Jul

Río	Capacidad total	Capacidad actual
▶ AGUAS LIMPIAS	18000000000	16000000000
▶ AGUAS VIVAS	8000000000	1000000000
▶ ALBERCOS	33000000000	18000000000
▶ ALBINA	6000000000	4000000000
▶ ARAGÓN	47000000000	22300000000
▶ ARANDA	21000000000	14000000000
▶ ARBA	6000000000	0
▶ ARGA	21000000000	17000000000
▶ ASMAT	10000000000	3000000000
▶ BARRANCO DE IP	5000000000	5000000000
▶ CALDARÉS	12000000000	10000000000
▶ CERTESCÁNS	16000000000	12000000000
▶ CINCA	82900000000	71300000000
▶ CIURANA	12000000000	4000000000
▶ EBRO	23040000000	18060000000
▶ ESCARRA	5000000000	2000000000
▶ ÉSERA	92000000000	65000000000
▶ FLAMISELL	32000000000	11000000000
▶ FLUMEN	13000000000	2000000000
▶ GÁLLEGO	96000000000	69000000000
▶ GUADALOPE	19700000000	14300000000
▶ GUADALUPE	19700000000	14300000000

### 10.2 Capacidad total vs capacidad actual para embalse y año-mes

Oracle Business Intelligence Discoverer Desktop - [2.- Capacidad total vs capacidad actual en Hm3 para embalse y año-mes]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 11 B U

Elementos de Página: Año: 2001 Mes: Jul

	Capacidad total	Capacidad actual
Embalse		
▶ ALBINA	6000000000	4000000000
▶ ALCAÑIZ (ESTANCA)	7000000000	6000000000
▶ ALLOZ	65000000000	37000000000
▶ BACHIMAÑA (LAGO)	7000000000	6000000000
▶ BARASONA	92000000000	65000000000
▶ BASERCA	22000000000	20000000000
▶ BRAMATUERO ALTO	5000000000	4000000000
▶ BÚBAL	64000000000	39000000000
▶ CALANDA	54000000000	36000000000
▶ CAMARASA	113000000000	64000000000
▶ CANELLES	688000000000	560000000000
▶ CASPE	82000000000	63000000000
▶ CAVALLERS	16000000000	13000000000
▶ CERTESCÁNS	16000000000	12000000000
▶ CIURANA	12000000000	4000000000
▶ CUEVA FORADADA	29000000000	2000000000
▶ EBRO	540000000000	353000000000
▶ ESCALES	152000000000	137000000000
▶ ESCARRA	5000000000	2000000000
▶ EUGUI	21000000000	17000000000
▶ GONZALEZ LACASA	33000000000	18000000000

### 10.3 Capacidad total vs capacidad actual por comunidad autónoma y año-mes

Oracle Business Intelligence Discoverer Desktop - [3.- Capacidad total vs capacidad actual en Hm3 por comunidad autonoma y año-mes]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 11 B U

Elementos de Página: Año: 2001 Mes: Jul

	Capacidad total	Capacidad actual
C.A.		
▶ ARAGÓN	4268000000000	3302000000000
▶ CANTABRIA	540000000000	353000000000
▶ CASTILLA LEON	9000000000	6000000000
▶ CATALUÑA	771000000000	588000000000
▶ LA RIOJA	101000000000	54000000000
▶ NAVARRA	570000000000	277000000000
▶ PAÍS VASCO	245000000000	186000000000





### 10.6 Aportación (Entrada-Salida) por río, para embalse y año-mes; comparativa con la media anual de aportación del embalse

Oracle Business Intelligence Discoverer Desktop - [6.- Aportación (Entrada-Salida) en Hm3 por río, para embalse y año-mes.DIS]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 11 B U

Aportación (Entrada-Salida) por río, para embalse y año-mes; comparativa con la media anual de aportación del embalse

Elementos de Página: Año: 2001 Mes: Jul

Río	Embalse	Aportación	Aportación media
> AGUAS LIMPIAS		0	0
> RESPOMUSO		0	0
> AGUAS VIVAS		0	0
> MONEVA		0	0
> ALBERCOS		-2200000000	-2200000000
> GONZALEZ LACASA		-2200000000	-2200000000
> ALBINA		0	0
> ALBINA		0	0
> ARAGÓN		-26300000000	-26300000000
> YESA		-26300000000	-26300000000
> ARANDA		-11000000000	-11000000000
> MAIDEVERA		-11000000000	-11000000000
> ARBA		0	0
> SAN BARTOLOMÉ		0	0
> ARGA		-5000000000	-5000000000
> EUGUI		-5000000000	-5000000000
> ASMAT		0	0
> GUIAMETS		0	0
> BARRANCO DE IP		0	0
> IP		0	0
> CALDARÉS		0	0

### 10.7 Entrada vs Salida media agregada anual por comunidad autónoma

Oracle Business Intelligence Discoverer Desktop - [7.- Entrada vs Salida media agregada anual por comunidad autónoma]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 11 B U

Entrada vs Salida media agregada anual por comunidad autónoma

Elementos de Página: Año: 2001

C.A.	Entrada media agregada	Salida media agregada
> C.A.		
> ARAGÓN	4,894444444444444444444444	5,336666666666666666666667
> CANTABRIA	2,166666666666666666666667	11,95
> CASTILLA LEON	0,016666666666666666666667	0,033333333333333333333333
> CATALUÑA	7,531111111111111111111111	7,943333333333333333333333
> LA RIOJA	0,666666666666666666666667	1,65
> NAVARRA	2,225	4,258333333333333333333333
> PAÍS VASCO	0	0

### 10.8 Aportación (Entrada-Salida) por río, embalse y mes; comparativa con la media anual de aportación del embalse

Oracle Business Intelligence Discoverer Desktop - [8.- Aportación (Entrada-Salida) en Hm3 por río, embalse y mes]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 11 B U

Aportación (Entrada-Salida) por río, embalse y mes; comparativa con la media anual de aportación del embalse

Elementos de Página: Mes: Ene

Río	Embalse	Aportación	Aportación media anual
▶ AGUAS LIMPIAS		0	0,00
	RESPOMUSO	0	0,00
▶ AGUAS VIVAS		-100000000	-25000000,00
	MONEVA	-100000000	-25000000,00
▶ ALBERCOS		3100000000	775000000,00
	GONZALEZ LACASA	3100000000	775000000,00
▶ ALBINA		0	0,00
	ALBINA	0	0,00
▶ ARAGÓN		6700000000	1675000000,00
	YESA	6700000000	1675000000,00
▶ ARANDA		700000000	175000000,00
	MAIDEVERA	700000000	175000000,00
▶ ARBA		0	0,00
	SAN BARTOLOMÉ	0	0,00
▶ ARGA		900000000	225000000,00
	EUGUI	900000000	225000000,00
▶ ASMAT		-100000000	-25000000,00
	GUIAMETS	-100000000	-25000000,00
▶ BARRANCO DE IP		0	0,00
	IP	0	0,00
▶ CALDARÉS		0	0,00

### 10.9 Agua diaria en reserva disponible en toda la confederación hidrográfica por año-mes

Oracle Business Intelligence Discoverer Desktop - [9.- Agua diaria en reserva disponible en toda la confederación hidrográfica por año-mes]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

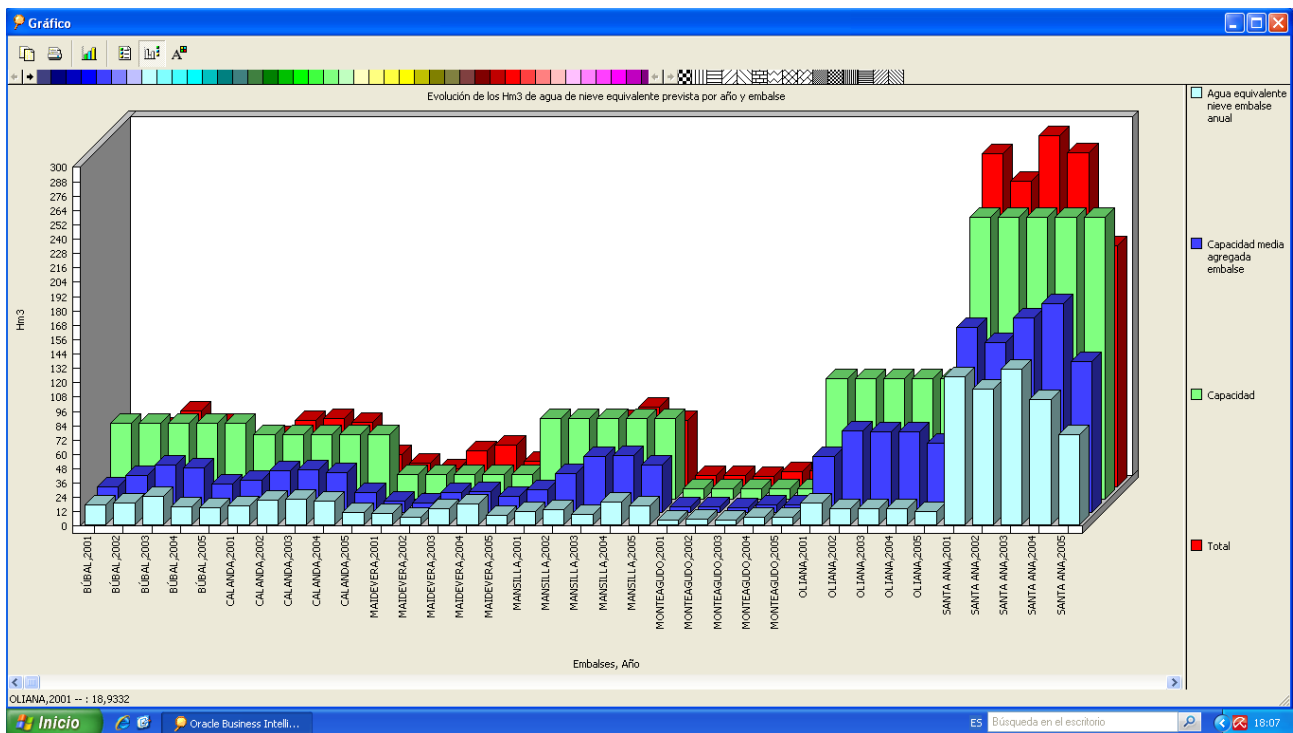
Tahoma 11 B U

Agua diaria en reserva disponible en toda la confederación hidrográfica por año-mes

Elementos de Página: Mes: Jul Embalse: <Todo>

Año	Reserva diaria
▶ 2001	153741935483,87096774193548387097
▶ 2002	127225806451,6129032258064516129
▶ 2003	146516129032,25806451612903225806
▶ 2004	162290322580,64516129032258064516
▶ 2005	119612903225,80645161290322580645

## 10.10 Gráfico de la evolución de los Hm<sup>3</sup> de agua de nieve equivalente prevista por año y embalse



Los **posibles** casos de desbordamiento son aquellos en los que la barra roja que representa el total de agua (suma del agua equivalente de nieve anual del embalse más **la capacidad media agregada del embalse**, de ahí que sea tan sólo una posibilidad) sobresalen por encima de la barra verde, que representa la capacidad total del embalse. En este gráfico los datos son mostrados en Hm<sup>3</sup>.

## 11. Predicción de periodos de sequía a partir del data mart desarrollado

Aunque el estado actual de la meteorología no permite predecir las sequías, si es cierto que se pueden mitigar. Una técnica que se utiliza para tratar de predecir periodos de sequía es el análisis de datos de años hidrológicos de los que se dispone.

Por lo tanto, se podría pensar que se aproximan periodos de sequía cuando la evolución del total de agua embalsada disminuye en el tiempo. Así, cuando se aprecie una disminución del agua en los embalses de una comunidad autónoma o si la estimación de nieve y la cantidad de agua embalsada escasea, es posible que se esté iniciando un periodo de sequía o quizás no. Para determinar entonces si se trata de uno de estos periodos tan perjudiciales, se pueden utilizar los informes ya desarrollados que se enumeran a continuación:

### 3.- Capacidad total vs capacidad actual por comunidad autónoma y año-mes:

Oracle Business Intelligence Discoverer Desktop - [3.- Capacidad total vs capacidad actual en Hm3 por comunidad autonoma y año-mes]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 11 B i U

Elementos de Página: Año: 2002 Mes: <Todo>

C.A.	Capacidad total	Capacidad actual
ARAGÓN	51216	34976
CANTABRIA	6480	2381
CASTILLA LEON	108	58
CATALUÑA	9252	6818
LA RIOJA	1212	567
NAVARRA	6840	2517
PAÍS VASCO	2940	1906

Este informe permite comparar la capacidad actual de los embalses de las distintas comunidades autónomas con su capacidad total (por año y/o mes), con lo que se puede observar el estado de los embalses en un momento determinado.

Oracle Business Intelligence Discoverer Desktop - [3.- Capacidad total vs capacidad actual en Hm3 por comunidad autonoma y año-mes]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 11 B i U

Elementos de Página: Año: 2002 Mes: <Todo>

C.A.	Provincia	Embalse	Capacidad total	Capacidad actual
ARAGÓN			51216	34976
	Huesca		28416	17908
	Teruel		1956	1261
	Zaragoza		20844	15807
CANTABRIA			6480	2381
	Cantabria		6480	2381
		EBRO	6480	2381
CASTILLA LEON			108	58
	Soria		108	58
CATALUÑA			9252	6818
	Lérida		6468	4304
	Tarragona		2784	2514
LA RIOJA			1212	567
	La Rioja		1212	567
NAVARRA			6840	2517
	Navarra		6840	2517
PAÍS VASCO			2940	1906
	Álava		2940	1906

También se pueden ir detallando los datos por provincias e incluso por embalses tal y como se muestra en la imagen anterior.

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

Oracle Business Intelligence Discoverer Desktop - [4. - Comparativa anual de la evolución por comunidad autónoma de la capacidad agregada de todos sus embalses

Elementos de Página:

C.A.	Capacidad					
	Año	2001	2002	2003	2004	2005
ARAGÓN		17483	34976	40330	39515	29249
CANTABRIA		1348	2381	3367	3502	4425
CASTILLA LEON		32	58	53	79	83
CATALUÑA		2894	6818	7024	6773	6348
LA RIOJA		172	567	828	861	716
NAVARRA		819	2517	4544	4244	3119
PAÍS VASCO		934	1906	2095	2209	2191

Este informe, permite comparar la evolución a través de los años, que sufre la capacidad de todos los embalses de una comunidad autónoma.

Oracle Business Intelligence Discoverer Desktop - [4. - Comparativa anual de la evolución por comunidad autónoma de la capacidad agregada de todos sus embalses

Elementos de Página:

C.A.	Provincia	Embalse	Capacidad					
			Año	2001	2002	2003	2004	2005
ARAGÓN				17483	34976	40330	39515	29249
	Huesca			9565	17908	21493	20902	13013
	Teruel			510	1261	1406	1440	959
	Zaragoza			7408	15807	17431	17173	15277
CANTABRIA				1348	2381	3367	3502	4425
	Cantabria			1348	2381	3367	3502	4425
CASTILLA LEON				32	58	53	79	83
	Soria			32	58	53	79	83
CATALUÑA				2894	6818	7024	6773	6348
	Lérida			1638	4304	4426	4109	3779
	Tarragona			1256	2514	2598	2664	2569
LA RIOJA				172	567	828	861	716
	La Rioja			172	567	828	861	716
		GONZALEZ LACASA		55	172	259	283	235
		MANSILLA		117	395	569	578	481
NAVARRA				819	2517	4544	4244	3119
	Navarra			819	2517	4544	4244	3119
PAÍS VASCO				934	1906	2095	2209	2191
	Álava			934	1906	2095	2209	2191

Al igual que ocurría con el informe anterior, se pueden ir detallando los datos por provincias y embalses tal y como se muestra en la imagen anterior.



fácil contrastar el agua de nieve prevista con la cantidad de agua del embalse para determinar si se trata o no de un periodo de sequía, simplemente se suma el agua equivalente de nieve del embalse y la capacidad media agregada del embalse, y esta suma se compara con la capacidad del embalse.

Por último, aclarar que el análisis de cualquiera de los informes desarrollados durante el proyecto puede servir para ayudar a predecir periodos de sequía, pero pensamos que los destacados en este apartado, pueden resultar especialmente útiles en dicha tarea.

## 12. Conclusiones

Tras la realización de este trabajo fin de carrera, queda latente la necesidad y el auge de los almacenes de datos y herramientas de bussiness intelligence, ya que permiten a las organizaciones extraer de sus sistemas de información, información funcional decisiva para cualquier área de negocio, sin la necesidad de conocimientos técnicos en tecnología de la información y con un coste temporal mínimo, aumentando la productividad y la eficiencia de sus procesos.

Esta es la principal conclusión y más fundamental que extraigo, tras la finalización de este trabajo de almacenes de datos, que ha dado un nuevo enfoque al conocimiento que tengo en materia de bases de datos.

## 13. Hardware y software utilizado

### 13.1 Hardware

Máquina virtual de VirtualBox (2 procesadores a 1,6 GHz, 512 Mb de RAM, 2 discos duros con 6 y 4 GB respectivamente, y Windows XP Profesional SP3) proporcionada por el cliente e instalada en un servidor con las siguientes especificaciones:

- Microprocesador Intel Core i7 i7-720QM a 1,6 GHz con 6 MB de caché de nivel 2.
- 4 GB de memoria RAM DDR3 a 1.333 MHz (2 x 2 GB SODIMM Hynix).
- Unidad de disco duro SATA de 640 GB (5400 rpm).
- Interfaz de red Realtek PCIe Gigabit Ethernet.

### 13.2 Software

- Sistema operativo Windows 7 Home Premium de 64 bits.
- Base de datos Oracle 10g Express Edition, preinstalada en la máquina virtual.

- Herramienta de Bussiness Intelligence Oracle Bussiness Intelligence Tools 10g (Discoverer Administrator y Discoverer Desktop), preinstalado en la máquina virtual.
- Paquete ofimático Microsoft Office Professional Plus 2010.
- Herramienta IDE (Entorno de desarrollo integrado) Oracle SQL Developer 2.1.1.64 (64 bits).
- Herramienta CASE (Ingeniería de Software asistida por computadora) UML Pad.
- Herramientas de edición de video Camtasia Studio 7 y WinFF.

## 14. Bibliografía

### 14.1 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.
- **Introducción a los Sistemas de Bases de Datos (Séptima edición).** C.J. Date.

### 14.2 Manuales

- **Oracle Database Administrator's Guide 10g Release 2 (10.2).** **B14231-01.** June 2005.
- **Oracle Database SQL Reference 10g Release 2 (10.2).** **B14200-01.** June 2005.
- **Oracle Database PL/SQL User's Guide and Reference 10g Release 2 (10.2).** **B14261-01.** June 2005.
- **Oracle Bussiness Intelligence Concepts Guide 10g Release 2 (10.1.2.1).** **B16378-01.** July 2005.
- **Oracle Bussines Intelligence Discoverer Administration Guide 10g Release 2 (10.1.2.1).** **B13916-04.** July 2005.
- **Oracle Bussines Intelligence Discoverer Desktop User's Guide 10g Release 2 (10.1.2.1) for Windows.** **B13917-03.** July 2005.



### ***14.3 Online***

- <http://www.inmoncif.com>
- <http://www.rkimball.com>
- <http://www.dwreview.com>
- <http://www.1keydata.com/datawarehousing/datawarehouse.html>
- <http://www.intranetjournal.com/features/datawarehousing.html>
- [http://es.wikipedia.org/wiki/Almac%C3%A9n\\_de\\_datos](http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos)
- <http://www.chebro.es>