
Memoria del Proyecto

Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales



Alumno:

Pablo Alayeto García

Estudiante de Ingeniería Técnica Informática de Gestión

Consultor:

Xavier Plaza Sierra

Fecha: 10/06/2012



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

RESUMEN

Este documento constituye la memoria del trabajo de fin de carrera (TFC) de Ingeniería Técnica Informática de Gestión, en el área de Almacenes de Datos. En la misma se describe el plan de trabajo, el análisis y la implementación de un almacén de datos para el análisis del sistema de prestaciones sociales.

El cliente que nos solicita la implementación del almacén de datos es el Organismo de Análisis del Departamento de Trabajo (OADT), y para ello nos facilita una serie de ficheros Excel con la información en bruto obtenida de varias fuentes de datos.

A partir de esta información inicial y una serie de requerimientos funcionales, diseñaremos el cubo OLAP que albergará estos datos y nos permitirá trabajar de manera eficiente y ordenada con la información, que posteriormente explotaremos con herramientas de *Bussiness Intelligence*, AtlasSBI en nuestro caso.

Una vez volcada la información y preparado el entorno de informes y estadísticas solicitadas, el cliente podrá realizar un estudio de la situación actual e histórica del sistema de prestaciones sociales, así como la evolución de éstas.

PALABRAS CLAVE

Almacén de datos, Data Warehouse, OLAP, Oracle, análisis multidimensional, ETL, SQL, PL/SQL, AtlasSBI, SQL*Loader.



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

TABLA DE CONTENIDOS

1. INTRODUCCIÓN	5
1.1 Justificación del TFC y contexto en el cual se desarrolla: Punto de partida y aportación del TFC.	5
1.2 Objetivos del proyecto	5
1.3 Enfoque y método seguido	6
1.4 Planificación del proyecto	7
1.5 Productos obtenidos	10
1.6 Breve descripción del resto de capítulos	11
2. ANÁLISIS	12
2.1 Análisis de datos de Origen	12
2.2 Diagramas de casos de uso	13
2.3 Diseño Conceptual	14
3. DISEÑO	17
3.1 Diagrama de la arquitectura de Software	17
3.2 Diagrama de la arquitectura de Hardware	17
3.3 Diseño Físico	18
3.4 Procesos ETL	21
3.5 Diseño y descripción de los informes creados	24
4. INFORMES PREPARADOS	28
4.1 Total retribuciones	29
4.2 Total retención	30
4.3 Retribución media	31
4.4 Porcentaje de retención medio	32
4.5 Número de retribuciones medias	33
4.6 Porcentaje de población por segmento	34
4.7 Número de trabajadores / Número de perceptores no activos	35
4.8 Número de trabajadores / Habitantes	36
4.9 Número de trabajadores / Número de personas activas	37
4.10 Salarios totales / Total prestaciones	38



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

4.11	Proyecciones	39
5.	CONCLUSIONES	40
6.	LÍNEAS DE EVOLUCIÓN FUTURAS	40
7.	GLOSARIO	41
8.	BIBLIOGRAFÍA Y REFERENCIAS	41
9.	ANEXOS	42
9.1	Anexo 1. Scripts de creación de base de datos	42
9.2	Anexo 2. Script de transformación de datos volcados	47
9.3	Anexo 3. Script de volcado de datos en tablas definitivas	55
9.4	Anexo 4. Origen de datos de las tablas de los informes en Atlas SBI	57
9.5	Anexo 5. Código fuente de fichero por lotes “Creación de tablas.bat”	63
9.6	Anexo 5. Código fuente de fichero por lotes “Creación de tablas.bat”	63



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

1. INTRODUCCIÓN

1.1 JUSTIFICACIÓN DEL TFC Y CONTEXTO EN EL CUAL SE DESARROLLA: PUNTO DE PARTIDA Y APORTACIÓN DEL TFC.

El presente proyecto consiste en el análisis, diseño e implementación de un Almacén de datos (*Data Warehouse*) para el Organismo de Análisis del Departamento de Trabajo (OADT).

Mediante este Almacén de datos se pretende analizar la sostenibilidad del sistema de prestaciones sociales, mediante la explotación de la información recogida de los distintos organismos que la gestionan.

El almacén de datos nos permitirá cruzar esta información de manera sencilla y eficaz, y disponer en una misma plataforma de toda la información necesaria para hacer los distintos análisis que el OADT requiere.

Los datos se reciben de diferentes organismos debido a la gestión distribuida de las competencias sociales y a la LOPD. El TFC permitirá explotar esta información de manera conjunta.

1.2 OBJETIVOS DEL PROYECTO

1.2.1 OBJETIVOS GENERALES DEL PROYECTO

Los objetivos generales de este proyecto son:

- Adquirir experiencia en el diseño, construcción y explotación de un almacén de datos:
 - Diseño de la base de datos.
 - Preparación de volcados de datos.
 - Elaboración de informes y estadísticas para el usuario (explotación de los datos).
 - Análisis de proyecciones a futuro.

- Fomentar las habilidades para analizar, planificar y gestionar un proyecto:
 - Analizar y estudiar los requerimientos del cliente.
 - Estudio de la viabilidad de las distintas posibilidades de implementación.
 - Planificación de las tareas a realizar en el tiempo, para la implementación y documentación la opción elegida.
 - Implementación del almacén de datos.
 - Documentación del proyecto
 - Presentación del proyecto.



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

1.2.2 OBJETIVOS ESPECÍFICOS DEL PROYECTO

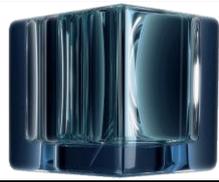
Los objetivos específicos del proyecto son los requerimientos que nos ha trasladado el cliente, es decir:

- Disponer de un almacén de datos que aglutine toda la información que nos facilita.
- Tener una serie de informes con los que explotar la información y poder realizar un análisis de los datos (los informes se podrán mostrar agregados, por año, por comunidad autónoma, por tipo de receptor y por tipo de retribución):
 - o Total retribución
 - o Total retención
 - o Retribución media
 - o % de retención medio (importe retenido respecto del total percibido)
 - o Número de retribuciones medias (si una persona recibe pensión y salario contará como dos)
 - o % de población por segmento (Asalariado, Pensionista, Desempleado con prestación)
 - o Salarios sobre el total de prestaciones de un año determinado
 - o Número de trabajadores/Número de perceptores no activos (Determina la sostenibilidad del sistema)
 - o Número de trabajadores/Habitantes
 - o Número de trabajadores/Número de personas activas
 - o Salarios totales/Total Prestaciones
 - o Proyección de los indicadores:
 - “Num Trabajadores / Num perceptores no activos”
 - “Num trabajadores / Población”

1.3 ENFOQUE Y MÉTODO SEGUIDO

Las fases que se han seguido para la realización del proyecto son las siguientes:

- Análisis
 - o Lectura del caso
 - o Revisión y lectura de la bibliografía
 - o Preparación del plan de trabajo y análisis preliminar
- Definición de los requerimientos
 - o Análisis detallado de los requerimientos de cliente



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

- Diseño
 - o Diseño conceptual
 - o Diseño técnico
 - o Diseño físico
- Implementación
 - o Creación de base de datos
 - o Volcado de datos recibidos por cliente
 - o Transformación de datos volcados
 - o Creación de informes solicitados por cliente

1.4 PLANIFICACIÓN DEL PROYECTO

1.4.1 HITOS Y FECHAS CLAVE

La planificación del proyecto está fuertemente vinculada a los requerimientos de entregas de las distintas PECs que componen la evaluación continua. A partir de esas fechas clave trazaremos el resto de la planificación desglosando cada una de las subtarefas que componen las PECs.

Los principales hitos de seguimiento del proyecto son los siguientes:

Hito	Inicio	Fin
PEC 1 - Plan de trabajo y análisis preliminar de los requerimientos	01/03/2012	16/03/2012
PEC 2 - Análisis de requerimientos y diseño técnico	17/03/2012	20/04/2012
PEC 3 - Implementación	21/04/2012	25/05/2012
Entrega Final	26/05/2012	13/06/2012
Defensa del trabajo (Debate)	25/06/2012	28/06/2012

1.4.2 PLANIFICACIÓN DE TAREAS

En la fase de Implementación se ha planeado el desarrollo de mejoras o nuevas funcionalidades sobre los requerimientos básicos solicitados en el enunciado. En caso de sufrir retrasos esta tarea será cancelada y se centrarán los esfuerzos en la depuración de las funcionalidades básicas.

La dedicación semanal estimada del proyecto será de 10 horas, por lo que cada día de planificación correspondería a ~1,5 horas. En caso de ajustes de plazos, estos esfuerzos semanales pueden llegar a 30 horas/semana, ya que mi disponibilidad por motivos laborales es reducida.

	Memoria del Proyecto	Fecha: 10/06/2012
	Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales	

El desglose de tareas quedaría de la siguiente manera:

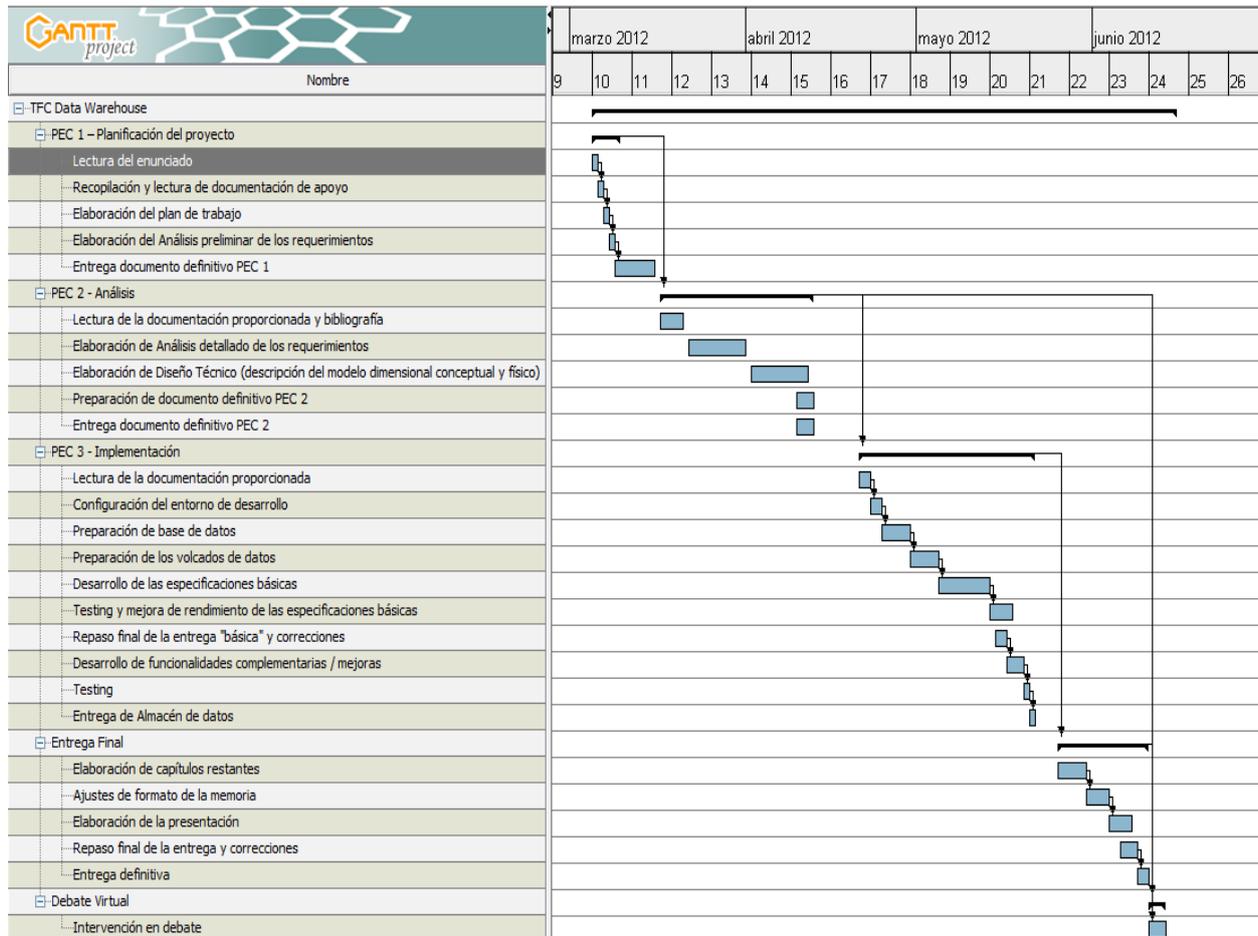
ID TAREA	NOMBRE DE LA TAREA	FECHA INICIO	FECHA FIN	Duración (días)
0	TFC Data Warehouse	05/03/2012	28/06/2012	115
1	PEC 1 – Planificación del proyecto	05/03/2012	10/03/2012	5
1.1	Lectura del enunciado	05/03/2012	06/03/2012	1
1.2	Recopilación y lectura de documentación de apoyo	06/03/2012	07/03/2012	1
1.3	Elaboración del plan de trabajo	07/03/2012	08/03/2012	1
1.4	Elaboración del Análisis preliminar de los requerimientos	08/03/2012	09/03/2012	1
1.5	Entrega documento definitivo PEC 1	09/03/2012	16/03/2012	7
2	PEC 2 - Análisis	17/03/2012	20/04/2012	34
2.1	Lectura de la documentación proporcionada y bibliografía	17/03/2012	21/03/2012	4
2.2	Elaboración de Análisis detallado de los requerimientos	22/03/2012	01/04/2012	10
2.3	Elaboración de Diseño Técnico (descripción del modelo dimensional conceptual y físico)	02/04/2012	12/04/2012	10
2.4	Preparación de documento definitivo PEC 2	13/04/2012	16/04/2012	3
2.5	Entrega documento definitivo PEC 2	17/04/2012	20/04/2012	3
3	PEC 3 - Implementación	21/04/2012	25/05/2012	34
3.1	Lectura de la documentación proporcionada	21/04/2012	23/04/2012	2
3.2	Configuración del entorno de desarrollo	23/04/2012	25/04/2012	2
3.3	Preparación de base de datos	25/04/2012	30/04/2012	5
3.4	Preparación de los volcados de datos	30/04/2012	05/05/2012	5
3.5	Desarrollo de las especificaciones básicas	05/05/2012	14/05/2012	9
3.6	Testing y mejora de rendimiento de las especificaciones básicas	14/05/2012	18/05/2012	4
3.7	Repaso final de la entrega "básica" y correcciones	18/05/2012	20/05/2012	2
3.8	Desarrollo de funcionalidades complementarias / mejoras	20/05/2012	23/05/2012	3
3.9	Testing	23/05/2012	24/05/2012	1
3.10	Entrega de Almacén de datos	24/05/2012	25/05/2012	1
4	Entrega Final	26/05/2012	28/06/2012	33
4.1	Elaboración de capítulos restantes	26/05/2012	31/05/2012	5
4.2	Ajustes de formato de la memoria	31/05/2012	03/06/2012	3
4.3	Elaboración de la presentación	03/06/2012	07/06/2012	4
4.4	Repaso final de la entrega y correcciones	07/06/2012	10/06/2012	3
4.5	Entrega definitiva	10/06/2012	13/06/2012	3
5	Debate Virtual	25/06/2012	28/06/2012	3
5.1	Intervención en debate	25/06/2012	28/06/2012	3



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

1.4.3 DIAGRAMA DE GANTT

A continuación se visualizan las distintas tareas que componen el proyecto, distribuidas temporalmente dentro de un diagrama de Gantt.



1.4.4 RIESGOS E INCIDENCIAS

Los posibles riesgos o incidencias que pueden alterar la planificación del proyecto son:

- Problemas de disponibilidad (Enfermedad, viaje no previsto..)
- Problemas Técnicos

Para problemas del primer tipo, (a no ser que se trate de una enfermedad prolongada que impida la actividad), con la diferencia entre la dedicación media y la dedicación máxima (20 horas/semana), habría más que suficiente para recuperar las horas perdidas.

Para reducir la incidencia del segundo tipo de problemática, realizaremos copias de seguridad de los datos en un repositorio ubicado en otra máquina, y otro repositorio online con gestor de cambios.

También se realizará una copia de la imagen del disco con la configuración de desarrollo de manera que una avería de nuestra máquina no nos impida trabajar.

	Memoria del Proyecto	Fecha: 10/06/2012
	Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales	

1.5 PRODUCTOS OBTENIDOS

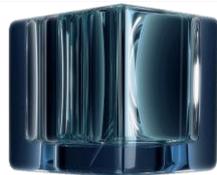
Durante la elaboración del proyecto se han generado los siguientes productos:

- **Documento PEC1.** En este documento se definía el plan de trabajo, así como un análisis preliminar de los requisitos.
- **Documento PEC2.** Análisis detallado y Diseño conceptual y físico del sistema.
- **Documento PEC3.** Informe de la implementación del proceso de carga, del almacén de datos y de los informes desarrollados en la herramienta AtlasSBI
- **Memoria Final.** El presente documento, en el que se detalla todas las tareas realizadas para la elaboración del proyecto.
- **Presentación del Proyecto.** Video en el que se describen las tareas realizadas y se realiza una demostración del producto.
- **Script de creación del almacén de datos.** Consulta SQL en el que se crean las tablas, relaciones, secuencias y triggers que forman el almacén de datos.
- **Scripts de volcado y transformación de datos en el almacén.** Compuesto por archivos de lotes que lanza la ejecución de todos los subprocesos de extracción, consulta SQL de transformación de datos volcados y consulta SQL de volcado en tabla de hechos.
- **Informes solicitados por cliente dentro la herramienta Atlas SBI.** Documentos analíticos creados en herramienta AtlasSBI.

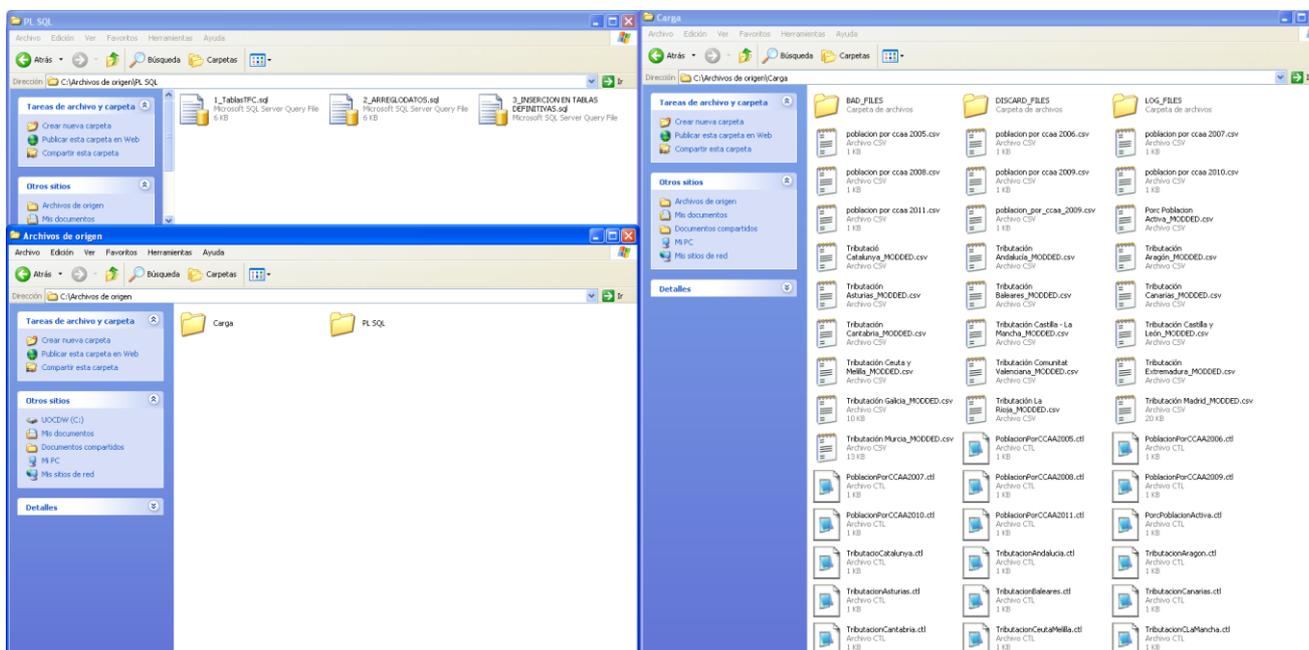
Los ficheros necesarios para la creación de la BD y los procesos de carga y transformación se encuentran en la siguiente ruta de la máquina virtual: **C:\Archivos de origen**

Dentro de esta ruta tenemos:

- **Carga.** Ficheros asociados a la carga en base de datos con SQL LOADER
 - o **BAD_FILES:** En esta carpeta tenemos aquellos registros que por la especificación de nuestros archivos de control no han sido volcados en la base de datos.
 - o **DISCARD_FILES:** Ficheros con registros rechazados por no cumplir criterio de carga definido.
 - o **LOG_FILES:** Ficheros de log de las distintas ejecuciones realizadas para preparar la carga de datos.
 - o **Directorio raíz:** en C:\Archivos de origen\Carga disponemos de los ficheros de origen, así como de los ficheros de control para realizar la carga.
- **PL SQL.** Aquí disponemos de las consultas SQL para:
 - o Crear las distintas tablas de la base de datos.
 - o Transformación de datos.
 - o Carga a tablas definitivas.



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales



1.6 BREVE DESCRIPCIÓN DEL RESTO DE CAPÍTULOS

A continuación en el capítulo 2 (Análisis) se muestran los casos de uso y el diagrama del modelo conceptual. Posteriormente, en el capítulo 3 (Diseño), se puede ver la arquitectura del sistema y el diseño de base de datos. En este capítulo también se detallarán cada uno de los informes generados en la herramienta AtlasSBI.

En el capítulo 4 visualizaremos las capturas de pantalla de cada uno de los informes y gráficas implementados en los documentos analíticos de AtlasSBI.

En los capítulos 5 y 6 tendremos las conclusiones y líneas de evolución futura respectivamente.

La memoria finaliza con el glosario de términos técnicos y los anexos.



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

2. ANÁLISIS

2.1 ANÁLISIS DE DATOS DE ORIGEN

Los datos vienen almacenados en ficheros Excel proporcionados por los distintos organismos que los gestionan. Existen tres tipos distintos de fichero:

- **Población por CCAA xxxx:** Un fichero por cada uno de los años con el número de habitantes por CCAA con los siguientes campos:
 - Comunidad autónoma (incluyendo fila con totales): Muestra la comunidad autónoma para la que se proporciona el dato.
 - Varones: número de varones por región
 - Mujeres: número de mujeres por región

En este caso, los ficheros csv utilizan el carácter ‘,’ como separador de columnas.

- **Porc Poblacion Activa:** Fichero con el porcentaje de población activa para cada comunidad autónoma y años con los siguientes campos:
 - Comunidad autónoma (incluyendo fila con totales): Muestra la comunidad autónoma para la que se proporciona el dato.
 - % de población activa (ambos sexos) para un año determinado.

Este fichero csv utiliza el carácter ‘,’ como separador de columnas, y recoge datos de 2005 a 2010.

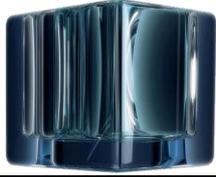
- **Tributación XXXXXXXX:** Muestra los siguientes datos:
 - Personas
 - Retribuciones
 - Retribución media anual
 - Retenciones
 - Tipo medio de retención

Para los siguientes colectivos:

- Asalariados
- Pensionistas
- Desempleados
- Asalariados y pensionistas
- Asalariados y desempleados
- Pensionistas y desempleados
- Asalariados, pensionistas y desempleados

Y tipos de prestación:

- TOTAL
- SALARIOS
- PENSIONES
- DESEMPLEO



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

Estos ficheros csv no utilizan un carácter homogéneo de separación de columnas.

Las tareas previas al volcado serán:

- Homogeneización de los caracteres de separación de los Exceles.

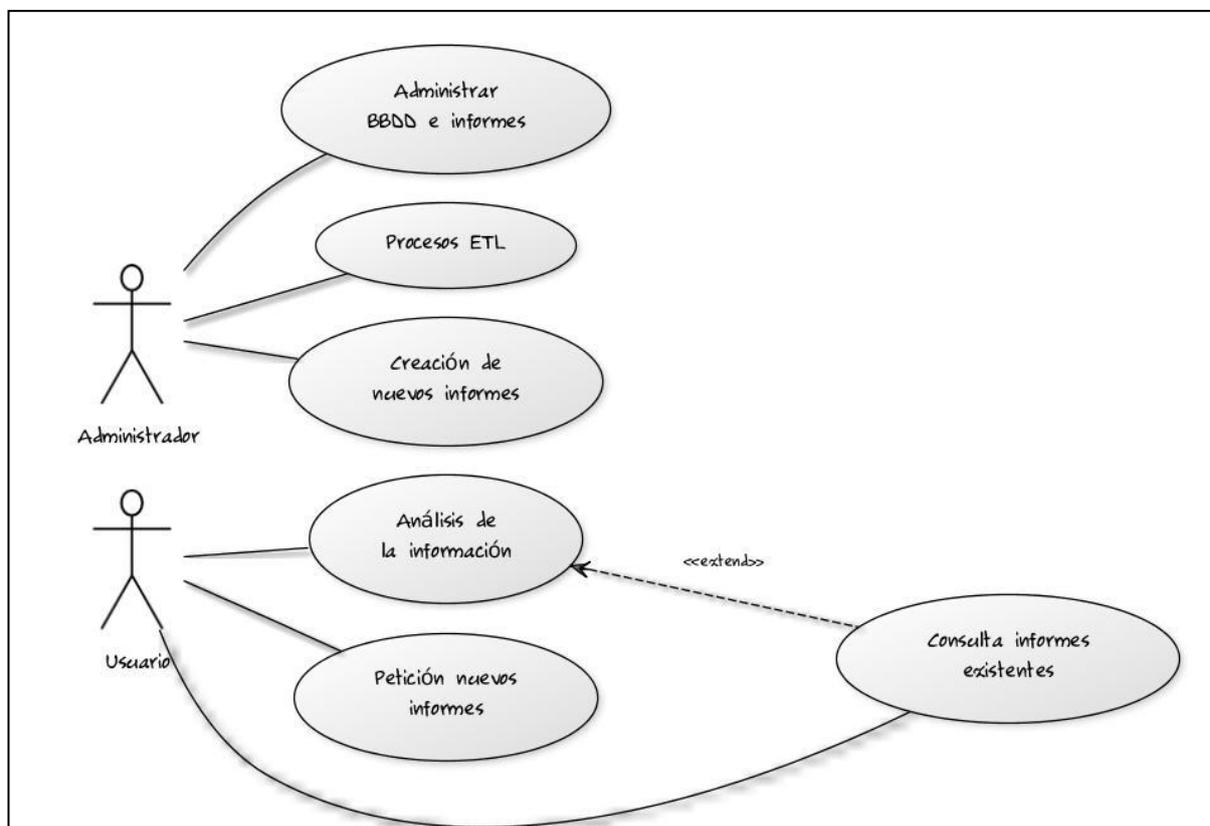
Las tareas de transformación de datos posteriores al volcado serán:

- Unificar las medidas en las que se expresan los distintos valores (se decide transformar los valores medios en brutos).
- Calcular el número de habitantes como media de los valores del 1 de enero del año en curso con el 1 de enero del año siguiente.

2.2 DIAGRAMAS DE CASOS DE USO

Este diagrama define cómo los distintos usuarios interactuarán con el sistema. Podemos identificar dos clases de roles principales:

- Administrador: Se encargará de realizar los procesos ETL, de gestionar los accesos a base de datos e informes y de implementar los nuevos informes que solicite el cliente.
- Usuario: Será quien consulte los informes existentes, analice la información extraída y solicite nuevos informes.



	Memoria del Proyecto	Fecha: 10/06/2012
	Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales	

2.3 DISEÑO CONCEPTUAL

En esta sección se detallan las fases del proceso de Diseño conceptual de nuestro *DataWarehouse*, hasta la construcción del modelo de estrella que soportará las especificaciones solicitadas por cliente.

2.3.1 IDENTIFICAR EL HECHO

El primer paso para la obtención de nuestro modelo de estrella es identificar el hecho que constituye el núcleo del modelo.

Para ello debemos analizar el caso, tener claro los datos que queremos explotar y las necesidades del cliente. Traslado esto a nuestra casuística, los datos que queremos obtener en los informes, y por lo tanto nuestro hecho, son las retribuciones obtenidas en una temporalidad determinada (año), clasificados por comunidad autónoma, tipo de perceptor y tipo de retribución.

2.3.2 ENCONTRAR LA GRANULARIDAD ADECUADA

En esta fase debemos identificar el nivel de detalle que deberán tener nuestras celdas que conformarán nuestro cubo. Esta es una fase crítica ya que, en caso de no definir correctamente el detalle de las celdas eligiendo un nivel de detalle demasiado profundo, provocará que nuestras consultas a base de datos sean demasiado costosas reduciendo considerablemente la usabilidad de nuestro cubo.

Teniendo en cuenta ésto, revisando la granularidad con que nos facilitan los datos en los ficheros csv, parece una granularidad adecuada y que no debería darnos problemas de rendimiento a la hora de ejecutar nuestras consultas.

En caso de que las fases posteriores de este análisis indicaran lo contrario, deberíamos modificar esta granularidad por otra más adecuada.

2.3.3 ESCOGER LAS DIMENSIONES A EMPLEAR EN EL ANÁLISIS

Las dimensiones en los cubos de datos, son aquellas informaciones complementarias a los hechos, necesarias para presentar los datos al usuario. En nuestro caso estas informaciones complementarias (dimensiones) serían:

- Periodo temporal (Año)
- Zona geográfica (Comunidad autónoma)
- Tipo de retribución
- Tipo de perceptor

	Memoria del Proyecto	Fecha: 10/06/2012
	Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales	

2.3.4 ESTABLECER LOS ATRIBUTOS DE CADA DIMENSIÓN

Una vez tenemos identificadas las dimensiones, debemos determinar qué atributos tiene cada una de esas dimensiones y cuáles nos interesa tener en cuenta para nuestro caso.

- Periodo temporal: La manera en que se nos facilita la información es por años, por lo que en principio el atributo que definirá esta dimensión será el propio año.
 - Año
- Zona geográfica: la dimensión Zona en nuestro caso viene informada únicamente por Comunidad Autónoma.
 - Comunidad autónoma
- Tipo de retribución:
 - Tipo. Valores posibles: Salarios, pensiones y desempleo.
- Tipo de perceptor:
 - Tipo. Valores posibles: Asalariado; pensionista; desempleado; asalariados y pensionistas; asalariados y desempleados; pensionistas y desempleados; asalariados, pensionistas y desempleados.
- Población:
 - N° habitantes
 - Sexo habitantes. Valores posibles: Hombre, Mujer
 - % activos

2.3.5 IDENTIFICAR LAS MEDIDAS A EMPLEAR

Las medidas a emplear, serán todos aquellos datos que por su carácter variable tendremos que incluir en nuestra tabla de hechos.

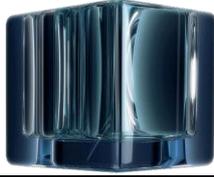
Para nuestro caso hemos considerado las siguientes medidas:

- Cantidad total de retribución
- n° de Perceptores
- Retención total

2.3.6 DEFINIR LAS CELDAS

En este paso definimos cuales de las posibles celdas que resultan del análisis, pueden ser consideradas interesantes para nuestro modelo (y por tanto han de ser almacenadas) y cuáles de ellas se pueden considerar derivadas de otras y por tanto no será necesario almacenarlas.

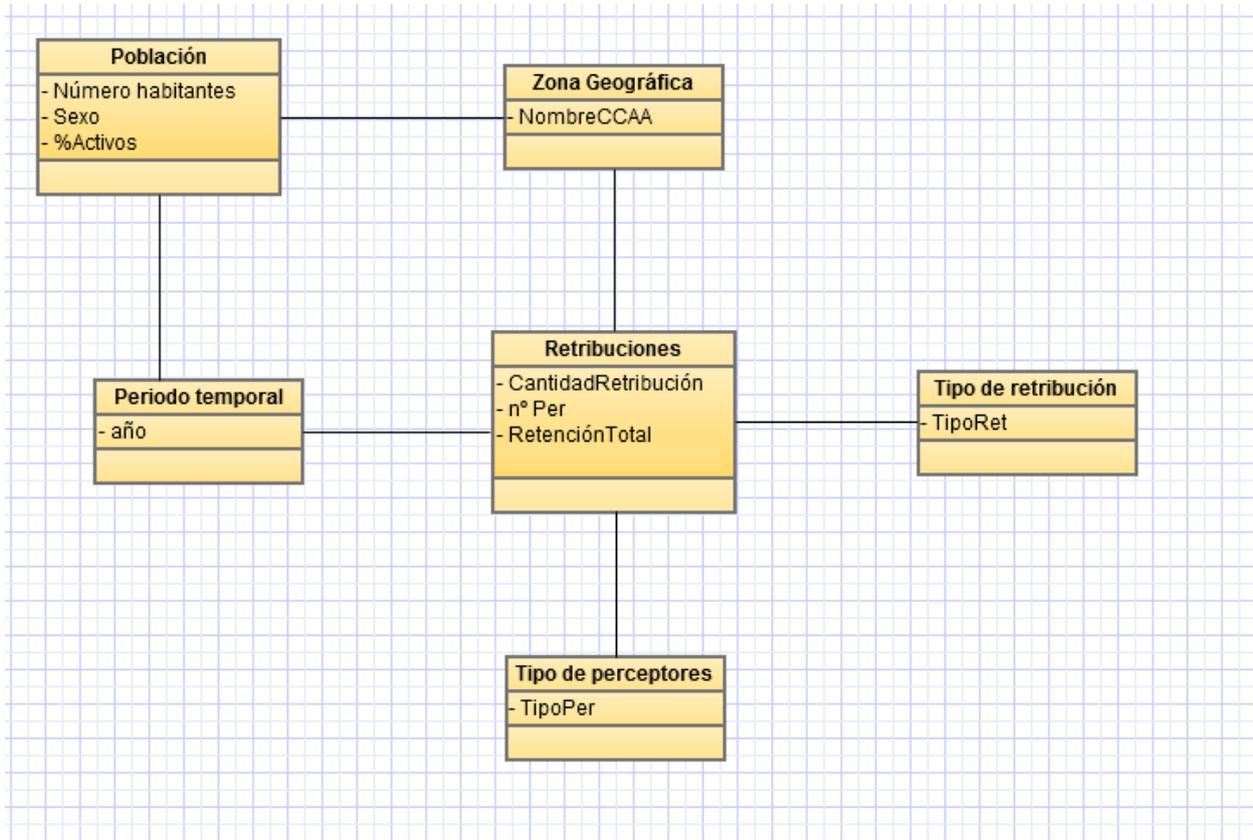
En nuestro caso solo tenemos una celda que refleje el hecho analizado. Usaremos una celda que representará los datos de una retribución en un año dado para una Comunidad Autónoma.

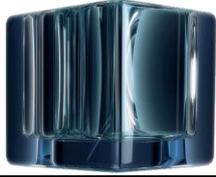


Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

2.3.7 APROXIMACIÓN AL MODELO

Tras los pasos descritos anteriormente, una aproximación al modelo conceptual podría ser la siguiente:

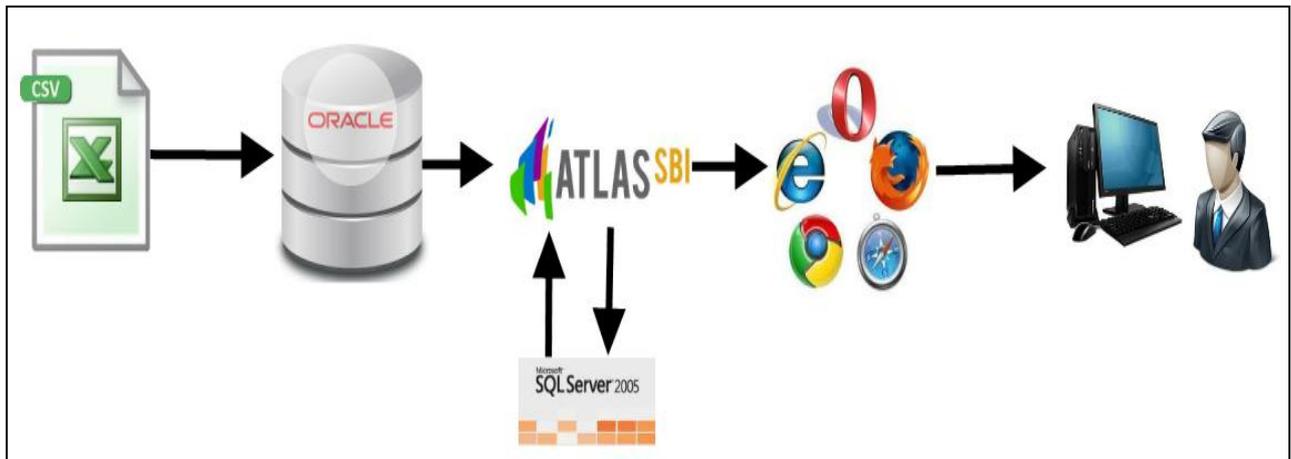




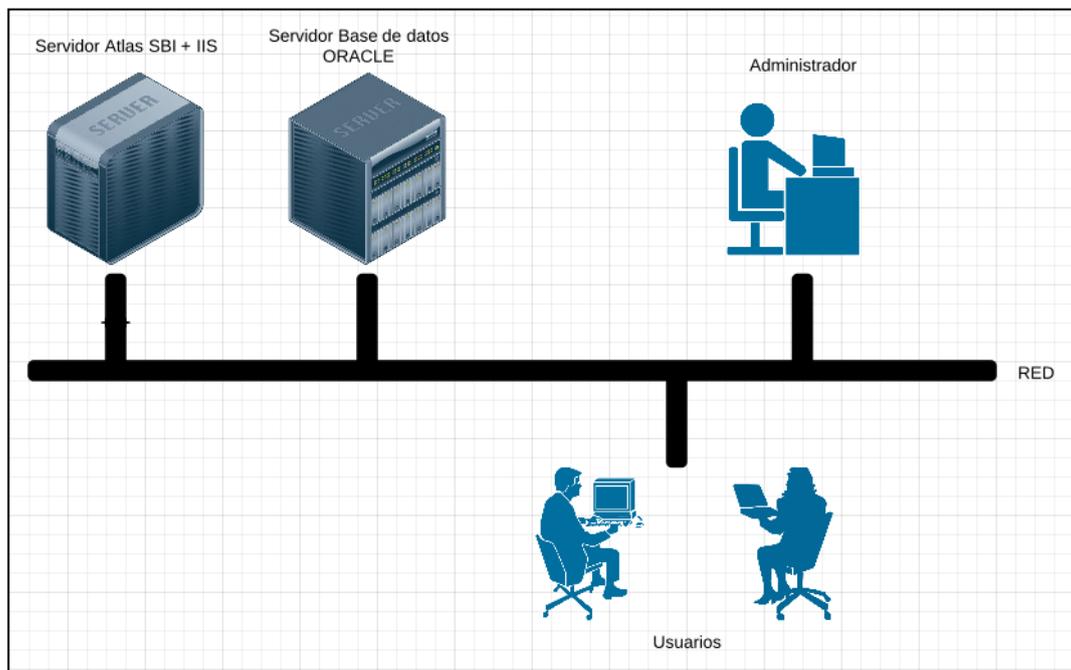
Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

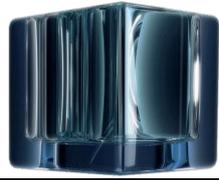
3. DISEÑO

3.1 DIAGRAMA DE LA ARQUITECTURA DE SOFTWARE



3.2 DIAGRAMA DE LA ARQUITECTURA DE HARDWARE





Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

3.3 DISEÑO FÍSICO

A continuación se describen las tablas del modelo de datos (también se incluye la definición de las tablas auxiliares utilizadas para los procesos ETL).

3.3.1 TABLA DE HECHOS

Nombre de Campo	Tipo de campo	Tamaño	Dec	Nullable	Descripción
CantidadRetribucion	NUMBER	30	2	No	
numPerceptores	NUMBER	10	0	No	
Retencion	NUMBER	24	0	No	
<u>IDPERIODOTEMPORAL</u>	NUMBER	5	0	No	Clave foránea de tabla de PeriodoTemporal
<u>IDCCAA</u>	NUMBER	5	0	No	Clave foránea de tabla de ZonaGeografica
<u>IDRET</u>	NUMBER	5	0	No	Clave foránea de tabla de TipoRetribuciones
<u>IDTIOPER</u>	NUMBER	5	0	No	Clave foránea de tabla de TipoPerceptores

3.3.2 PERIODOTEMPORAL

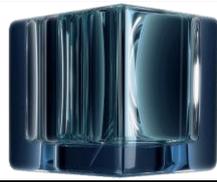
Nombre de Campo	Tipo de campo	Size	Dec	Nullable	Descripción
<u>IDPERIODOTEMPORAL</u>	NUMBER	5	0	No	Clave primaria, código autonumérico
anyo	VARCHAR	20	0	No	Literal del año

3.3.3 ZONAGEOGRAFICA

Nombre de Campo	Tipo de campo	Size	Dec	Nullable	Descripción
<u>IDCCAA</u>	NUMBER	5	0	No	Clave primaria, código autonumérico
nombreccaa	VARCHAR	100	0	No	Literal de la comunidad autónoma

3.3.4 TIPORETRIBUCIONES

Nombre de Campo	Tipo de campo	Size	Dec	Nullable	Default
<u>IDRET</u>	NUMBER	5	0	No	Clave primaria, código autonumérico
nombretipoRet	VARCHAR	100	0	No	Literal del nombre del fichero



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

3.3.5 TIPOPERCEPTORES

Nombre de Campo	Tipo de campo	Size	Dec	Nullable	Descripción
IDTIPOPER	NUMBER	5	0	No	Clave primaria, código autonumérico
nombretipoPer	VARCHAR	100	0	No	Literal del nombre del perceptor

3.3.6 HABITANTES

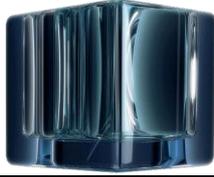
Nombre de Campo	Tipo de campo	Size	Dec	Nullable	Descripción
IDHAB	NUMBER	5	0	No	Clave primaria, código autonumérico
numerohab	NUMERO	10	0	Sí	
sexo	VARCHAR	1	0	No	
IDCCAA	NUMBER	5	0	No	Clave foránea con tabla de ZonaGeografica
IDPERIODOTEMPORAL	NUMBER	5	0	No	Clave foránea con tabla periodoTemporal
porcActivos	NUMBER	5	2	No	

3.3.7 TMP_POBLACIONACTIVA

Nombre de Campo	Tipo de campo	Size	Dec	Nullable	Default
año	VARCHAR	20	0	No	Utilizamos un campo de tipo texto por si en algún momento nos pasaran cadenas de texto en vez de números.
ccaa	VARCHAR	100	0	No	
porcentaje	NUMBER	5	2	No	

3.3.8 TMP_POBLACION

Nombre de Campo	Tipo de campo	Size	Dec	Nullable	Descripción
año	VARCHAR	20	0	No	
ccaa	VARCHAR	100	0	No	
Hombres	NUMBER	10	0	No	
Mujeres	NUMBER	10	0	No	

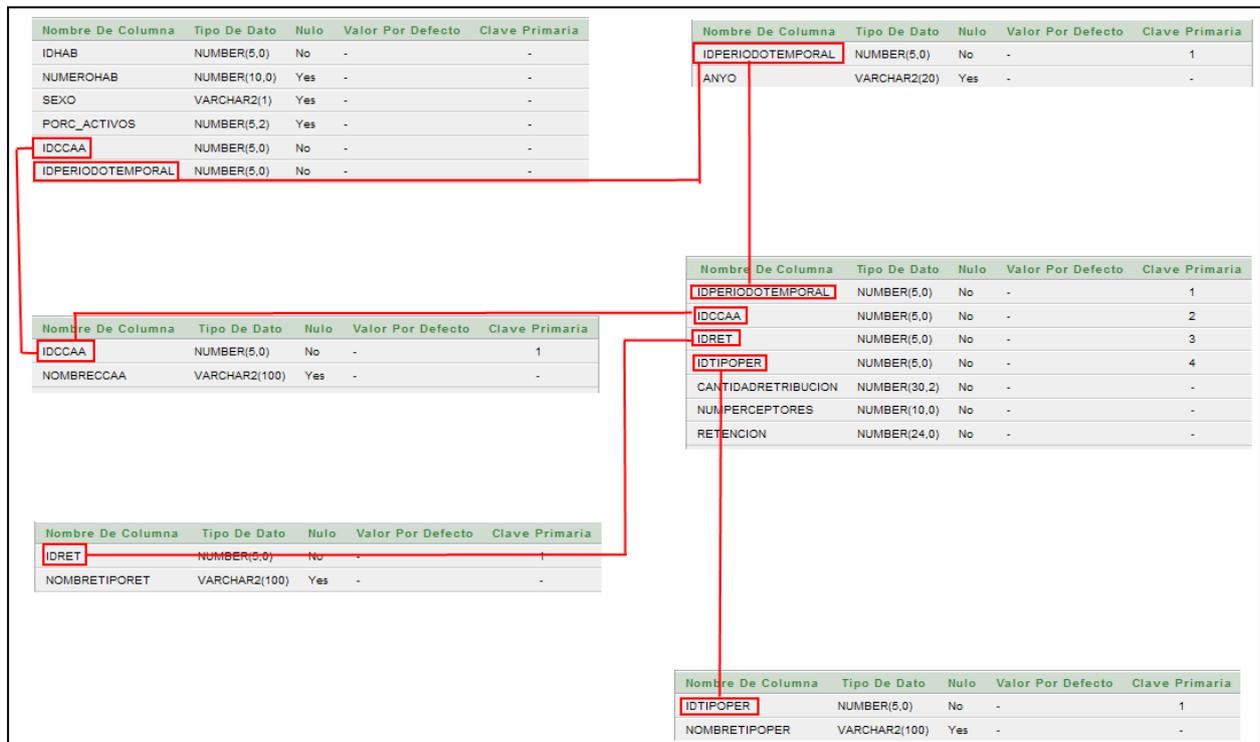


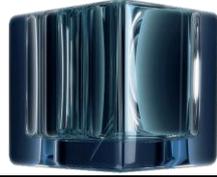
Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

3.3.9 TMP_TRIBUTACION

Nombre de Campo	Tipo de campo	Size	Dec	Nullable	Descripción
año	VARCHAR	20	0	No	
ccaa	VARCHAR	100	0	No	
tipoRet	VARCHAR	100	0	No	Lo cargamos en modo texto para posteriormente cargar la tabla de Tipos Retribución
tipoPer	VARCHAR	100	0	No	Lo cargamos en modo texto para posteriormente cargar la tabla de Tipos Perceptores
CantidadRetribucion	NUMBER	24	0	No	
NumeroPer	NUMBER	10	0	No	
cantidadRet	NUMBER	24	0	No	

3.3.10 DIAGRAMA FÍSICO DE RELACIÓN ENTRE TABLAS





Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

3.4 PROCESOS ETL

Podemos dividir a los procesos ETL en 3 subprocesos:

3.4.1 PROCESOS DE EXTRACCIÓN

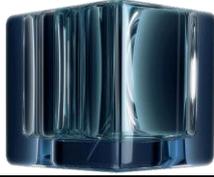
El proceso de extracción toma como fuente de datos los ficheros CSV que nos proporcionan, exportándolos a tablas temporales donde realizaremos el tratamiento de datos previo al volcado en las tablas finales.

Previa a la exportación a la base de datos mediante la herramienta SQL Loader debemos hacer algunos cambios en el formato de los ficheros de origen para que sean correctamente cargados.

- Ficheros de población por ccaa: no sufren alteraciones
- Fichero Porc Población activa: Se modifica para dejar en tres columnas independientes:
 - o CCAA
 - o Porcentaje
 - o Año
- Ficheros de tributación: Se realizan las siguientes modificaciones:
 - o Llevar los años a todas las columnas
 - o Trasponer todo el contenido del fichero
 - o Replicar la estructura de filas para dejar únicamente las siguientes columnas: Año, Perceptor, CCAA, TipoRetribución, Personas, Cantidad retribución y Cantidad retención.

Una vez modificados los ficheros, se crean las tablas de base de datos mediante Script (1_TablasTFC.sql VER ANEXO 1) y se procede al volcado en las tablas temporales mediante la herramienta SQL Loader, a través de los ficheros de control.

Ejemplo de fichero de control:



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
OPTIONS (SKIP=1)
LOAD DATA
INFILE 'Tributación Canarias_MODALDED.csv'
BADFILE 'c:\Archivos de origen\Carga\BAD_FILES\badTributacionCanarias.bad'
DISCARDFILE 'c:\Archivos de origen\Carga\DISCARD_FILES\dscTributacionCanarias.dsc'
APPEND
INTO TABLE TMP_TRIBUTACION
FIELDS TERMINATED BY ';'
optionally enclosed by '"'
TRAILING NULLCOLS
(
  ANYO "ltrim(rtrim(:ANYO))",
  TIPOPER "ltrim(rtrim(:TIPOPER))" ,
  CCAA "ltrim(rtrim(:CCAA))",
  TIPORET "ltrim(rtrim(:TIPORET))",
  NUMEROPER "REPLACE(REPLACE(:NUMEROPER, '..', '0,0'),' ','')",
  CANTIDADRETRIBUCION "REPLACE(REPLACE(:CANTIDADRETRIBUCION, '..', '0,0'),' ','')",
  CANTIDADRET "REPLACE(REPLACE(:CANTIDADRET, '..', '0,0'),' ','')"
)
```

Ejemplo de fichero de log:

"Total de registros lógicos ignorados: 1
Total de registros lógicos leídos: 170
Total de registros lógicos rechazados: 2
Total de registros lógicos desechados: 0

La ejecución empezó en Lun May 21 01:05:08 2012

La ejecución terminó en Lun May 21 01:05:08 2012"

Ejemplo de comando de ejecución de SQL Loader:

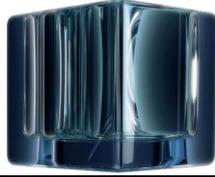
```
sqlldr tfc/tfc control=c:\archivos de origen\carga\poblacionporccaa2009.ctl log=c:\archivos de origen\carga\poblacionporccaa2009.log
```

3.4.2 PROCESOS DE TRANSFORMACIÓN

Para realizar el proceso de transformación, una vez tenemos la información de los ficheros CSV cargada en las tablas temporales, corregiremos mediante script de forma masiva aquellos datos erróneos que debemos arreglar antes de volcar en las tablas finales.

El script (**2_ARREGLODATOS.sql VER ANEXO 2**) realizará las siguientes tareas:

- Homogeneización de datos en campo "Tipo de perceptor"
- Homogeneización de nombres de Comunidades Autónomas y arreglo de datos en aquellas que



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

tienen un nombre compuesto

- Actualización de los datos de habitantes, haciendo la media entre el 1 de Enero del año actual y el 1 de Enero del año siguiente, tal como indica el enunciado.
- Homogeneización de datos de retribuciones para aquellas Comunidades Autónomas que nos envían valores medios en vez de valores totales. (Madrid, Valencia y Murcia)
- Arreglo de problemática detectada en las retribuciones y retenciones desde el año 2006 en adelante. Se multiplican por 1000 las retribuciones y retenciones en las comunidades autónomas en las que se detecta el problema.

3.4.3 PROCESOS DE CARGA

Preparamos scripts de carga de datos tomando como origen de datos las tablas temporales, cargando primero las tablas de parametrizaciones (tipos de retribución, tipos de perceptores, comunidades autónomas, años) y posteriormente la tabla de hechos.

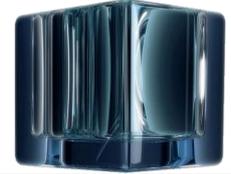
(Script **3_INSERTION EN TABLAS DEFINITIVAS.sql VER ANEXO 3**)

En la carga de los datos parametrizados (años, comunidades autónomas) se ha decidido volcar únicamente aquellas parametrizaciones para las que tenemos TODOS los datos.

De esta modo nuestro *Data Warehouse* tendrá integridad de datos y coherencia, no permitiendo extraer información de años para los que tenemos la información incompleta, o de Comunidades Autónomas que no nos han facilitado su información retributiva.

3.4.4 PASOS A SEGUIR PARA EJECUCIÓN DE PROCESOS ETL

- Ejecución de archivo por lotes "*Creación de tablas.bat*". Este archivo por lotes ejecuta el script de creación de tablas, relaciones entre ellas y crea las secuencias y disparadores para poder gestionar campos autonuméricos.
- Ejecución de archivo por lotes "*CargaTablasTemporales_Completo.bat*". Este archivo por lotes realiza la ejecución de todos los procesos de carga de SQL*Loader, y posteriormente ejecuta los scripts de arreglo de datos y carga de datos en tablas definitivas.

	Memoria del Proyecto	Fecha: 10/06/2012
Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales		

3.5 DISEÑO Y DESCRIPCIÓN DE LOS INFORMES CREADOS

3.5.1 TOTAL RETRIBUCIONES

En este documento analítico mostramos por una parte la tabla-informe con los datos solicitados, en este caso las retribuciones por zona, año, tipo de perceptor y tipo de retribución, y apoyándonos en una de las propiedades de las tablas de Atlas SBI mostramos el sumatorio total de retribuciones.

De este modo siempre tenemos la referencia del total de retribución según los filtros seleccionados.

Además de la tabla de datos, mostramos las siguientes gráficas:

- **Retribuciones or CCAA.** Diagrama de barras en el que visualizamos la distribución de retribuciones por Comunidad autónoma según los filtros elegidos en la tabla.
- **Retribuciones por Año.** En este gráfico vemos la evolución de las retribuciones a lo largo del tiempo.
- **Retribuciones por tipo de Retribución.** Se muestra mediante un gráfico tipo tarta la distribución de las retribuciones según el tipo de retribución (Salario, Prestación por desempleo y Pensiones).
- **Retribuciones por Perceptor.** Diagrama de barras en el que visualizamos la distribución de retribuciones según el tipo de perceptor.

3.5.2 TOTAL RETENCIÓN

En este documento analítico mostramos por una parte la tabla-informe con los datos solicitados, en este caso las retenciones por zona, año, tipo de perceptor y tipo de retribución, y apoyándonos en una de las propiedades de las tablas de Atlas SBI mostramos el sumatorio total de retenciones.

De este modo siempre tenemos la referencia del total de retención según los filtros seleccionados.

Además de la tabla de datos, mostramos las siguientes gráficas:

- **Retenciones por CCAA.** Diagrama de barras en el que visualizamos la distribución de retenciones por Comunidad autónoma según los filtros elegidos en la tabla.
- **Retenciones por Año.** En este gráfico vemos la evolución de las retenciones a lo largo del tiempo.
- **Retenciones por tipo de Retribución.** Se muestra mediante un gráfico tipo tarta la distribución de las retenciones según el tipo de retribución (Salario, Prestación por desempleo y Pensiones).
- **Retenciones por Perceptor.** Diagrama de barras en el que visualizamos la distribución de retribuciones según el tipo de perceptor.

	Memoria del Proyecto	Fecha: 10/06/2012
Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales		

3.5.3 RETRIBUCIÓN MEDIA

En este documento analítico mostramos por una parte la tabla-informe con los datos solicitados, en este caso las retribuciones medias por zona, año, tipo de perceptor y tipo de retribución. Crearemos una columna calculada para sacar la retribución media de cada uno de los registros, y nos apoyaremos en una de las propiedades de las tablas de Atlas SBI para mostrar la retribución media global según los filtros elegidos en la tabla.

Además de la tabla de datos, mostramos las siguientes gráficas:

- **Retribución media por CCAA.** Diagrama de barras en el que visualizamos la distribución de retribuciones medias por Comunidad autónoma según los filtros elegidos en la tabla.
- **Retribución media por Año.** En este gráfico vemos la evolución de las retribuciones medias a lo largo del tiempo.
- **Retribución media por tipo de Retribución.** Se muestra mediante un gráfico tipo tarta la distribución de las retribuciones medias según el tipo de retribución (Salario, Prestación por desempleo y Pensiones).
- **Retribución media por Perceptor.** Diagrama de barras en el que visualizamos la distribución de retribuciones medias según el tipo de perceptor.

3.5.4 PORCENTAJE DE RETENCIÓN MEDIO

En este documento analítico mostramos por una parte la tabla-informe con los datos solicitados, en este caso el porcentaje de retención medio por zona, año, tipo de perceptor y tipo de retribución. Crearemos una columna calculada para sacar el porcentaje de retención medio de cada uno de los registros, y nos apoyaremos en una de las propiedades de las tablas de Atlas SBI para mostrar el porcentaje de retención medio global según los filtros elegidos en la tabla.

Además de la tabla de datos, mostramos las siguientes gráficas:

- **Porcentaje de retención medio por CCAA.** Diagrama de barras en el que visualizamos la distribución del porcentaje de retención medio medias por Comunidad autónoma según los filtros elegidos en la tabla.
- **Porcentaje de retención medio por Año.** En este gráfico vemos la evolución del porcentaje de retención medio a lo largo del tiempo.
- **Porcentaje de retención medio por tipo de Retribución.** Se muestra mediante un gráfico tipo tarta la distribución del porcentaje de retención medio según el tipo de retribución (Salario, Prestación por desempleo y Pensiones).
- **Porcentaje de retención medio por Perceptor.** Diagrama de barras en el que visualizamos el porcentaje de retención medio según el tipo de perceptor.

	Memoria del Proyecto	Fecha: 10/06/2012
Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales		

3.5.5 NÚMERO DE RETRIBUCIONES MEDIAS

En este documento analítico mostramos por una parte la tabla-informe con los datos solicitados, en este caso el número de retribuciones medias por zona, año, tipo de perceptor y tipo de retribución. Para ello hemos creado una columna auxiliar donde calcularemos mediante SQL el número de percepciones que tiene cada registro, y dividiremos este número de percepciones entre el número total de perceptores para conocer el número de retribuciones medias del registro.

Apoyándonos en una de las propiedades de las tablas de Atlas SBI mostramos la media global de retribuciones medisa de todos los registros.

Además de la tabla de datos, mostramos las siguientes gráficas:

- Número de retribuciones medio por CCAA. Diagrama de barras en el que visualizamos la distribución del número de retribuciones medio por Comunidad autónoma según los filtros elegidos en la tabla.
- Número de retribuciones medio por Año. En este gráfico vemos la evolución del número de retribuciones medio a lo largo del tiempo.
- Número de retribuciones medio por tipo de Retribución. Se muestra mediante un gráfico tipo tarta la distribución del número de retribuciones medio según el tipo de retribución (Salario, Prestación por desempleo y Pensiones).
- Número de retribuciones medio por Perceptor. Diagrama de barras en el que visualizamos la distribución del número de retribuciones medio según el tipo de perceptor.

3.5.6 PORCENTAJE DE POBLACIÓN POR SEGMENTO

En este documento analítico, nos apoyamos en una tabla para mostrar los totales y porcentajes de población por segmento, según los criterios seleccionados en el filtro.

A través de diagramas de barras desglosamos estos totales en:

- Porcentaje de población por segmento por CCAA
- Porcentaje de población por segmento por Año

Para este informe, hemos considerado como Asalariado, Pensionista y Desempleado con prestación a cualquier persona que haya recibido ese tipo de retribución.

3.5.7 NÚMERO DE TRABAJADORES / NÚMERO DE PERCEPTORES NO ACTIVOS

En este documento analítico, nos apoyamos en una tabla para mostrar los totales y porcentajes de trabajadores y de perceptores no activos, según los criterios seleccionados en el filtro.

A través de diagramas de barras desglosamos estos totales en:

- Trabajadores / Perceptores no activos por CCAA
- Trabajadores / Perceptores no activos por Año

Hemos tomado como perceptores no activos al conjunto de pensionistas + desempleados con prestación.

	Memoria del Proyecto	Fecha: 10/06/2012
Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales		

3.5.8 *NÚMERO DE TRABAJADORES / HABITANTES*

En este documento analítico, nos apoyamos en una tabla para mostrar los totales y porcentajes de trabajadores por habitantes, según los criterios seleccionados en el filtro.

A través de diagramas de barras desglosamos estos totales en:

- Porcentaje de Trabajadores/Habitantes por CCAA
- Porcentaje de Trabajadores/Habitantes por Año

3.5.9 *NÚMERO DE TRABAJADORES / NÚMERO DE PERSONAS ACTIVAS*

En este documento analítico, nos apoyamos en una tabla para mostrar los totales y porcentajes de trabajadores por personas activas, según los criterios seleccionados en el filtro.

A través de diagramas de barras desglosamos estos totales en:

- Porcentaje de Trabajadores/Personas activas por segmento por CCAA
- Porcentaje de Trabajadores/Personas activas por Año

3.5.10 *SALARIOS TOTALES / TOTAL PRESTACIONES*

En este documento analítico, nos apoyamos en una tabla para mostrar los totales y porcentajes de prestaciones por los salarios totales, según los criterios seleccionados en el filtro.

A través de diagramas de barras desglosamos estos totales en:

- Porcentaje de Prestaciones/Salarios por CCAA
- Porcentaje de Prestaciones/Salarios por Año

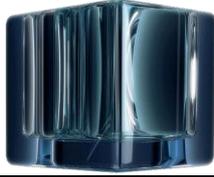
3.5.11 *PROYECCIONES*

En la primera proyección, calculamos mediante una regresión lineal, el año en que se igualaría el número de trabajadores con el número de pensionistas+desempleados con prestación.

El año de colapso sería cercano a 2013, que es cuando la proporción sería 1:1.

En la segunda proyección, calculamos el año en que dos trabajadores mantendrían a un desempleado con prestación.

En este caso el año de colapso también será 2013, que es el año en que la proporción alcanzará el 2:1.



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

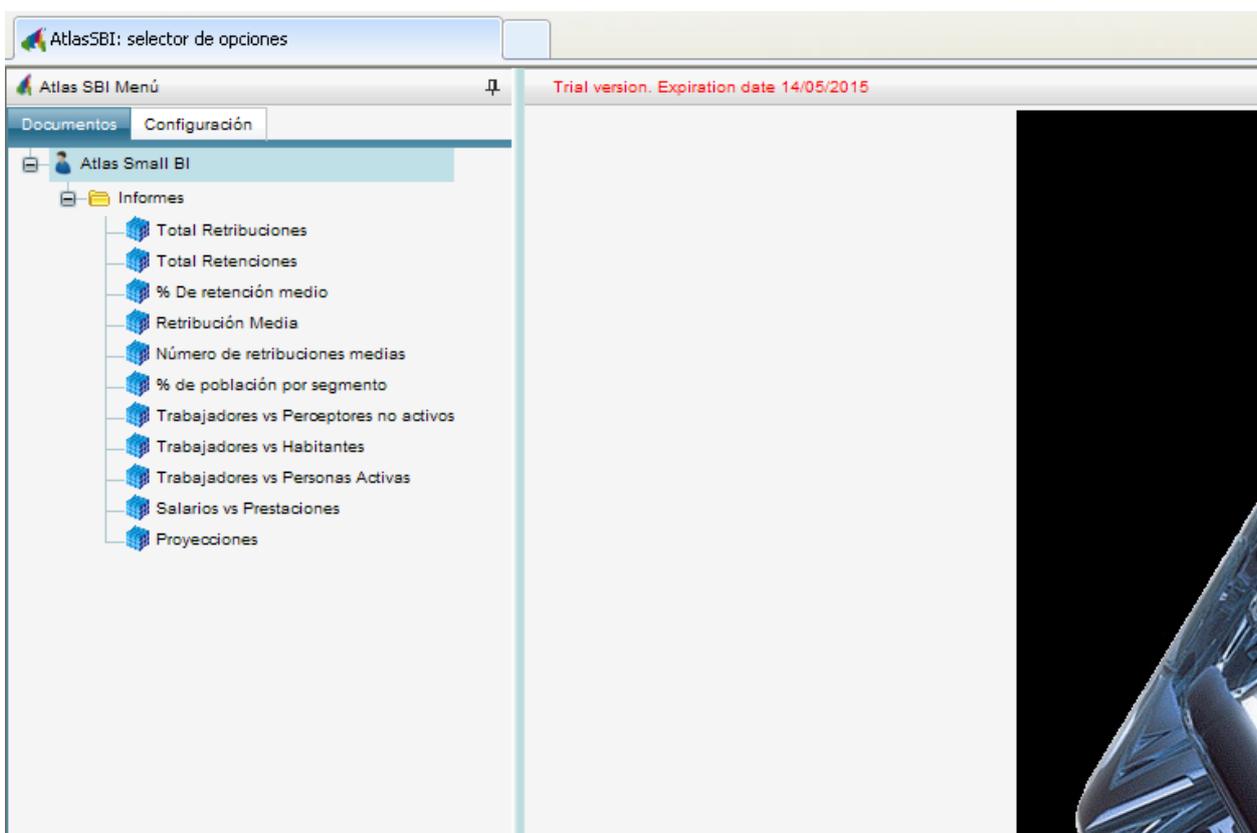
4. INFORMES PREPARADOS

Para acceder a los informes dentro de la herramienta Atlas SBI, deberemos acceder a la siguiente URL:

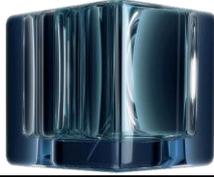
<http://localhost/atlasSBI>

Una vez en la página de login, introduciremos el usuario (atlassbi/atlassbi)

Al logearnos aparecerá el menú, donde podremos acceder a cada uno de los informes desarrollados.



(EN EL ANEXO 4 SE ADJUNTAN LAS CONSULTAS UTILIZADAS PARA CADA UNO DE LOS INFORMES)



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

4.1 TOTAL RETRIBUCIONES

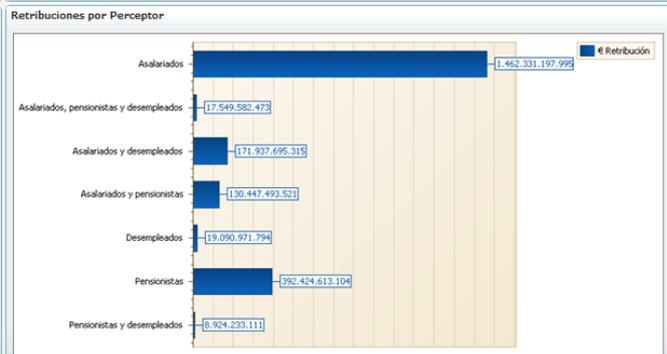
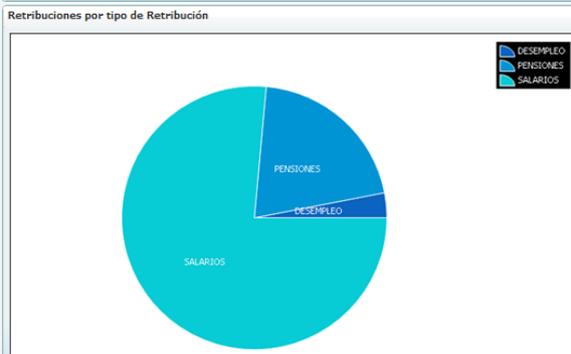
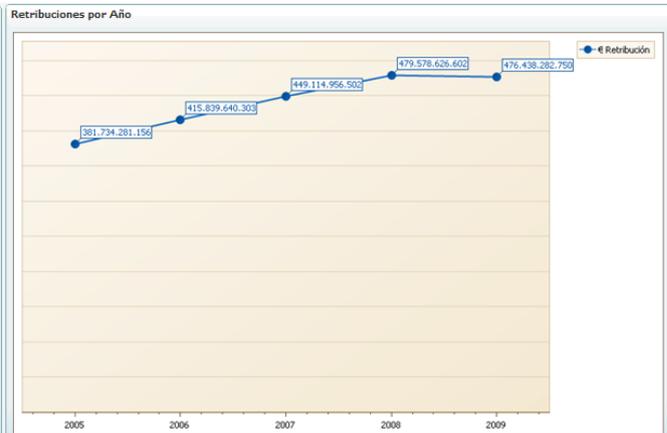
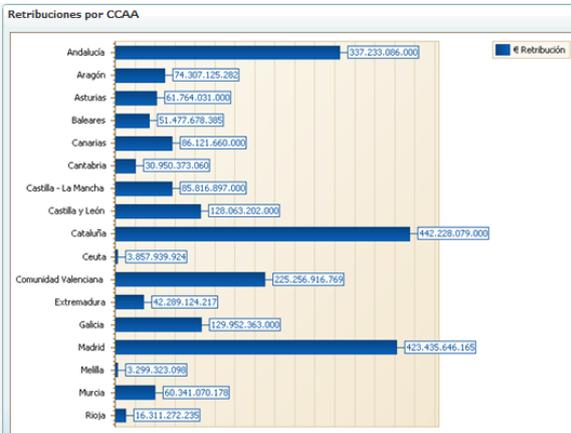
Este informe muestra de forma agregada toda la información de retribuciones, así como el total de € de Retribución para el filtro seleccionado (sumatorio).

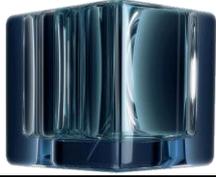
Informe de retribuciones por zona, año, tipo de perceptor y tipo de retribución

Año: Comunidad autónoma: Tipo de retribución: Tipo de perceptor:

NOMBRECCAA	NUMPERCEPTORES	ANYO	NOMBRETIPOPER	NOMBRETIPRET	CANTIDADRETRIBUCION
Baleares	128.602	2009	Asalariados y desempleados	SALARIOS	1.257.440.000 €
Baleares	0	2009	Pensionistas y desempleados	SALARIOS	0 €
Baleares	11.742	2009	Asalariados, pensionistas y desempleados	SALARIOS	100.447.000 €
Baleares	0	2005	Asalariados	PENSIONES	0 €
Baleares	137.673	2005	Pensionistas	PENSIONES	1.280.000.000 €
Baleares	0	2005	Desempleados	PENSIONES	0 €
Baleares	23.368	2005	Asalariados y pensionistas	PENSIONES	178.000.000 €
Baleares	0	2005	Asalariados y desempleados	PENSIONES	0 €
Baleares	1.915	2005	Pensionistas y desempleados	PENSIONES	13.400.000 €
Baleares	8.002	2005	Asalariados, pensionistas y desempleados	PENSIONES	31.400.000 €
Baleares	0	2006	Asalariados	PENSIONES	0 €
Baleares	139.771	2006	Pensionistas	PENSIONES	1.396.770.000 €
Baleares	0	2006	Desempleados	PENSIONES	0 €
Baleares	23.554	2006	Asalariados y pensionistas	PENSIONES	173.592.000 €
Baleares	0	2006	Asalariados y desempleados	PENSIONES	0 €
					2.202.705.787.313 €

Página 1 of 119 (1785 elementos)





Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

4.2 TOTAL RETENCIÓN

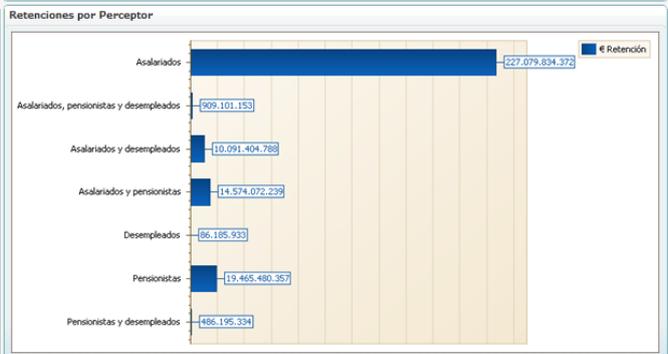
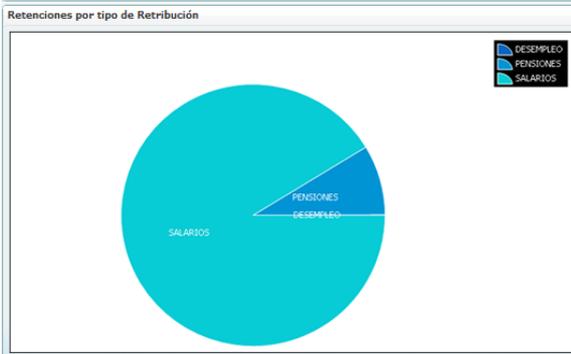
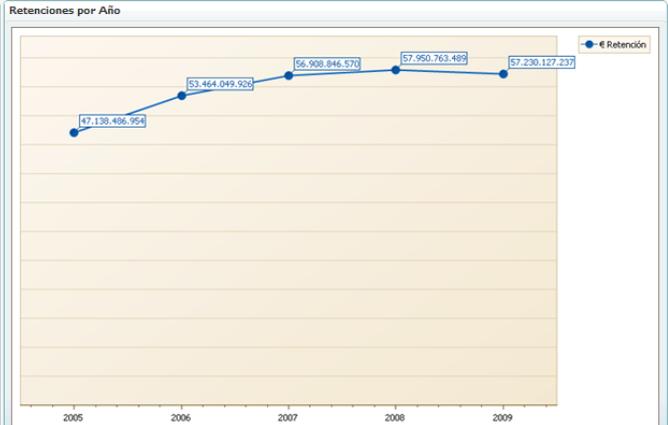
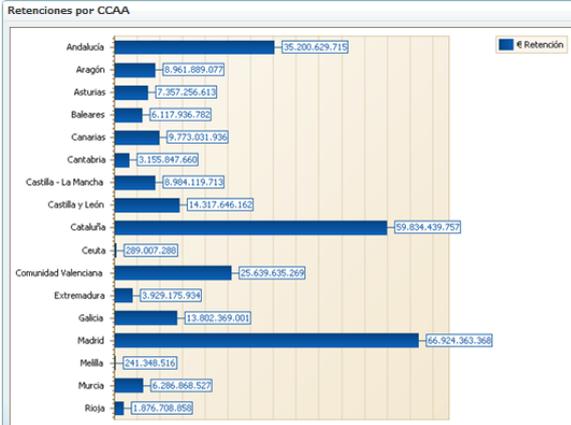
Se ha vuelto a optar por la solución del sumatorio al final de las retenciones para conocer el total de retenciones de los parámetros filtrados.

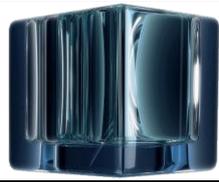
Informe donde se muestran todas las retenciones según los parámetros del filtro

Año: 2007 | Comunidad autónoma: Madrid | Tipo de retribución: -Todas- | Tipo de perceptor: -Todos-

NOMBRECCAA	NUMPERCEPTORES	ANYO	NOMBRETIPOPER	NOMBRETIPORET	RETENCIONES
Madrid	2.607.238	2007	Asalariados	SALARIOS	12.164.433.902 €
Madrid	0	2007	Pensionistas	SALARIOS	0 €
Madrid	0	2007	Desempleados	SALARIOS	0 €
Madrid	180.289	2007	Asalariados y pensionistas	SALARIOS	576.050.398 €
Madrid	301.713	2007	Asalariados y desempleados	SALARIOS	394.278.548 €
Madrid	0	2007	Pensionistas y desempleados	SALARIOS	0 €
Madrid	24.852	2007	Asalariados, pensionistas y desempleados	SALARIOS	22.349.404 €
Madrid	0	2007	Asalariados	PENSIONES	0 €
Madrid	853.862	2007	Pensionistas	PENSIONES	824.053.678 €
Madrid	0	2007	Desempleados	PENSIONES	0 €
Madrid	180.289	2007	Asalariados y pensionistas	PENSIONES	150.350.209 €
Madrid	0	2007	Asalariados y desempleados	PENSIONES	0 €
Madrid	16.219	2007	Pensionistas y desempleados	PENSIONES	22.588.201 €
Madrid	24.852	2007	Asalariados, pensionistas y desempleados	PENSIONES	12.547.278 €
Madrid	0	2007	Asalariados	DESEMPLEO	0 €
					14.171.410.457 €

Página 1 of 2 (21 elementos)





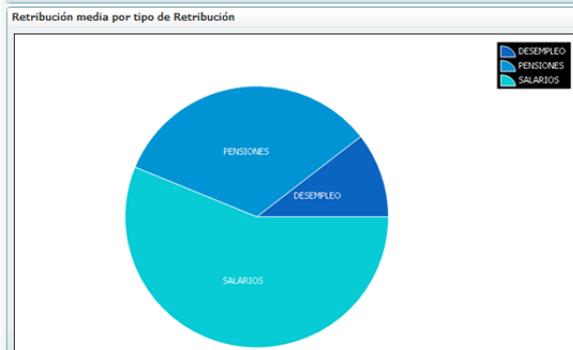
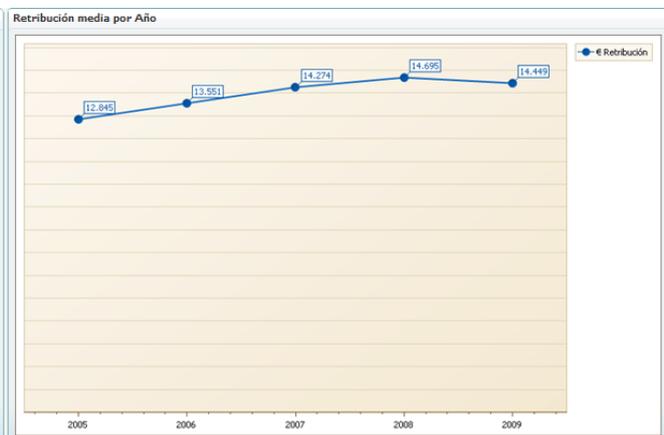
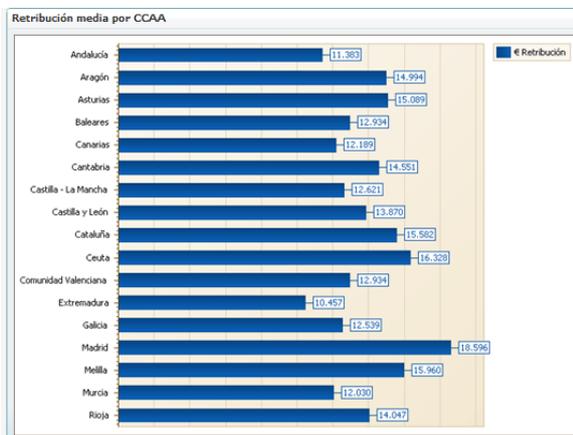
Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

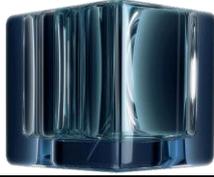
4.3 RETRIBUCIÓN MEDIA

A través de una columna calculada se calculan las retribuciones medias.

Año: -Todos- Comunidad autónoma: Cataluña Tipo de retribución: SALARIOS Tipo de perceptor: Asalariados

NOMBRECAA	NUMPERCEPTORES	ANYO	NOMBRETIPOPER	NOMBRETIPORET	CANTIDADRETRIBUCION	Retribucion Media
Cataluña	2.852.672	2005	Asalariados	SALARIOS	54.800.000.000,00 €	19.210,06 €
Cataluña	2.960.667	2006	Asalariados	SALARIOS	59.400.000.000,00 €	20.063,05 €
Cataluña	2.865.044	2007	Asalariados	SALARIOS	63.000.000.000,00 €	21.989,19 €
Cataluña	2.713.401	2008	Asalariados	SALARIOS	64.200.000.000,00 €	23.660,34 €
Cataluña	2.420.484	2009	Asalariados	SALARIOS	59.100.000.000,00 €	24.416,60 €
Sum=13812268					300.500.000.000,00 €	21.756,02 €





Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

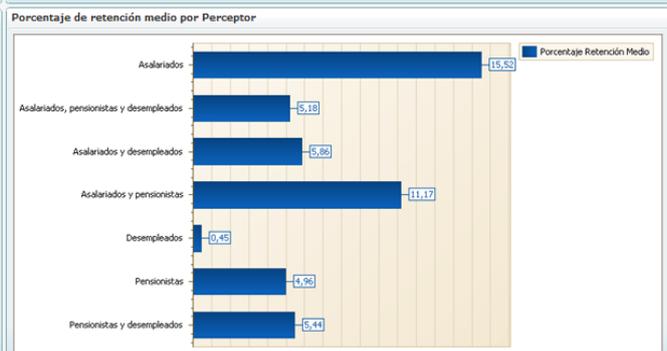
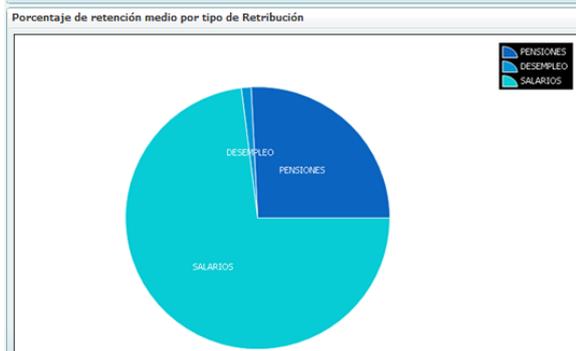
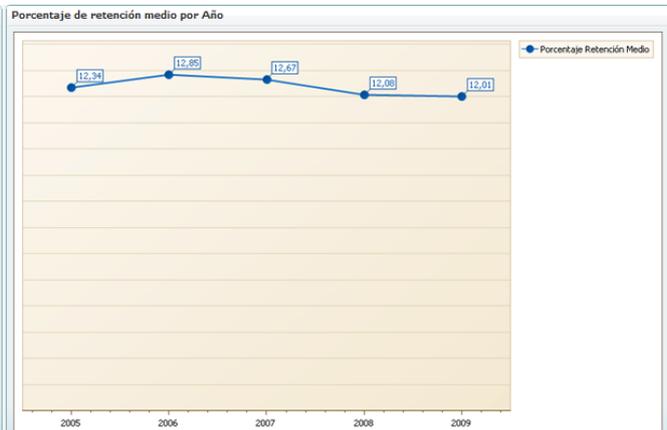
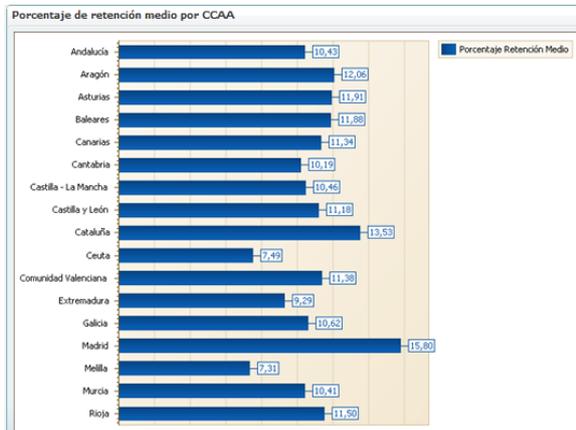
4.4 PORCENTAJE DE RETENCIÓN MEDIO

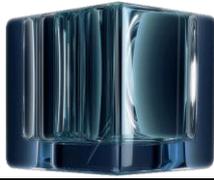
Calculamos el porcentaje de retención media a partir de una columna calculada.

Informe donde se muestra el porcentaje de retención medio según los parámetros del filtro

Año: Todos- Comunidad autónoma: Galicia Tipo de retribución: SALARIOS Tipo de perceptor: Asalariados

NOMBRECAA	NUMPERCEPTORES	ANYO	NOMBRETIPOPER	NOMBRETIPORET	Cantidad Retribución	Cantidad Retención	Retencion Media
Galicia	857.976	2005	Asalariados	SALARIOS	14000000000	1930000000	13,79%
Galicia	890.074	2006	Asalariados	SALARIOS	15300000000	2186926000	14,29%
Galicia	910.527	2007	Asalariados	SALARIOS	16700000000	2398415000	14,36%
Galicia	871.559	2008	Asalariados	SALARIOS	17200000000	2333460000	13,57%
Galicia	798.875	2009	Asalariados	SALARIOS	16600000000	2349494000	14,15%
					79.800.000.000,00 €	11.198.295.000,00 €	14,03%





Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

4.5 NÚMERO DE RETRIBUCIONES MEDIAS

Calculamos una variable auxiliar (número de percepciones) como resultado de multiplicar el número de perceptores por el número de retribuciones que perciben.

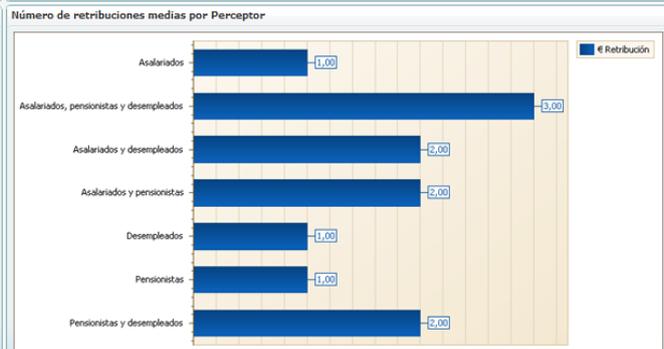
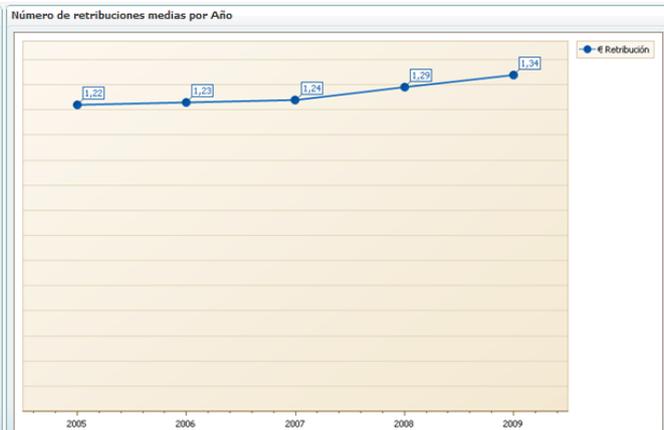
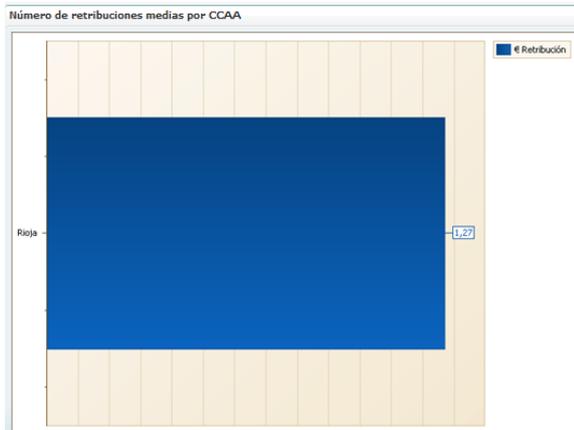
A partir de ese dato, podemos crear una columna calculada que nos proporcione la información.

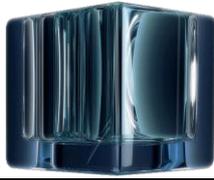
Informe donde se muestra el número de retribuciones medias según los parámetros del filtro

Año: -Todos- Comunidad autónoma: Rioja Tipo de retribución: -Todas- Tipo de perceptor: -Todos-

NOMBRECCAA	ANYO	NOMBRETIPORET	NOMBRETIPOPER	NUMPERCEPTORES	NUMPERCEPCIONES	Número de Retribuciones Medias
Rioja	2009	SALARIOS	Desempleados	0	0	0
Rioja	2009	SALARIOS	Asalariados y pensionistas	9.490	18980	2
Rioja	2009	SALARIOS	Asalariados y desempleados	25.765	51530	2
Rioja	2009	SALARIOS	Pensionistas y desempleados	0	0	0
Rioja	2009	SALARIOS	Asalariados, pensionistas y desempleados	2.108	6324	3
Rioja	2005	PENSIONES	Asalariados y pensionistas	7.316	14632	2
Rioja	2005	PENSIONES	Asalariados y desempleados	0	0	0
Rioja	2005	PENSIONES	Pensionistas y desempleados	619	1238	2
Rioja	2005	PENSIONES	Asalariados, pensionistas y desempleados	1.127	3381	3
Rioja	2006	PENSIONES	Asalariados	0	0	0
Rioja	2006	PENSIONES	Pensionistas	56.680	56680	1
Rioja	2006	PENSIONES	Desempleados	0	0	0
Rioja	2006	PENSIONES	Asalariados y pensionistas	7.595	15190	2
Rioja	2006	PENSIONES	Asalariados y desempleados	0	0	0
Rioja	2006	PENSIONES	Pensionistas y desempleados	652	1304	2
				1.161.196,00	1.478.490,00	1,27

Página 1 of 7 (105 elementos)

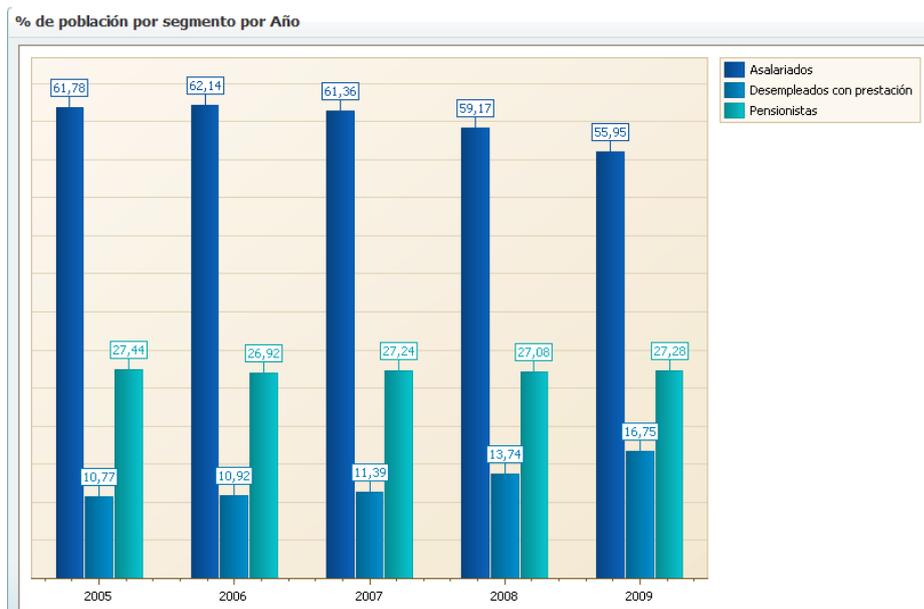
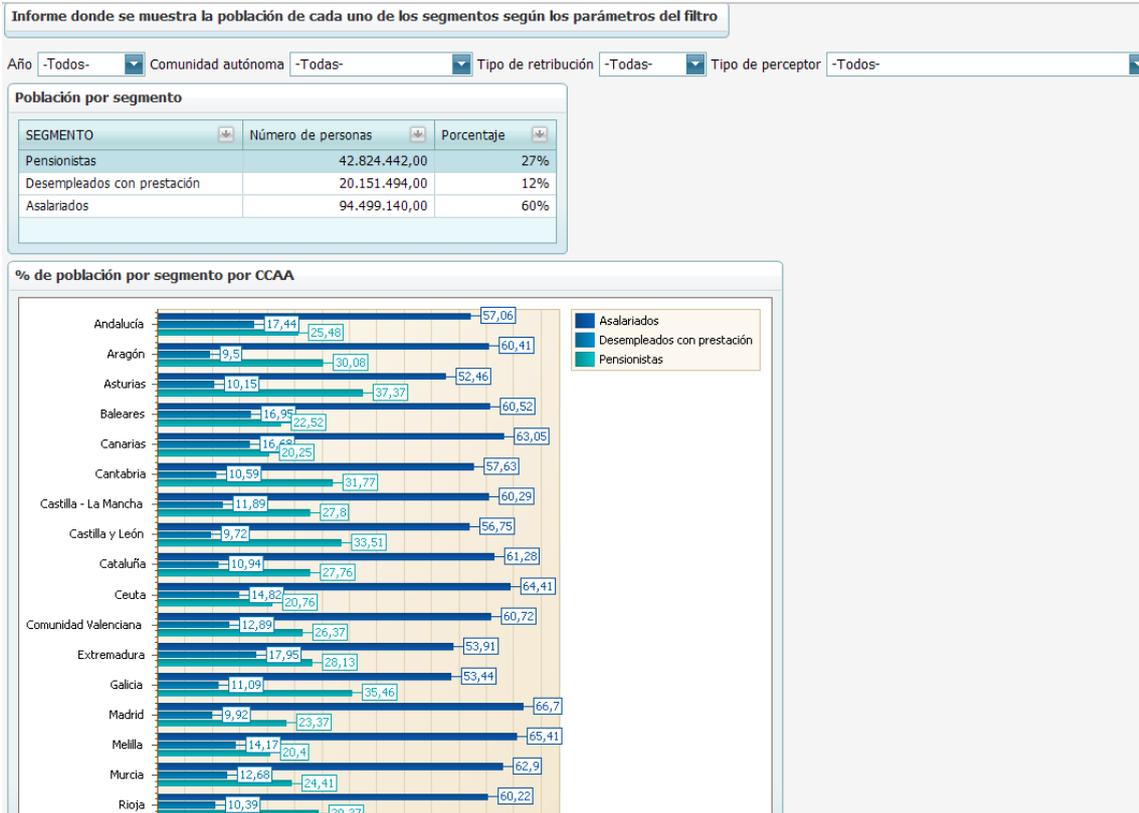


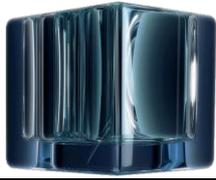


Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

4.6 PORCENTAJE DE POBLACIÓN POR SEGMENTO

Se agrupan las poblaciones por tipo de retribución que han percibido. Un individuo que haya percibido las tres clases de retribuciones, contabilizará en las tres bolsas.





Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

4.7 NÚMERO DE TRABAJADORES / NÚMERO DE PERCEPTORES NO ACTIVOS

En este informe no aplica el filtro por tipo de retribución y tipo de perceptor.

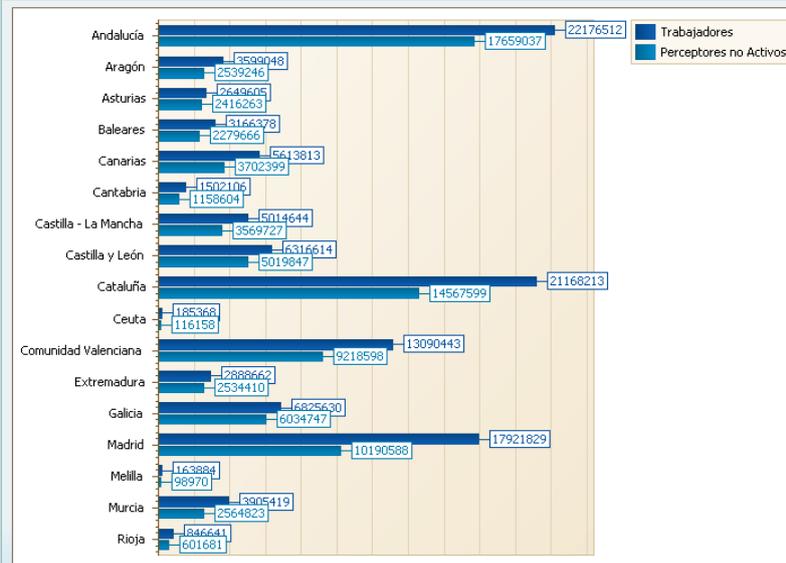
Informe donde se muestra la comparativa entre trabajadores y perceptores no activos según los parámetros del filtro

Año -Todos- Comunidad autónoma -Todas-

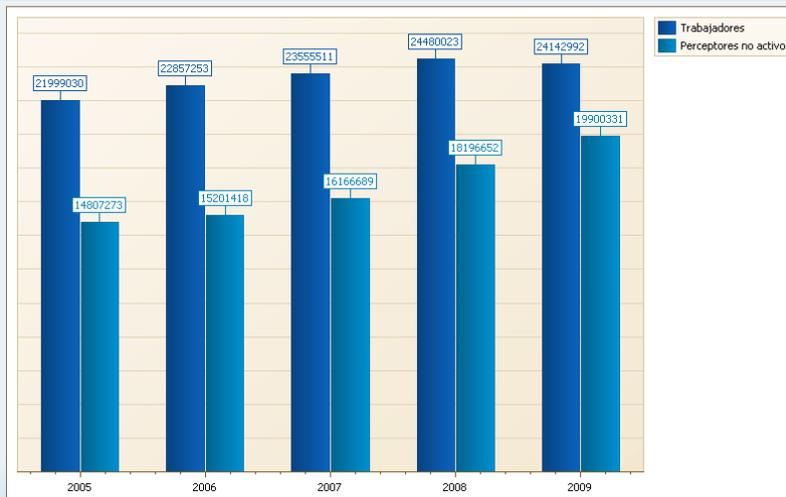
Trabajadores / Perceptores no activos

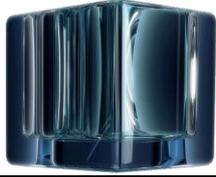
TRABAJADORES	Porcentaje Trabajadores	PERCEPTORES NO ACTIVOS	Porcentaje Perceptores No Activos
117.034.809,00	58,14%	84272363	41,86%

Trabajadores vs Perceptores de ayudas por CCAA



Trabajadores vs Perceptores de ayudas por CCAA

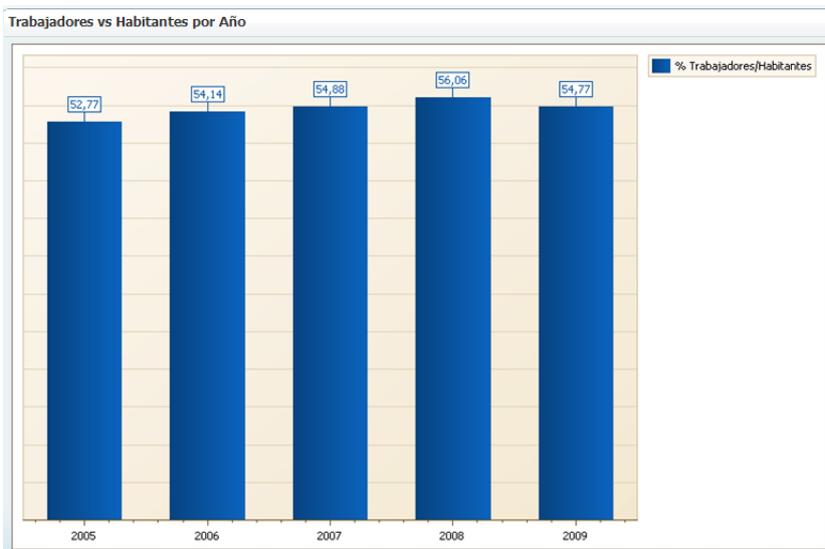


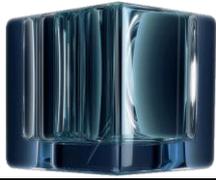


Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

4.8 NÚMERO DE TRABAJADORES / HABITANTES

En este informe no aplica el filtro por tipo de retribución y tipo de perceptor. En la tabla mostramos los totales según el filtro seleccionado, y en las gráficas distribuimos esos resultados por Comunidad Autónoma y por Año.





Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

4.9 NÚMERO DE TRABAJADORES / NÚMERO DE PERSONAS ACTIVAS

En este informe no aplica el filtro por tipo de retribución y tipo de perceptor. En la tabla mostramos los totales según el filtro seleccionado, y en las gráficas distribuimos esos resultados por Comunidad Autónoma y por Año.

Informe donde se muestra la comparativa entre trabajadores y personas activas según los parámetros del filtro

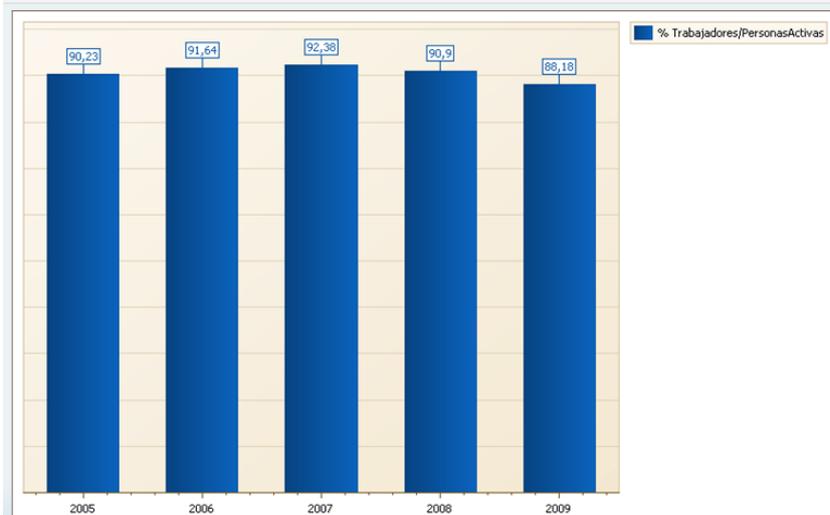
Año: -Todos- Comunidad autónoma: Madrid

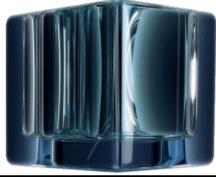
TRABAJADORES	ACTIVOS	Trabajadores/Activos
17.921.829	19.771.352	90,65%

Trabajadores vs Personas activas por CCAA



Trabajadores vs Personas activas por Año

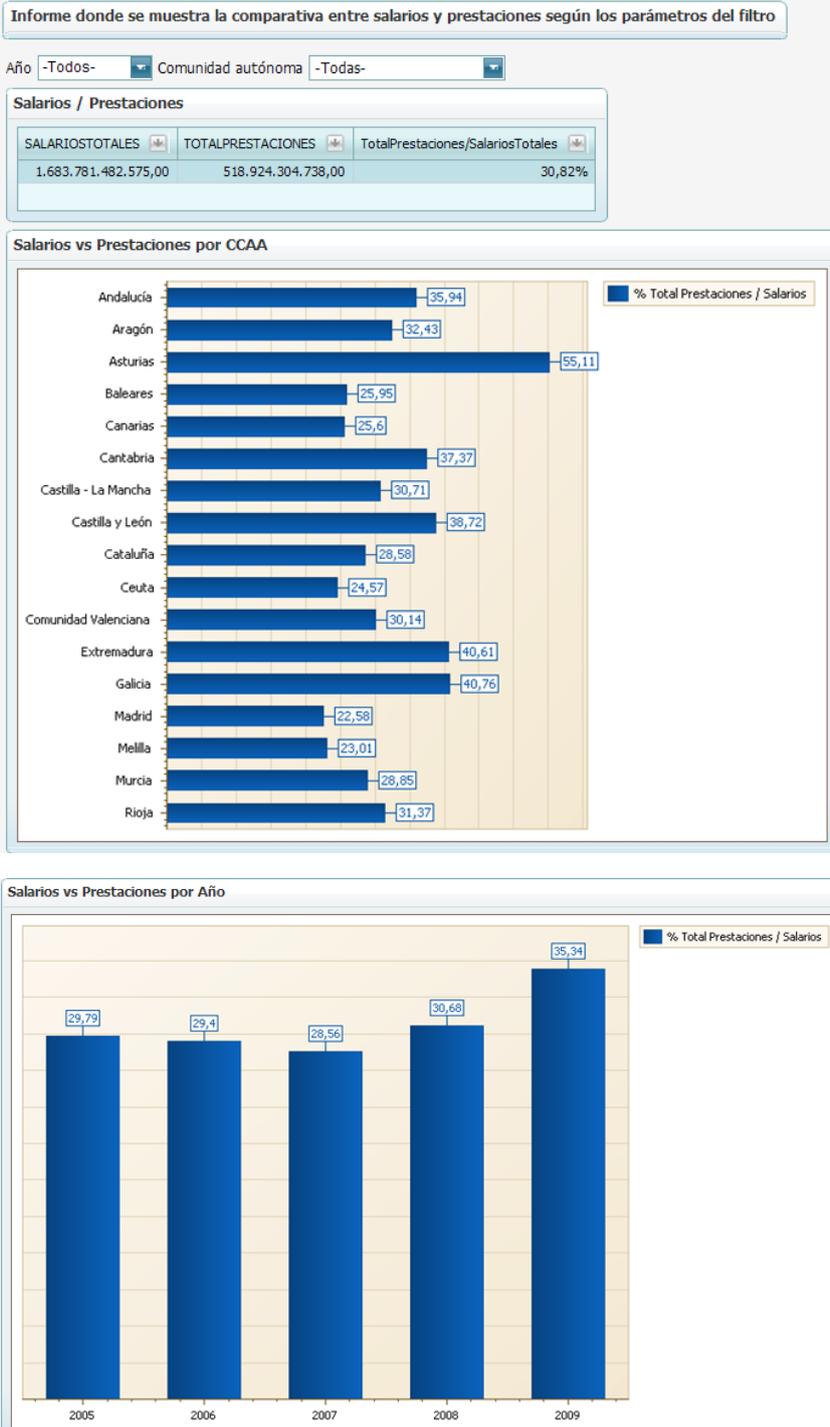


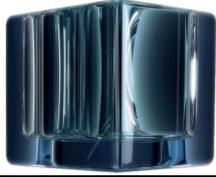


Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

4.10 SALARIOS TOTALES / TOTAL PRESTACIONES

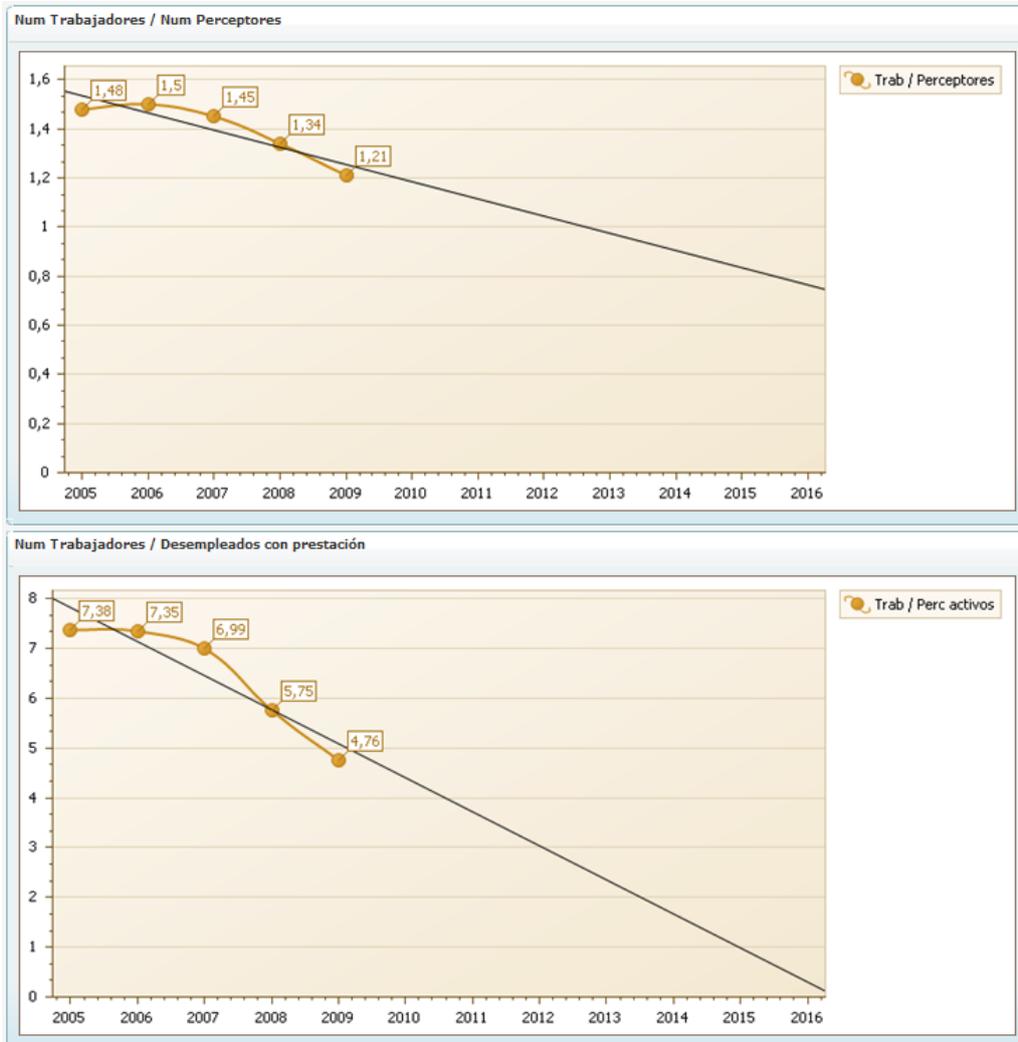
En este informe no aplica el filtro por tipo de retribución y tipo de perceptor. En la tabla mostramos los totales según el filtro seleccionado, y en las gráficas distribuimos esos resultados por Comunidad Autónoma y por Año.





Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

4.11 PROYECCIONES



	Memoria del Proyecto	Fecha: 10/06/2012
Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales		

5. CONCLUSIONES

Tras la realización del proyecto se ha llegado a las siguientes conclusiones:

- Se han cumplido los objetivos generales propuestos, ya que:
 - o He adquirido experiencia en el diseño, construcción y explotación de un almacén de datos.
 - o He desarrollado áreas técnicas en las que no tenía tan apenas conocimientos, como son el *Business Intelligence*, la carga masiva de datos y la gestión de bases de datos Oracle.
 - o El proyecto, debido a la gran cantidad de tecnologías y conocimientos que abarca, ha sido muy enriquecedor y ha cumplido mis expectativas iniciales y espero que me sea de gran utilidad en el futuro inmediato.
- Se han cumplido los requerimientos del cliente, creando el almacén de datos solicitado así como los informes y proyecciones para la explotación de la información.
- He descubierto el valor que pueden llegar a aportar las herramientas de *Business Intelligence* en el entorno empresarial para la toma de decisiones y el pronóstico de escenarios futuros.
- En cuanto a la dinámica de trabajo, creo que debería haber dedicado más tiempo a las entregas parciales ya que en ciertos aspectos no he llegado a la calidad esperada y deseada y ha provocado que la carga de trabajo en la entrega final haya sido mayor a la estimada.

6. LÍNEAS DE EVOLUCIÓN FUTURAS

La línea de evolución más clara del proyecto es la ampliación del número de Comunidades Autónomas gestionado en el almacén de datos, incorporando los datos de Navarra y País Vasco.

Esta línea de evolución permitiría tener una visión global y real de la situación de todo el estado español, y no de una parte sesgada del mismo. La única consecuencia negativa sería que al tener un régimen foral propio habría datos muy heterogéneos con respecto al resto, pero en términos de globalidad creo que sería enriquecedor para el proyecto.

Otra línea de evolución sería la incorporación de nuevas variables a nuestro almacén de datos, que nos permitieran hacer una análisis de cómo se distribuyen las retribuciones según otros factures no contemplados en el proyecto inicial como podrían ser:

- Edad de la población
- Sexo (conocemos el número total de hombres y mujeres por CCAA pero no lo tenemos vinculado a las retribuciones)
- Nivel de estudios



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

Por último, y en el contexto de crisis global en el que nos encontramos, otra línea de evolución sería el llevar nuestro almacén de datos a nivel europeo, consiguiendo la información retributiva de los países de la zona euro y de ese modo analizar el estado de situación de cada uno de ellos, la evolución y sobre todo líneas de intervención futuras para conseguir aplicar políticas que nos permitan mejorar la situación en la que nos encontramos.

7. GLOSARIO

Almacén de datos: conjunto de datos orientado a una temática concreta, integrado, variable en el tiempo y no volátil, y que da soporte a la toma de decisiones.

Procesos ETL: Procesos de extracción, transformación y carga de datos que permiten extraer la información de fuentes heterogéneas de datos y cargarla en una base de datos.

SQL: lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones en estas.

PL/SQL: lenguaje de programación que soporta las consultas SQL y además incluye el manejo de variables, estructuras modulares...

Business Intelligence: inteligencia empresarial, conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos.

8. BIBLIOGRAFÍA Y REFERENCIAS

“*Building the Data Warehouse*”, William H. Inmon
“*The Data Warehouse Toolkit*” Ralph Kimball, Margy Ross

Guía de instalación de Atlas SBI

Guía Web de uso de Atlas SBI

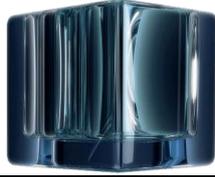
Otros enlaces consultados:

<http://es.wikipedia.org>

www.uoc.edu Biblioteca/Repositorio O2

<http://yuml.me/diagram/usecase/draw> (Creación de diagramas UML online)

<http://www.glify.com> (Creación de diagramas UML online)



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

9. ANEXOS

9.1 ANEXO 1. SCRIPTS DE CREACIÓN DE BASE DE DATOS

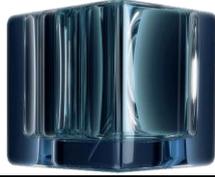
```
CREATE TABLE PERIODOTEMPORAL (  
    IDPERIODOTEMPORAL NUMBER(5,0) NOT NULL ENABLE,  
    ANYO VARCHAR2(20),  
    CONSTRAINT PK_PERIODOTEMPORAL PRIMARY KEY (IDPERIODOTEMPORAL) ENABLE  
);
```

```
CREATE TABLE ZONAGEOGRAFICA (  
    IDCCAA NUMBER(5,0) NOT NULL ENABLE,  
    NOMBRECCAA VARCHAR2(100),  
    CONSTRAINT PK_ZONAGEOGRAFICA PRIMARY KEY (IDCCAA) ENABLE  
);
```

```
CREATE TABLE TIPORETRIBUCIONES (  
    IDRET NUMBER(5,0) NOT NULL ENABLE,  
    NOMBRETIPORET VARCHAR2(100),  
    CONSTRAINT PK_TIPORETRIBUCIONES PRIMARY KEY (IDRET) ENABLE  
);
```

```
CREATE TABLE TIPOPERCEPTORES (  
    IDTIPOPER NUMBER(5,0) NOT NULL ENABLE,  
    NOMBRETIPOPER VARCHAR2(100),  
    CONSTRAINT PK_TIPOPERCEPTORES PRIMARY KEY (IDTIPOPER) ENABLE  
);
```

```
CREATE TABLE HABITANTES (  
    IDHAB NUMBER(5,0) NOT NULL ENABLE,  
    NUMEROHAB NUMBER(10,0),  
    SEXO VARCHAR2(1),
```



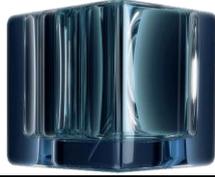
Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
PORC_ACTIVOS NUMBER (5,2),  
IDCCAA NUMBER(5,0) NOT NULL ENABLE,  
IDPERIODOTEMPORAL NUMBER(5,0) NOT NULL ENABLE,  
    CONSTRAINT FK_HABITANTES_PERIODOTEMPORAL foreign key (IDPERIODOTEMPORAL)  
REFERENCES PERIODOTEMPORAL (IDPERIODOTEMPORAL) ENABLE,  
CONSTRAINT FK_HABITANTES_ZONAGEOGRAFICA foreign key (IDCCAA)  
REFERENCES ZONAGEOGRAFICA (IDCCAA) ENABLE  
);
```

```
CREATE TABLE TMP_POBLACIONACTIVA (  
    anyo VARCHAR2(20) NOT NULL ENABLE,  
    ccaa VARCHAR2(100) NOT NULL ENABLE,  
    porcentaje NUMBER (5,2) NOT NULL ENABLE  
);
```

```
CREATE TABLE TMP_POBLACION (  
    anyo VARCHAR2(20) NOT NULL ENABLE,  
    ccaa VARCHAR2(100) NOT NULL ENABLE,  
    hombres NUMBER (10,0) NOT NULL ENABLE,  
    mujeres NUMBER (10,0) NOT NULL ENABLE  
);
```

```
CREATE TABLE TMP_TRIBUTACION (  
    anyo VARCHAR2(20) NOT NULL ENABLE,  
    ccaa VARCHAR2(100) NOT NULL ENABLE,  
    tiporet VARCHAR2(100) NOT NULL ENABLE,  
    tipoper VARCHAR2(100) NOT NULL ENABLE,  
    cantidadretribucion NUMBER (24,0) NOT NULL ENABLE,  
    numeroper NUMBER (10,0) NOT NULL ENABLE,  
    cantidadret NUMBER (24,0) NOT NULL ENABLE  
);
```

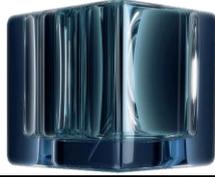


Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
CREATE TABLE HECHOS (  
    cantidadretribucion NUMBER(30,2) NOT NULL ENABLE,  
    numperceptores NUMBER (10,0) NOT NULL ENABLE,  
    retencion NUMBER (24,0) NOT NULL ENABLE,  
    idperiodotemporal NUMBER(5,0) NOT NULL ENABLE,  
    idcaa NUMBER (5,0) NOT NULL ENABLE,  
    idret NUMBER (5,0) NOT NULL ENABLE,  
    idtipoper NUMBER (5,0) NOT NULL ENABLE,  
    CONSTRAINT PK_HECHOS PRIMARY KEY (idperiodotemporal,idcaa,idret,idtipoper) ENABLE,  
    CONSTRAINT FK_HECHOS2PERIODOTEMPORAL foreign key (IDPERIODOTEMPORAL)  
        REFERENCES PERIODOTEMPORAL (IDPERIODOTEMPORAL) ENABLE,  
    CONSTRAINT FK_HECHOS2ZONAGEOGRAFICA foreign key (IDCCAA)  
        REFERENCES ZONAGEOGRAFICA (IDCCAA) ENABLE,  
    CONSTRAINT FK_HECHOS2TIPORETRIBUCIONES foreign key (IDRET)  
        REFERENCES TIPORETRIBUCIONES (IDRET) ENABLE,  
    CONSTRAINT FK_HECHOS2TIPOPERCEPTORES foreign key (IDTIPOPER)  
        REFERENCES TIPOPERCEPTORES (IDTIPOPER) ENABLE  
);
```

```
CREATE SEQUENCE SEQ_PERIODOTEMPORAL  
MINVALUE 0 MAXVALUE 9999999 INCREMENT BY 1  
START WITH 1 NOCACHE ORDER NOCYCLE ;
```

```
CREATE SEQUENCE SEQ_ZONAGEOGRAFICA  
MINVALUE 0 MAXVALUE 9999999 INCREMENT BY 1  
START WITH 1 NOCACHE ORDER NOCYCLE ;
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
CREATE SEQUENCE SEQ_TIPORETRIBUCIONES  
MINVALUE 0 MAXVALUE 9999999 INCREMENT BY 1  
START WITH 1 NOCACHE ORDER NOCYCLE ;
```

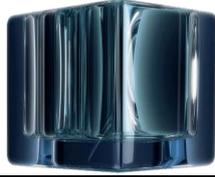
```
CREATE SEQUENCE SEQ_TIPOPERCEPTORES  
MINVALUE 0 MAXVALUE 9999999 INCREMENT BY 1  
START WITH 1 NOCACHE ORDER NOCYCLE ;
```

```
CREATE SEQUENCE SEQ_HABITANTES  
MINVALUE 0 MAXVALUE 9999999 INCREMENT BY 1  
START WITH 1 NOCACHE ORDER NOCYCLE ;
```

```
commit;
```

```
create or replace trigger AUTO_PERIODOTEMPORAL_SEQ before insert on PERIODOTEMPORAL  
for each row begin  
  if inserting then  
    if :NEW.IDPERIODOTEMPORAL is null then  
      select SEQ_PERIODOTEMPORAL.nextval into :NEW.IDPERIODOTEMPORAL  
      from dual;  
    end if;  
  end if;  
end;  
/
```

```
create or replace trigger AUTO_ZONAGEOGRAFICA_SEQ before insert on ZONAGEOGRAFICA  
for each row begin  
  if inserting then  
    if :NEW.IDCCAA is null then  
      select SEQ_ZONAGEOGRAFICA.nextval into :NEW.IDCCAA
```



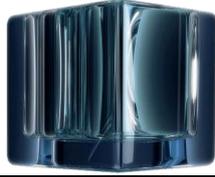
Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
        from dual;
    end if;
end if;
end;
/

create or replace trigger AUTO_TIPOPERCEPTORES_SEQ before insert on TIPOPERCEPTORES
for each row begin
    if inserting then
        if :NEW.IDTIOPER is null then
            select SEQ_TIPOPERCEPTORES.nextval into :NEW.IDTIOPER
            from dual;
        end if;
    end if;
end;
/

create or replace trigger AUTO_TIPORETRIBUCIONES_SEQ before insert on TIPORETRIBUCIONES
for each row begin
    if inserting then
        if :NEW.IDRET is null then
            select SEQ_TIPORETRIBUCIONES.nextval into :NEW.IDRET
            from dual;
        end if;
    end if;
end;
/

create or replace trigger AUTO_HABITANTES_SEQ before insert on HABITANTES
for each row begin
    if inserting then
        if :NEW.IDHAB is null then
            select SEQ_HABITANTES.nextval into :NEW.IDHAB
            from dual;
        end if;
    end if;
end;
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
end if;  
end;
```

9.2 ANEXO 2. SCRIPT DE TRANSFORMACIÓN DE DATOS VOLCADOS

```
UPDATE TMP_TRIBUTACION  
SET TIOPER ='Asalariados, pensionistas y desempleados'  
WHERE TIOPER ='Asalariados pensionistas y desempleados';  
COMMIT;
```

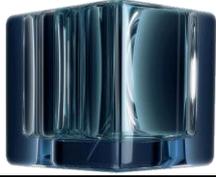
```
UPDATE TMP_TRIBUTACION  
SET CCAA ='Cataluña'  
WHERE CCAA ='Catalunya';  
COMMIT;
```

```
UPDATE TMP_POBLACION  
SET CCAA ='Cataluña'  
WHERE CCAA ='Catalunya';  
COMMIT;
```

```
UPDATE TMP_POBLACIONACTIVA  
SET CCAA ='Cataluña'  
WHERE CCAA ='Catalunya';  
COMMIT;
```

```
UPDATE TMP_TRIBUTACION  
SET CCAA ='Asturias'  
WHERE CCAA ='Principado de Asturias';  
COMMIT;
```

```
UPDATE TMP_POBLACION  
SET CCAA ='Asturias'
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
WHERE CCAA ='Principado de Asturias';
```

```
COMMIT;
```

```
UPDATE TMP_POBLACIONACTIVA
```

```
SET CCAA ='Asturias'
```

```
WHERE CCAA ='Principado de Asturias';
```

```
COMMIT;
```

```
UPDATE TMP_TRIBUTACION
```

```
SET CCAA ='Comunidad Valenciana'
```

```
WHERE CCAA ='Comunitat Valenciana';
```

```
COMMIT;
```

```
UPDATE TMP_POBLACION
```

```
SET CCAA ='Comunidad Valenciana'
```

```
WHERE CCAA ='Comunitat Valenciana';
```

```
COMMIT;
```

```
UPDATE TMP_POBLACIONACTIVA
```

```
SET CCAA ='Comunidad Valenciana'
```

```
WHERE CCAA ='Comunitat Valenciana';
```

```
COMMIT;
```

```
UPDATE TMP_TRIBUTACION
```

```
SET CCAA ='Comunidad Valenciana'
```

```
WHERE CCAA ='COMUNITAT VALENCIANA';
```

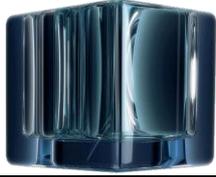
```
COMMIT;
```

```
UPDATE TMP_POBLACION
```

```
SET CCAA ='Comunidad Valenciana'
```

```
WHERE CCAA ='COMUNITAT VALENCIANA';
```

```
COMMIT;
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
UPDATE TMP_POBLACIONACTIVA  
  
SET CCAA ='Comunidad Valenciana'  
  
WHERE CCAA ='COMUNITAT VALENCIANA';  
  
COMMIT;
```

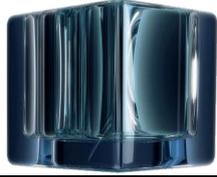
```
UPDATE TMP_TRIBUTACION  
  
SET CCAA ='Melilla'  
  
WHERE CCAA ='Ciudad autónoma de Melilla';  
  
COMMIT;
```

```
UPDATE TMP_POBLACION  
  
SET CCAA ='Melilla'  
  
WHERE CCAA ='Ciudad autónoma de Melilla';  
  
COMMIT;
```

```
UPDATE TMP_POBLACIONACTIVA  
  
SET CCAA ='Melilla'  
  
WHERE CCAA ='Ciudad autónoma de Melilla';  
  
COMMIT;
```

```
UPDATE TMP_TRIBUTACION  
  
SET CCAA ='Ceuta'  
  
WHERE CCAA ='Ciudad autónoma de Ceuta';  
  
COMMIT;
```

```
UPDATE TMP_POBLACION  
  
SET CCAA ='Ceuta'  
  
WHERE CCAA ='Ciudad autónoma de Ceuta';  
  
COMMIT;
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
UPDATE TMP_POBLACIONACTIVA  
  
SET CCAA ='Ceuta'  
  
WHERE CCAA ='Ciudad autónoma de Ceuta';  
  
COMMIT;
```

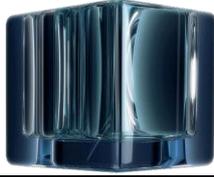
```
UPDATE TMP_TRIBUTACION  
  
SET CCAA ='Baleares'  
  
WHERE CCAA ='Baleares' OR CCAA='Illes Balears';  
  
COMMIT;
```

```
UPDATE TMP_POBLACION  
  
SET CCAA ='Baleares'  
  
WHERE CCAA ='Baleares' OR CCAA='Illes Balears';  
  
COMMIT;
```

```
UPDATE TMP_POBLACIONACTIVA  
  
SET CCAA ='Baleares'  
  
WHERE CCAA ='Baleares' OR CCAA='Illes Balears';  
  
COMMIT;
```

```
UPDATE TMP_TRIBUTACION  
  
SET CCAA ='Madrid'  
  
WHERE CCAA LIKE 'Comunidad de Madrid';  
  
COMMIT;
```

```
UPDATE TMP_POBLACION  
  
SET CCAA ='Madrid'  
  
WHERE CCAA LIKE 'Comunidad de Madrid';  
  
COMMIT;
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
UPDATE TMP_POBLACIONACTIVA
SET CCAA ='Madrid'
WHERE CCAA LIKE 'Comunidad de Madrid';
COMMIT;
```

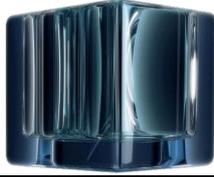
```
UPDATE TMP_TRIBUTACION
SET CCAA ='Madrid'
WHERE CCAA LIKE 'COMUNIDAD DE MADRID';
COMMIT;
```

```
UPDATE TMP_POBLACION
SET CCAA ='Madrid'
WHERE CCAA LIKE 'COMUNIDAD DE MADRID';
COMMIT;
```

```
UPDATE TMP_POBLACIONACTIVA
SET CCAA ='Madrid'
WHERE CCAA LIKE 'COMUNIDAD DE MADRID';
COMMIT;
```

```
UPDATE TMP_POBLACION
SET ccaa=SUBSTR(
CCAA,
1,
INSTR(CCAA,' ')-1 )
WHERE ccaa like '% (%)';
```

```
UPDATE TMP_POBLACIONACTIVA
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

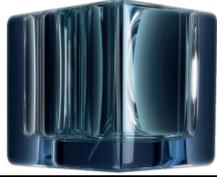
```
SET ccaa=SUBSTR(  
  CCAA,  
  1,  
  INSTR(CCAA,'(')-1 )  
WHERE ccaa like '% (%)';
```

```
UPDATE TMP_TRIBUTACION  
  
SET ccaa=SUBSTR(  
  CCAA,  
  1,  
  INSTR(CCAA,'(')-1 )  
WHERE ccaa like '% (%)';
```

```
UPDATE TMP_POBLACION  
  
SET ccaa=SUBSTR(  
  CCAA,  
  1,  
  INSTR(CCAA,',' )-1 )  
WHERE ccaa like '%, %';
```

```
UPDATE TMP_POBLACIONACTIVA  
  
SET ccaa=SUBSTR(  
  CCAA,  
  1,  
  INSTR(CCAA,',' )-1 )  
WHERE ccaa like '%, %';
```

```
UPDATE TMP_TRIBUTACION  
  
SET ccaa=SUBSTR(  
  CCAA,  
  1,  
  INSTR(CCAA,',' )-1 )
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
WHERE ccaa like '%: %';
```

```
UPDATE TMP_POBLACION
```

```
SET CCAA=trim(ccaa);
```

```
UPDATE TMP_POBLACIONACTIVA
```

```
SET CCAA=trim(ccaa);
```

```
UPDATE TMP_TRIBUTACION
```

```
SET CCAA=trim(ccaa);
```

```
--Actualizar habitantes segun media de año actual + año proximo
```

```
CREATE GLOBAL TEMPORARY TABLE aux_table_poblacion (
```

```
zone VARCHAR2(100),
```

```
year NUMBER,
```

```
newhab NUMBER(10,2)
```

```
) ON COMMIT PRESERVE ROWS;
```

```
INSERT INTO aux_table_poblacion (zone,year,newhab)
```

```
SELECT T1.CCAA,T1.ANYO,(T1.HOMBRES+NVL((SELECT T2.HOMBRES FROM TMP_POBLACION T2 WHERE  
T2.CCAA=T1.CCAA AND T2.ANYO=(T1.ANYO+1)),T1.HOMBRES))/2
```

```
FROM TMP_POBLACION T1;
```

```
UPDATE TMP_POBLACION
```

```
SET TMP_POBLACION.HOMBRES=aux_table_poblacion.newhab
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
FROM TMP_POBLACION INNER JOIN aux_table_poblacion ON ccaa=zone and anyo=year;
```

```
UPDATE TMP_POBLACION
```

```
SET HOMBRES= (SELECT newhab from aux_table_poblacion where aux_table_poblacion.zone=tmp_poblacion.ccaa and  
aux_table_poblacion.year=tmp_poblacion.anyo)
```

```
DELETE FROM aux_table_poblacion;
```

```
INSERT INTO aux_table_poblacion (zone,year,newhab)
```

```
SELECT T1.CCAA,T1.ANYO,(T1.MUJERES+NVL((SELECT T2.MUJERES FROM TMP_POBLACION T2 WHERE T2.CCAA=T1.CCAA  
AND T2.ANYO=(T1.ANYO+1)),T1.MUJERES))/2
```

```
FROM TMP_POBLACION T1;
```

```
UPDATE TMP_POBLACION
```

```
SET MUJERES= (SELECT newhab from aux_table_poblacion where aux_table_poblacion.zone=tmp_poblacion.ccaa and  
aux_table_poblacion.year=tmp_poblacion.anyo)
```

```
--Homogeneización de datos en casos que nos pasan valores medios:
```

```
UPDATE TMP_TRIBUTACION
```

```
SET CANTIDADRETRIBUCION=CANTIDADRETRIBUCION*NUMEROPER WHERE UPPER(CCAA) LIKE '%MADRID%' OR  
UPPER(CCAA) LIKE '%VALENCIA%' OR UPPER(CCAA) LIKE '%MURCIA%';
```

```
UPDATE TMP_TRIBUTACION
```

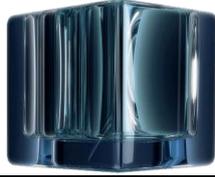
```
SET CANTIDADRET=CANTIDADRETRIBUCION*CANTIDADRET/100
```

```
WHERE UPPER(CCAA) LIKE '%MADRID%' OR UPPER(CCAA) LIKE '%VALENCIA%' OR UPPER(CCAA) LIKE '%MURCIA%';
```

```
--Arreglo de datos de retribuciones de años superiores a 2005
```

```
UPDATE TMP_TRIBUTACION
```

```
SET CANTIDADRETRIBUCION=CANTIDADRETRIBUCION*1000
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
WHERE ANYO>2005 AND UPPER(CCAA) NOT LIKE '%MADRID%' AND UPPER(CCAA) NOT LIKE '%VALENCIA%' AND  
UPPER(CCAA) NOT LIKE '%MURCIA%';
```

```
UPDATE TMP_TRIBUTACION
```

```
SET CANTIDADRET=CANTIDADRET*1000
```

```
WHERE ANYO>2005 AND UPPER(CCAA) NOT LIKE '%MADRID%' AND UPPER(CCAA) NOT LIKE '%VALENCIA%' AND  
UPPER(CCAA) NOT LIKE '%MURCIA%';
```

9.3 ANEXO 3. SCRIPT DE VOLCADO DE DATOS EN TABLAS DEFINITIVAS

```
--Insercion de años (tomamos sólo los comunes)
```

```
INSERT INTO PERIODOTEMPORAL (ANYO)
```

```
SELECT DISTINCT ANYO FROM TMP_POBLACION
```

```
WHERE ANYO IN (
```

```
SELECT DISTINCT ANYO FROM TMP_POBLACIONACTIVA
```

```
)
```

```
AND ANYO IN (
```

```
SELECT DISTINCT ANYO FROM TMP_TRIBUTACION
```

```
)
```

```
ORDER BY ANYO
```

```
--Insercion de ccaas (tomamos sólo los comunes)
```

```
INSERT INTO ZONAGEOGRAFICA (NOMBRECCAA)
```

```
SELECT DISTINCT CCAA FROM TMP_POBLACION
```

```
WHERE LOWER(CCAA) IN (
```

```
SELECT DISTINCT LOWER(CCAA) FROM TMP_POBLACIONACTIVA
```

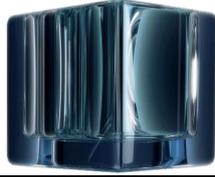
```
)
```

```
AND LOWER(CCAA) IN (
```

```
SELECT DISTINCT LOWER(CCAA) FROM TMP_TRIBUTACION
```

```
)
```

```
ORDER BY CCAA
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

--Insercion de habitantes (cualquier vale, con LEFTS+WHERE o INNERS)

```
INSERT INTO HABITANTES (NUMEROHAB,SEXO,PORC_ACTIVOS,IDCCAA ,IDPERIODOTEMPORAL)
```

```
SELECT HOMBRES,'H',PORCENTAJE,IDCCAA,IDPERIODOTEMPORAL
```

```
FROM TMP_POBLACION
```

```
LEFT JOIN TMP_POBLACIONACTIVA ON TMP_POBLACION.ANYO=TMP_POBLACIONACTIVA.ANYO AND  
TMP_POBLACION.CCAA=TMP_POBLACIONACTIVA.CCAA
```

```
LEFT JOIN ZONAGEOGRAFICA ON TMP_POBLACION.CCAA=ZONAGEOGRAFICA.NOMBRECCAA
```

```
LEFT JOIN PERIODOTEMPORAL ON TMP_POBLACION.ANYO=PERIODOTEMPORAL.ANYO
```

```
WHERE IDCCAA is not null and idperiodotemporal is not null;
```

```
INSERT INTO HABITANTES (NUMEROHAB,SEXO,PORC_ACTIVOS,IDCCAA ,IDPERIODOTEMPORAL)
```

```
SELECT HOMBRES,'H',PORCENTAJE,IDCCAA,IDPERIODOTEMPORAL
```

```
FROM TMP_POBLACION
```

```
INNER JOIN TMP_POBLACIONACTIVA ON TMP_POBLACION.ANYO=TMP_POBLACIONACTIVA.ANYO AND  
TMP_POBLACION.CCAA=TMP_POBLACIONACTIVA.CCAA
```

```
INNER JOIN ZONAGEOGRAFICA ON TMP_POBLACION.CCAA=ZONAGEOGRAFICA.NOMBRECCAA
```

```
INNER JOIN PERIODOTEMPORAL ON TMP_POBLACION.ANYO=PERIODOTEMPORAL.ANYO
```

```
INSERT INTO HABITANTES (NUMEROHAB,SEXO,PORC_ACTIVOS,IDCCAA ,IDPERIODOTEMPORAL)
```

```
SELECT MUJERES,'M',PORCENTAJE,IDCCAA,IDPERIODOTEMPORAL
```

```
FROM TMP_POBLACION
```

```
INNER JOIN TMP_POBLACIONACTIVA ON TMP_POBLACION.ANYO=TMP_POBLACIONACTIVA.ANYO AND  
TMP_POBLACION.CCAA=TMP_POBLACIONACTIVA.CCAA
```

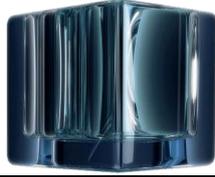
```
INNER JOIN ZONAGEOGRAFICA ON TMP_POBLACION.CCAA=ZONAGEOGRAFICA.NOMBRECCAA
```

```
INNER JOIN PERIODOTEMPORAL ON TMP_POBLACION.ANYO=PERIODOTEMPORAL.ANYO
```

--Insercion en tabla de hechos

--Ponemos el Lower porque algunas no las cogía por Case Sensitive

```
INSERT INTO HECHOS (IDPERIODOTEMPORAL,IDCCAA ,IDRET ,IDTIPOPER  
,CANTIDADRETRIBUCION,NUMPERCEPTORES ,RETENCION)
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
SELECT IDPERIODOTEMPORAL, IDCCAA, IDRET, IDTIPOPER, CANTIDADRETRIBUCION, NUMEROPER, CANTIDADRET
FROM TMP_TRIBUTACION

INNER JOIN ZONAGEOGRAFICA ON
LOWER(TMP_TRIBUTACION.CCAA)=LOWER(ZONAGEOGRAFICA.NOMBRECCAA)

INNER JOIN PERIODOTEMPORAL ON LOWER(TMP_TRIBUTACION.ANYO)=LOWER(PERIODOTEMPORAL.ANYO)

INNER JOIN TIPOPERCEPTORES ON
LOWER(TMP_TRIBUTACION.TIPOPER)=LOWER(TIPOPERCEPTORES.NOMBRETIPOPER)

INNER JOIN TIPORETRIBUCIONES ON
LOWER(TMP_TRIBUTACION.TIPORET)=LOWER(TIPORETRIBUCIONES.NOMBRETIPORET)
```

9.4 ANEXO 4. ORIGEN DE DATOS DE LAS TABLAS DE LOS INFORMES EN ATLAS SBI

6.1 TOTAL RETRIBUCIONES

```
SELECT
ZONAGEOGRAFICA.NOMBRECCAA, PERIODOTEMPORAL.ANYO, TIPOPERCEPTORES.NOMBRETIPOPER, TIPORETRIBUCIONES.N
OMBRETIPORET, NUMPERCEPTORES, TO_NUMBER(CANTIDADRETRIBUCION)

FROM HECHOS

LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCCAA=HECHOS.IDCCAA

LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=HECHOS.IDPERIODOTEMPORAL

LEFT JOIN TIPOPERCEPTORES ON TIPOPERCEPTORES.IDTIPOPER=HECHOS.IDTIPOPER

LEFT JOIN TIPORETRIBUCIONES ON TIPORETRIBUCIONES.IDRET=HECHOS.IDRET

WHERE ($[ControlParameter:ANYO]$=IDPERIODOTEMPORAL OR $[ControlParameter:ANYO]$=0)

AND ($[ControlParameter:CCAA]$=HECHOS.IDCCAA OR $[ControlParameter:CCAA]$=0)

AND ($[ControlParameter:TIPORET]$=IDRET OR $[ControlParameter:TIPORET]$=0)

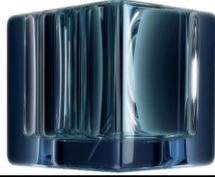
AND ($[ControlParameter:TIPOPER]$=IDTIPOPER OR $[ControlParameter:TIPOPER]$=0)
```

6.2 TOTAL RETENCIÓN

```
SELECT
ZONAGEOGRAFICA.NOMBRECCAA, PERIODOTEMPORAL.ANYO, TIPOPERCEPTORES.NOMBRETIPOPER, TIPORETRIBUCIONES.N
OMBRETIPORET, NUMPERCEPTORES, RETENCION

FROM HECHOS

LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCCAA=HECHOS.IDCCAA
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

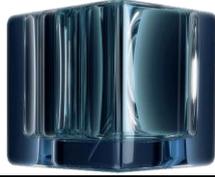
```
LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=HECHOS.IDPERIODOTEMPORAL
LEFT JOIN TIPOPERCEPTORES ON TIPOPERCEPTORES.IDTIPOPER=HECHOS.IDTIPOPER
LEFT JOIN TIPORETRIBUCIONES ON TIPORETRIBUCIONES.IDRET=HECHOS.IDRET
WHERE ([ControlParameter:ANYO]$=IDPERIODOTEMPORAL OR [ControlParameter:ANYO]$=0)
AND ([ControlParameter:CCAA]$=HECHOS.IDCCAA OR [ControlParameter:CCAA]$=0)
AND ([ControlParameter:TIPORET]$=IDRET OR [ControlParameter:TIPORET]$=0)
AND ([ControlParameter:TIPOPER]$=IDTIPOPER OR [ControlParameter:TIPOPER]$=0)
```

6.3 RETRIBUCIÓN MEDIA

```
SELECT
ZONAGEOGRAFICA.NOMBRECCAA,PERIODOTEMPORAL.ANYO,TIPOPERCEPTORES.NOMBRETIPOPER,TIPORETRIBUCIONES.N
OMBRETIPORET,NUMPERCEPTORES,CANTIDADRETRIBUCION
FROM HECHOS
LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCCAA=HECHOS.IDCCAA
LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=HECHOS.IDPERIODOTEMPORAL
LEFT JOIN TIPOPERCEPTORES ON TIPOPERCEPTORES.IDTIPOPER=HECHOS.IDTIPOPER
LEFT JOIN TIPORETRIBUCIONES ON TIPORETRIBUCIONES.IDRET=HECHOS.IDRET
WHERE ([ControlParameter:ANYO]$=IDPERIODOTEMPORAL OR [ControlParameter:ANYO]$=0)
AND ([ControlParameter:CCAA]$=HECHOS.IDCCAA OR [ControlParameter:CCAA]$=0)
AND ([ControlParameter:TIPORET]$=IDRET OR [ControlParameter:TIPORET]$=0)
AND ([ControlParameter:TIPOPER]$=IDTIPOPER OR [ControlParameter:TIPOPER]$=0)
```

6.4 PORCENTAJE DE RETENCIÓN MEDIO

```
SELECT
ZONAGEOGRAFICA.NOMBRECCAA,PERIODOTEMPORAL.ANYO,TIPOPERCEPTORES.NOMBRETIPOPER,TIPORETRIBUCIONES.N
OMBRETIPORET,NUMPERCEPTORES,CANTIDADRETRIBUCION,RETENCION
FROM HECHOS
LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCCAA=HECHOS.IDCCAA
LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=HECHOS.IDPERIODOTEMPORAL
LEFT JOIN TIPOPERCEPTORES ON TIPOPERCEPTORES.IDTIPOPER=HECHOS.IDTIPOPER
LEFT JOIN TIPORETRIBUCIONES ON TIPORETRIBUCIONES.IDRET=HECHOS.IDRET
WHERE ([ControlParameter:ANYO]$=IDPERIODOTEMPORAL OR [ControlParameter:ANYO]$=0)
AND ([ControlParameter:CCAA]$=HECHOS.IDCCAA OR [ControlParameter:CCAA]$=0)
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

AND (\$[ControlParameter:TIPORET]\$=IDRET OR \$[ControlParameter:TIPORET]\$=0)

AND (\$[ControlParameter:TIPOPER]\$=IDTIPOPER OR \$[ControlParameter:TIPOPER]\$=0)

6.5 NÚMERO DE RETRIBUCIONES MEDIAS

SELECT

ZONAGEOGRAFICA.NOMBRECAA,PERIODOTEMPORAL.ANYO,TIPOPERCEPTORES.NOMBRETIPOPER,TIPORETRIBUCIONES.NOMBRETIPORET,NUMPERCEPTORES,

(case

when LOWER(TIPOPERCEPTORES.NOMBRETIPOPER) LIKE 'asalariados' OR LOWER(TIPOPERCEPTORES.NOMBRETIPOPER) LIKE 'desempleados' OR LOWER(TIPOPERCEPTORES.NOMBRETIPOPER) LIKE 'pensionistas' then NUMPERCEPTORES

when LOWER(TIPOPERCEPTORES.NOMBRETIPOPER) LIKE 'asalariados y desempleados' OR LOWER(TIPOPERCEPTORES.NOMBRETIPOPER) LIKE 'pensionistas y desempleados' OR LOWER(TIPOPERCEPTORES.NOMBRETIPOPER) LIKE 'asalariados y pensionistas' then NUMPERCEPTORES*2

when LOWER(TIPOPERCEPTORES.NOMBRETIPOPER) LIKE 'asalariados, pensionistas y desempleados' then NUMPERCEPTORES*3

else 0

end) as numpercepciones

FROM HECHOS

LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCAA=HECHOS.IDCAA

LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=HECHOS.IDPERIODOTEMPORAL

LEFT JOIN TIPOPERCEPTORES ON TIPOPERCEPTORES.IDTIPOPER=HECHOS.IDTIPOPER

LEFT JOIN TIPORETRIBUCIONES ON TIPORETRIBUCIONES.IDRET=HECHOS.IDRET

WHERE (\$[ControlParameter:ANYO]\$=IDPERIODOTEMPORAL OR \$[ControlParameter:ANYO]\$=0)

AND (\$[ControlParameter:CAA]\$=HECHOS.IDCAA OR \$[ControlParameter:CAA]\$=0)

AND (\$[ControlParameter:TIPORET]\$=IDRET OR \$[ControlParameter:TIPORET]\$=0)

AND (\$[ControlParameter:TIPOPER]\$=IDTIPOPER OR \$[ControlParameter:TIPOPER]\$=0)

6.6 PORCENTAJE DE POBLACIÓN POR SEGMENTO

SELECT

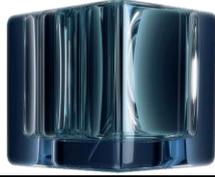
(CASE

WHEN UPPER(NOMBRETIPORET)='PENSIONES' THEN 'Pensionistas'

WHEN UPPER(NOMBRETIPORET)='SALARIOS' THEN 'Asalariados'

WHEN UPPER(NOMBRETIPORET)='DESEMPLEO' THEN 'Desempleados con prestación'

END) AS Segmento

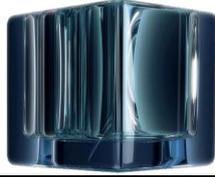


Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
,SUM(NUMPERCEPTORES)
,TRUNC(SUM(NUMPERCEPTORES)/(SELECT SUM(NUMPERCEPTORES)
FROM HECHOS
LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCCAA=HECHOS.IDCCAA
LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=HECHOS.IDPERIODOTEMPORAL
LEFT JOIN TIPORETRIBUCIONES ON TIPORETRIBUCIONES.IDRET=HECHOS.IDRET
LEFT JOIN TIPOPERCEPTORES ON TIPOPERCEPTORES.IDTIPOPER=HECHOS.IDTIPOPER
WHERE ($[ControlParameter:ANYO]$=IDPERIODOTEMPORAL OR $[ControlParameter:ANYO]$=0)
AND ($[ControlParameter:CCAA]$=HECHOS.IDCCAA OR $[ControlParameter:CCAA]$=0)
AND ($[ControlParameter:TIPORET]$=IDRET OR $[ControlParameter:TIPORET]$=0)
AND ($[ControlParameter:TIPOPER]$=IDTIPOPER OR $[ControlParameter:TIPOPER]$=0)
),2)
FROM HECHOS
LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCCAA=HECHOS.IDCCAA
LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=HECHOS.IDPERIODOTEMPORAL
LEFT JOIN TIPORETRIBUCIONES ON TIPORETRIBUCIONES.IDRET=HECHOS.IDRET
LEFT JOIN TIPOPERCEPTORES ON TIPOPERCEPTORES.IDTIPOPER=HECHOS.IDTIPOPER
WHERE ($[ControlParameter:ANYO]$=IDPERIODOTEMPORAL OR $[ControlParameter:ANYO]$=0)
AND ($[ControlParameter:CCAA]$=HECHOS.IDCCAA OR $[ControlParameter:CCAA]$=0)
AND ($[ControlParameter:TIPORET]$=IDRET OR $[ControlParameter:TIPORET]$=0)
AND ($[ControlParameter:TIPOPER]$=IDTIPOPER OR $[ControlParameter:TIPOPER]$=0)
GROUP BY TIPORETRIBUCIONES.NOMBRETIPORET
```

6.7 NÚMERO DE TRABAJADORES / NÚMERO DE PERCEPTORES NO ACTIVOS

```
SELECT SUM(T1.NUMPERCEPTORES) AS Trabajadores,
(SELECT SUM(T2.NUMPERCEPTORES)
FROM HECHOS T2
LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCCAA=T2.IDCCAA
LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=T2.IDPERIODOTEMPORAL
WHERE ($[ControlParameter:ANYO]$=T2.IDPERIODOTEMPORAL OR $[ControlParameter:ANYO]$=0)
AND ($[ControlParameter:CCAA]$=T2.IDCCAA OR $[ControlParameter:CCAA]$=0)
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
AND T2.IDTIPOPER IN (SELECT TIPOPERCEPTORES.IDTIPOPER FROM TIPOPERCEPTORES WHERE UPPER(NOMBRETIPOPER) LIKE '%PENSION%' OR UPPER(NOMBRETIPOPER) LIKE '%DESEMPLE%')
```

```
GROUP BY 2
```

```
) AS PerceptoresNoActivos
```

```
FROM HECHOS T1
```

```
LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCCAA=T1.IDCCAA
```

```
LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=T1.IDPERIODOTEMPORAL
```

```
WHERE ($[ControlParameter:ANYO]$=T1.IDPERIODOTEMPORAL OR $[ControlParameter:ANYO]$=0)
```

```
AND ($[ControlParameter:CCAA]$=T1.IDCCAA OR $[ControlParameter:CCAA]$=0)
```

```
AND T1.IDTIPOPER IN (SELECT TIPOPERCEPTORES.IDTIPOPER FROM TIPOPERCEPTORES WHERE UPPER(NOMBRETIPOPER) LIKE '%SALARI%')
```

```
GROUP BY 1
```

6.8 NÚMERO DE TRABAJADORES / HABITANTES

```
SELECT SUM(T1.NUMPERCEPTORES) AS Trabajadores,
```

```
(
```

```
SELECT SUM(NUMEROHAB) FROM HABITANTES WHERE ($[ControlParameter:ANYO]$=HABITANTES.IDPERIODOTEMPORAL OR $[ControlParameter:ANYO]$=0)
```

```
AND ($[ControlParameter:CCAA]$=HABITANTES.IDCCAA OR $[ControlParameter:CCAA]$=0)
```

```
GROUP BY 2
```

```
) AS HABITANTES
```

```
FROM HECHOS T1
```

```
LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCCAA=T1.IDCCAA
```

```
LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=T1.IDPERIODOTEMPORAL
```

```
--LEFT JOIN TIPORETRIBUCIONES ON TIPORETRIBUCIONES.IDRET=HECHOS.IDRET
```

```
--LEFT JOIN TIPOPERCEPTORES ON TIPOPERCEPTORES.IDTIPOPER=HECHOS.IDTIPOPER
```

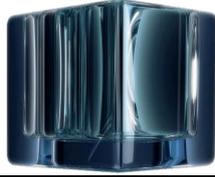
```
WHERE ($[ControlParameter:ANYO]$=T1.IDPERIODOTEMPORAL OR $[ControlParameter:ANYO]$=0)
```

```
AND ($[ControlParameter:CCAA]$=T1.IDCCAA OR $[ControlParameter:CCAA]$=0)
```

```
AND T1.IDTIPOPER IN (SELECT TIPOPERCEPTORES.IDTIPOPER FROM TIPOPERCEPTORES WHERE UPPER(NOMBRETIPOPER) LIKE '%SALARI%')
```

```
GROUP BY 1
```

6.9 NÚMERO DE TRABAJADORES / NÚMERO DE PERSONAS ACTIVAS



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
SELECT SUM(T1.NUMPERCEPTORES) AS Trabajadores,
(
SELECT      TRUNC(SUM(NUMEROHAB*PORC_ACTIVOS/100),2)      FROM      HABITANTES      WHERE
($[ControlParameter:ANYO]$=HABITANTES.IDPERIODOTEMPORAL OR $[ControlParameter:ANYO]$=0)
AND ($[ControlParameter:CCAA]$=HABITANTES.IDCCAA OR $[ControlParameter:CCAA]$=0)
GROUP BY 2
) AS ACTIVOS
FROM HECHOS T1
LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCCAA=T1.IDCCAA
LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=T1.IDPERIODOTEMPORAL
WHERE ($[ControlParameter:ANYO]$=T1.IDPERIODOTEMPORAL OR $[ControlParameter:ANYO]$=0)
AND ($[ControlParameter:CCAA]$=T1.IDCCAA OR $[ControlParameter:CCAA]$=0)
AND T1.IDTIPOPER IN (SELECT TIPOPERCEPTORES.IDTIPOPER FROM TIPOPERCEPTORES WHERE UPPER(NOMBRETIPOPER)
LIKE '%SALARI%')
GROUP BY 1
```

6.10 SALARIOS TOTALES / TOTAL PRESTACIONES

```
SELECT SUM(T1.CANTIDADRETRIBUCION) AS SalariosTotales,
(SELECT SUM(T2.CANTIDADRETRIBUCION)
FROM HECHOS T2
LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCCAA=T2.IDCCAA
LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=T2.IDPERIODOTEMPORAL
WHERE ($[ControlParameter:ANYO]$=T2.IDPERIODOTEMPORAL OR $[ControlParameter:ANYO]$=0)
AND ($[ControlParameter:CCAA]$=T2.IDCCAA OR $[ControlParameter:CCAA]$=0)
AND T2.IDRET IN (SELECT TIPORETRIBUCIONES.IDRET FROM TIPORETRIBUCIONES WHERE UPPER(NOMBRETIPORET) LIKE
'%DESEMP%' OR UPPER(NOMBRETIPORET) LIKE '%PENSION%')
GROUP BY 2
) AS TotalPrestaciones
FROM HECHOS T1
LEFT JOIN ZONAGEOGRAFICA ON ZONAGEOGRAFICA.IDCCAA=T1.IDCCAA
LEFT JOIN PERIODOTEMPORAL ON PERIODOTEMPORAL.IDPERIODOTEMPORAL=T1.IDPERIODOTEMPORAL
WHERE ($[ControlParameter:ANYO]$=T1.IDPERIODOTEMPORAL OR $[ControlParameter:ANYO]$=0)
AND ($[ControlParameter:CCAA]$=T1.IDCCAA OR $[ControlParameter:CCAA]$=0)
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

```
AND T1.IDRET IN (SELECT TIPORETRIBUCIONES.IDRET FROM TIPORETRIBUCIONES WHERE UPPER(NOMBRETIPORET) LIKE '%SALARI%')
```

```
GROUP BY 1
```

9.5 ANEXO 5. CÓDIGO FUENTE DE FICHERO POR LOTES “CREACIÓN DE TABLAS.BAT”

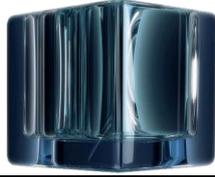
```
cls
echo off
echo -----
echo Creación de tablas y tablas temporales Oracle
pause

SQLPLUS TFC/TFC @1_TablasTFC.sql
echo.
pause

echo Pulse INTRO para finalizar.
pause >nul
GOTO endpgm
:FINAL
echo.
echo *** ERROR. Proceso ABORTADO.
pause
:ENDPGM
```

9.6 ANEXO 5. CÓDIGO FUENTE DE FICHERO POR LOTES “CREACIÓN DE TABLAS.BAT”

```
cls
echo off
echo -----
echo Volcado en tablas temporales Oracle
pause
```



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

DEL bad.bad*

DEL dsc.dsc*

SQLLDR USERID=TFC/TFC CONTROL=TributacioCatalunya.ctl LOG=TributacioCatalunya.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionAndalucia.ctl LOG=TributacionAndalucia.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionAragon.ctl LOG=TributacionAragon.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionAsturias.ctl LOG=TributacionAsturias.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionBalears.ctl LOG=TributacionBalears.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionCanarias.ctl LOG=TributacionCanarias.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionCantabria.ctl LOG=TributacionCantabria.log

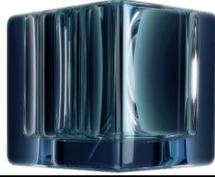
echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionCeutaMelilla.ctl LOG=TributacionCeutaMelilla.log

echo.



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionCLaMancha.ctl LOG=TributacionCLaMancha.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionCLeon.ctl LOG=TributacionCLeon.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionCValenciana.ctl LOG=TributacionCValenciana.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionExtremadura.ctl LOG=TributacionExtremadura.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionGalicia.ctl LOG=TributacionGalicia.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionLaRioja.ctl LOG=TributacionLaRioja.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionMadrid.ctl LOG=TributacionMadrid.log

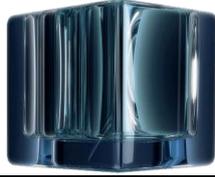
echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=TributacionMurcia.ctl LOG=TributacionMurcia.log

echo.



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=PorcPoblacionActiva.ctl LOG=PorcPoblacionActiva.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=PoblacionPorCCAA2005.ctl LOG=PoblacionPorCCAA2005.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=PoblacionPorCCAA2006.ctl LOG=PoblacionPorCCAA2006.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=PoblacionPorCCAA2007.ctl LOG=PoblacionPorCCAA2007.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=PoblacionPorCCAA2008.ctl LOG=PoblacionPorCCAA2008.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=PoblacionPorCCAA2009.ctl LOG=PoblacionPorCCAA2009.log

echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=PoblacionPorCCAA2010.ctl LOG=PoblacionPorCCAA2010.log

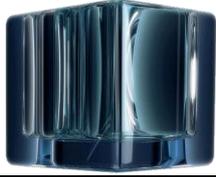
echo.

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLLDR USERID=TFC/TFC CONTROL=PoblacionPorCCAA2011.ctl LOG=PoblacionPorCCAA2011.log

echo.



Construcción y explotación de un almacén de datos para el análisis del sistema de prestaciones sociales

*echo *** OK. Carga de tabla temporal realizada correctamente.*

pause

SQLPLUS TFC / TFC @2_ARREGLODATOS.sql

echo.

pause

SQLPLUS TFC / TFC @3_INSERTION EN TABLAS DEFINITIVAS.sql

echo.

pause

echo Pulse INTRO para finalizar.

pause >nul

GOTO endpgm

:FINAL

echo.

*echo *** ERROR. Proceso ABORTADO.*

pause

:ENDPGM