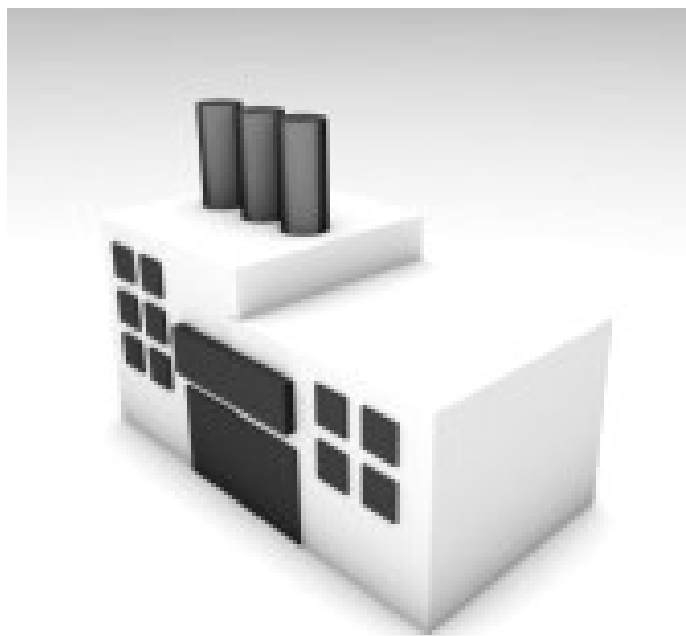


Construcció i explotació d'un magatzem de dades per a l'anàlisi del sistema de prestacions socials



Memòria del projecte

**Alumne: Carlos Toribio Martínez
Consultor: Xavier Plaza Sierra**

**Enginyeria Tècnica d'Informàtica de Sistemes
Juny 2012**

Resumen

Davant la situació de crisi actual i el constant increment del nombre de persones en atur, l' *Organismo de Análisis del Departamento de Trabajo* (OADT) ha decidit fer un estudi per determinar la sostenibilitat del sistema de prestacions socials i conèixer l'evolució de aquestes prestacions.

El problema amb el que es troba és que, degut a la gestió distribuïda de les competències socials, així com el compliment de la llei de protecció de dades, que impedeix accedir a informació que gestionen altres organismes, no té accés a totes les dades que necessita.

Per aquest motiu ha sol·licitat a les comunitats autònomes (CCAA), exceptuant el País Basc i Navarra que tenen un règim foral propi, informació sobre prestacions (Salaris, Pensions, Atur, etc), les retencions realitzades i el nombre de persones que rep cada prestació.

Tanmateix, ha sol·licitat informació al *INE* sobre les dades de població i persones en actiu per CCAA i any.

Amb tota aquesta informació, l'OADT ha decidit encarregar-nos com a consultoria externa independent, la creació d'un magatzem de dades.

Índex

Resumen.....	1
1.-Introducció.....	6
1.1.-Objectius.....	7
1.2.-Anàlisi de requeriments.....	8
1.2.1.-Anàlisi inicial.....	8
1.2.2.-Tractament de les dades aportades.....	8
1.2.3.-Possible visió del model de dades.....	9
1.3.-Planificació.....	10
1.3.1.-Planificació general.....	10
1.3.2.-Planificació específica.....	10
1.3.3.-Diagrama de Gantt.....	11
1.4.-Incidències i riscos.....	13
1.4.1.-Motius familiars.....	13
1.4.2.-Motius laborals.....	13
1.4.3.-Motius varis.....	13
1.5.-Material necessari.....	14
1.5.1.-Coneixements previs.....	14
1.5.2.-Requeriments de maquinari i programari.....	14
2.-Anàlisi i disseny.....	15
2.1.-Descripció.....	15
2.2.-Informes a realitzar.....	15
2.3.-Dades d'origen.....	15
2.4.-Especificacions tècniques.....	19
2.4.1.-Arquitectura de programari.....	20
2.4.2.-Arquitectura de maquinari.....	20
2.5.-Elements per a l'anàlisi multidimensional.....	21
2.6.-Disseny conceptual.....	22
2.6.1.-Triar el Fet.....	22
2.6.2.-Trobar el grànul escaient.....	22
2.6.3.-Escollir les dimensions que s'utilitzaran en l'anàlisi.....	23
2.6.4.-Trobar els atributs de cada dimensió.....	23
2.6.5.-Distingir entre descriptors i jerarquies d'agregació.....	23
2.6.6.-Decidir quines són les mesures que interessin.....	23
2.6.7.-Definir Cel·les.....	23
2.6.8.-Explicitar les restriccions d'integritat.....	23
2.6.9.-Estudiar la viabilitat.....	24
2.7.-Model Conceptual.....	24
2.8.-Disseny físic.....	25
2.9.-Procediment ETL.....	26
3.-Implementació.....	28
3.1.-Accés al producte.....	28
3.2.-Procés de carrega.....	31

3.3.-Informes.....	38
3.3.1.-Informe total retribucions i retencions.....	39
3.3.2.-Informe retribució mitjana i percentatge de retenció mig.....	40
3.3.3.-Informe sobre nombre de treballadors respecte altre factor.....	41
3.3.4.-Informe sobre dades estatals.....	43
3.3.5.-Informe sobre total salaris sobre la resta de prestacions.....	44
3.3.6.-Informe sobre dades per comunitat autònoma.....	45
3.3.7.-Informe sobre projeccions dades treballadors.....	46
3.3.8.-Informe sobre projeccions dades treballadors respecte població.....	47
4.-Conclusions.....	48
5.-Línies de evolució futura.....	48
6.-Glossari.....	48
7.-Bibliografia.....	49

Índex de figures

<i>Figura 1: Possible visió del model de dades.....</i>	<i>9</i>
<i>Figura 2: Diagrama de Gantt (vista general).....</i>	<i>11</i>
<i>Figura 3: Diagrama de Gantt (vista Pac 1).</i>	<i>11</i>
<i>Figura 4: Diagrama de Gantt (vista Pac 2).</i>	<i>12</i>
<i>Figura 5: Diagrama de Gantt (vista Pac 3).</i>	<i>12</i>
<i>Figura 6: Diagrama de Gantt (vista Lliurament final).....</i>	<i>12</i>
<i>Figura 7: Diagrama de Gantt (vista Debat virtual).....</i>	<i>13</i>
<i>Figura 8: Vista arxiu dades origen població per comunitat autònoma.....</i>	<i>16</i>
<i>Figura 9: Vista arxiu dades origen percentatge població activa per comunitat autònoma.....</i>	<i>17</i>
<i>Figura 10: Vista arxiu dades origen tributació comunitat autònoma (vista 1).....</i>	<i>18</i>
<i>Figura 11: Vista arxiu dades origen tributació comunitat autònoma (vista 2).....</i>	<i>18</i>
<i>Figura 12: Vista arxiu dades origen tributació comunitat autònoma (vista 3).....</i>	<i>19</i>
<i>Figura 13: Vista esquema procés creació magatzem de dades.....</i>	<i>20</i>
<i>Figura 14: Vista esquema procés disseny conceptual.....</i>	<i>22</i>
<i>Figura 15: Dimensions del magatzem de dades.....</i>	<i>23</i>
<i>Figura 16: Model conceptual.....</i>	<i>24</i>
<i>Figura 17:Taula dimensió Població.....</i>	<i>25</i>
<i>Figura 18:Taula de fet Evolució de prestacions.....</i>	<i>25</i>
<i>Figura 19:Taula de dimensió Data.....</i>	<i>25</i>
<i>Figura 20:Taula de dimensió Tipus de perceptor.....</i>	<i>25</i>
<i>Figura 21:Taula Grup de retribució.....</i>	<i>25</i>
<i>Figura 22:Taula Percentatge de població activa.....</i>	<i>26</i>
<i>Figura 23: Vista escriptori màquina virtual.....</i>	<i>28</i>
<i>Figura 24: Vista plana d'entrada Oracle10g EX.....</i>	<i>28</i>
<i>Figura 25: Entorn de treball Oracle10g EX (vista 1)</i>	<i>29</i>
<i>Figura 26: Entorn de treball Oracle10g EX (vista 2)</i>	<i>29</i>
<i>Figura 27: Atlas SBI Studio (Dades connexió amb magatzem de dades)</i>	<i>30</i>
<i>Figura 28: Vista plana d'entrada Atlas SBI Server (vista 1)</i>	<i>30</i>
<i>Figura 29: Vista plana d'entrada Atlas SBI Server (vista 2)</i>	<i>30</i>
<i>Figura 30: Vista escriptori màquina virtual.....</i>	<i>31</i>
<i>Figura 31: Vista carpeta ETL.....</i>	<i>33</i>
<i>Figura 32: Vista arxiu de control de carrega per SQL Loader (vista 1).....</i>	<i>33</i>
<i>Figura 33: Vista arxiu de control de carrega per SQL Loader (vista 2).....</i>	<i>34</i>
<i>Figura 34: Vista arxiu de control de carrega per SQL Loader (vista 3).....</i>	<i>35</i>
<i>Figura 35: Vista arxiu procés normalització de dades.....</i>	<i>36</i>
<i>Figura 36: Vista arxiu de automatització de ETL.....</i>	<i>38</i>
<i>Figura 37: Vista menú informes en Atlas SBI Server.....</i>	<i>38</i>
<i>Figura 38: Vista informe Total retribucions i retencions estatal en Atlas SBI Server.....</i>	<i>39</i>
<i>Figura 39: Vista informe Total retribucions per CCAA en Atlas SBI Server.....</i>	<i>39</i>
<i>Figura 40: Vista informe Total retencions per CCAA en Atlas SBI Server.....</i>	<i>40</i>
<i>Figura 41: Vista informe retribució mitjana i retenció mitjana estatal en Atlas SBI Server.....</i>	<i>40</i>
<i>Figura 42: Vista informe retribució mitjana i retenció mitjana per CCAA en Atlas SBI Server.....</i>	<i>41</i>
<i>Figura 43: Vista informe Ratios treballadors estatals en Atlas SBI Server.....</i>	<i>41</i>
<i>Figura 44: Vista informe Ratios treballadors per CCAA en Atlas SBI Server.....</i>	<i>42</i>
<i>Figura 45: Vista informe Datos Estatales a Atlas SBI Server.....</i>	<i>43</i>
<i>Figura 46: Vista informe Total salaris sobre la resta de prestacions a Atlas SBI Server.....</i>	<i>44</i>
<i>Figura 47: Vista informe dades CCAA a Atlas SBI Server.....</i>	<i>45</i>
<i>Figura 48: Vista informe projecció treballadors respecte percept. no actius estatal a Atlas SBI Server.....</i>	<i>46</i>

Figura 49: Vista informe projecció treballadors respecte percept. no actius CCAA a Atlas SBI Server...46
Figura 50: Vista informe projecció treballadors respecte població estatal a Atlas SBI Server.....47
Figura 51: Vista informe projecció treballadors respecte població per CCAA a Atlas SBI Server.....47

1.-Introducció

Aquest projecte consisteix en la construcció d'un magatzem de dades, com a eina de treball per als empleats del Organisme d'Anàlisi del Departament de Treball (OADT).

Degut a la crisi actual i el constant increment de persones en atur l'OADT vol conèixer la evolució de les prestacions i la sostenibilitat del sistema.

El fet que la informació estigui distribuïda per les comunitats autònomes implica que no es pot treballar de forma unificada, i no es pot fer cap anàlisi rigorós.

El magatzem de dades que creem ha de permetre la recollida de tota aquesta informació i la seva posterior explotació amb eines de *Business Intelligence*.

1.1.-Objectius

Dins dels objectius de la Universitat Oberta de Catalunya envers aquest treball de fi de carrera, comentar que és un treball eminentment pràctic i vinculat a l'exercici professional de la informàtica, a on es requereix, l'anàlisi de tècniques existents per projectar la base de dades d'un magatzem de dades, orientant el disseny a un magatzem de dades, creant dimensions, atributs i fets necessaris, així com considerant factors del tipus desnormalització de taules, inclusió d'informació agregada o historicació de la informació, entre d'altres.

D'altra banda com objectius personals que vull assolir amb la realització d'aquest treball de fi de carrera, comentar que vull familiaritzar-me i aprofundir en el disseny i construcció de magatzems de dades, bases de dades i eines OLAP.

Com a objectius pel nostre client podem ressenyar la solució al problema de dades escampades, amb formats diferents, lo que les fa impossibles de gestionar, els quals volem solucionar amb la creació d'un magatzem de dades, mitjançant el qual es pugui obtenir com a mínim la següent informació:

- ✚ Total retribució
- ✚ Total retenció
- ✚ Retribució mitjana
- ✚ Percentatge de retenció mig
- ✚ Numero de retribucions mig
- ✚ Percentatge de població per segment
- ✚ Nombre de treballadors / Nombre de perceptors
- ✚ Nombre de treballadors / Nombre habitants
- ✚ Nombre de treballadors / Nombre persones actives
- ✚ Salaris totals / Total prestacions

Tota aquesta informació es proporcionarà a nivell d'any i s'haurà de poder consultar de forma agregada, per comunitat autònoma, tipus de perceptor i tipus de retribució.

A més s'ha de proporcionar un conjunt predefinit d'informes on es mostri la informació sol·licitada i qualsevol altre que pugui ser útil al OADT.

També se'ns demana fer una projecció dels indicadors "Nombre Treballadors / Nombre Perceptors no actius" i "Nombre Treballadors / Població" per determinar el punt de col·lapse, és a dir, el moment en que dos treballadors mantinguin un perceptor, o s'iguali el nombre de treballadors i població perceptora, a on població perceptora s'aconsegueix sumant població no activa més perceptors de prestacions.

1.2.-Anàlisi de requeriments.

1.2.1.-Anàlisi inicial

L'OADT ha decidit encarregar-nos la creació i explotació d'un magatzem de dades per obtenir una sèrie de dades, per lo qual ens ha facilitat uns arxius amb extensió csv. Tenim tres tipus d'arxius:

📁 Població per CCAA i any que va des de 2005 fins a 2011.

- Tenim set arxius.
- Cadascú te els valors per un any.
- Les separacions de camps estan fetes amb “,” en tots els arxius.
- Fent una primera ullada hi ha sobre informació, podem obviar totals.
- La informació que hi ha és el número de homes, número de dones per any.

📁 Tributació per CCAA

- Tenim setze arxius.
- Hi ha diferents formats en quant a anys de les dades.
- Les separacions de camps estan fetes en diferents formats:
- Tenim sobre informació en algunes comunitats a les que apareixen totals.
- Tenim diferents tipus de retribució en les comunitats autònomes, que haurem de unificar.

📁 Percentatge de població activa.

- Tenim un arxiu.
- Te valors des de l'any 2005 a 2010.
- Les separacions de camps estan fetes amb “,” a l'arxiu.
- Fent una primera ullada, podem obviar totals.
- La informació que hi ha és percentatge població activa dels dos sexes junts per any i per província.

1.2.2.-Tractament de les dades aportades

Com s'ha pogut veure en l'apartat anterior s'ha de fer un treball molt acurat per adequar les dades a les nostres necessitats en quant a format, unificació, manera de visualització i filtrat. Després d'aquest treball s'han d'introduir aquestes dades en la base de dades Oracle.

1.2.3.-Possible visió del model de dades

Com a possible visió del model de dades tenim.

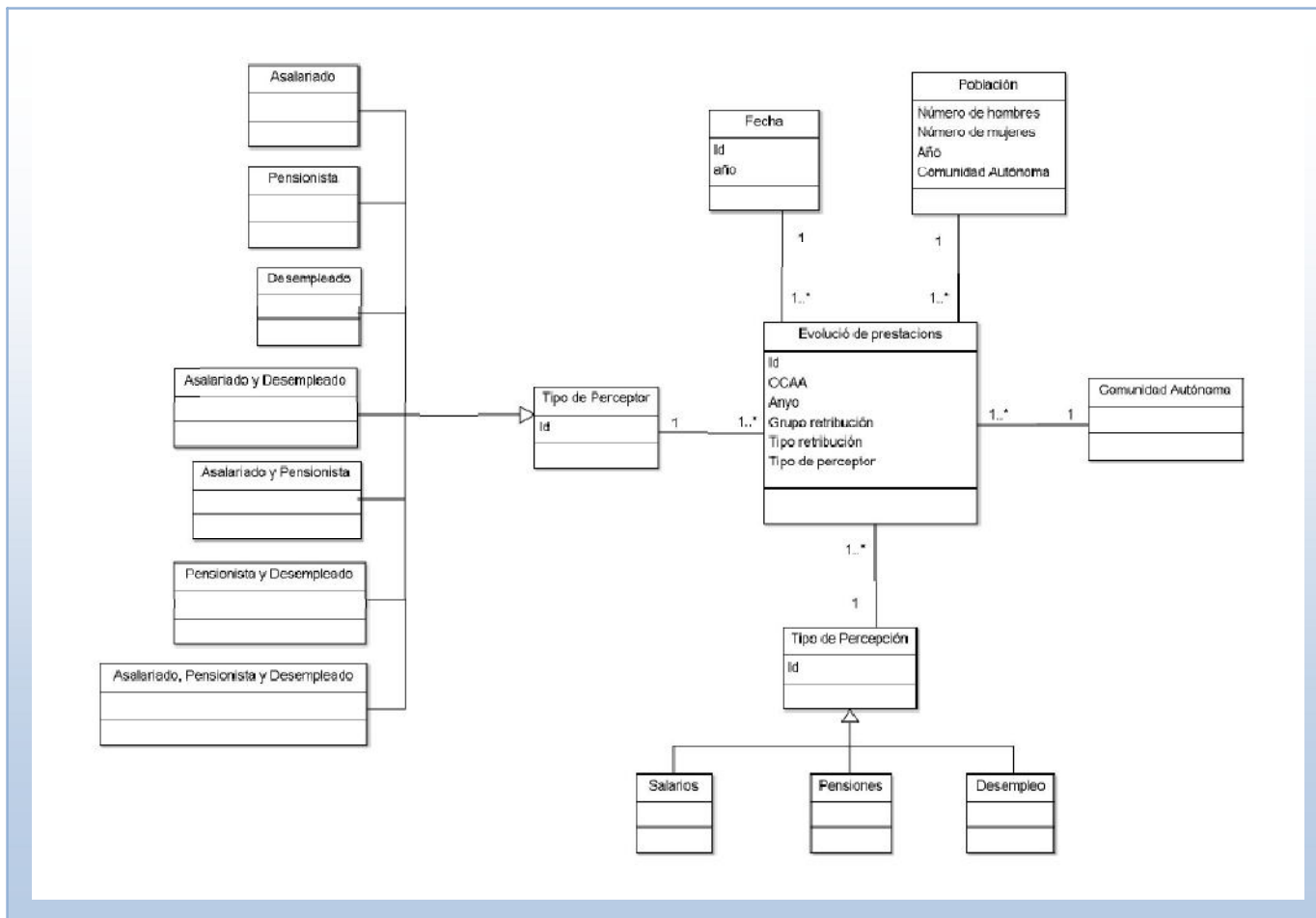


Figura 1: Possible visió del model de dades.

A on tenim el fet "Evolució de prestacions" i les dimensions:

- 📊 Tipus de percepció.
- 📊 Tipus de perceptor.
- 📊 Comunitat autònoma.
- 📊 Data.
- 📊 Població.

1.3.-Planificació

1.3.1.-Planificació general

Al pla docent de l'assignatura s'han publicat les següents dates clau pel lliurament de les diferents parts de que consta aquest TFC de magatzem de dades.

Títol	Inici	Lliurament	Qualificació
PAC 1 (Pla de Treball)	01/03/2012	16/03/2012	23/03/2012
PAC 2 (Anàlisi i Disseny)	17/03/2012	20/04/2012	27/04/2012
PAC 3 (Implementació)	21/04/2012	25/05/2012	01/06/2012
Lliurament Final	26/05/2012	13/06/2012	02/07/2012
Debat virtual	25/06/2012	28/06/2012	--

Cada lliurament comporta una sèrie de tasques que es veuen amb més detall en els subapartats següents.

1.3.2.-Planificació específica

Es divideix cada part o lliurament en tasques.

Num.Tasca	Descripció tasca	Data inici	Data final	Duració
1	PAC 1: Fer un anàlisi preliminar de requeriments i el pla de treball.	01/03/2012	16/03/2012	16
1.1	Descarrega de informació de UOC	02/03/2012	02/03/2012	1
1.2	Lectura documentació TFC	03/03/2012	03/03/2012	1
1.3	Lectura mòdul 1 TFC	05/03/2012	05/03/2012	1
1.4	Recerca informació magatzems de dades	08/03/2012	09/03/2012	2
1.5	Anàlisi preliminar de requeriments	10/03/2012	11/03/2012	2
1.6	Consulta treballs anys anteriors	12/03/2012	12/03/2012	1
1.7	Instal·lació i aprenentatge de l'aplicació Openproj	13/03/2012	13/03/2012	1
1.8	Preparació del document "Pla de Treball"	14/03/2012	16/03/2012	3
2	PAC 2: Fer anàlisi de requeriments, disseny conceptual, tècnic i anàlisi del procés ETL	17/03/2012	20/04/2012	33
2.1	Lectura ràpida apunts Bases de Dades 1 , Bases de dades 2 i Models Multidimensionals de dades.	19/03/2012	21/03/2012	3
2.2	Descarregar i instal·lar màquina virtual	22/03/2012	22/03/2012	1
2.3	Instal·lació Atlas SBI i Sqldeveloper a MV	23/03/2012	23/03/2012	1
2.4	Proves, aprenentatge i pràctica amb Oracle, Atlas SBI i Sqldeveloper	24/03/2012	26/03/2012	3
2.5	Anàlisi de requeriments	27/03/2012	31/03/2012	5
2.6	Realització del Model de Dades	03/04/2012	06/04/2012	4
2.7	Disseny procedimental	08/04/2012	10/04/2012	4
2.8	Revisió documentació	12/04/2012	12/04/2012	1
2.9	Realització documentació PAC 2	13/04/2012	15/04/2012	3
2.10	Entrega PAC 2	20/04/2012	20/04/2012	1
3	PAC 3: Implementació de procés ETL, Base de dades, Model de negoci i informes.	21/04/2012	25/05/2012	35
3.1	Construcció del magatzem de dades	23/04/2012	26/04/2012	3
3.2	Proves de carrega i explotació	28/04/2012	02/05/2012	5
3.3	Anàlisi informació	06/05/2012	07/05/2012	2
3.4	Construcció informe	08/05/2012	10/05/2012	3
3.5	Revisió resultats i possibles modificacions	13/05/2012	15/05/2012	3
3.6	Realització documentació PAC 3	16/05/2012	18/05/2012	3

3.7	Entrega PAC 3	25/05/2012	25/05/2012	1
4	Lliurament final: Entrega de les parts que conformen el treball de fi de carrera.	26/05/2012	13/06/2012	20
4.1	Preparació Producte per entrega	26/05/2012	29/05/2012	4
4.2	Preparació Memòria	01/06/2012	08/06/2012	8
4.3	Preparació Presentació Virtual	09/06/2012	12/06/2012	4
4.4	Entrega de Lliurament final	13/06/2012	13/06/2012	1
5	Debat virtual.	25/06/2012	28/06/2012	4

1.3.3.-Diagrama de Gantt

Amb les dades de l'apartat anterior podem fer el següent diagrama de Gantt, a on, per poder apreciar els detalls millor es fa una impressió de pantalla general, després, de cada Pac i entrega final amb més detall. Com es pot veure tenim les dates clau i la consecució de tasques a fer.

Primer ho veiem d'una forma més general amb les divisions naturals que es fa per Pacs i entrega final.

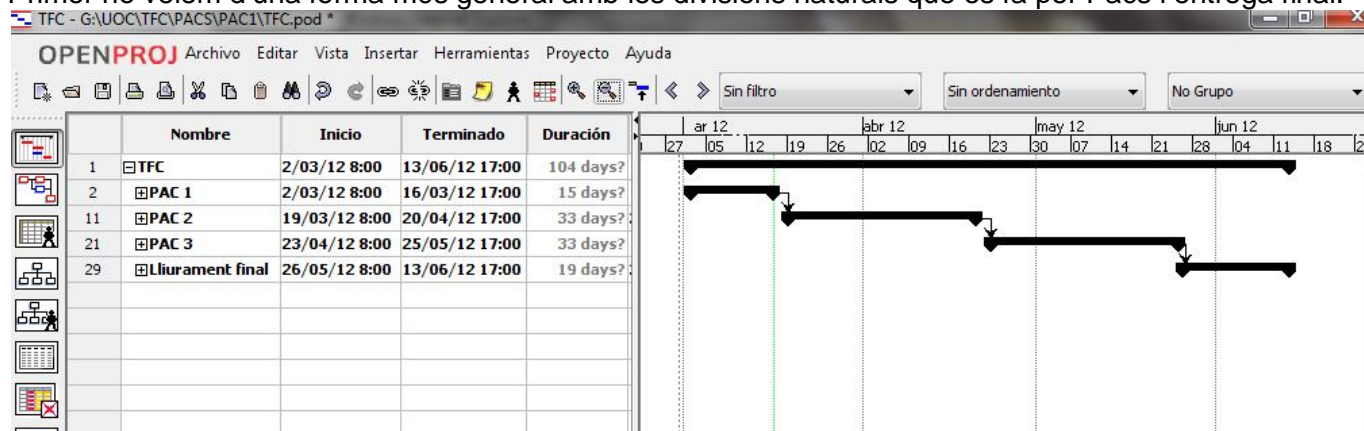


Figura 2: Diagrama de Gantt (vista general).

Ara veiem Pac 1.

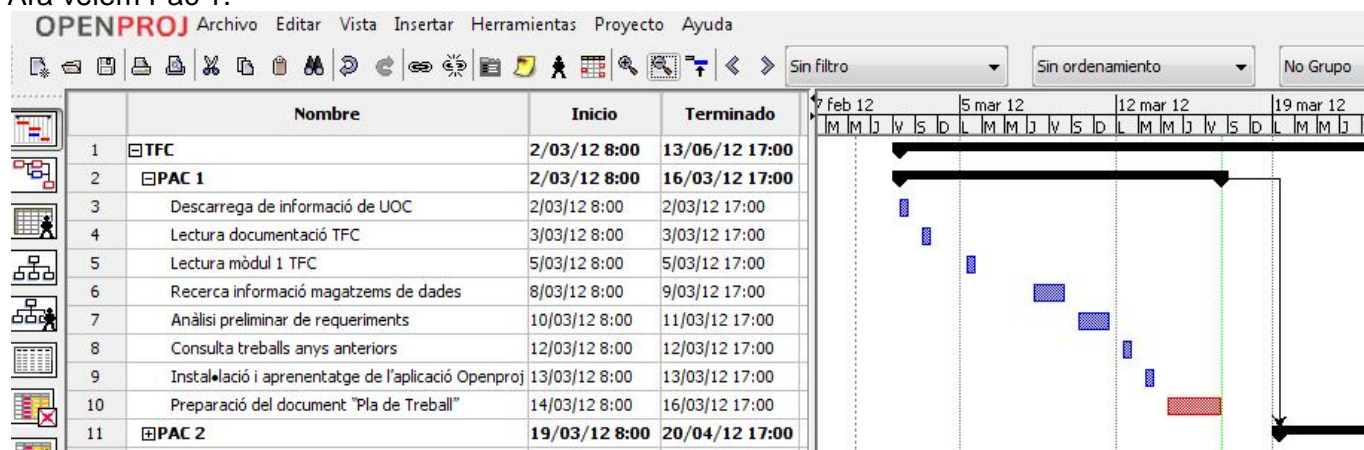


Figura 3: Diagrama de Gantt (vista Pac 1).

Pac 2.

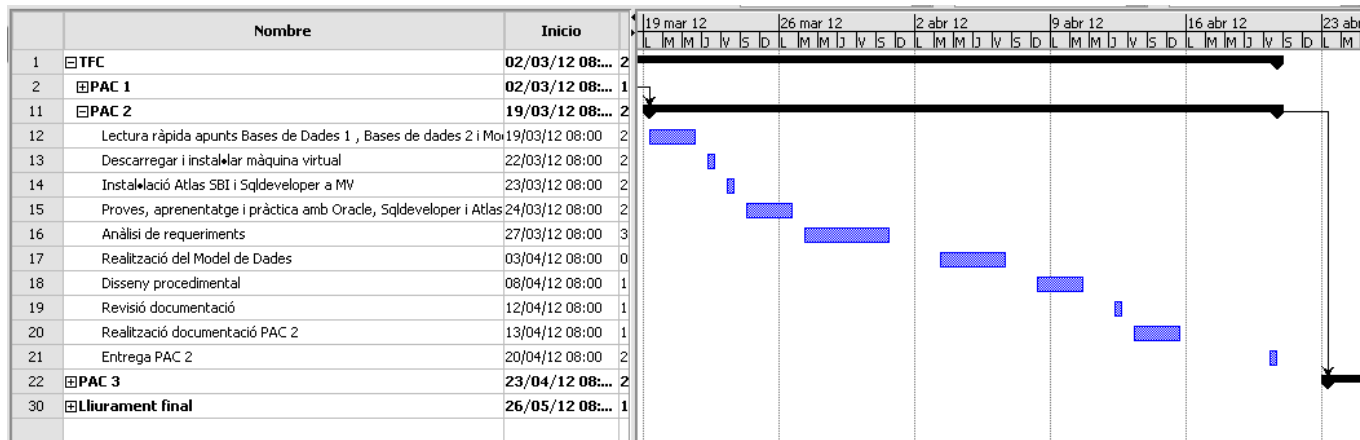


Figura 4: Diagrama de Gantt (vista Pac 2).

Pac 3.

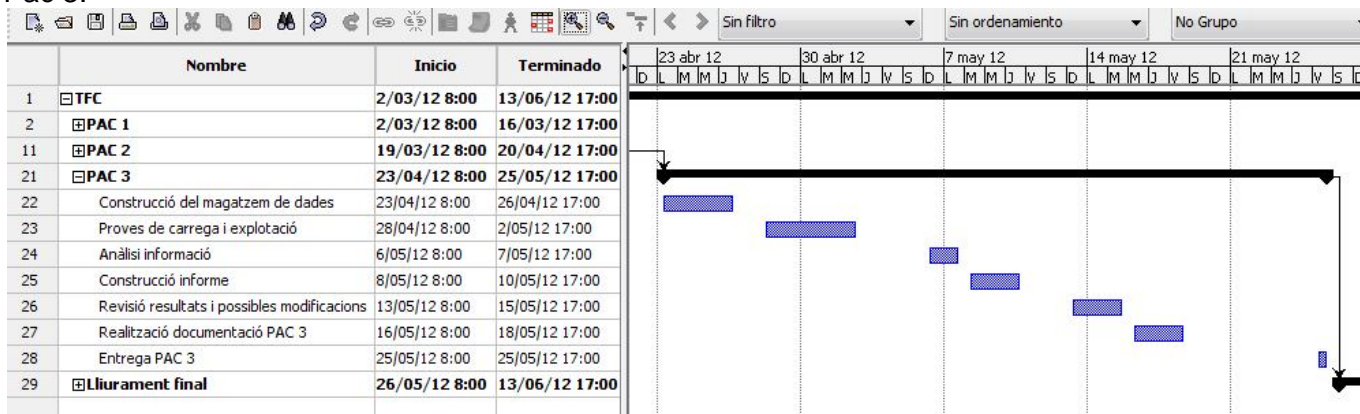


Figura 5: Diagrama de Gantt (vista Pac 3).

Lliurament final.

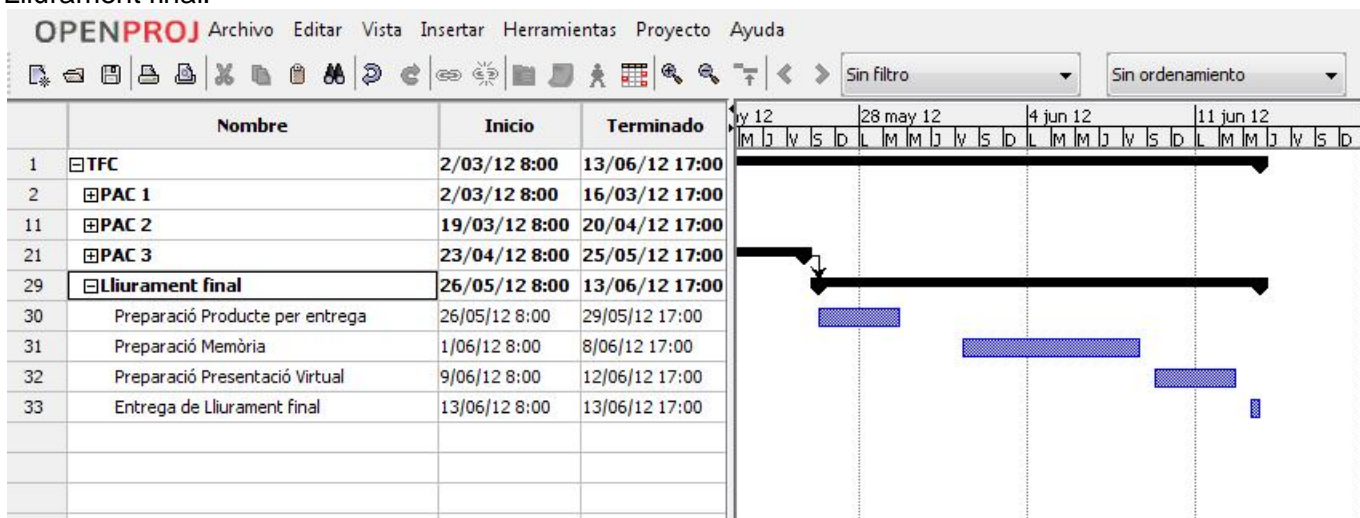


Figura 6: Diagrama de Gantt (vista Lliurament final).

I el debat virtual.

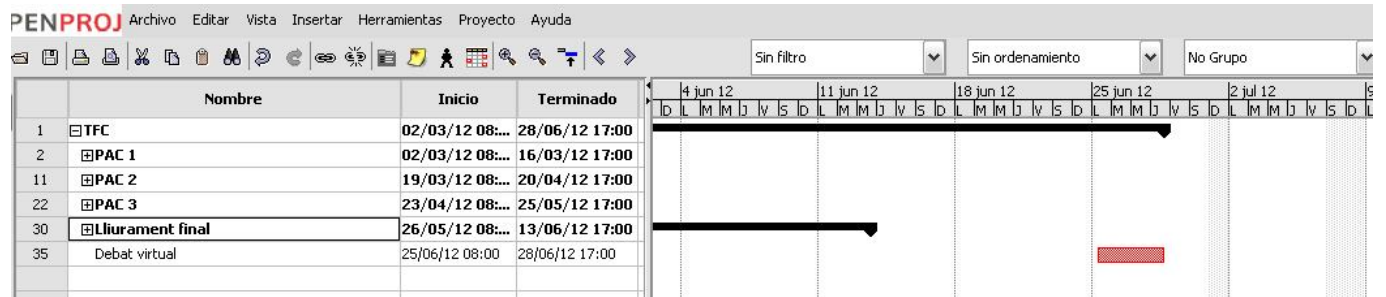


Figura 7: Diagrama de Gantt (vista Debat virtual).

1.4.-Incidències i riscos

1.4.1.-Motius familiars.

Com a pare divorciat amb custòdia compartida del meu fill de 5 anys tinc uns requeriments de temps i dedicació que canvien per períodes setmanals, és a dir, normalment puc aprofitar més hores per estudiar i avançar en el treball la setmana que no tinc el meu fill, aquest és un handicap amb el que ja fa anys que hi soc, per tant normalment es carrega més de feina aquesta setmana.

Puntualment i per necessitats imperioses podria avançar la feina a casa dels avis mentre cuiden del meu fill.

1.4.2.-Motius laborals.

Treball ja fa anys al sector de manteniment industrial, en una empresa gran, amb lo que les puntes de treball s'absorbeixen entre diferents companys, però si hi ha algun problema important en alguna màquina és necessari estar més hores al lloc de treball, anar caps de setmana i tenir una implicació i flexibilitat bastant més gran que de costum.

Normalment no passa això, i si fos cas s'haurien de retocar algunes dates de la previsió que s'han fet de manera bondadosa, o en el pitjor dels casos que em cobreixin els companys.

1.4.3.-Motius varis.

La meua forma de treballar és sobre disc dur extern, tinc instal·lat el programari necessari en diversos ordinadors, com el portàtil de casa i el punt de treball a la feina. Així vaig d'un lloc a l'altre amb la informació a sobre.

Com a seguretat al portàtil de casa tinc còpia dels documents que faig servir, que, a més hi són al disc dur extern, per tant s'haurà d'anar actualitzant aquesta còpia.

Si per qualsevol motiu s'espatllés el punt de treball, de casa o de la feina, llavors sempre em queda l'altre, per una urgència mentre arreglo l'espatllat. Com a última solució també puc fer servir el punt de treball de la meua parella.

1.5.-Material necessari.

1.5.1.-Coneixements previs.

Com a coneixements previs per la realització del TFC es demana:

- ✚ Conceptes generals de bases de dades.
- ✚ Conceptes bàsics d' HTML.
- ✚ Coneixement del llenguatge PL/SQL:

Aquests coneixements previs s'han adquirit cursant les assignatures de Bases de Dades 1 i Bases de dades 2.

1.5.2.-Requeriments de maquinari i programari.

Com a requeriment de maquinari tenim un punt de treball de la UOC.

Com a requeriment de programari tenim:

- ✚ Sistema operatiu Windows XP
- ✚ Oracle com gestor de bases de dades.
- ✚ PL/SQL de Oracle per la programació.
- ✚ Atlas SBI per la construcció d'informes.
- ✚ OpenProj per la planificació de les tasques a dur a terme.
- ✚ Oracle VM VirtualBox per allotjar màquina virtual amb el programari de dalt.

2.-Anàlisi i disseny

2.1.-Descripció

L' Organismo de Análisis del Departamento de Trabajo (OADT) ha decidit fer un estudi per determinar la sostenibilitat del sistema de prestacions socials i conèixer la seva evolució, lo que ha promogut que ens demanin la construcció d'un magatzem de dades per obtenir aquesta informació mitjançant una sèrie d'informes amb unes dades específiques.

2.2.-Informes a realitzar

Els informes que ens demanen i hauríem implementar són:

- ✚ Dintre d'una temporalitat a nivell d'any i amb la possibilitat de consultar de forma agregada, per comunitat autònoma, tipus de perceptor i tipus de retribució.
 - Total retribució
 - Total retenció
 - Retribució mitjana
 - Percentatge de retenció mig
 - Numero de retribucions mig
 - Percentatge de població per segment
 - Nombre de treballadors / Nombre de perceptors
 - Nombre de treballadors / Nombre habitants
 - Nombre de treballadors / Nombre persones actives
 - Salaries totals / Total prestacions
- ✚ Un conjunt predefinit d'informes on es mostri la informació sol·licitada i qualsevol i qualsevol altre que pugui ser útil per al OADT.
- ✚ Una projecció d'indicadors "Nombre treballadors / Nombre de perceptors no actius" i "Nombre de treballadors / Població" per determinar el punt de col·lapse del sistema, que arribarà al moment en que dos treballadors mantinguin un perceptor, o s'iguali el nombre de treballadors i població perceptora (no actius més perceptors de prestacions socials).

2.3.-Dades d'origen

El OADT ens proporciona tota la informació en una sèrie de fitxers que es poden dividir en tres tipus:

- ✚ Població per comunitat autònoma <any>: un arxiu per cadascun dels anys amb el nombre d'habitants per comunitat autònoma.
- ✚ Percentatge de població activa: un únic arxiu amb el percentatge de població activa per cada comunitat autònoma i any.
- ✚ Tributació <comunitat autònoma>: un arxiu de cada comunitat autònoma amb la informació del nombre de persones, el total de retribucions, les retencions per any i tipus de percepció.

Ens adverteixen que degut a que la informació s'ha extret de diferents sistemes hi ha tres formats de separacions diferents:

- ✚ Separat per coma.
- ✚ Separat per tabulador.
- ✚ Separat per punt i coma.

A més, ens indiquen que, tot i tenir una estructura semblant, no tots els arxius donen la mateixa informació, havent comunitats autònomes que proporcionen dades en valors absoluts i d'altres valors mitjos, pel que s'haurà d'unificar.

També ens demanen que per que les dades siguin més realistes, el nombre d'habitants el calcularem com a mitjana de valors a 1 de gener de l'any i l'1 de gener de l'any següent.

Si importem a un full d'Excel un arxiu del tipus població per comunitat autònoma i any podem veure

	A	B	C	D
1		Varones	Mujeres	
2	Total	2,18E+14	2,23E+14	
3	Andalucía	3889605.0	3960194.0	
4	Aragón	632148.0	636879.0	
5	Asturias (Principado de)	516329.0	560306.0	
6	Balears (Illes)	492642.0	490489.0	
7	Canarias	988230.0	980050.0	
8	Cantabria	274797.0	287512.0	
9	Castilla y León	1240450.0	1270399.0	
10	Castilla - La Mancha	950976.0	943691.0	
11	Cataluña	3468235.0	3526971.0	
12	Comunidad Valenciana	2332283.0	2360166.0	
13	Extremadura	539225.0	544654.0	
14	Galícia	1330703.0	1431495.0	
15	Madrid (Comunidad de)	2887718.0	3076425.0	
16	Murcia (Región de)	677049.0	658743.0	
17	Navarra (Comunidad Foral de)	296587.0	296885.0	
18	País Vasco	1040525.0	1084321.0	
19	Rioja (La)	151730.0	149354.0	
20	Ciudad autónoma de Ceuta	38315.0	36961.0	
21	Ciudad autónoma de Melilla	33322.0	32166.0	
22				

Figura 8: Vista arxiu dades origen població per comunitat autònoma.

Cada línia conté la següent informació:

- 📌 Comunitat autònoma: nom de la comunitat autònoma com pot ser Extremadura, Galícia.
- 📌 Varones: valor numèric que informa de nombre de habitants barons a aquesta CCAA.
- 📌 Mujeres: valor numèric que informa de nombre de habitants fèmines a aquesta CCAA.

D'aquest tipus de arxiu tots venen amb el mateix format que és separat per comes.

Si importem a un full d'Excel un arxiu del tipus percentatge de població activa tenim

	A	B	C	D	E	F	G
1		Total					
2		2010	2009	2008	2007	2006	2005
3	Ambos sexes						
4	Total Nacional	60.0	59.94	59.8	58.92	58.32	57.35
5	Andalucía	58.73	58.29	57.56	56.25	55.32	54.31
6	Aragón	58.16	58.95	59.3	58.81	56.59	56.57
7	Asturias; Principado de	51.65	51.96	52.68	50.53	49.85	48.14
8	Balears; Illes	66.01	65.62	64.76	64.07	64.11	62.71
9	Canarias	62.1	62.14	60.93	61.04	61.02	59.38
10	Cantabria	55.91	56.78	56.93	56.22	55.32	54.71
11	Castilla y León	55.14	54.55	54.53	53.66	53.09	52.47
12	Castilla - La Mancha	57.84	57.69	57.09	56.02	55.03	54.02
13	Cataluña	62.81	62.67	63.19	62.5	62.17	61.27
14	Comunitat Valenciana	60.47	61.31	61.13	60.13	59.62	58.64
15	Extremadura	54.61	53.81	53.49	52.85	51.58	51.63
16	Galicia	54.58	55.21	55.07	54.27	53.61	53.04
17	Madrid; Comunidad de	65.1	64.78	64.58	63.89	63.57	62.38
18	Murcia; Región de	62.26	61.6	61.79	60.45	58.98	57.84
19	Navarra; Comunidad Foral de	60.11	60.55	61.15	60.82	60.67	59.6
20	País Vasco	57.59	57.38	58.06	57.63	58.07	56.92
21	Rioja; La	59.52	60.0	59.92	58.95	59.48	59.41
22	Ceuta	55.06	57.06	55.46	51.59	52.79	58.29
23	Melilla	52.32	52.86	53.34	54.09	52.5	51.18

Figura 9: Vista arxiu dades origen percentatge població activa per comunitat autònoma.

Cada línia conté la següent informació:

- ✚ Comunitat autònoma: nom de la comunitat autònoma com pot ser Extremadura, Galicia.
- ✚ Percentatge població activa en 2010: valor numèric que informa de percentatge de població activa tant homes com dones a aquesta CCAA en l'any 2010.
- ✚ Percentatge població activa en 2009: valor numèric que informa de percentatge de població activa tant homes com dones a aquesta CCAA en l'any 2009.
- ✚ Percentatge població activa en 2008: valor numèric que informa de percentatge de població activa tant homes com dones a aquesta CCAA en l'any 2008.
- ✚ Percentatge població activa en 2007: valor numèric que informa de percentatge de població activa tant homes com dones a aquesta CCAA en l'any 2007.
- ✚ Percentatge població activa en 2006: valor numèric que informa de percentatge de població activa tant homes com dones a aquesta CCAA en l'any 2006.
- ✚ Percentatge població activa en 2005: valor numèric que informa de percentatge de població activa tant homes com dones a aquesta CCAA en l'any 2005.

D'aquest tipus de arxiu, l'únic que hi ha, te el format que és separat per comes.

Si importem a fulls d'Excel arxius diferents del tipus tributació comunitat autònoma tenim

	A	B	C	D	E	F	G	H	I	J	K	
1		2003							2004			
2		Asalariados	Pensionistas	Desempleados	Asalariados y pensionistas	Asalariados y desempleados	Pensionistas y desempleados	Asalariados, pensionistas y desempleados	Asalariados	Pensionistas	Desempleados	
3	Catalunya											
4	TOTAL											
5	Personas	2674405.0	1231158.0	85040.0	193995.0	305706.0	18377.0	28842.0	2691461.0	1212935.0	92821.0	
6	Retribuciones	4,82E+20	1,15E+19	3,78E+16	3,35E+18	3,61E+18	2,24E+16	3,70E+16	5,09E+20	1,20E+20	4,65E+16	
7	Retribución media anual	18025.0	9348.0	4446.0	17264.0	11804.0	12187.0	12812.0	18919.0	9896.0	5013.0	
8	Retenciones	7,59E+17	5,26E+16	3558272.0	3,38E+16	2,63E+16	9917197.0	1,97E+13	8,21E+18	5,84E+16	4482673.0	
9	Tipo medio de retención	15.75	4.57	0.94	10.08	7.29	4.43	5.32	16.13	4.86	0.96	
10	SALARIOS											
11	Personas	2674405.0	193995.0	305706.0	..	28842.0	2691461.0	
12	Retribuciones	4,82E+20	2,09E+18	2,80E+18	..	1,75E+16	5,09E+20	
13	Retribución media anual	18025.0	10793.0	9154.0	..	6083.0	18919.0	
14	Retenciones	7,59E+17	2,70E+16	2,60E+16	..	1,28E+13	8,21E+18	
15	Tipo medio de retención	15.75	12.9	9.29	..	7.32	16.13	
16	PENSIONES											
17	Personas	..	1231158.0	..	193995.0	..	18377.0	28842.0	..	1212935.0	..	
18	Retribuciones	..	1,15E+19	..	1,26E+18	..	1,57E+16	1,17E+16	..	1,20E+20	..	
19	Retribución media anual	..	9348.0	..	6472.0	..	8529.0	4047.0	..	9896.0	..	
20	Retenciones	..	5,26E+16	..	6,74E+14	..	9546636.0	6655047.0	..	5,84E+16	..	
21	Tipo medio de retención	..	4.57	..	5.37	..	6.09	5.7	..	4.86	..	
22	DESEMPLEO											
23	Personas	85040.0	..	305706.0	18377.0	28842.0	92821.0	
24	Retribuciones	3,78E+16	..	8,10E+16	6,72E+14	7,73E+14	4,65E+16	
25	Retribución media anual	4446.0	..	2651.0	3658.0	2682.0	5013.0	
26	Retenciones	3558272.0	..	3015382.0	370561.0	175743.0	4482673.0	
27	Tipo medio de retención	0.94	..	0.37	0.55	0.23	0.96	

Figura 10: Vista arxiu dades origen tributació comunitat autònoma (vista 1).

Tributació de la comunitat autònoma de Catalunya

	A	B	C	D	E	F	G	H	I	J	K	L
1		2003							2004			
2		Todos	Asalariados	Pensionistas	Desempleados	Asalariados y pensionistas	Asalariados y desempleado	Pensionistas y desempleados	Asalariados, pensionistas y desempleados	Todos	Asalariados	Pensionistas
3	ANDALUCÍA											
4	TOTAL											
5	Personas	4377354.0	2240713.0	1152687.0	162728.0	150622.0	595713.0	24326.0	50565.0	4453530.0	2277466.0	1181807.0
6	Retribuciones	4,94E+20	3,17E+20	9,47E+18	5,74E+16	2,23E+18	4,74E+18	2,25E+16	4,66E+16	5,33E+18	3,41E+20	1,03E+20
7	Retenciones	4,77E+18	4,06E+18	3,30E+16	2007424.0	1,79E+16	1,79E+16	7402295.0	1,64E+14	5,32E+18	4,51E+18	3,71E+15
8	SALARIOS											
9	Personas	3037613.0	2240713.0	150622.0	595713.0	..	50565.0	3084332.0	2277466.0	..
10	Retribuciones	3,65E+20	3,17E+20	1,32E+18	3,38E+18	..	2,03E+16	3,94E+20	3,41E+20	..
11	Retenciones	4,38E+16	4,06E+18	1,34E+16	1,78E+16	..	9091016.0	4,88E+18	4,51E+18	..
12	PENSIONES											
13	Personas	1378200.0	..	1152687.0	..	150622.0	..	24326.0	50565.0	1411950.0	..	1181807.0
14	Retribuciones	1,07E+20	..	9,47E+18	..	9,12E+16	..	1,51E+16	1,53E+16	1,16E+20	..	1,03E+20
15	Retenciones	3,89E+16	..	3,30E+16	..	4,48E+14	..	7190966.0	7185241.0	4,35E+16	..	3,71E+15
16	DESEMPLEO											
17	Personas	833332.0	162728.0	..	595713.0	24326.0	50565.0	841063.0
18	Retribuciones	2,12E+18	5,74E+16	..	1,36E+18	7,33E+14	1,09E+16	2,25E+18
19	Retenciones	3837454.0	2007424.0	..	1495875.0	211329.0	122826.0	3896724.0

Figura 11: Vista arxiu dades origen tributació comunitat autònoma (vista 2).

Tributació de la comunitat autònoma de Andalusia

	A	B	C	D	E	F	G	H	I	J	K	
1		2004							2005			
2		Asalariados	Pensionistas	Desempleados	Asalariados y pensionistas	Asalariados y desempleados	Pensionistas y desempleados	Asalariados, pensionistas y desempleados	Asalariados	Pensionistas	Desempleados	
3	ARAGÓN											
4	SALARIOS											
5	Personas	449678.0	29107.0	56298.0	..	3983.0	483231.0	
6	Retribuciones	7,97E+18	3,42E+16	6,48E+15	..	2,97E+14	8,81E+18	
7	Retribución media anual	17730.0	11762.0	11504.0	..	7460.0	18221.0	
8	Retenciones	1,17E+18	4,24E+14	6,46E+14	..	2146536.0	1,32E+18	
9	Tipo medio de retención	14.69	12.39	9.98	..	7.22	15.0	
10	PENSIONES											
11	Personas	..	249490.0	..	29107.0	..	2239.0	3983.0	..	249295.0	..	
12	Retribuciones	..	2,41E+18	..	2,42E+16	..	2,61E+14	2,88E+14	..	2,55E+18	..	
13	Retribución media anual	..	9679.0	..	8300.0	..	11644.0	7243.0	..	10212.0	..	
14	Retenciones	..	1,11E+16	..	1,49E+14	..	1908554.0	2660808.0	..	1,26E+16	..	
15	Tipo medio de retención	..	4.62	..	6.18	..	7.32	9.22	..	4.93	..	
16	DESEMPLEO											
17	Personas	9813.0	..	56298.0	2239.0	3983.0	10733.0	
18	Retribuciones	4,59E+14	..	1,16E+16	8303615.0	9833341.0	4,92E+14	
19	Retribución media anual	4681.0	..	2054.0	3709.0	2469.0	4580.0	
20	Retenciones	475955.0	..	398518.0	50686.0	23341.0	492856.0	

Figura 12: Vista arxiu dades origen tributació comunitat autònoma (vista 3).

Tributació de la comunitat autònoma de Aragó

Com podem veure tenim dades diferents als diferents arxius, d'altra banda és un tipus de taula anomenada estadística a la que les dades es llegeixen de forma posicional, és a dir, la dada que hi ha en una cel·la és la conjunció de fila, on tenim el tipus de retribució, i columna, on tenim el tipus de perceptor i l'any.

2.4.-Especificacions tècniques

Per dur a terme el projecte i garantir el seu correcte funcionament és necessari com a mínim que el sistema operatiu instal·lat sigui Microsoft Windows XP Professional, amb les necessitats intrínseques que això porta com pot ser un mínim de memòria RAM i espai de disc.

2.4.1.-Arquitectura de programari

El programari fet servir al projecte és:

- ✚ Oracle 10g Express Edition: per la gestió de la base de dades.
- ✚ SQL*Loader: per la carrega de dades a les diferents taules que es faran a partir dels arxius de dades d'origen aportats per l'OADT.
- ✚ SQL Developer: per la construcció i proves de processos de llenguatge estructurat pel procés ETL.
- ✚ Atlas SBI Studio: per la fabricació d'informes.
- ✚ Atlas SBI Server: per la explotació dels informes i magatzem de dades.

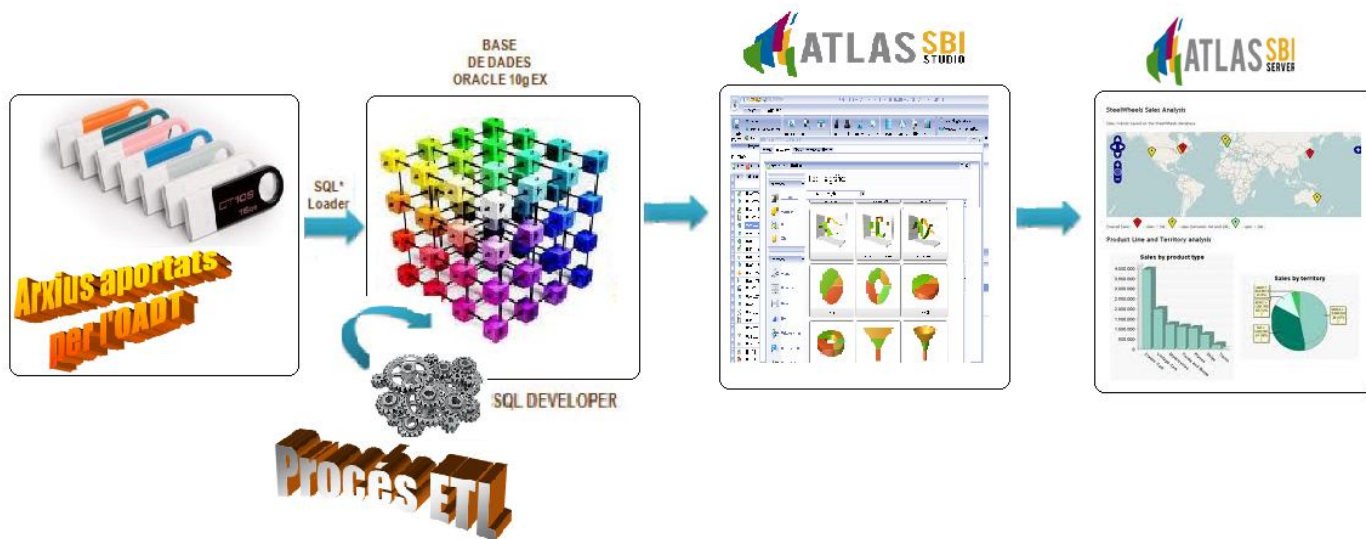


Figura 13: Vista esquema procés creació magatzem de dades.

2.4.2.-Arquitectura de maquinari

La opció que es presenta és un servidor per la base de dades i un servidor web per Atlas SBI, així es podrà accedir des de qualsevol punt amb accés a Internet per poder fer consultes, anàlisi, visualitzar informes, o quelcom que sigui necessari.

2.5.-Elements per a l'anàlisi multidimensional

Dels requeriments de l'OADT, juntament amb les dades aportades per la construcció del magatzem de dades podem fer una primera aproximació als elements que tenim per fer l'anàlisi multidimensional, que són:

- ✚ **Fet:** considerarem com a fet la *evolució del sistema de prestacions*.
- ✚ **Mesures:** Considerarem com a mesures les retribucions i retencions que reben els tipus de perceptors a l'any en les diferents CCAA, nombre de persones per CCAA i any.
- ✚ **Dimensions:** considerarem les següents dimensions:
 - Data: període de temps en anys del que disposem d'informació.
 - Comunitat autònoma: comunitat autònoma de la que disposem d'informació.
 - Tipus de perceptor: tipus de persona que rep la percepció.
 - Tipus de retribució: percepció que es dona a un perceptor.
 - Població: número de homes i dones en una comunitat autònoma i any.
- ✚ **Descriptors:** per cada dimensió tenim els següents atributs o descriptors:
 - Data: any.
 - Comunitat autònoma: nom.
 - Tipus de perceptor: nom.
 - Tipus de retribució: nom.
 - Població: número de homes, número de dones, any, comunitat autònoma.

2.6.-Disseny conceptual

Per fer el disseny del cub multidimensional ho farem en nou etapes o fases, en les que cadascuna puja un esglauó en la tasca.

A la imatge de sota es pot veure un esquema de seqüenciació per la creació del cub multidimensional.

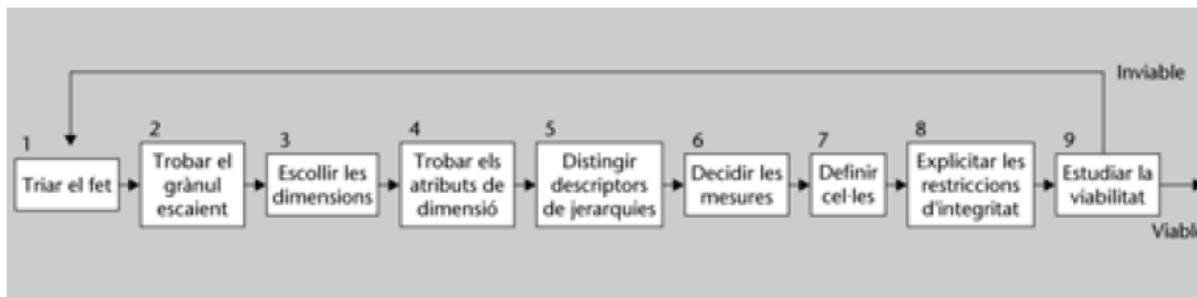


Figura 14: Vista esquema procés disseny conceptual.

2.6.1.-Triar el Fet

Com a primer pas hem d'identificar el fet que constituirà el nucli del nostre model. En aquest projecte ens demanen crear un magatzem de dades per determinar la sostenibilitat del sistema de prestacions socials i conèixer l'evolució de aquestes prestacions, per això ens donen una sèrie de fitxers de dades a on trobem poblacions per comunitat autònoma i any, percentatge de població activa i any, tributació comunitat autònoma i any.

L'OADT ens demana:

- ✚ Total retribució
- ✚ Total retenció
- ✚ Retribució mitjana
- ✚ Percentatge de retenció mig
- ✚ Numero de retribucions mig
- ✚ Percentatge de població per segment
- ✚ Nombre de treballadors / Nombre de perceptors
- ✚ Nombre de treballadors / Nombre habitants
- ✚ Nombre de treballadors / Nombre persones actives
- ✚ Salariis totals / Total prestacions

I tota aquesta informació l'haurem de proporcionar a nivell d'any i de forma agregada per comunitat autònoma, tipus de perceptor i tipus de retribució.

És per això que triarem com a fet "evolució del sistema de prestacions".

2.6.2.-Trobar el grànul escaient

El grànul és l'individu últim que volem analitzar, la Cel·la més petita que volem tenir disponible, és per això que un grànul mal triat pot no deixar veure tot el detall d'informació que es vol veure, o per una altra banda malbaratar espai i fer el magatzem de dades sigui impossible de fer treballar pel gran volum de dades de que disposa.

Les dades que ens faciliten comprenen diferents períodes anuals segons la comunitat autònoma i segons tipus de perceptor, però en tots, les dades van separades anualment.

Per tant la separació en períodes de temps serà anual, per comunitat autònoma i per tipus de perceptor, aquesta és la granularitat.

2.6.3.-Escollir les dimensions que s'utilitzaran en l'anàlisi

Al nostre projecte identifiquem les següents dimensions:

- ✚ **Comunitat autònoma:** comunitats autònomes que han aportat dades
- ✚ **Tipus de perceptor:** tipus de persona a la que va dirigida la percepció, a on tenim assalariat, pensionista, aturat, assalariat i aturat, assalariat i pensionista, pensionista i aturat, i assalariat, pensionista i aturat.
- ✚ **Tipus de retribució:** percepció entre les que tenim salaris, pensions i atur.
- ✚ **Data:** període de temps en anys de que disposem d'informació.
- ✚ **Població:** número de homes i dones en una comunitat autònoma.

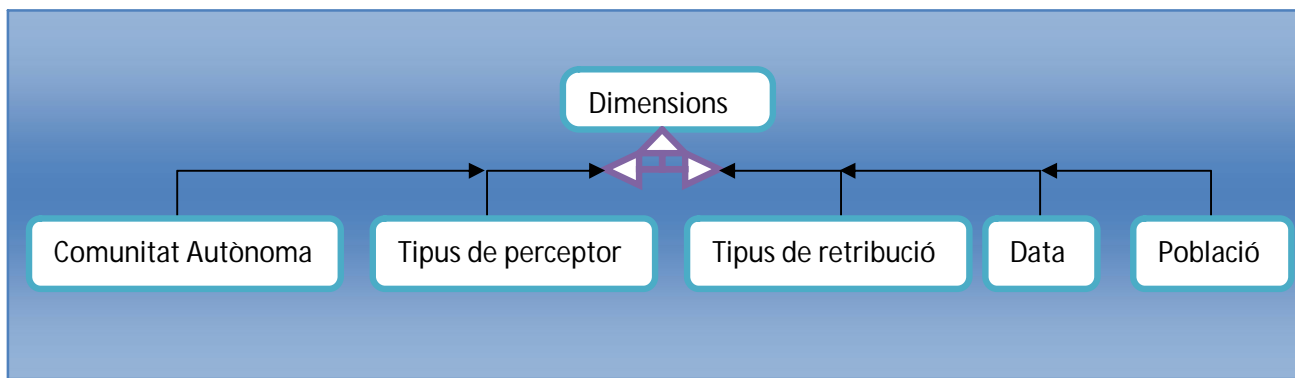


Figura 15: Dimensions del magatzem de dades.

2.6.4.-Trobar els atributs de cada dimensió

Un cop tenim identificades les dimensions ara podem trobar els atributs d'aquestes.

- ✚ **Comunitat autònoma:** tindria com atribut nom (Catalunya, Madrid, Galicia).
- ✚ **Tipus de perceptor:** tindria com atribut nom (assalariat, pensionista, aturat).
- ✚ **Tipus de retribució:** tindria com atribut nom (salaris, pensions).
- ✚ **Data:** com que les dades venen agrupades en anys s'agafarà aquest atribut.
- ✚ **Població:** tindria com atributs número d'homes, número de dones, comunitat autònoma i any.

2.6.5.-Distingir entre descriptors i jerarquies d'agregació

Entre els atributs que hi ha en una dimensió podem trobar dos tipus, els que fem servir per agrupar o els que fem servir per seleccionar.

En les dimensions obtingudes crec que el que tenim són atributs descriptors, és a dir per seleccionar informació.

2.6.6.-Decidir quines són les mesures que interessin

Considerarem com a mesures les retribucions i retencions que reben els tipus de perceptors a l'any en les diferents CCAA, nombre de persones per CCAA i any.

2.6.7.-Definir Cel·les

La Cel·la que es fa servir és la mínima unitat d'informació que es guardarà, i en el nostre cas tindrem el tipus de retribució d'un tipus de perceptor d'una comunitat autònoma en un any determinat.

2.6.8.-Explicar les restriccions d'integritat

El conjunt de dimensions mencionades en apartats anteriors, després d'haver definit el gràdul dona lloc a una Base, i substituint dimensions s'obtenen altres Bases en l'espai.

2.6.9.-Estudiar la viabilitat

Per fer l'estudi de viabilitat estímem l'espai que ocuparan totes les dades emmagatzemades en el sistema.

Partim de uns arxius inicials de tres tipus on:

- ✚ Arxiu del tipus població per comunitat autònoma i any, a on tenim 7 diferents, cadascun dels quals te 20 registres, i amb una longitud de registre de 256 bytes tenim 35840 bytes.
- ✚ Arxiu de percentatge de població activa que té 20 registres, i amb una longitud de registre de 256 bytes tenim 5120 bytes.
- ✚ Arxius diferents del tipus tributació comunitat autònoma, a on tenim 16 diferents, cadascun dels quals té 160 registres, i amb una longitud de registre de 256 bytes tenim 655.360 bytes.

Així tenim que el volum de dades és la suma dels tres apartats anteriors, o sigui 35.840 bytes més 5.120 bytes més 655.360 bytes, el que fa uns 0,7MB, lo que no és un tamany gran ni molt menys.

El rendiment amb un volum de dades així no ens ha de preocupar, així que crec que és viable el projecte.

2.7.-Model Conceptual

Partint de l'anàlisi feta en els punts anteriors sabem que necessitem una taula per al fet, i una taula per cada dimensió de les que formen el model.

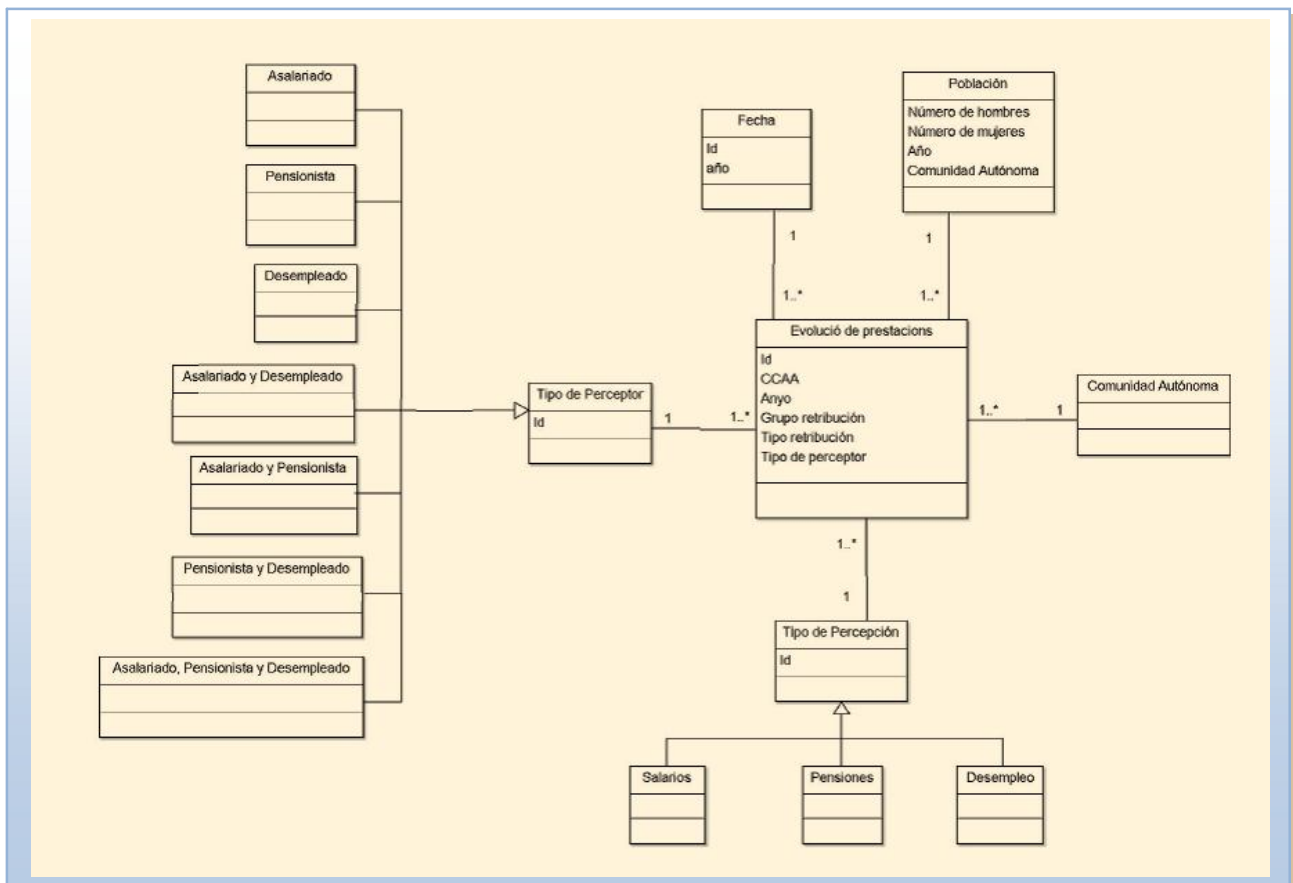


Figura 16: Model conceptual.

2.8.-Disseny físic

Un cop tenim les relacions que componen la nostra base de dades fem el disseny físic. Les taules que s'obtenen són les següents:

Taula de dimensió **Població**

```
CREATE TABLE " T_POBLACION_CCAA "  
(  
  "ID"          VARCHAR (10 BYTE) NOT NULL ENABLE,  
  "ANYO"        VARCHAR (10 BYTE),  
  "CCAA"        CHAR (40 BYTE),  
  "NUM_HOMBRES" NUMBER (30,2),  
  "NUM_MUJERES" NUMBER (30,2),  
  CONSTRAINT PK_T_POBLACION_CCAA PRIMARY KEY (ID) ENABLE  
);
```

Figura 17:Taula dimensió Població.

Taula de fet **Evolució de prestacions**

```
CREATE TABLE "T_TRIBUT_CCAA"  
(  
  "ID"          VARCHAR (10 BYTE) NOT NULL ENABLE,  
  "CCAA"        CHAR (40 BYTE),  
  "ANYO"        VARCHAR (30 BYTE),  
  "GRUPO_RETRIB" VARCHAR (40 BYTE),  
  "TIPOUS_RETRIB" VARCHAR (30 BYTE),  
  "ASALARIADOS" NUMBER (30,2),  
  "PENSIONISTAS" NUMBER (30,2),  
  "DESEMPLEADOS" NUMBER (30,2),  
  "ASAL_Y_PENSI" NUMBER (30,2),  
  "ASAL_Y_DESEMP" NUMBER (30,2),  
  "PENSI_Y_DESEM" NUMBER (30,2),  
  "ASAL_PENSI_Y_DESEMP" NUMBER (30,2),  
  CONSTRAINT PK_T_TRIBUT_CCAA PRIMARY KEY (ID) ENABLE  
);
```

Figura 18:Taula de fet Evolució de prestacions.

Taula de dimensió **Data**

```
CREATE TABLE "T_ANYOS"  
(  
  "ID"          VARCHAR (10 BYTE) NOT NULL ENABLE,  
  "ANYO"        VARCHAR (10 BYTE),  
  CONSTRAINT PK_T_ANYOS PRIMARY KEY (ID) ENABLE  
);
```

Figura 19:Taula de dimensió Data.

Taula de dimensió **Tipus de perceptor**

```
CREATE TABLE "T_TIPO_PERCEPTOR"  
(  
  "ID"          VARCHAR (10 BYTE) NOT NULL ENABLE,  
  "TIPO_PERCEPTOR" VARCHAR (30 BYTE),  
  CONSTRAINT PK_T_TIPO_PERCEPTOR PRIMARY KEY (ID) ENABLE  
);
```

Figura 20:Taula de dimensió Tipus de perceptor.

Taula **Grup de retribució**

```
CREATE TABLE "T_GRUPO_RETRIBUCION"  
(  
  "ID"          VARCHAR (10 BYTE) NOT NULL ENABLE,  
  "GRUPO_RETRIBUCION" VARCHAR (30 BYTE),  
  CONSTRAINT PK_T_GRUPO_RETRIBUCION PRIMARY KEY (ID)  
ENABLE  
);
```

Figura 21:Taula Grup de retribució.

Taula Percentatge de població activa

```
CREATE TABLE "T_PORC_POB_ACT_CCAA"  
(  
  "ID"          VARCHAR (10 BYTE) NOT NULL ENABLE,  
  "CCAA"       CHAR (40 BYTE),  
  "ANYO_2010"  NUMBER (30,2),  
  "ANYO_2009"  NUMBER (30,2),  
  "ANYO_2008"  NUMBER (30,2),  
  "ANYO_2007"  NUMBER (30,2),  
  "ANYO_2006"  NUMBER (30,2),  
  "ANYO_2005"  NUMBER (30,2),  
  CONSTRAINT PK_T_PORC_POB_ACT_CCAA PRIMARY KEY (ID) ENABLE  
);
```

Figura 22: Taula Percentatge de població activa.

2.9.-Procediment ETL

Aquest és el procés de manipulació de les dades d'origen consistent en l'extracció de les mateixes dels seus arxius d'origen per portar-les cap a les taules de la base de dades que conformarà el magatzem de dades, la seva transformació per tal d'adaptar-los a les necessitats analitzades i la seva carrega en les taules del magatzem de dades.

En la extracció, haurem de portar les dades d'origen cap a unes taules temporals creades per aquest propòsit, i aquest transport es farà mitjançant els arxius de control de carrega que fa servir SQL Loader, als quals fent servir els paràmetres adequats introduïrem les dades de tots els arxius de població per comunitat autònoma i any en una única taula, les dades del arxiu de percentatge de població activa les introduïrem en una altra taula temporal, i les dades dels arxius de tributació per comunitat autònoma les introduïrem en sis taules temporals fent les agrupacions que s'expliquen en transformació. Per tenir un control dels possibles errors s'haurà de encaminar tot arxiu de log o dades mal carregades cap a una carpeta amb aquest nom i l'arxiu portarà el nom de la comunitat autònoma.

En la transformació començarem unificant els noms de les comunitats autònomes a la taula temporal que tingui les dades de població per CCAA, canviarem els punts decimals per comes i els registres buits per valors zero, a més sumarem els valors de Ceuta y Melilla per tenir-los com un bloc. També unificarem els noms de les comunitats autònomes a la taula temporal que tingui les dades de percentatge de població activa, així com canviar els punts decimals per comes i posar ceros a on hi hagi registre nul. Finalment, a les taules temporals on carreguem les dades de tributació per comunitat autònoma haurem de deixar com tipus de percepció retribucions i retencions, a més de número de persones, haurem de calcular aquestes dades en algunes taules com les que porten les dades de Madrid o Murcia. A continuació detallarem les agrupacions de dades de CCAA en les taules temporals i les modificacions que s'han de fer a aquestes taules per unificar:

- Taula temporal per allotjar les dades provinents dels arxius de tributació de les comunitats autònomes de Catalunya, Castilla la Manxa i La Rioja, que porten els mateixos tipus de retribució i per tant el treball que farem per una comunitat ens valdrà per les tres, que serà treure total de tipus de percepció, treure retribució mitjana i tipus mitja de retenció, ja que no els necessitem en la taula de fet, i arribat el cas són calculables de cara als informes, transformarem els valors numèrics en format científic a nombres naturals i canviarem els punts decimals per comes.
- Taula temporal per allotjar les dades provinents dels arxius de tributació de les comunitats autònomes de Andalusia, Astúries i Canàries que porten els mateixos tipus de retribució, aquí el treball que farem serà treure total de tipus de percepció, transformarem els valors numèrics en format científic a nombres naturals i canviarem els punts decimals per comes.

- ✚ Taula temporal per allotjar les dades provinents del arxiu de tributació de la comunitat autònoma de Madrid, aquí el treball que farem serà treure total de tipus de percepció, transformarem els valors numèrics en format científic a nombres naturals i canviarem els punts decimals per comes, a més calcularem retribucions i retencions amb les dades que aporta l'arxiu per després esborrar-les ja que els valor mitjans no els necessitem.
- ✚ Taula temporal per allotjar les dades provinents dels arxius de tributació de les comunitats autònomes de Aragó, Balears, Cantabria i Extremadura que porten els mateixos tipus de retribució, aquí el treball que farem serà treure els valors mitjans de tipus de percepció, transformarem els valors numèrics en format científic a nombres naturals i canviarem els punts decimals per comes.
- ✚ Taula temporal per allotjar les dades provinents dels arxius de tributació de les comunitats autònomes de Castilla i Lleó i Galicia, aquí transformarem els valors numèrics en format científic a nombres naturals i canviarem els punts decimals per comes.
- ✚ Taula temporal per allotjar les dades provinents dels arxius de tributació de les comunitats autònomes de Comunitat Valenciana i Murcia, aquí transformarem els valors numèrics en format científic a nombres naturals i canviarem els punts decimals per comes, a més calcularem retribucions i retencions amb les dades que aporta l'arxiu per després esborrar-les ja que els valor mitjans no els necessitem.
- ✚ Taula temporal per allotjar les dades provinents del arxiu de tributació de Ceuta y Melilla, aquí calcularem la suma de Ceuta més Melilla dels diferents tipus de percepció, esborrant els valors mitjans, transformarem els valors numèrics en format científic a nombres naturals i canviarem els punts decimals per comes.

Per la carrega haurem de tenir en compte que les dades no s'han de sobre escriure, ja que és un magatzem de dades, i han d'acumular-se per tant seguirem el procés següent:

- ✚ Passar les dades dels arxius de origen a taules temporals mitjançant SQL Loader.
- ✚ Transformar les dades segons l'apartat anterior.
- ✚ Fer la inserció en les taules definitives i esborrar les taules temporals.

3.-Implementació

Un cop finalitzada la fase de disseny fem la implementació, a on crearem el procés de carrega automatitzat, que crearà la base de dades del magatzem de dades amb els arxius aportats per l'OADT, i amb aquest magatzem de dades podrem crear els informes que ens han demanat amb Atlas SBI Studio per després poder-los visualitzar i explotar amb Atlas SBI Server.

3.1.-Accés al producte

A l'escriptori de la MV tenim les icones que necessitem per accedir al nostre producte.

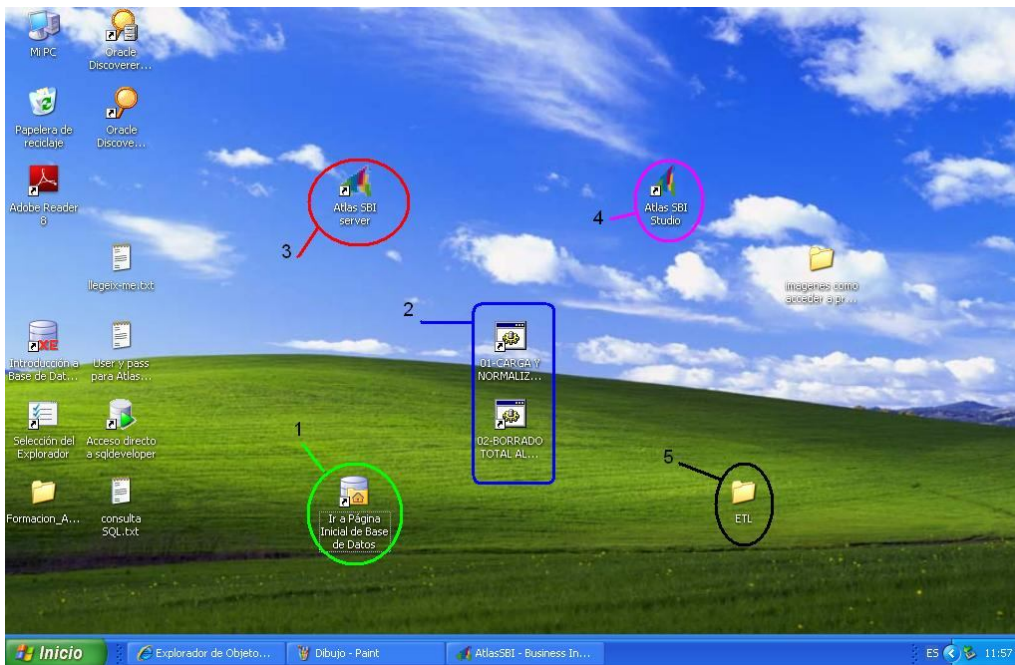


Figura 23: Vista escriptori màquina virtual.

Per accedir a la base de dades que hem dipositat a ORACLE pitgem a la icona de l'escriptori (1) i ens apareix la plana d'inici:



Figura 24: Vista plana d'entrada Oracle 10g EX.

- usuari: system
- password: uoc

Un cop aquí podem accedir a les taules que hem creat

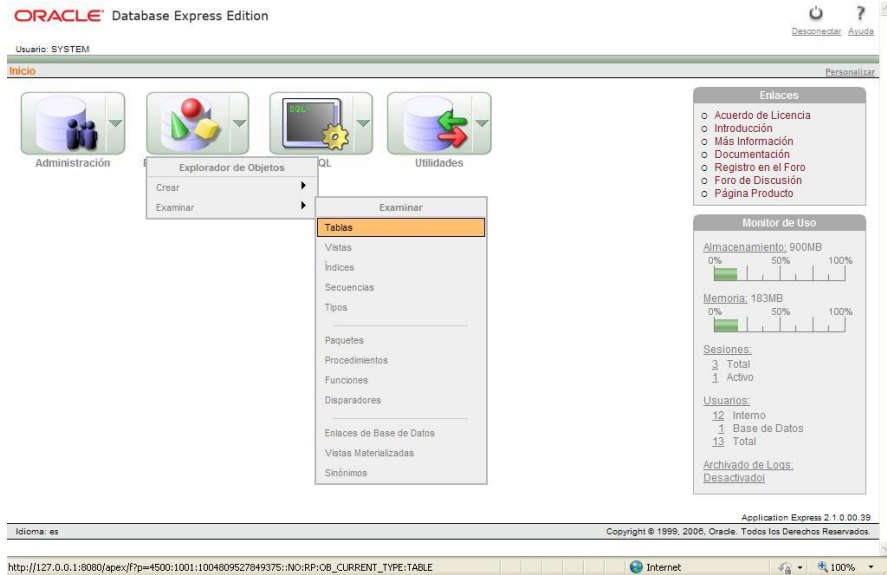


Figura 25: Entorn de treball Oracle 10g EX (vista 1).

I veure el contingut de qualsevol de les taules

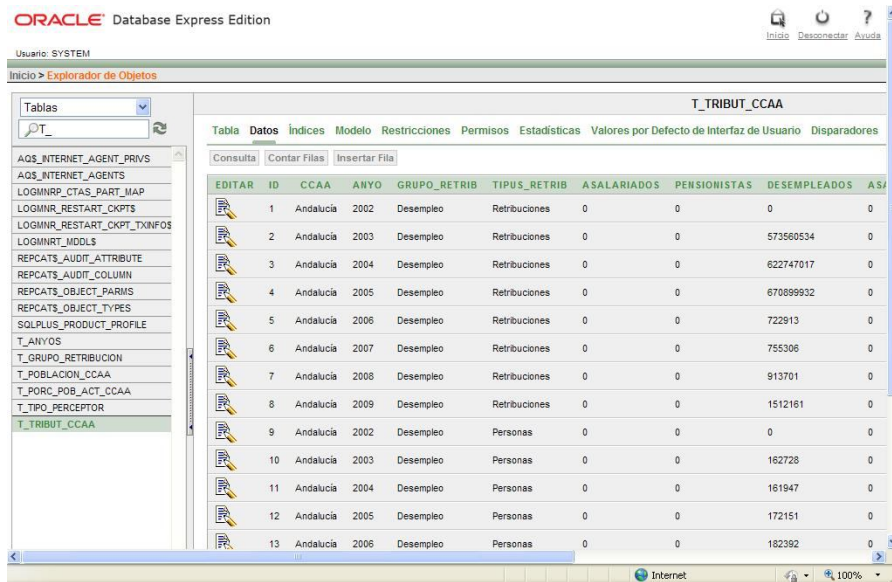


Figura 26: Entorn de treball Oracle 10g EX (vista 2).

Per crear els informes s'ha fet servir el programari Atlas SBI Studio (4), a on s'ha creat un projecte, i un dels paràmetres necessaris és la creació d'una connexió amb la base de dades per poder fer les consultes. Com es veu abaix tenim les següents dades de connexió des de Atlas SBI Studio a la base de dades.

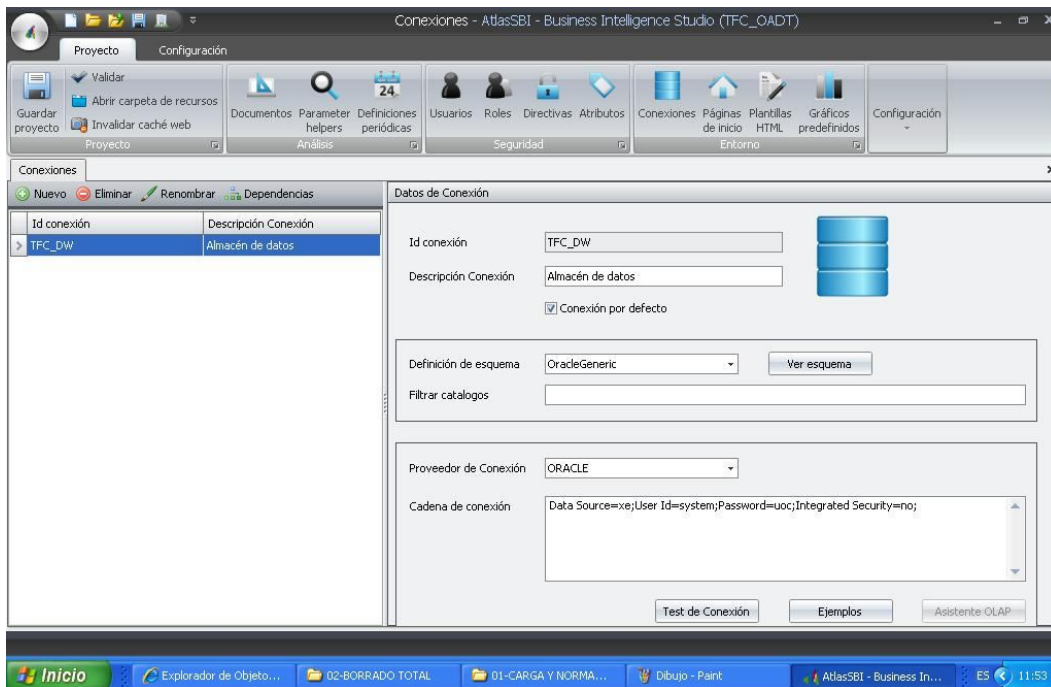


Figura 27: Atlas SBI Studio (Dades connexió amb magatzem de dades).

Per accedir als informes a Atlas SBI server hem de pitjar a la icona de l'escriptori (3) i ens surt la plana d'inici, com que s'han creat dos rols i un usuari per cada rol, cadascun haurà d'entrar amb el seu usuari i password:

- 🚦 Rol OADT
- 🚦 Usuari:user2
- 🚦 Password:uoc



Figura 28: Vista plana d'entrada Atlas SBI Server (vista 1).

- 🚦 Rol OADT_Extern
- 🚦 Usuari:user1
- 🚦 Password:uoc



Figura 29: Vista plana d'entrada Atlas SBI Server (vista 2).

A partir d'aquí user2 tindrà accés a tots els informes , i user1 tindrà accés limitat.

Com apunt adicional comentar que s'ha automatitzat el procés d'extracció transformació i carrega de les dades tal com ens les portava el client, i a l'escriptori tenim dues icones (2) que ens serveixen, una per crear el nostre magatzem de dades fent tot el procés ETL i l'altre per esborrar el nostre magatzem de dades. A l'escriptori també podem observar la carpeta ETL (5) que és a on tenim guardada tota la informació, processos per lots, arxius de comandes sql etc, per la creació del nostre magatzem de dades.

3.2.-Procés de carrega

Per fer el procediment d'extracció transformació i carrega s'ha pensat en crear unes taules temporals per ficar les dades aportades per l'OADT, un cop aquí fer les transformacions necessàries per unificar dades, i transferir-les a les taules definitives de la base de dades per poder consultar-les. Tot això es vol automatitzar, és per això que es crea un arxiu de treball per lots a on s'ajunta tot el procés, que expliquem a continuació.

A l'escriptori de la màquina virtual tenim una carpeta anomenada ETL (5) que és a on s'ubiquen tots els arxius necessaris per fer aquesta tasca, i també a l'escriptori tenim els arxius de treball per lots (2) "01-CARGA Y NORMALIZACION DATOS OADT.bat" i "02-BORRADO TOTAL ALMACEN DATOS.bat", amb el primer fem la carrega i normalització inicial de les dades a la base de dades, i amb el segon fem un esborrat total, per si és necessari en alguna situació.

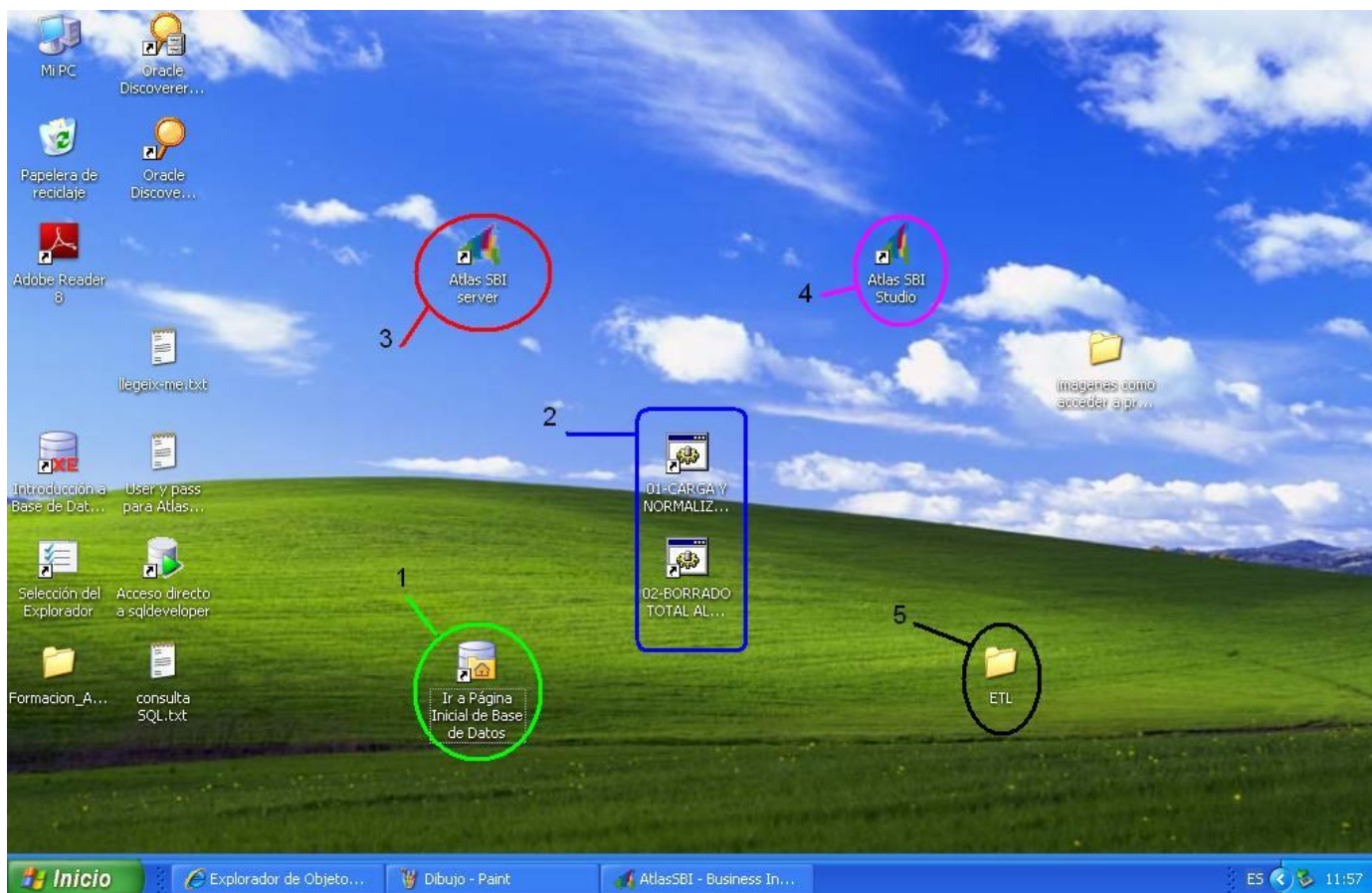


Figura 30: Vista escriptori màquina virtual.

Dintre de la primera carrega i normalització tenim una seqüència d'esdeveniments que fan possible aquest procés.

La primera tasca és crear totes les taules que necessitem pel procés, tant les temporals com les definitives, ho farem amb l'arxiu "*Crea_t_temp.sql*" que ho trobem ubicat a "*ETL\01-CARGA Y NORMALIZACION DE DATOS\1*", i es crearan les següents taules:

- ✚ T: Taula temporal per allotjar les dades provinents dels arxius de població per comunitat autònoma i any.
- ✚ T_POBLACION_CCAA: Taula definitiva per allotjar les dades tractades i unificades de la taula T.
- ✚ T2: Taula temporal per acollir les dades provinents de l'arxiu percentatge de població activa i any.
- ✚ T_PORC_POB_ACT_CCAA: Taula definitiva per allotjar les dades tractades i unificades de la taula T2.
- ✚ T3: Taula temporal per allotjar les dades provinents dels arxius de tributació de les comunitats autònomes de Catalunya, Castilla la Manxa i La Rioja, que porten els mateixos tipus de retribució i per tant el treball que farem per una comunitat ens valdrà per les tres.
- ✚ T4: Taula temporal per allotjar les dades provinents dels arxius de tributació de les comunitats autònomes de Andalusia, Astúries i Canàries que porten els mateixos tipus de retribució.
- ✚ T5: Taula temporal per allotjar les dades provinents del arxiu de tributació de la comunitat autònoma de Madrid.
- ✚ T5NUM: Taula temporal creada a partir de les dades transformades i unificades de T5.
- ✚ T6: Taula temporal per allotjar les dades provinents dels arxius de tributació de les comunitats autònomes de Aragó, Balears, Cantabria i Extremadura que porten els mateixos tipus de retribució.
- ✚ T7: Taula temporal per allotjar les dades provinents dels arxius de tributació de les comunitats autònomes de Castilla i Lleó i Galicia.
- ✚ T8: Taula temporal per allotjar les dades provinents dels arxius de tributació de les comunitats autònomes de Comunitat Valenciana i Murcia.
- ✚ T8NUM: Taula temporal creada a partir de les dades transformades i unificades de T8.
- ✚ T9: Taula temporal per allotjar les dades provinents del arxiu de tributació de Ceuta y Melilla.
- ✚ T_TRIBUTACION: Taula temporal creada a partir de les dades transformades i unificades de les taules T3, T4, T5NUM, T6, T7, T8NUM, T9.
- ✚ T_TRIBUT_CCAA: Taula definitiva per allotjar les dades tractades i unificades de T_TRIBUTACION, aquesta serà la taula de fet.
- ✚ T_ANYOS: Taula definitiva de dimensió Data.
- ✚ T_TIPO_PERCEPTOR: Taula definitiva de la dimensió Tipus de perceptor.
- ✚ T_GRUPO_RETRIBUCION: Taula definitiva de la dimensió tipus de retribució.

Un cop hem creat les taules passem a omplir les temporals, tasca que farem fent servir SQL Loader i els seus arxius de control mitjançant els quals podem carregar les dades a les taules fent servir una sèrie de paràmetres.

Degut a les moltes diferències entre arxius s'ha optat per tenir un arxiu de control per cada arxiu de dades aportat per l'OADT, i centralitzar les dades descartades, les dades mal carregades, i els arxius log de les operacions realitzades a les carpetes "CARGADOS DESCARTES", "CARGADOS MAL" i "LOG" respectivament, també tenim a la carpeta "Datos Tributación" les dades d'origen.

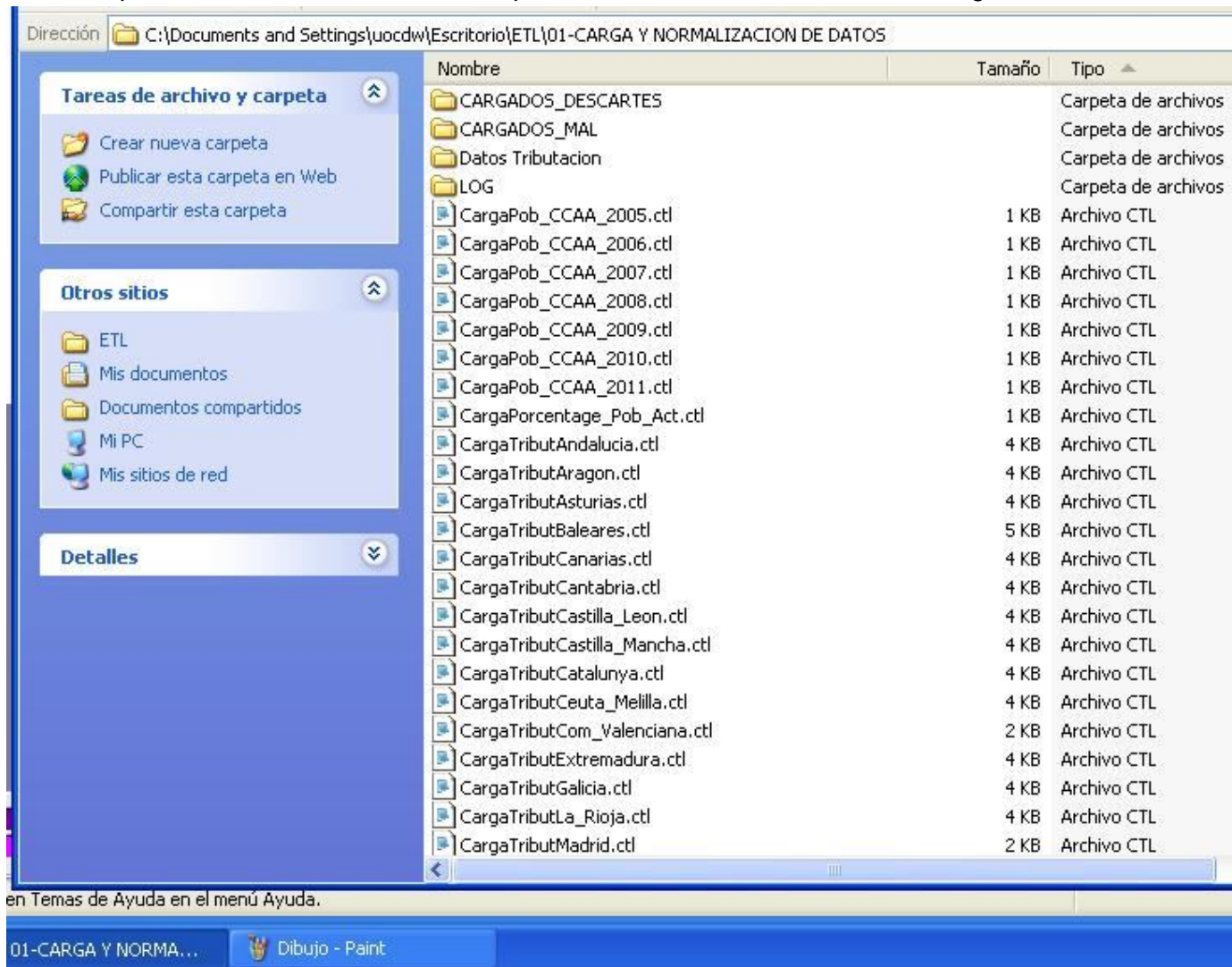


Figura 31: Vista carpeta ETL.

Dels diferents arxius de control de carrega de població per comunitat autònoma i any presentem aquest a tall d'exemple ja que els demés són pràcticament iguals, amb la diferència que al carregar dades cap a la taula posem a l'atribut "ANYO" l'any corresponent a les dades.

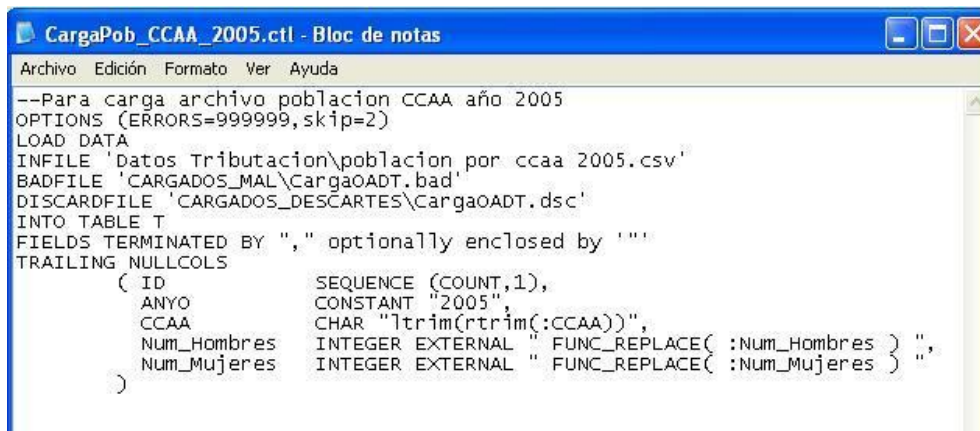
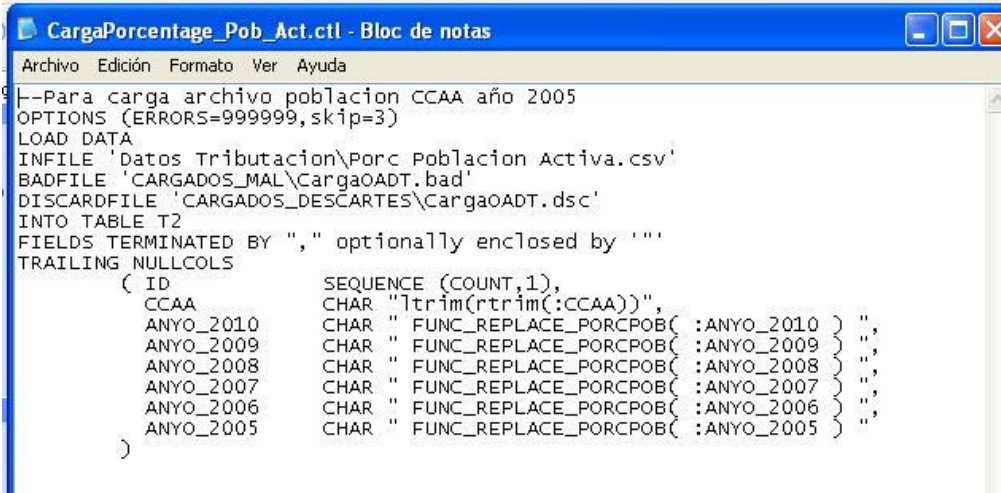


Figura 32: Vista arxiu de control de carrega per SQL Loader (vista 1).

Per explicar una mica que és el que fem amb aquest arxiu, ens saltem les dos primeres línies de l'arxiu de font de dades amb la comanda "skip=2", li diem les rutes de recollida de dades, on deu deixar les dades mal carregades, i les dades descartades amb les comandes INFILE, BADFILE i DISCARDFILE. Li diem a quina taula ha de ficar les dades amb INTO TABLE i finalment li fem un esquema de la taula. En fer l'esquema de la taula hem de dir-li quin tipus d'atribut estem carregant, i podem cridar a funcions o procediments en el moment de la carrega com es pot veure per exemple a "Num_Hombres" a on fem una crida a FUNC_REPLACE, funció que transforma les dades si venen en format científic, treu els dos punts que porten alguns camps, i canvia el punt decimal per una coma.

Per carregar l'arxiu de percentatge de població activa per comunitat tenim el seu arxiu de control a on fem les mateixes operacions que en l'explicat anteriorment.



```
Archivo Edición Formato Ver Ayuda
--Para carga archivo poblacion CCAA año 2005
OPTIONS (ERRORS=999999,skip=3)
LOAD DATA
INFILE 'Datos Tributacion\Porc Poblacion Activa.csv'
BADFILE 'CARGADOS_MAL\CargaoADT.bad'
DISCARDFILE 'CARGADOS_DESCARTES\CargaoADT.dsc'
INTO TABLE T2
FIELDS TERMINATED BY "," optionally enclosed by '"'
TRAILING NULLCOLS
( ID          SEQUENCE (COUNT,1),
  CCAA        CHAR "trim(rtrim(:CCAA))",
  ANYO_2010   CHAR " FUNC_REPLACE_PORCPOB( :ANYO_2010 ) ",
  ANYO_2009   CHAR " FUNC_REPLACE_PORCPOB( :ANYO_2009 ) ",
  ANYO_2008   CHAR " FUNC_REPLACE_PORCPOB( :ANYO_2008 ) ",
  ANYO_2007   CHAR " FUNC_REPLACE_PORCPOB( :ANYO_2007 ) ",
  ANYO_2006   CHAR " FUNC_REPLACE_PORCPOB( :ANYO_2006 ) ",
  ANYO_2005   CHAR " FUNC_REPLACE_PORCPOB( :ANYO_2005 ) "
)
```

Figura 33: Vista arxiu de control de carrega per SQL Loader (vista 2).

Acte seguit carregarem els arxius de tributació de comunitat autònoma tenint per fer-lo possible tants arxius de control com arxius de dades.

```

CargaTributAndalucia.ctl - Bloc de notas
-----
-- Cargamos datos tributacion Andalucia que vienen
-- Periodo: 2003-2009
-- Tipo retribucion:Total,Salario, Pension, Desempleo
-- Param. Retrib: Personas, retrib,retenc
-----

OPTIONS (ERRORS=999999,skip=4)
LOAD DATA
INFILE 'Datos Tributacion\Tributaci4n Andalucia.csv'
BADFILE 'CARGADOS_MAL\CargaOADT.bad'
DISCARDFILE 'CARGADOS_DESCARTES\CargaOADT.dsc'
APPEND
INTO TABLE T4
FIELDS TERMINATED BY "," optionally enclosed by '"'
TRAILING NULLCOLS
( ID SEQUENCE (COUNT,1),
  CCAA CONSTANT "Andalucia",
  GRUPO_RETRIB CONSTANT "-",
  TIPUS_RETRIB CHAR ,
  tpa_2 CONSTANT "",
  tpp_2 CONSTANT "",
  tpd_2 CONSTANT "",
  tpap_2 CONSTANT "",
  tpad_2 CONSTANT "",
  tppd_2 CONSTANT "",
  tpapd_2 CONSTANT "",
  a2 FILLER,
  tpa_3 DECIMAL EXTERNAL " FUNC_REPLACE( :tpa_3 ) ",
  tpp_3 DECIMAL EXTERNAL " FUNC_REPLACE( :tpp_3 ) ",
  tpd_3 DECIMAL EXTERNAL " FUNC_REPLACE( :tpd_3 ) ",
  tpap_3 DECIMAL EXTERNAL " FUNC_REPLACE( :tpap_3 ) ",
  tpad_3 DECIMAL EXTERNAL " FUNC_REPLACE( :tpad_3 ) ",
  tppd_3 DECIMAL EXTERNAL " FUNC_REPLACE( :tppd_3 ) ",
  tpapd_3 DECIMAL EXTERNAL " FUNC_REPLACE( :tpapd_3 ) ",
  a3 FILLER,
  tpa_4 DECIMAL EXTERNAL " FUNC_REPLACE( :tpa_4 ) ",
  tpp_4 DECIMAL EXTERNAL " FUNC_REPLACE( :tpp_4 ) ",
  tpd_4 DECIMAL EXTERNAL " FUNC_REPLACE( :tpd_4 ) ",
  tpap_4 DECIMAL EXTERNAL " FUNC_REPLACE( :tpap_4 ) ",
  tpad_4 DECIMAL EXTERNAL " FUNC_REPLACE( :tpad_4 ) ",

```

Figura 34: Vista arxius de control de carrega per SQL Loader (vista 3).

A tots els arxius de control farem l'esquema de la taula temporal començant des de l'any 2002 i acabant a l'any 2009, ja que hi ha força varietat en els anys amb dades aportats per les comunitats autònomes, per exemple el que es veu més amunt és l'arxius de control de Andalusia, que aporta dades des de 2003 fins 2009 Omplirem amb espais la falta de dades, i ens saltarem l'atribut "Total" amb el tipus d'atribut a l'esquema de la taula "FILLER".

calcula el valor retenció a partir de número de persones, tipus mitja de retenció i retribució mitjana, i *"FUNC_MULTIPlica"*, que multiplica dos valors, essent capaç de canalitzar les excepcions. Al procediment carreguem en tres cursors els registres a on el tipus de retribució és retribució mitjana, tipus mitja de retenció i número de persones per fer dos recorreguts, el primer en el que calcularem retribucions dels diferents tipus de retribucions, i el segon en el que calcularem retencions dels diferents tipus de retribucions. Haurem insertat a la taula T5 els registres necessaris per acollir els càlculs anteriors, i per finalitzar esborrarem els registres que portin valors mitjans.

Per la taula T6, a la que tenim les dades de tributació de les comunitats autònomes de Aragó, Balears, Cantabria i Extremadura hem fet un procediment anomenat *"PR_NORM_T6"* al que actualitzem als registres el tipus de retribució posant salari, pensió i atur, esborrem els registres buits i també esborrem els registres que porten dades de valors mitjans.

Per la taula T7, a la que tenim les dades de tributació de les comunitats autònomes de Castilla i Lleó i Galicia, hem fet un procediment anomenat *"PR_NORM_T7"* al que actualitzem als registres el tipus de retribució posant salari, pensió i atur, esborrem els registres buits.

Per la taula T8, a la que tenim les dades de tributació de la Comunitat Valenciana i Murcia, hem creat un procediment anomenat *"PR_NORM_T8"*, i fem la mateixa operació que a T5, calcular retribucions i retencions per tots els tipus de retribucions, esborrar registres buits, actualitzar els registres de tipus de retribució i finalment esborrar els registres que tinguin dades mitjanes.

Per la taula T9, a la que tenim les dades de tributació de Ceuta y Melilla, hem creat un procediment anomenat *"PR_NORM_T9"*, a on actualitzem als registres el tipus de retribució posant salari, pensió i atur, esborrem els registres buits, sumem els mateixos registres de Ceuta amb els de Melilla i finalment esborrem els sumands.

Creem seqüències i triggers per ajuntar els registres de les taules temporals en una i en fer la inserció en la nova taula portin un número identificador únic i s'inserti un valor 0 en els camps buits. Ajuntem de T3 a T9 en la taula *"T_TRIBUTACION"*.

I per la taula T_TRIBUTACION, que acabem d'omplir amb les dades de tributació de totes les comunitats autònomes, hem creat un procediment anomenat *"PR_CONV_T_TRIB_CON_ANYO"*, al que farem la darrera transformació, i deixarem els atributs de la taula de fet.

Un cop hem finalitzat aquesta fase ja solament ens queda esborrar les taules temporals per així deixar solament les definitives, i això ho fem amb l'arxiu *"BORRA_TABLAS_TEMPORALES.sql"*, ubicat a la mateixa ruta que els anteriors.

Com s'ha comentat al principi de l'apartat aquest procés s'ha automatitzat en un arxiu de treball per lots anomenat "01-CARGA Y NORMALIZACION DATOS OADT.bat", el qual ens anirà demanant que pitgem una tecla per avançar en les fases i a la mateixa vegada ens anirà informant per pantalla de les operacions que esta realitzant.

```

01-CARGA Y NORMALIZACION DATOS OADT.bat - Bloc de notas
Archivo Edición Formato Ver Ayuda
@ECHO OFF
color F0
ECHO
ECHO          #####
ECHO          #  CARGA DE DATOS PROCEDENTES DE #
ECHO          #          OADT                    #
ECHO          #####
ECHO.

REM      Arrancamos SQLPLUS y creamos una tabla que sera temporal para efectuar
REM      cargas de datos

exit|sqlplus system/uoc@XE @Crea_t_temp.sql

cls
@ECHO OFF
color F0
ECHO
ECHO          #####
ECHO          #  CARGA DE DATOS PROCEDENTES DE #
ECHO          #          OADT                    #
ECHO          #####
ECHO.

REM      Cargamos los registros de cada fichero de Poblacion por CCAA y año
REM      en tabla creada anteriormente

sqlldr system/uoc CONTROL=CargaPob_CCAA_2005.ct1 LOG=LOG\logfilePOBCCAA2005.1

cls
@ECHO OFF
color F0
ECHO
ECHO          #####
ECHO          #  CARGA DE DATOS PROCEDENTES DE #
ECHO          #          OADT                    #
ECHO          #####
ECHO.
    
```

Figura 36: Vista arxiu de automatització de ETL.

3.3.-Informes

A continuació presento els informes demanats per l'OADT. S'ha creat un menú a l'Atlas SBI Server pel rol OADT a on agrupem els informes per estatals i per comunitats autònomes, com podrà veure's més endavant a la carpeta "Informes estatals" tenim els informes amb les dades consolidades de totes les comunitats que han aportat, i a la carpeta "Informes por CCAA" tenim els informes a on podem triar la comunitat autònoma.

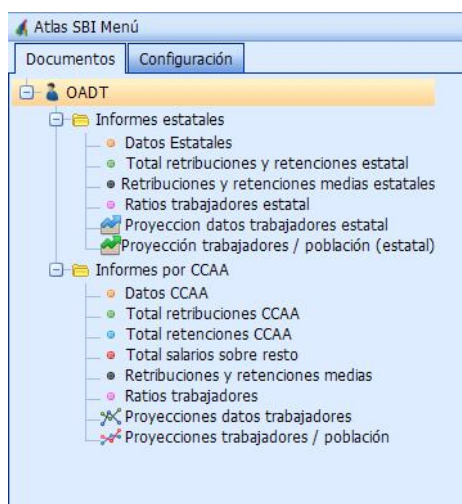


Figura 37: Vista menú informes a Atlas SBI Server.

3.3.1.-Informe total retribucions i retencions

A la carpeta “*Informes estatales*” tenim l'informe “*Retribuciones y retenciones estatales*” que respon a la demanda del client quan demana total retribució i total retenció, en aquest informe veiem les dades conjuntes de totes les comunitats autònomes i podem triar l'any que volem veure, en els gràfics tenim les dades en euros i en percentatge.

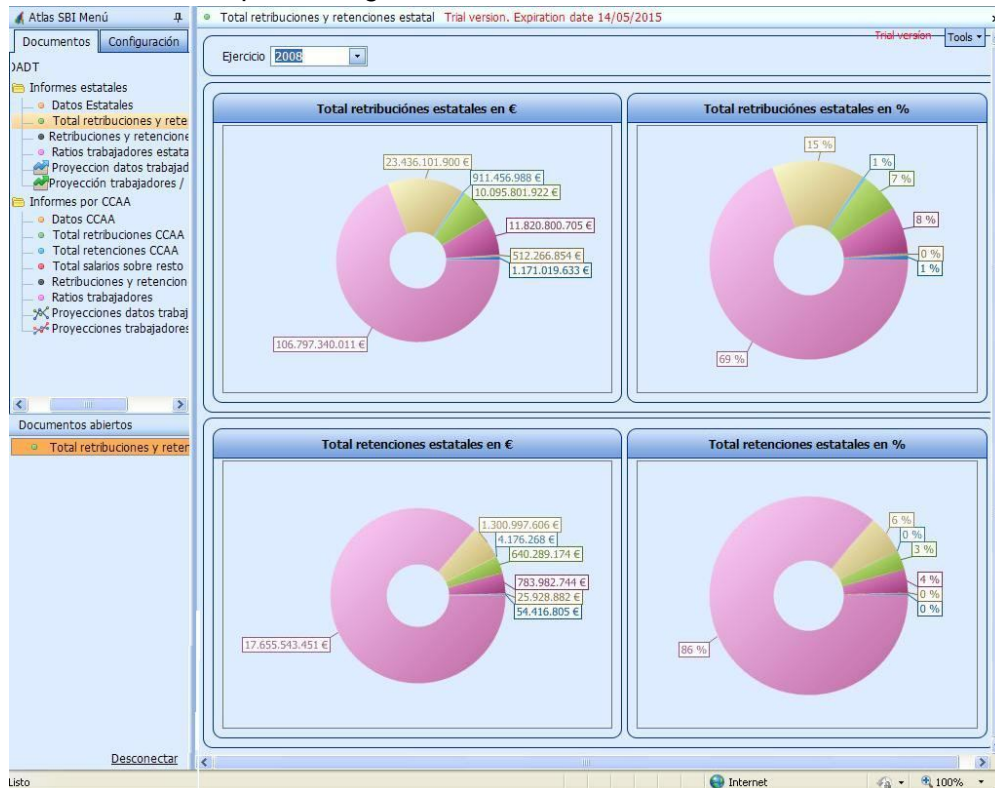


Figura 38: Vista informe Total retribucions i retencions estatal a Atlas SBI Server.

La mateixa informació per CCAA la trobem al menú “*Informes por CCAA*” en dos informes, el primer a on podem veure total retribucions.



Figura 39: Vista informe Total retribucions per CCAA en Atlas SBI Server.

I el segon, total retencions.

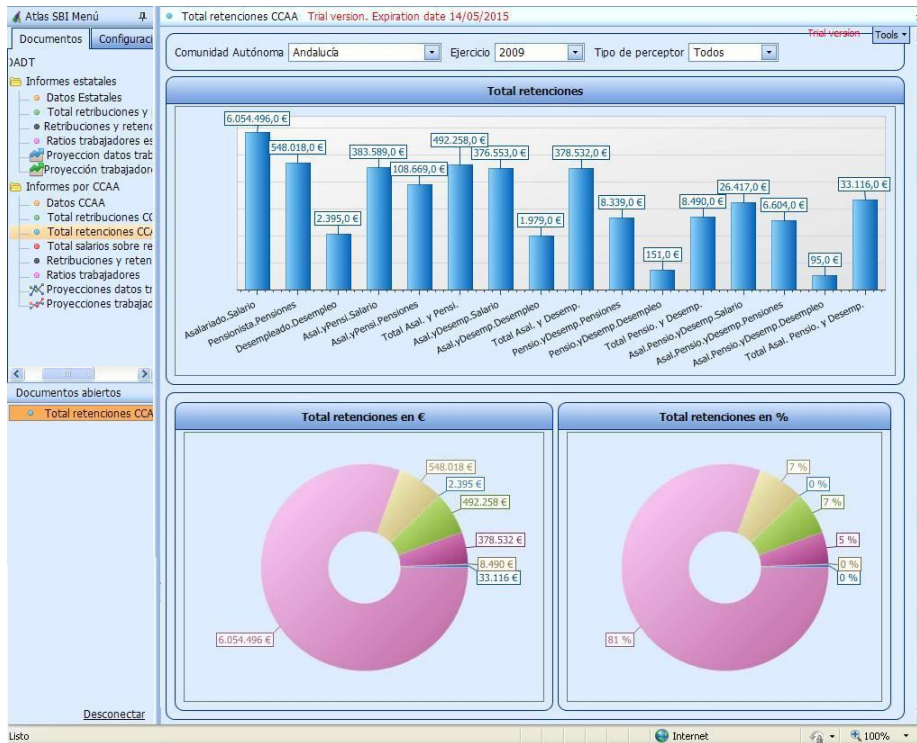


Figura 40: Vista informe Total retencions per CCAA en Atlas SBI Server.

En aquests dos últims informes tenim els combos de dalt a on podem triar la comunitat autònoma i l'any dels que volem la informació.

3.3.2.-Informe retribució mitjana i percentatge de retenció mig

A la carpeta "Informes estatales" tenim l'informe "Retribuciones y retenciones medias estatales" que respon a la demanda del client quan demana retribució mitjana i percentatge de retenció mig, en aquest informe veiem les dades conjuntes de totes les comunitats autònomes i podem triar l'any que volem veure, en els gràfics de barres tenim les dades en euros i en percentatge.

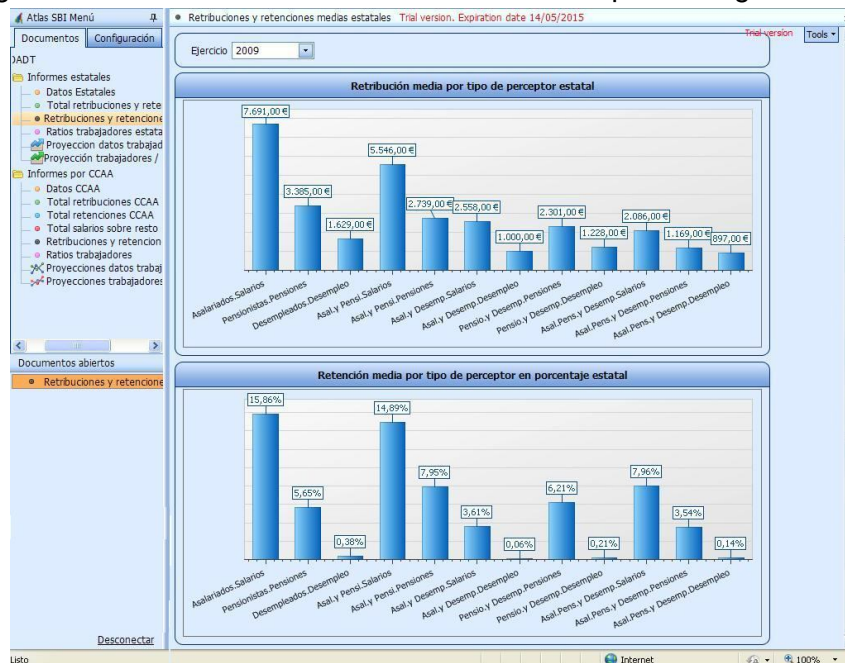


Figura 41: Vista informe retribució mitjana i retenció mitjana estatal en Atlas SBI Server.

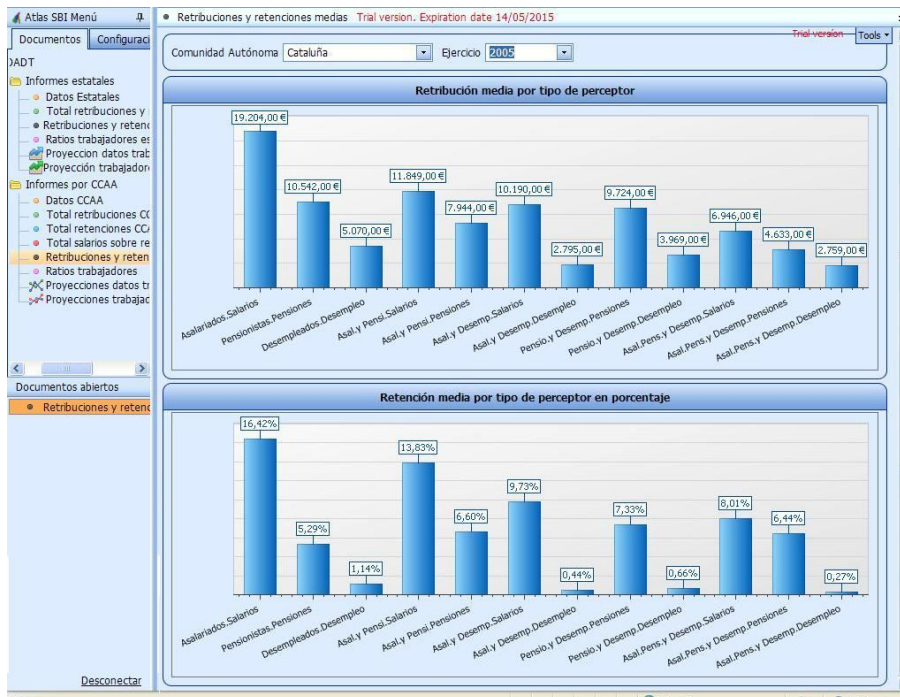


Figura 42: Vista informe retribució mitjana i retenció mitjana per CCAA en Atlas SBI Server.

En aquest informe com en els altres de les comunitats autònomes podem triar l'any i la comunitat de la que volem veure les dades.

3.3.3.-Informe sobre nombre de treballadors respecte altre factor.

A la carpeta "Informes estatales" tenim l'informe "Ratios trabajadores estatales" que respon a la demanda del client quan demana nombre de treballadors respecte nombre de perceptors, o respecte nombre d'habitants, o respecte nombre de persones actives, i també nombre de retribucions mig, en aquest informe veiem les dades conjuntes de totes les comunitats autònomes i podem triar l'any que volem veure.

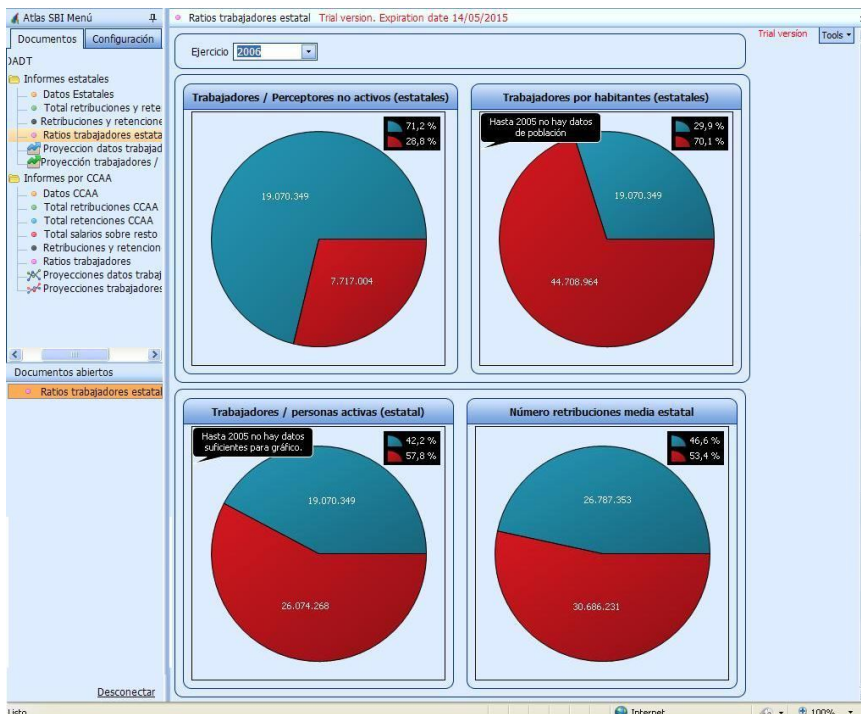


Figura 43: Vista informe Ratios treballadors estatals en Atlas SBI Server.

A la carpeta "Informes por CCAA" tenim el mateix informe enfocat a les comunitats autònomes per separat, a on podem escollir de quina volem obtenir les dades i de quin any.

Com que les dades aportades de població per comunitat autònoma són a partir de l'any 2005 tenim que hi ha gràfics que no donen dades per anys anteriors, així que s'ha optat per posar uns rètols avisant del fet als gràfics afectats.

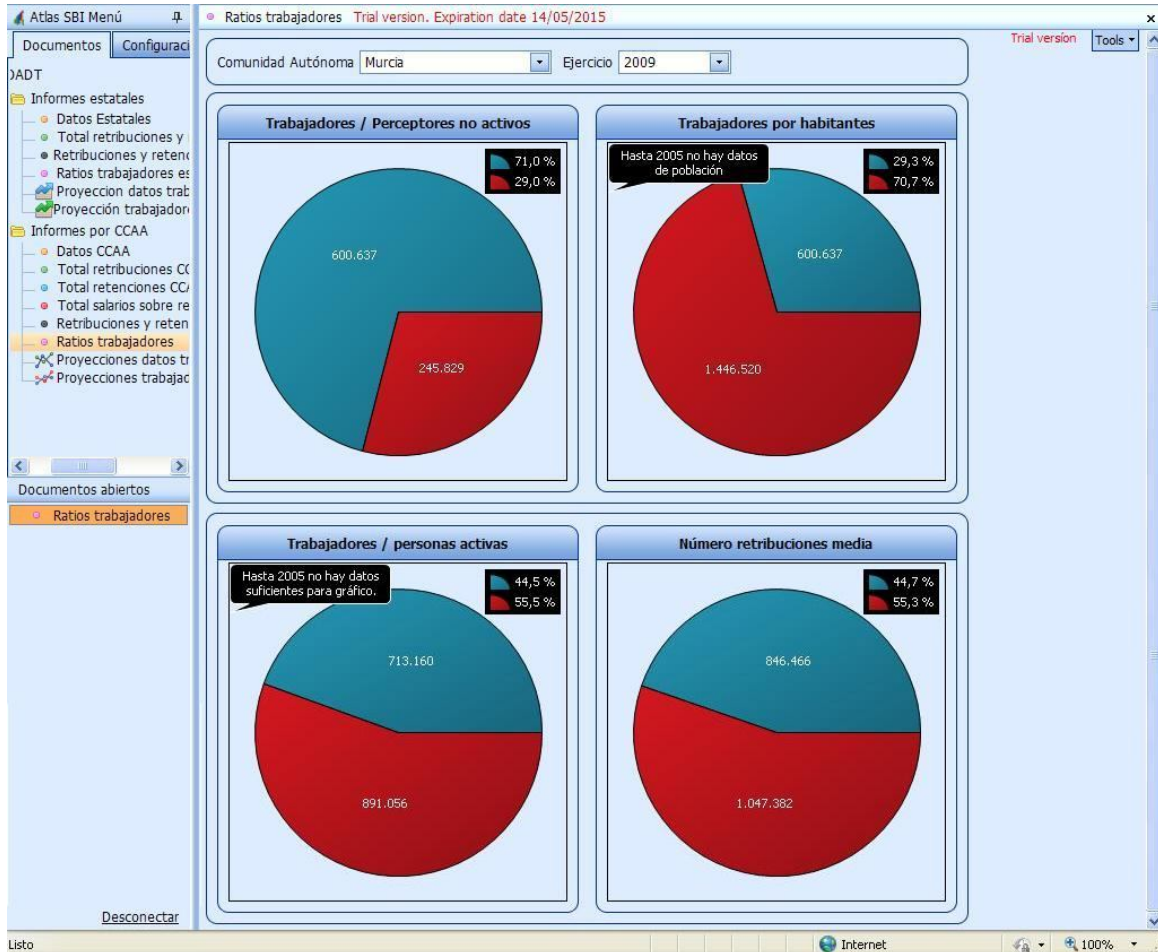


Figura 44: Vista informe Ratios treballadors per CCAA en Atlas SBI Server.

3.3.4.-Informe sobre dades estatals.

A la carpeta "Informes estatales" tenim l'informe "Datos Estatales" que respon a la demanda del client quan demana salari sobre el total de prestacions, o població per segment, en aquest informe veiem les dades conjuntes de totes les comunitats autònomes i podem triar l'any que volem veure.

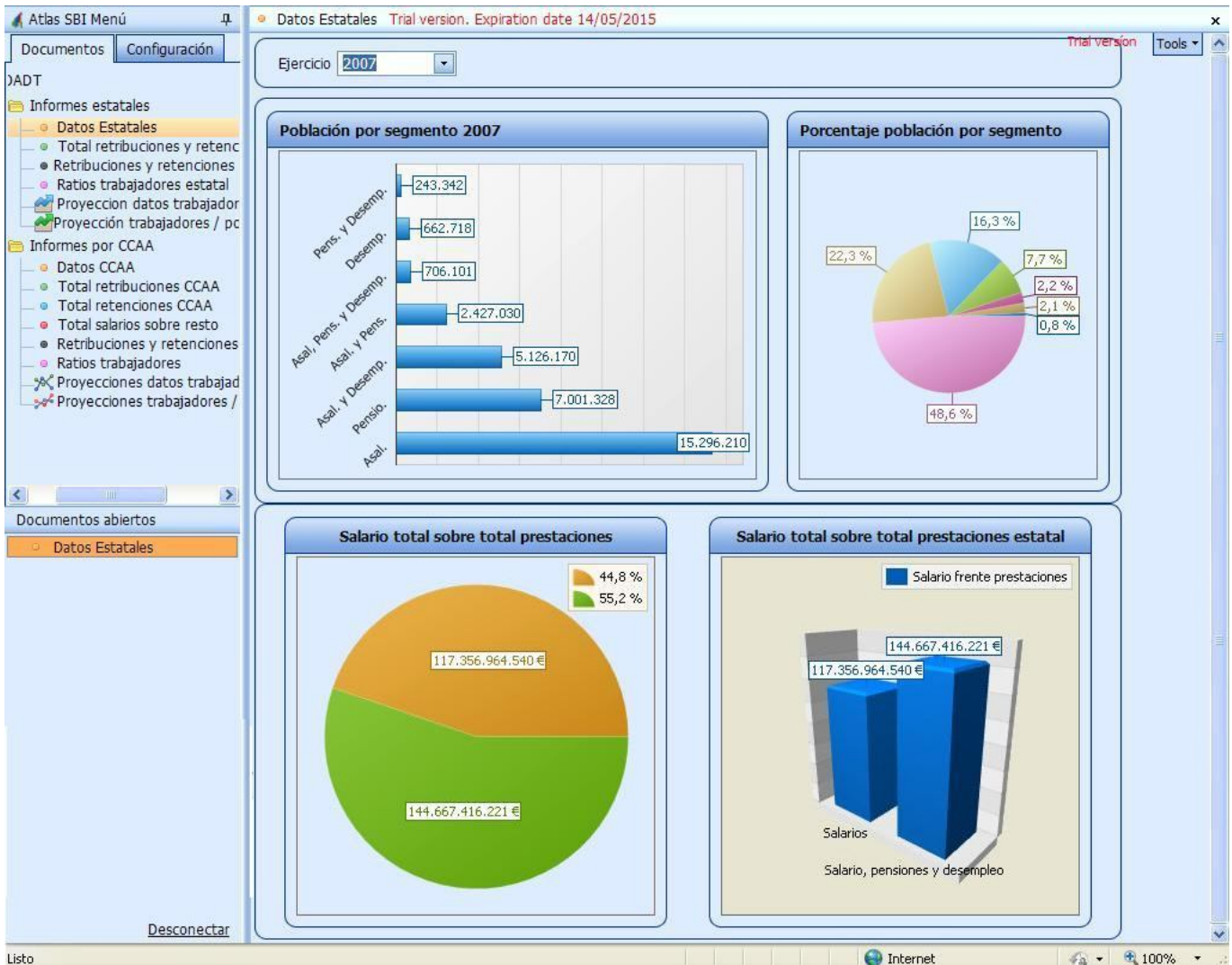


Figura 45: Vista informe Datos Estatales a Atlas SBI Server.

3.3.5.-Informe sobre total salaris sobre la resta de prestacions.

A la carpeta "Informes por CCAA" tenim l'informe "Total salaris sobre resto" que respon a la demanda del client quan demana salari sobre el total de prestacions, en aquest informe podem triar la comunitat autònoma i l'any del que volem veure la informació.

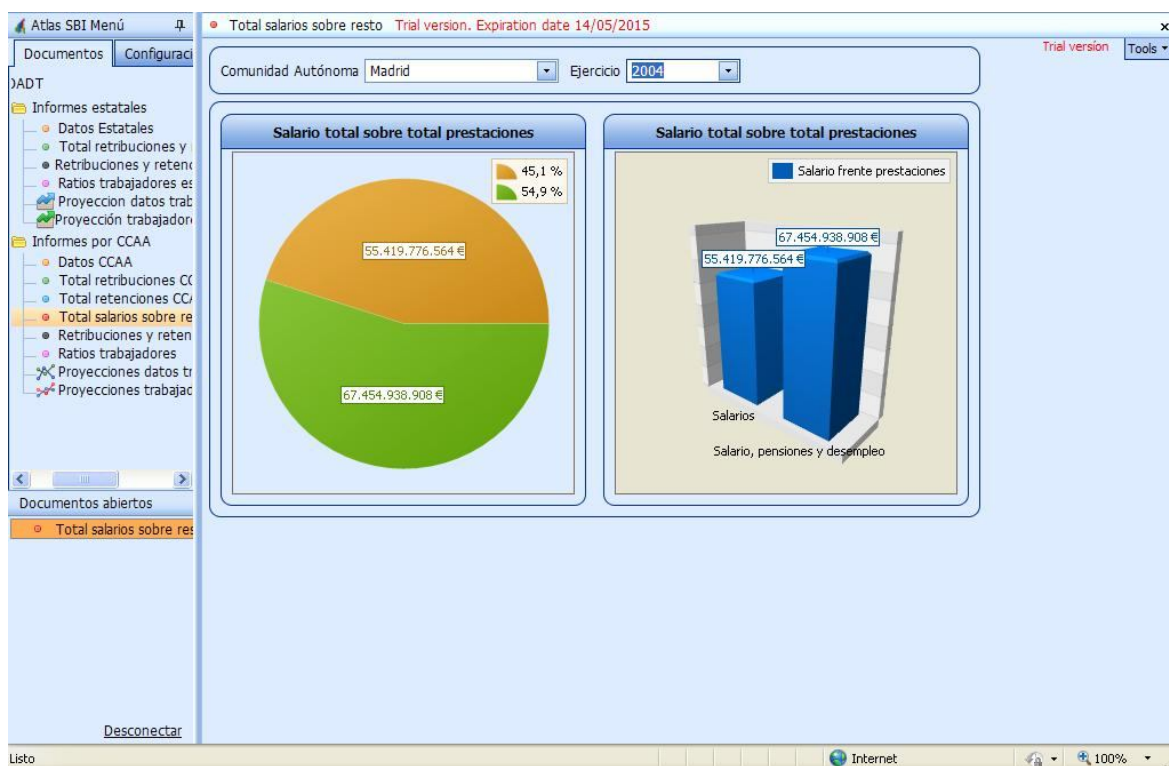


Figura 46: Vista informe Total salaris sobre la resta de prestacions a Atlas SBI Server.

3.3.6.-Informe sobre dades per comunitat autònoma.

A la carpeta "Informes por CCAA" tenim l'informe "Datos CCAA" que respon a la demanda del client quan demana percentatge de població per segment, aquesta informació ja la hem vist de manera estatal i ara la trobem per CCAA en aquest informe, a on podem triar la comunitat autònoma, l'any del que volem veure la informació i el tipus de percepció, a més tenim una taula que mostra les dades de tributació de la comunitat autònoma triada.

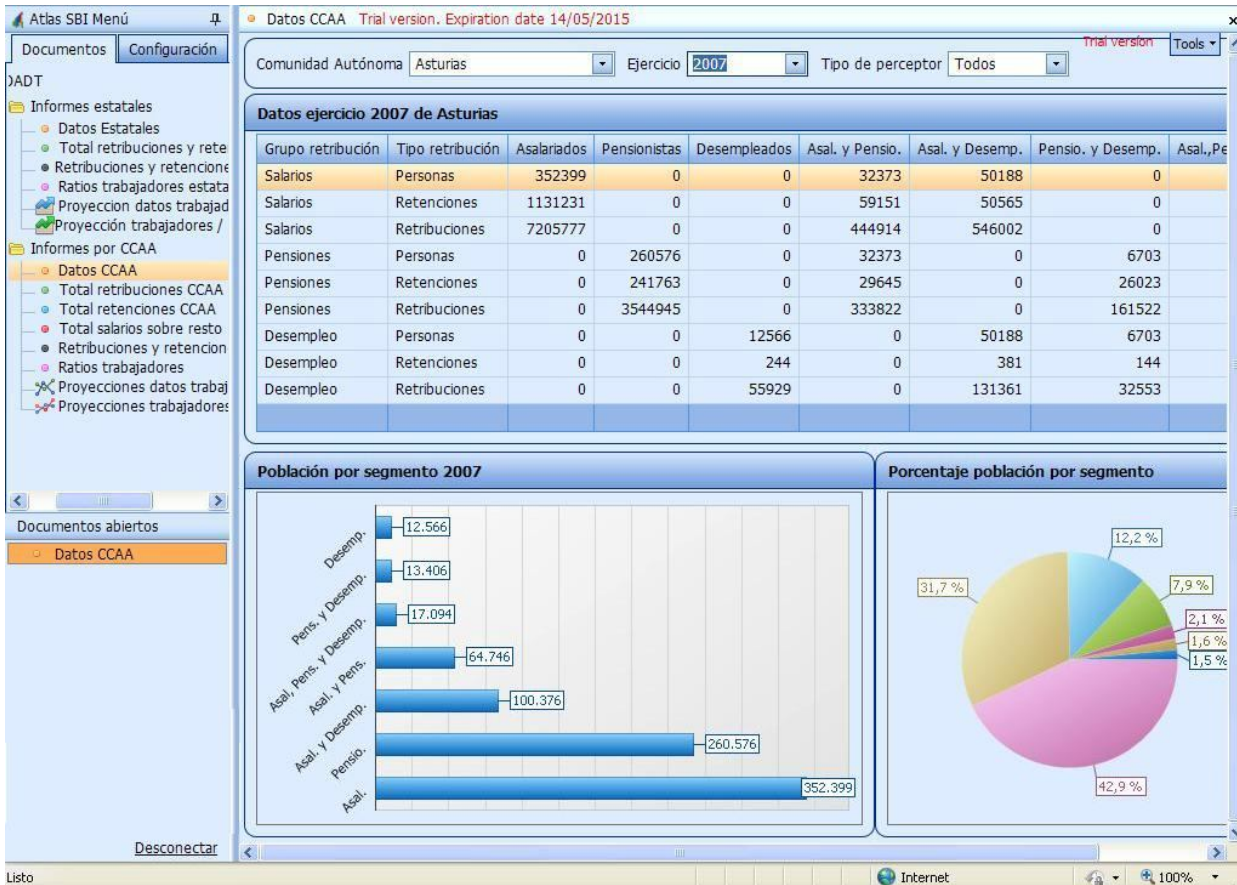


Figura 47: Vista informe dades CCAA a Atlas SBI Server.

3.3.7.-Informe sobre proyecciones de datos de trabajadores.

A la carpeta "Informes estatales" tenim l'informe "Proyección datos trabajadores estatal" que respon a la demanda del client quan demana una projecció d'indicadors treballadors respecte perceptors no actius, en aquest informe veiem les dades conjuntes de totes les comunitats autònomes, tenim els tres gràfics que aporten una visió des de tots els angles, i al gràfic de baix es poden observar les línies de tendència que ens aporten una informació molt valuosa de cap a on van encaminades les dades.



Figura 48: Vista informe proyección de trabajadores respecto a perceptores no activos estatales a Atlas SBI Server.

A la versió per CCAA podem triar la comunitat autònoma a observar.

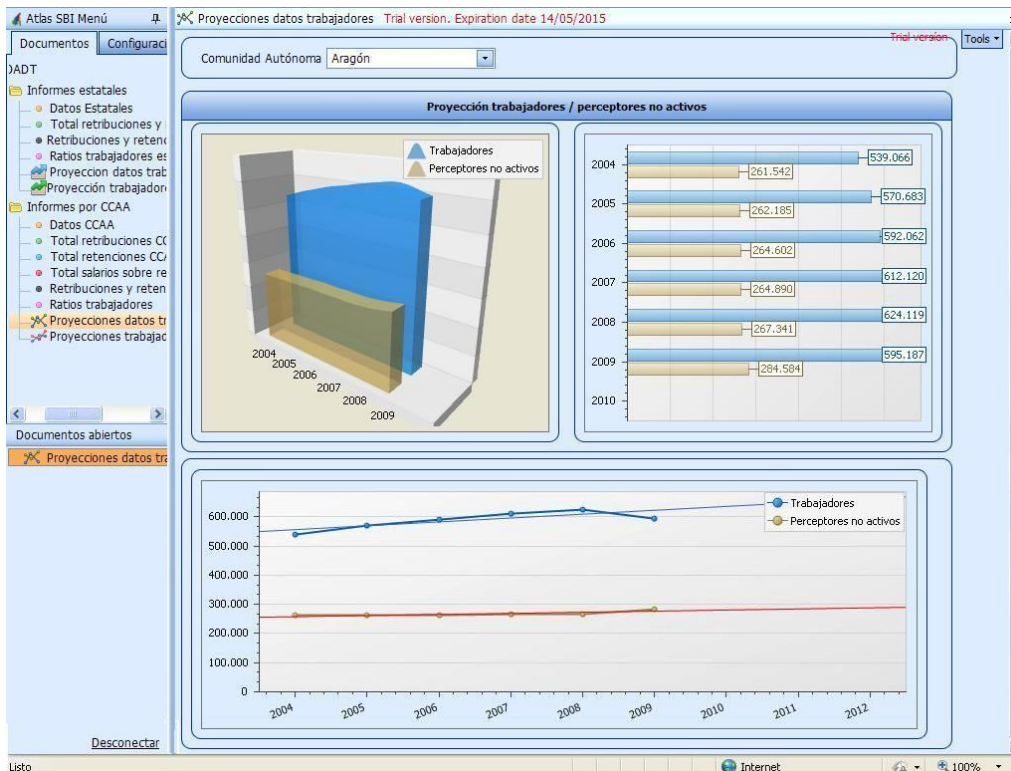


Figura 49: Vista informe proyección de trabajadores respecto a perceptores no activos CCAA a Atlas SBI Server.

3.3.8.-Informe sobre projeccions dades treballadors respecte població.

A la carpeta "Informes estatales" tenim l'informe "Proyección trabajadores / población (estatal)" que respon a la demanda del client quan demana una projecció d'indicadors treballadors respecte a població, en aquest informe veiem les dades conjuntes de totes les comunitats autònomes, al gràfic de baix es poden observar les línies de tendència que ens aporten una informació molt valuosa de cap a on van encaminades les dades com en l'informe anterior.

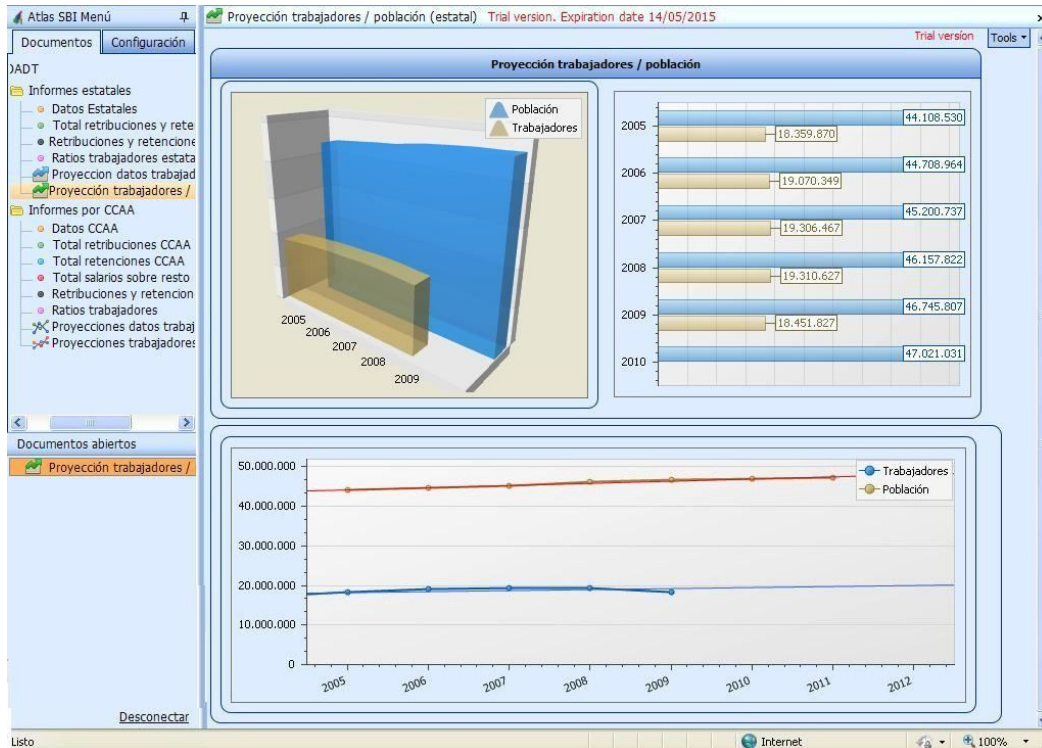


Figura 50: Vista informe projecció treballadors respecte població estatal a Atlas SBI Server.

A la versió per CCAA podem triar la comunitat autònoma a observar.

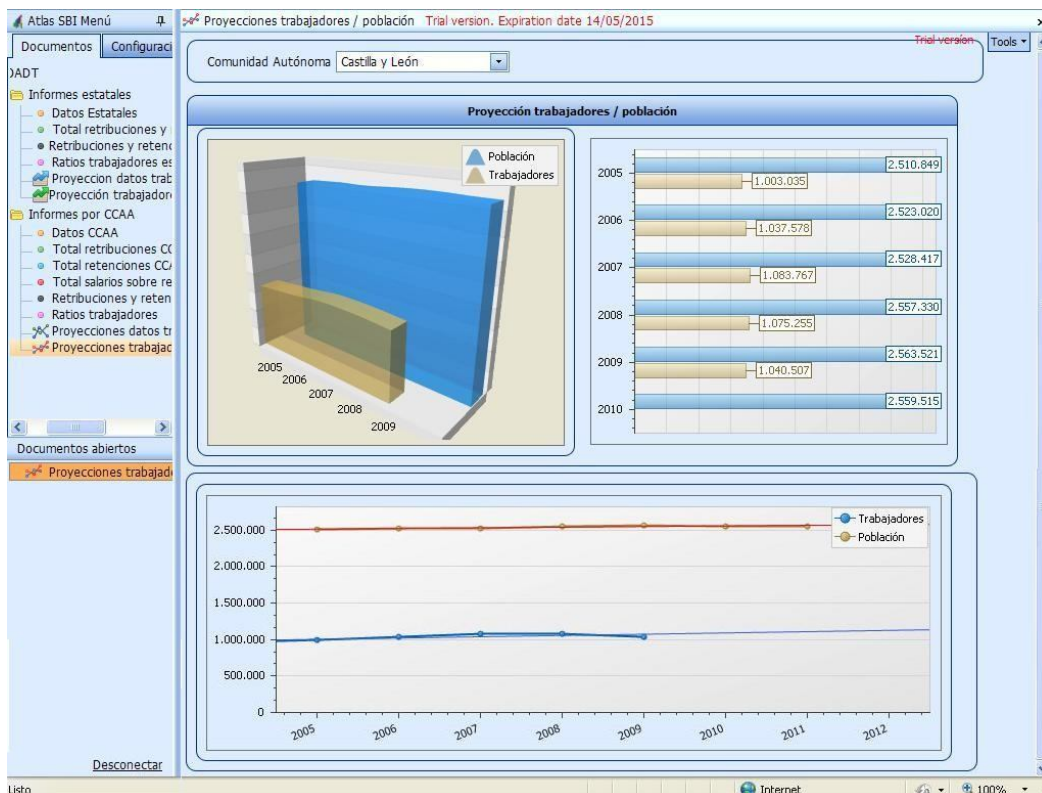


Figura 51: Vista informe projecció treballadors respecte població per CCAA a Atlas SBI Server.

4.-Conclusions

S'han assolit els objectius marcats al començament del projecte en quant a la temporalització.

S'ha realitzat el projecte d'un cas real de creació d'un magatzem de dades.

S'han pogut realitzar totes les consultes necessàries per la elaboració dels informes demanats.

S'ha entregat un producte final.

5.-Línies de evolució futura

Com a millora del producte es podrien treballar les comparacions entre dades estatals i dades per comunitat autònoma per fer uns rankings de paràmetres avaluables, com pot ser la CCAA que aporta més diners a l'estat, la CCAA que s'aprofita més de les prestacions.

Seguint amb la millora anterior també seria prou vistós la realització d'informes mitjançant mapes.

Una altra millora pot ser ampliar les dades aportades i així poder tenir operatius tots els informes del producte, per als primers anys, a on falten dades d'algunes CCAA o de poblacions.

6.-Glossari

Magatzem de dades. Repositori de dades que te la finalitat de recopilar, resumir i tractar grans volums de dades eficientment que existeixen en els sistemes de informació de les organitzacions, proporcionant un anàlisi de la informació en diferents perspectives o dimensions.

Business Intelligence. Conjunt d'eines i estratègies enfocades a la administració i creació de coneixement a partir de l'anàlisi de dades existents a la organització.

OLAP. *On-Line Analytical Processing.* Solució que es fa servir al camp de la intel·ligència empresarial que te com objectiu fer més ràpida la consulta en grans volums de dades.

ROLAP. *Relatinal On-Line Analytical Processing.* Implementació OLAP construïda sobre bases de dades relacionals.

Procés ETL. Procés per la extracció transformació i carrega de dades sobre un magatzem de dades.

Fet. En magatzems de dades representa el concepte que es vol analitzar.

Dimensions. Representen diferents perspectives per realitzar l'anàlisi del fet, permeten agrupar, filtrar i ordenar la informació.

PL / SQL. *Procedural Language Structured Query Language.* Es tracta d'una extensió de Oracle del estàndar *Structured Query Language (SQL)* en forma de llenguatge de programació per crear procediments i funcions que poden ser emmagatzemats com objectes de la base de dades.

SQL. *Structured Query Language.* Llenguatge de consulta estructurat utilitzat per crear, modificar i recuperar dades d'un sistema de base de dades relacional.

Model dimensional. Tipus de model de dades que es fa servir als magatzems de dades. En un model dimensional existeixen dos tipus de taules, les de dimensions i de fets.

7.-Bibliografia.

Llibres de consulta:

The Data Warehouse Toolkit. (2nd Edition). Ralph Kimball and Margy Ross.

Building the Data Warehouse (3rd Edition). W. H. Inmon.

The Data Warehouse ETL Toolkit. Joe Caserta and Ralph Kimball.

Enllaços d'interès a Internet:

<http://www.inmoncif.com>

<http://www.rkimball.com>

<http://www.dwreview.com>

<http://www.1keydata.com/datawarehousing/datawarehouse.html>

<http://www.intranetjournal.com/features/datawarehousing.html>

http://www.oracle.com/technology/docs/tech/sql_plus/index.html

<http://cisnet.baruch.cuny.edu/holowczak/oracle/sqlplus/>

<http://www.oracle.com/technetwork/index.html>

http://www.orafaq.com/wiki/SQL*Loader_FAQ

<http://www.pgconocimiento.com/Productos/AtlasSBI.html>