Implantació de l'eina de *Business Intelligence Pentaho Community 7.0* i aplicació pràctica per al disseny d'un quadre de comandament.

Luis Pérez Sánchez

Grau d'Enginyeria Informàtica

Consultor: Humberto Andrés Sanz

14/06/2017

GNU Free Documentation License (GNU FDL)

Copyright © 2017 Luis Pérez Sánchez.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
A copy of the license is included in the section entitled "GNU Free Documentation License".

FITXA DEL TREBALL FINAL

Títol del treball:	Implantació de l'eina de *Business Intelligence Pentaho Community 7.0* i aplicació pràctica per al disseny d'un quadre de comandament.
Nom de l'autor:	Luis Pérez Sánchez
Nom del consultor:	Humberto Andrés Sanz
Data de lliurament:	06/2017
Àrea del Treball Final:	Business Intelligence
Titulació:	Grau Enginyeria Informàtica

Resum del Treball:

La intel·ligència de negoci (BI) és un procés pel qual a través l'ús de les noves tecnologies permet analitzar dades i presentar informació útil per ajudar als executius, administradors d'empreses i altres usuaris finals a prendre decisions de negoci, d'aquí que sigui una estratègia cada cop més imprescindible per a les empreses de cara a que disposin d'un sistema més fiable que els hi faciliti la presa de decisions.

En aquest treball s'ha analitzat una de les solucions *open source* més populars entre els professionals que s'hi dediquen a aquesta àrea de la Informàtica, *Pentaho Community Edition 7.0*, i una aplicació pràctica d'aquesta solució en l'anàlisi d'un servei d'atenció al ciutadà concret com és una oficina de denúncies d'una administració pública local on es vol analitzar els dies amb més demandes del ciutadans, hores amb major afluència dels ciutadans i el tipus de delictes més denunciats.

Per a tal fi, es duran a terme les diferents etapes imprescindibles en tot procés de BI com són la extracció, transformació i càrrega de dades al magatzem de dades o data warehouse, disseny de consultes àgils que permetin el maneig de gran quantitat de dades i, en darrer lloc, la presentació de les dades en un quadre de comandament com a forma de visualització més intuïtiva i fàcilment comprensible.

***Abstract*:**

The business intelligence (BI) is a process through which the new technologies permit to analyze and present useful information to help the executives, managers of businesses and other users to finalize business decisions. This is followed by a strategy that is essential to every business in order to have a more reliable system that has facilitated the decision-making process.

In this regard, there is one of the most open source solutions among the professionals who work in the area of Information Technology, Pentaho Community Edition 7.0, and a practical application of this solution in the analysis of a local public serves as a service to the citizen like as complaints office of a local government where is necessary analyzing the days of more demands on citizens, hours busiest of citizens and type of crime more complainants.

For this purpose, will carry out the essential stages in any process of BI such as the extraction, transformation and loading of data into the data warehouse, and consultation of large volumes of data and finally the data presentation on a dashboard as the way of an intuitive and easily understood visualisation.

Paraules clau:

Business Intelligence, ETL(Entry – Tranform – Load) , Data Warehouse, OLAP (On line analisy and processing), Pentaho BI Server, Dashboard

Índex

1. Introducció 2
1.1 Context i justificació del Treball 4
1.2 Objectius del Treball 5
1.3 Enfocament i mètode seguit 6
1.4 Planificació del Treball 7
1.6 Breu descripció dels altres capítols de la memòria 8
2. Instal·lació i proves de programari 10
2.1 Plataforma BI (*pentaho-server-ce-7.0.0.0-25*) 12
2.2 *Report Designer* (*prd-ce-7.0.0.0-25*) 15
2.3 *Data Integration* (*pdi-ce-7.0.0.0-25*) (*Extraction Transformation Load*) ... 16
2.4 Motor *OLAP* (*psw-ce-7.0.0.0-25*) *Schema Workbench: Mondrian* 18
3. Disseny del *data warehouse* 20
4. Disseny dels processos ETL (*Entry - Transformation - Load*) 23
5. Disseny d'anàlisi *OLAP* (*On-Line Analytical Processing*) 32
6. Visualització del cub *OLAP* (*On-Line Analytical Processing*) 39
7. Disseny del quadre de comandament (*CDE Dashboard*) 43
7.1 *Layout Panel* 45
7.2 *Component Panel* 48
7.3 *Datasources Panel* 50
8. Conclusions 55
9. Glossari 57
10. Bibliografia 60

Llista de figures

Figura 1 Què és el Business Intelligence 5
Figura 2 Planificació del Treball de Fi de Grau 7
Figura 3: Descripció del hardware i sistema operatiu 10
Figura 4 Variable JAVA_HOME 11
Figura 5 Variable PENTAHO_JAVA_HOME 12
Figura 6 Ubicació dels arxius corresponent a la plataforma de BI 12
Figura 7 Inici de la plataforma BI 13
Figura 8 Versió *Pentaho BI server open source* 14
Figura 9 Pantalla inicial *Pentaho BI server 7.0* 14
Figura 10 Ubicació dels arxius corresponent a la utilitat *Report Designer* 15
Figura 11 Pantalla inicial *Pentaho Report Designer 7.0* 15
Figura 12 Entorn de disseny de *Pentaho Report Designer 7.0* 16
Figura 13 Ubicació dels arxius corresponent a la utilitat *Data Integration* 17
Figura 14 Pantalla inicial *Pentaho Data Integration 7.0* 17
Figura 15 Entorn de disseny de *Pentaho Data Integration 7.0* 18
Figura 16 Ubicació dels arxius de la utilitat *Schema Workbench(Mondrian)* 18
Figura 17 Entorn de disseny de *Schema Workbench* 19
Figura 18 *Data warehouse* implementat en *MySQL* 20
Figura 19 Representació gràfica d'un *data warehouse* 21
Figura 20 IDE de la base de dades *MySQL Workbench 6.3* 21
Figura 21 IDE de la base de dades *MySQL Workbench 6.3* 22
Figura 22 Creació d'una transformació 24
Figura 23 Operació d'entrada de dades 24
Figura 24 Selecció de l'arxiu d'*Excel* 25
Figura 25 Selecció de la fulla de l'arxiu d'*Excel* 25
Figura 26 Columnes detectades al full de l'arxiu d'*Excel* 26
Figura 27 Operació de sortida de dades 26
Figura 28 *Insert / Update* 27
Figura 29 Connexió a *MySQL* 28
Figura 30 Pantalla de creació de la taula denuncies 29
Figura 31 Resultat d'execució 29
Figura 32 Taula creada 30

Figura 33 Transformació de nom “*denuncies”* creada 30
Figura 34 Resultat de l’execució del procés de transformació 31
Figura 35 Execució del procés de transformació a la base de dades 31
Figura 36 Opció per definir la connexió 32
Figura 37 Paràmetres de la connexió a *MySQL* i comprovació 32
Figura 38 Creació nou esquema (*schema*) 33
Figura 39 Creació del cub “Denuncies” 33
Figura 40 Taula de fets “denuncies” 34
Figura 41 Dimensió “DiaSetmana” 35
Figura 42 Jerarquia per defecte 35
Figura 43 Nivell “DiaSetmana” 36
Figura 44 Mètrica “Num Denuncies” 37
Figura 45 Publicació del cub 37
Figura 46 Dades de connexió per publicar el cub 38
Figura 47 Publicació amb èxit 38
Figura 48 Accés a *Saiku Analytics* 39
Figura 49 Pantalla inicial de *Saiku Analytics* 40
Figura 50 Pantalla inicial de *Saiku Analytics* 40
Figura 51 Número de denuncies per dia de la setmana (2015-2016) 41
Figura 52 Emmagatzemat de la consulta. 41
Figura 53 Número de denuncies per hores del dia dels anys 2015 i 2016 42
Figura 54 Número de denuncies per delicte dels anys 2015 i 2016 42
Figura 55 Opció de creació d’un *Dashboard* 43
Figura 56 *Layout Panell* del *Dashboard* 44
Figura 57 *Components Panell* del *Dashboard* 44
Figura 58 Opció de creació d’un *Dashboard* 45
Figura 59 Opció de emmagatzemat d’un *Dashboard* 45
Figura 60 Elements *html* del *Layout Panell* 46
Figura 61 Format del *Layout Panell* 46
Figura 62 Etiqueta *html* del títol de la pàgina 47
Figura 63 Modificació de l’etiqueta *html* del títol de la pàgina 47
Figura 64 Pantalla de disseny del *Component Panell* 48
Figura 65 Elements on es mostrarà la informació. 48
Figura 66 Propietats de l’element *CCC Bar Chart DenunciasHora*. 49

Figura 67 Propietats de l'element *CCC Bar Chart DenunciasDias*. 49
Figura 68 Propietats de l'element *Table Component Delictes*. 50
Figura 69 Pantalla de disseny del *Datasources Panell* 50
Figura 70 Propietats de l'element *mdx over mondrianjndi ds_denDiaSetman*.51
Figura 71 Definició de la consulta MDX. 51
Figura 72 Propietats de l'element *mdx over mondrianjndi ds_denHores*. 52
Figura 73 Propietats de l'element *mdx over mondrianjndi ds_delictes*. 52
Figura 74 Vista preliminar del *Dashboard* (part superior) 52
Figura 75 Localització del quadre de comandament. 53
Figura 76 Resultat final del *Dashboard* (part superior) 53
Figura 77 Resultat final del *Dashboard* (part inferior) 54

1. Introducció

La necessitat de prendre decisions de negoci precises i de forma ràpida s'ha convertit en una de les claus perquè una empresa arribi a l'èxit. No obstant això, els sistemes d'informació tradicionals tals com la majoria dels programes de gestió, les aplicacions a mesura, i fins i tot els ERP(*Enterprise Resource Planning* o Planificació de recursos empresarials) més sofisticats presenten una estructura poc flexible per a aquesta finalitat .

Encara que el seu disseny s'adapta amb major o menor mesura per manejar les dades de l'empresa, no permet obtenir la informació dels mateixos, i molt menys extrapolar el coneixement emmagatzemat en el dia a dia de les bases de dades.

Les principals limitacions d'aquests sistemes són:

- Gran rigidesa a l'hora d'extreure dades.
- Necessitat de coneixements tècnics per a la generació de nous informes.
- Llargs temps de resposta.
- Gran disminució en el rendiment del sistema quan la consulta de la base de dades per generar informes o ràtios de negoci és la mateixa que la que suporta l'operatiu de l'empresa.
- Falta d'integració perquè moltes organitzacions disposen de múltiples sistemes d'informació i les seves bases de dades no solen estar integrades.
- Dades errònies, obsoletes o incompletes.
- Problemes per adequar la informació a càrrec de l'usuari.
- Absència d'informació històrica.

Per superar totes aquestes limitacions sorgeix el *Business Intelligence* recolzant-se en un conjunt d'eines tecnològiques que faciliten l'extracció, la

depuració, l'anàlisi i l'emmagatzematge de les dades generades en una organització per a la seva explotació directa (*reporting*, anàlisi OLAP, mineria de dades, etc.) o per a la seva anàlisi i conversió en coneixement com a suport a la presa de decisions sobre el negoci.

Per tant, es pot definir al *Business Intelligence* com el conjunt de processos tecnològics que transformen dades en coneixement per aconseguir una major competitivitat.

Així doncs, aquest Treball Final de Grau (TFG) es basarà en l'aplicació pràctica d'unes de les plataformes de codi obert que gaudeix de més popularitat entre els professionals de les TIC, *Pentaho Community Edition 7.0*.

Aquesta versió disposa de diverses eines que permeten el següent:

- Disseny d'un *data warehouse* o magatzem de dades que no és més que una col·lecció de dades orientada en un àmbit concret, integrat, no volàtil i variable en el temps.

- Disseny de processos ETL (*Extract, Transform and Load* - extreure, transformar i carregar informació): processos que permet moure gran quantitat de dades des de múltiples fonts transformar-los, depurar-los i carregar-los en altra base de dades, *data mart* o *data warehouse* per a analitzar o en altre sistema operacional per donar suport a un procés de negoci.

- Disseny d'anàlisi OLAP(*On line Analytical Processing* - Processament Analític en Línia): el seu objectiu és agilitzar la consulta de grans quantitats de dades. Per això fa ús d'estructures de dades diverses, normalment multidimensionals (Cubs OLAP), que contenen dades resumides de grans Bases de dades o Sistemes Transaccionals (OLTP). S'empren en informes de negocis de vendes, *marketing*, informes de direcció, mineria de dades i àrees similars.

- Disseny d'informes o transformar les dades en informació: es transformen totes les seves dades en informació significativa adaptada a

l'audiència amb informes en formats PDF, Excel, HTML, text, Rich-Text-File, XML i CSV. Aquests informes generats combinen fàcilment les dades de diverses fonts en un format llegible.

- Disseny de quadres de comandaments o *dashboards*: amb aquesta funcionalitat es permet visualitzar les dades amb panells personalitzats. *Pentaho* ofereix la imaginació, coneixements, eines i serveis per crear quadres de comandament personalitzats per donar suport visualitzacions de dades i dissenys d'interfície d'usuari per oferir una visió analítica en realitat.

1.1 Context i justificació del Treball

Actualment ens hem de situar en l'escenari d'una organització on es manegen gran volum de dades si bé no és un sistema totalment integrat donat que moltes d'aquestes dades es recopilen per mitjans diferents i en repositoris de dades diferents (*ORACLE, SLQ SERVER, MySQL, Excel, Microsoft Access, S390*, etc.) També a l'hora d'extreure informació d'aquests gran volums de dades es fa un disseny individual i un tractament arbitrari de les dades mitjançant l'ús d'eines ofimàtiques com l'*Excel* que amb la funcionalitat de les taules dinàmiques permet la creació de diverses gràfiques, si bé el seu manteniment és costós i no està exempt de males praxis quan al seu maneig i interpretació.

D'aquesta manera, una vegada analitzada la informació i la possibilitat que ens ofereix un sistema de *Business Intelligence*, es veu que el més adient és la seva implantació. Els principals motius són els següents:

- Es disposa d'un repositori de dades o *dataware house* per emmagatzemar totes aquelles dades provinents de diferents fonts;

- l'establiment de processos d'extracció, càrrega i transformació per a poder incorporar les dades de les diferents fonts al repositori habilitat i la seva transformació per a l'obtenció d'informació vàlida;

- la possibilitat d'obtenir coneixement d'aquestes dades que permeti a la direcció de l'organització prendre decisions basades en fets objectius i no de forma intuïtiva o per impulsos.

En definitiva, amb la implantació d'un sistema d'informació com *Pentaho Community Edition 7.0*, es vol innovar clarament en l'obtenció de coneixement de les diferents dades que es maneja a l'organització i que poden donar amb claredat les decisions a prendre per millorar en els processos de negoci.

Figura 1 Què és el Business Intelligence[1]

1.2 Objectius del Treball

L'objectiu principal del present treball consisteix en la implantació d'una eina de *Business Intelligence* com és *Pentaho Community Edition 7.0* i la seva aplicació pràctica per al disseny d'un quadre de comandament o *dashboard*.

Per a dur a terme la seva implantació és descarregarà del repositori (http://community.pentaho.com/) les següents utilitats:

- Plataforma BI (pentaho-server-ce-7.0.0.0-25): proveeix l'arquitectura i la infraestructura a la vegada.

- *Report Designer* (prd-ce-7.0.0.0-25): és una eina gràfica que genera informes de dades que provenen del motor d'integració de dades sense la necessitat de cap taules d'etapes intermèdies. Pot donar sortida als

[1] Business Intelligence (2009). *Qué es el Business Intelligence* [en línia] http://www.buyto.es/general-business-intelligence/que-es-y-para-que-sirve-el-business-intelligence [data de consulta: 14 de març de 2017]

seus informes en diversos formats, PDF, Excel, HTML, rich-text-file , XML i CSV.

- Data Integration (pdi-ce-7.0.0.0-25) (*Extraction Transformation Load*): amb l'aplicació gràfica *SPOON* permet dissenyar tasques automatitzades de transformació de dades per a ser processades per les eines de *Kettle* (PAN y KITCHEN)

- Motor OLAP (psw-ce-7.0.0.0-25) *Schema Workbench: Mondrian* , forma part del motor OLAP integrada en la *SUITE BI* de *PENTAHO*

D'altra banda, s'haurà d'establir i determinar el coneixement que es vol obtenir. Això implicarà a la vegada l'anàlisi de les dades que es disposen així com si aquestes permeten arribar a obtenir la informació per a prendre les decisions de millora de l'organització.

Per a aquest objectiu s'haurà de determinar totes aquelles dades necessàries per a arribar a l'obtenció de la informació i que es troben a diferents fonts de dades, recopilar-les al repositori a través dels processos d'extracció, transformació i càrrega per a poder-ne dissenyar els informes i els quadres de comandaments.

D'aquesta manera es passarà d'un model d'obtenció d'informació basat en taules dinàmiques d'Excel a un model més professional basat en les funcionalitats que proporciona una eina de *Business Intelligence* com *Pentaho Community Edition 7.0.*

1.3 Enfocament i mètode seguit

Bàsicament l'enfocament i la metodologia de treball serà la substitució del procediment d'extracció de les dades que s'obtenen en l'escenari actual d'una forma totalment artesanal mitjançant eines ofimàtiques amb l'eina de BI de *Pentaho Community Edition 7.0*.

D'aquesta manera al llarg del desenvolupament del projecte es mostrarà la forma d'obtenció de dades actuals i s'habilitarà una estratègia similar però fent ús de la metodologia del *Business Anàlisi*. És possible que durant la

implementació de l'extracció de dades sorgeixi nou coneixement que fins ara amb les eines que es feien servir eren impossible d'obtenir, per tant, es tractarà d'un treball en certa manera experimental també per la manca de coneixements profunds en l'ús d'aquesta solució.

Ara bé, s'ha de tenir present que la complexitat d'ús de les diferents funcionalitat i el seu elevat cost d'aprenentatge pot ocasionar que durant el decurs del present treball s'hagi de contemplar la revisió dels objectiu per tal d'adaptar-los al temps disponible la qual cosa pot esdevenir en una modificació de la programació establerta a l'apartat de la programació de les tasques.

1.4 Planificació del Treball

S'ha procedit a temporalitzar les diferents etapes requerides d'acord a la programació de les diferents Proves d'Avaluacions Continuades per a l'elaboració d'aquest Treball de Fi de Grau, tot respectant les dates clau de lliurament de les diferents activitats d'avaluació continuada i de la memòria final.

A continuació es detalla el diagrama de *Gantt* realitzat amb una solució *open source* anomenada *OpenProj*:

	Nombre	Duración	Inicio	Terminado
1	TFG – Business Intelligence	80 days	23/02/17 8:00	14/06/17 17:00
2	**PAC 1**	15 days	**23/02/17 8:00**	**15/03/17 17:00**
3	Cerca d'informació i obtenció de software	7 days	23/02/17 8:00	3/03/17 17:00
4	Context i justificació del treball	2 days	6/03/17 8:00	7/03/17 17:00
5	Objectius del treball	2 days	8/03/17 8:00	9/03/17 17:00
6	Enfocament i mètode seguit	2 days	10/03/17 8:00	13/03/17 17:00
7	Planificació del treball	2 days	14/03/17 8:00	15/03/17 17:00
8	**PAC 2**	25 days	**16/03/17 8:00**	**19/04/17 17:00**
9	Intal·lació i proves de software	12 days	16/03/17 8:00	31/03/17 17:00
10	Disseny del data warehouse	9 days	3/04/17 7:00	13/04/17 17:00
11	Disseny dels processos ETL	10 days	6/04/17 7:00	19/04/17 17:00
12	**PAC 3**	25 days	**20/04/17 8:00**	**24/05/17 17:00**
13	Disseny d'anàlisi OLAP	18 days	20/04/17 8:00	15/05/17 17:00
14	Visualització del cub OLAP	4 days	16/05/17 8:00	19/05/17 17:00
15	Disseny dels quadres de comandament	3 days	20/05/17 8:00	24/05/17 17:00
16	**LLIURAMENT MEMÒRIA**	15 days	**25/05/17 8:00**	**14/06/17 17:00**
17	Adaptació de les PACs a la plantilla del TFG	2 days	25/05/17 8:00	26/05/17 17:00
18	Revisió del contingut d'acord a l'AC de les PACs	10 days	29/05/17 8:00	9/06/17 17:00
19	Revisió del contingut d'acord als criteris d'avaluació	2 days	10/06/17 8:00	13/06/17 17:00
20	Lliurament del TFG	1 day	14/06/17 8:00	14/06/17 17:00

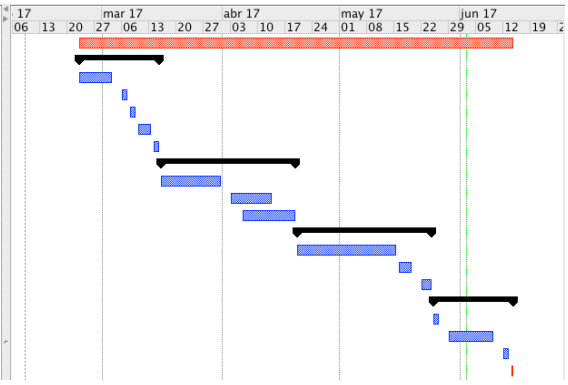


Figura 2 Planificació del Treball de Fi de Grau

Seguint la metodologia de la guia PMBOK (*Project Management Body of Knowledge*), les tasques es mostren en cascada. En un desenvolupament d'aquest tipus, es disposa d'una data final de desenvolupament orientativa i s'ha de tenir present que qualsevol desviament en les tasques pot afectar a

l'inici de la següent tasca. Això es posa de manifest quan s'observa que en aquest cas les tasques estan enllaçades de forma que s'ha d'haver realitzat la tasca prèvia per a començar la següent.

1.5 Breu sumari de productes obtinguts

Aquest treball no suposa el lliurament de cap producte de programari. Així doncs, la memòria final del Treball de Fi de Grau es confeccionarà amb els successius lliuraments dels resultats finals de cada una de les Proves d'Avaluació Continuada (PAC) numerades 1, 2 i 3 d'acord a la planificació establerta.

Finalment, una vegada que sigui avaluada positivament pel Tribunal Qualificador, es procedirà a la publicació al repositori d'O2 de la UOC.

1.6 Breu descripció dels altres capítols de la memòria

A l'apartat 2 es detallarà el procés d'instal·lació de la suite de *Business Intelligence*, *Pentaho Community Edition 7.0*. Es tracta d'una versió de codi obert del popular programari d'intel·ligència de negoci *Pentaho BI Suite 7.1*. Per a tal fi s'haurà de complir una sèrie de requisits tant de hardware com de software que es troben a la documentació del producte. Una vegada instal·lats tots els components que composen el producte, s'efectuaran els ajustaments necessaris de variables del sistema operatiu per tal de que les diferents eines funcionin adequadament.

Al 3er apartat s'aborda el disseny del *dataware house* o magatzem de dades. En el context de la intel·ligència de negoci, el magatzem de dades alberga una col·lecció de dades orientada a un determinat àmbit (empresa, organització, etc.), integrat, no volàtil i variable en el temps, que ajuda a la presa de decisions en l'entitat en la qual s'utilitza. En aquest sentit, en el present treball s'ha determinat que les dades s'emmagatzemin en un SGBD com és *MySQL Workbench 6.3*.

L'apartat 4 detalla el procés de recopilació de dades i la seva incorporació al magatzem de dades. Aquest procés permet a les organitzacions moure dades

des de múltiples fonts (bases de dades, fulles càlculs, fitxers de text pla, etc), reformatar-los així com netejar-los, i carregar-los en una altra base de dades o *data warehouse* per analitzar, o en un altre sistema operacional per recolzar un procés de negoci. D'aquí que aquest procés s'anomeni ETL (*Entry, transformation and load*).

Quant al disseny d'anàlisi OLAP (*on-line analytical processing*), el principal objectiu es l'obtenció d'una consulta àgil que permeti el maneig de grans quantitats de dades. Per a això s'utilitzen estructures multidimensionals que contenen dades resumides de grans bases de dades. En aquest cas, l'ús de l'eina *Schema WorkBench* permetrà el disseny de consultes que s'anomenen cubs com a forma de representació gràfica més intuïtiva.

Si l'anterior apartat es centrava en el disseny del cub *OLAP*, és precisament en aquest apartat on es duu a terme la visualització del cub dissenyat i si realment el resultat és el que en un principi s'havia determinat. D'aquesta manera les dades es presenten en un format de tipus taula on el producte cartesià de les files amb les columnes presenten dades referent precisament a aquest dos paràmetres.

Al darrer apartat pràctic es dissenya una de les característiques més en auge en els darrers temps que són el disseny del quadre de comandament (*CDE dashboard*). El quadre de comandament o *dashboard* és una representació gràfica dels principals indicadors (*KPI – Key Performance Indicator*) que intervenen en la consecució dels objectius de negoci, i que està orientat a la presa de decisions per optimitzar l'estratègia de l'organització. Es tracta doncs d'un element primordial per a la direcció de l'organització empresa que li permet saber en tot moment la consecució o no dels objectius establerts. És similar al quadre de comandament d'un vehicle que permet veure els principals indicadors del seu funcionament.

Finalment a l'apartat de conclusions es realitza una anàlisi personal de l'autor del present en el sentit d'exposar aspectes com són la facilitat d'instal·lació, ús de les diferents utilitats que composen la suite *open source* de Pentaho així com les possibilitats de millora de cara al futur.

2. Instal·lació i proves de programari

Segons la documentació de *Pentaho Community Edition 7.0*[2], els requisits d'instal·lació són els següents:

- Hardware(mínim): Windows
  - Processador de doble nucli
  - RAM 4 Gb
  - HD 100 Gb
- Software
  - Java JDK 1.8
  - MySQL
  - Controladors per a MySQL (JDBC Driver for MySQL)
  - Windows 7

Concretament, la màquina sobre la que s'instal·larà les diferents eines disposa de les següents característiques:

Figura 3: Descripció del hardware i sistema operatiu

[2] Pentaho.com (2017). *Components Reference* [en línia] https://help.pentaho.com/Documentation/7.0/0D0/160/000 [data de consulta: 16 de març de 2017]

A més, s'han de definir les següents variables d'entorn per a que la suite d'eines de *Pentaho* funcionin adequadament:

- Variable JAVA_HOME:
  - Nom de la variable: JAVA_HOME
  - Valor de la variable: C:\Program Files\Java\jdk1.8.0_121

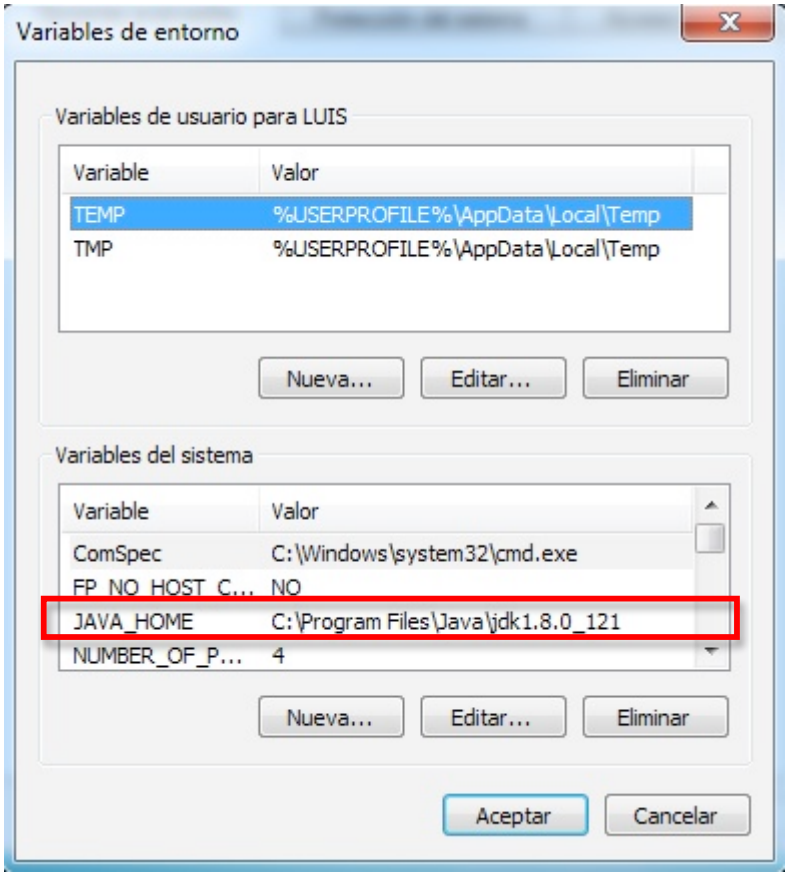


Figura 4 Variable JAVA_HOME

- Variable PATH: afegir la cadena ";%JAVA_HOME%"
- Afegir una nova variable de sistema PENTAHO_JAVA_HOME:
  - Nom de la variable: PENTAHO_JAVA_HOME
  - Valor de la variable: C:\Program Files\Java\jdk1.8.0_121

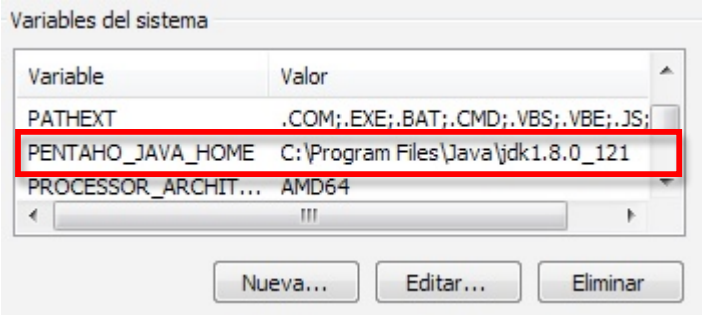


Figura 5 Variable PENTAHO_JAVA_HOME

D'acord als objectius establerts, la instal·lació del programari necessari per al desplegament de la versió *open source* de *Pentaho Community Edition 7.0* s'obtindrà del repositori (http://community.pentaho.com/) segons el següent:

2.1 Plataforma BI (*pentaho-server-ce-7.0.0.0-25*)

En la pàgina principal del portal *Pentaho Community Edition 7.0*, es selecciona l'opció *Download section* que permet seleccionar les funcionalitats a descarregar. En aquest cas, a la secció *Business Analitycs Platform* es tria l'opció *ALL OS* i es descarrega un arxiu comprimit de nom "pentaho-server-ce-7.zip". A continuació, es descomprimeix i la carpeta "pentaho-server" s'ubica al directori arrel "c:\":

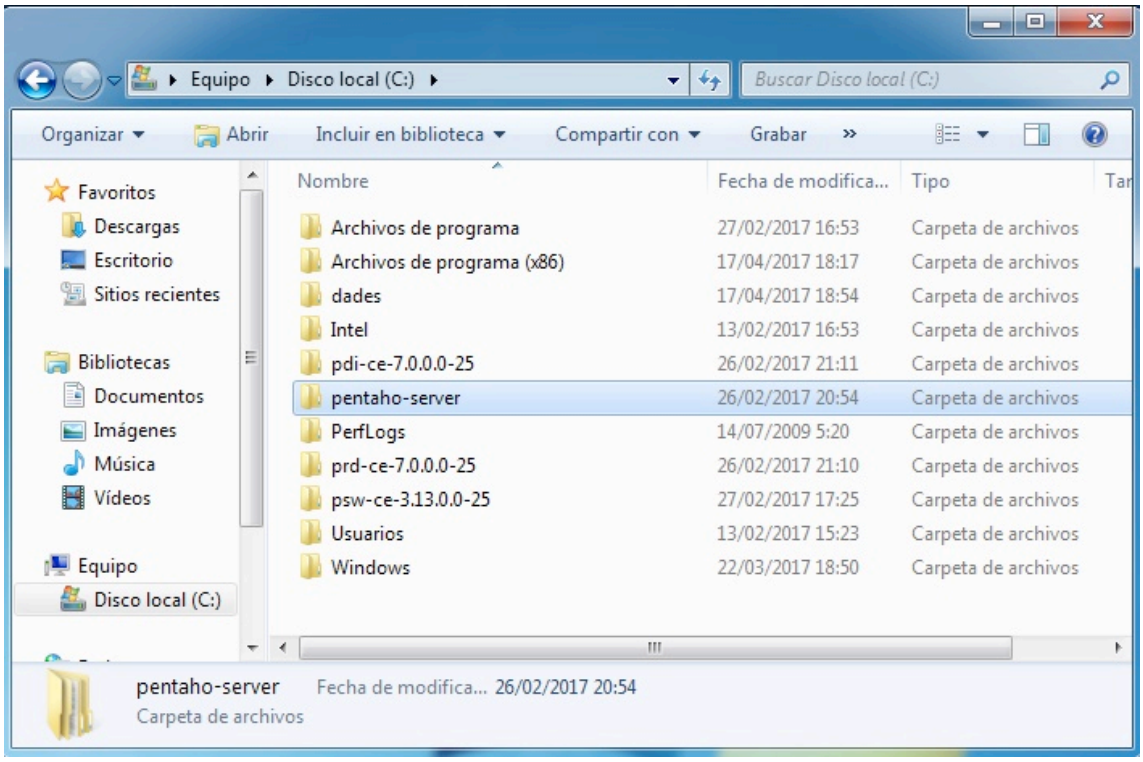


Figura 6 Ubicació dels arxius corresponent a la plataforma de BI

A continuació, s'executa l'arxiu de nom *start-pentaho.bat* i es mostra una finestra de la consola de comandes on es veu clarament que es van executant diferents arxius en la màquina virtual de Java.

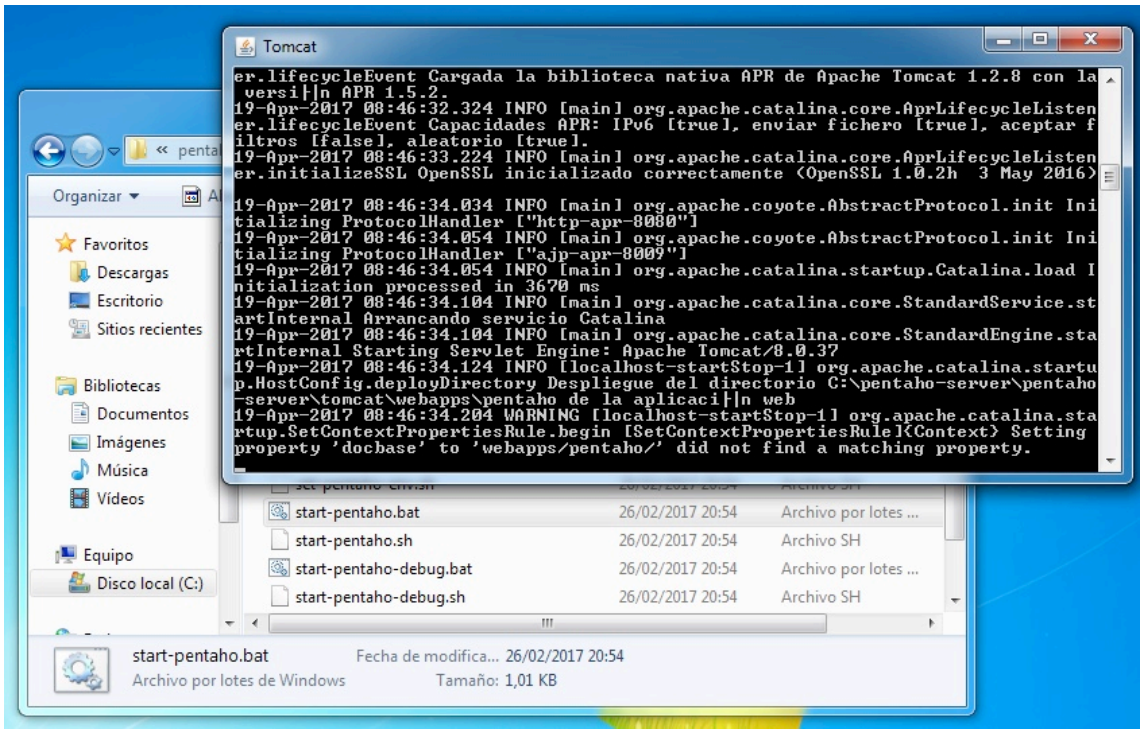


Figura 7 Inici de la plataforma BI

Passats uns minuts, apareix el missatge "*Server startup in ... ms*", que indica que la plataforma web ja està disponible i a través de l'adreça *http://localhost:8080* es podrà veure com es troba en funcionament:

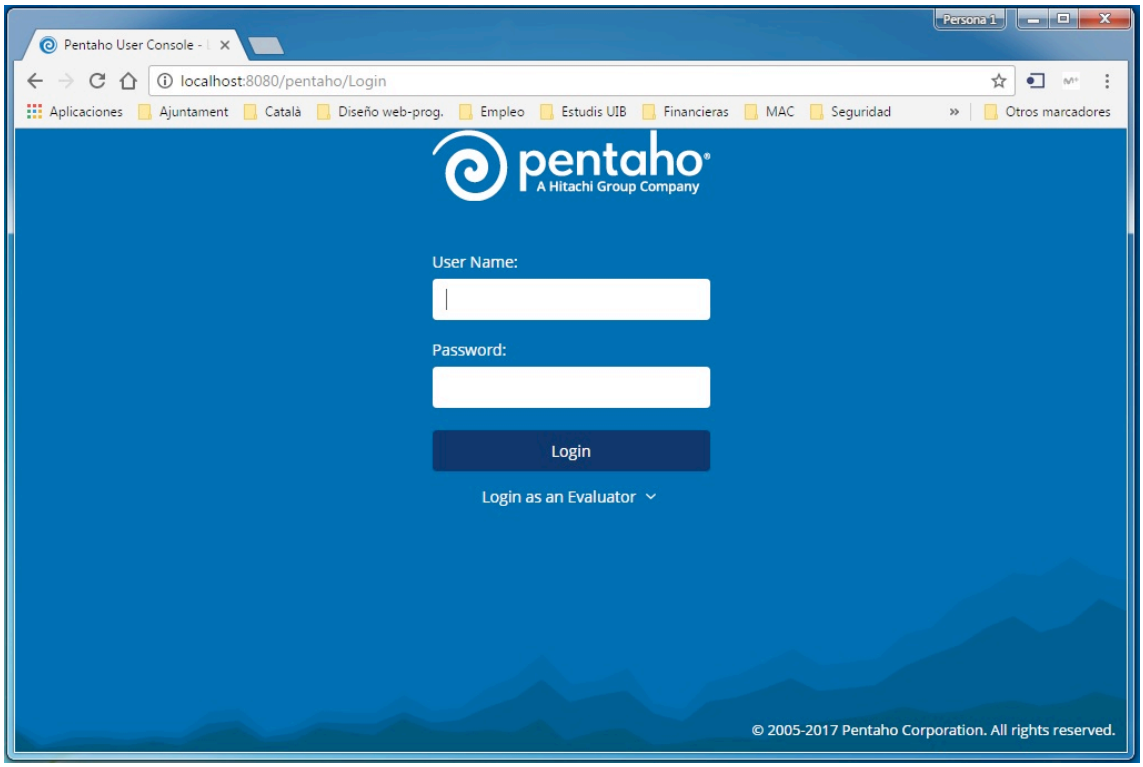


Figura 8 Versió *Pentaho BI server open source*

Finalment, es fa ús de l'opció *Login as an Evaluator* per accedir a la pantalla inicial de la plataforma.

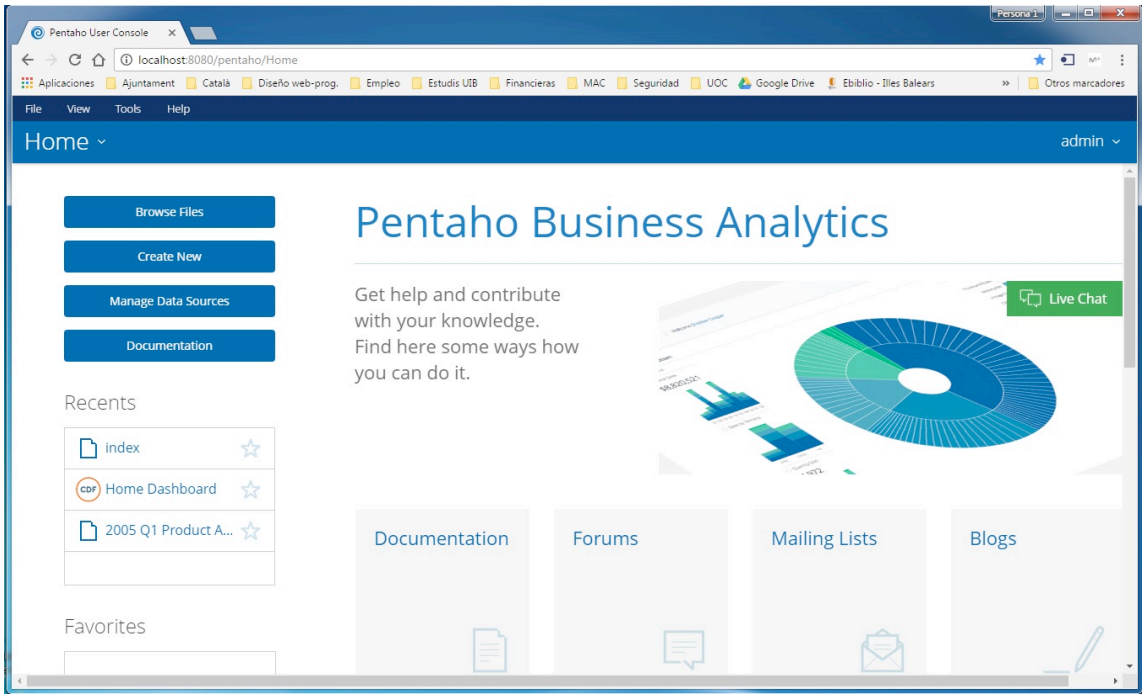


Figura 9 Pantalla inicial *Pentaho BI server 7.0*

2.2 *Report Designer* (*prd-ce-7.0.0.0-25*)

A l'igual que a l'apartat anterior, es selecciona en la pàgina principal del portal *Pentaho Community Edition 7.0* l'opció *Download section* i, en aquest cas, a la secció *Report Designer* es tria l'opció *Windows* i es descarrega un arxiu comprimit de nom "prd-ce-7.0.0.0-25.zip". A continuació, es descomprimeix i la carpeta "prd-ce-7.0.0.0-25" s'ubica al directori arrel "c:\":

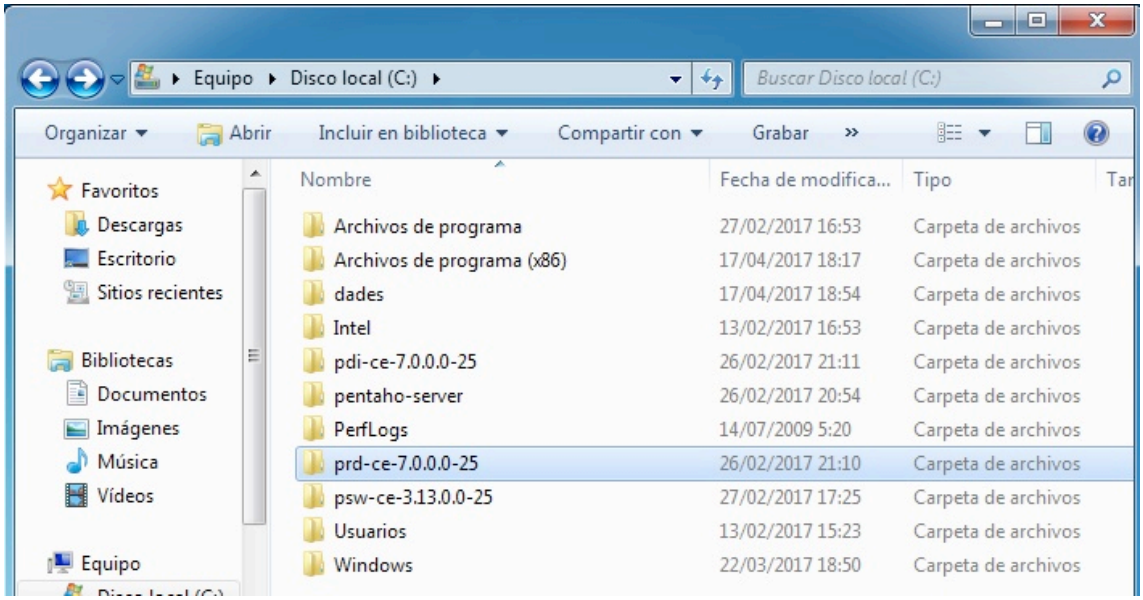


Figura 10 Ubicació dels arxius corresponent a la utilitat *Report Designer*

Acte seguit, s'executa l'arxiu de nom *report-designer.bat* i es mostra una finestra corresponent al dissenyador d'informes:

Figura 11 Pantalla inicial *Pentaho Report Designer 7.0*

Finalment, passats uns minuts, apareix la pantalla principal corresponent a l'IDE (*Integrated Development Environment* o entorn de desenvolupament integrat) de *Pentaho Report Designer*.

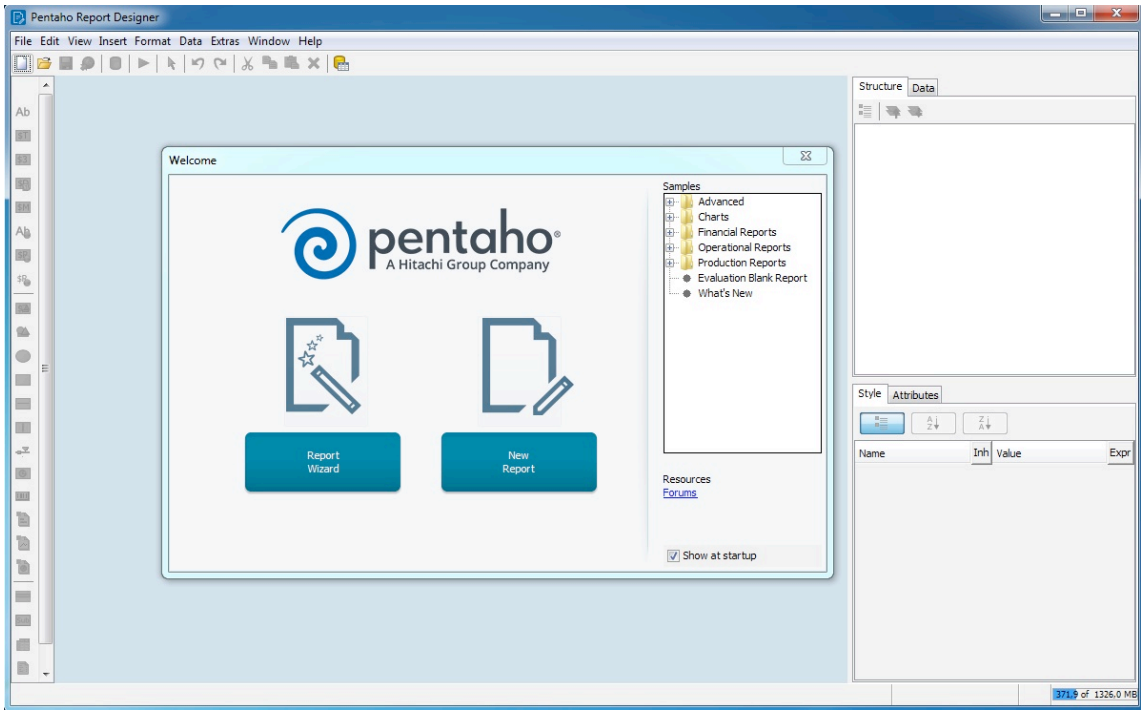


Figura 12 Entorn de disseny de *Pentaho Report Designer 7.0*

2.3 *Data Integration* (*pdi-ce-7.0.0.0-25*) (*Extraction Transformation Load*)

Seguint les mateixes indicacions anteriors, en aquest cas, després de l'opció *Download section*, es selecciona la secció *Data Integration* i es descarrega un arxiu comprimit de nom "pdi-ce-7.0.0.0-25.zip". A continuació, es descomprimeix i la carpeta "pdi-ce-7.0.0.0-25" s'ubica al directori arrel "c:\":

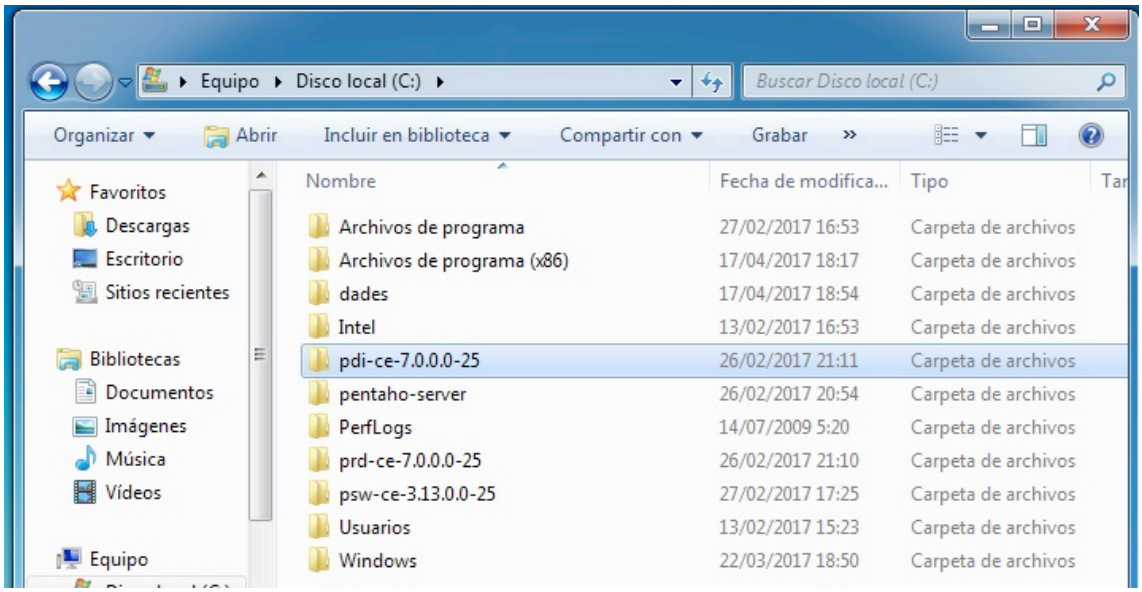


Figura 13 Ubicació dels arxius corresponent a la utilitat *Data Integration*

A l'interior de la carpeta "data-integration" s'executa l'arxiu *Spoon.bat* i apareix la finestra inicial corresponent a la versió de la utilitat que indica que s'està carregant el programari.

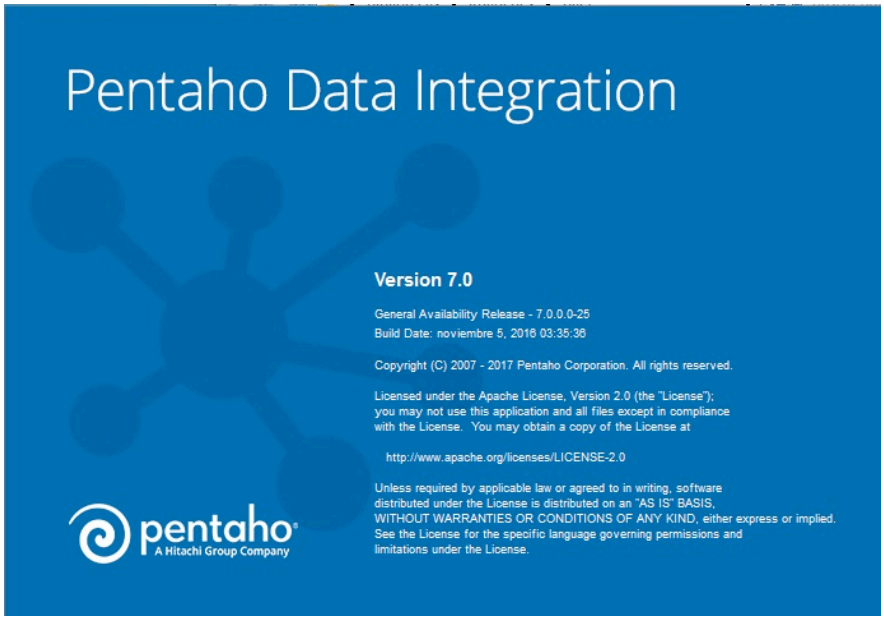


Figura 14 Pantalla inicial *Pentaho Data Integration 7.0*

Posteriorment, passats uns minuts, apareix la pantalla principal corresponent a l'IDE.

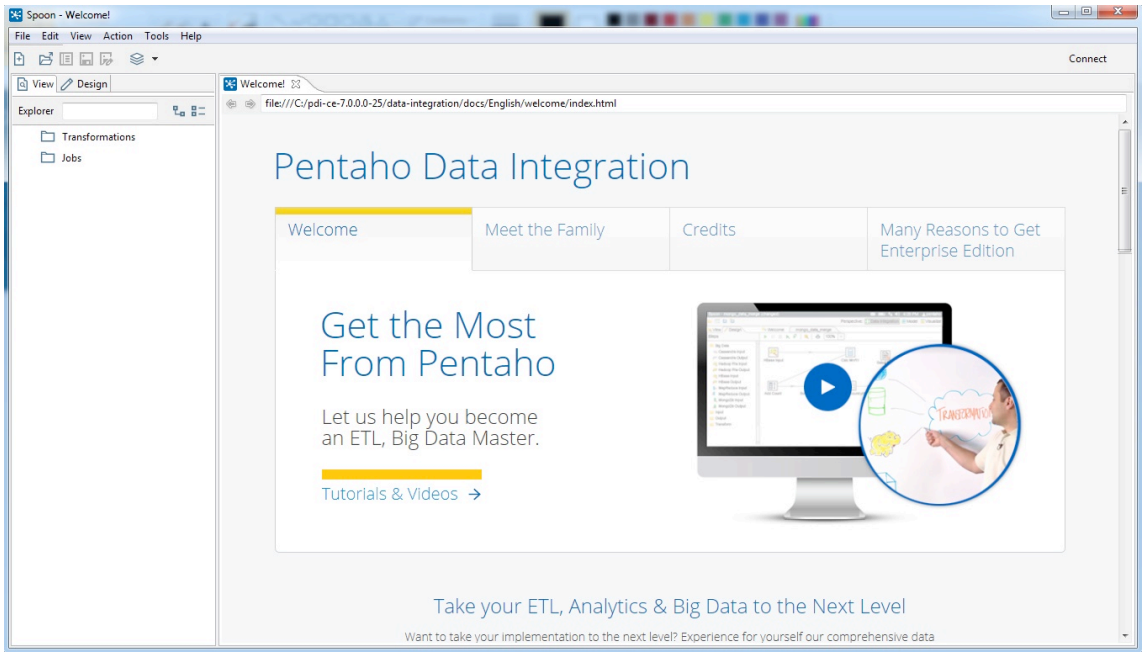


Figura 15 Entorn de disseny de *Pentaho Data Integration 7.0*

2.4 Motor *OLAP* (*psw-ce-7.0.0.0-25*) *Schema Workbench: Mondrian*

En darrer lloc, en aquest cat, es selecciona *Schema Workbench* i es descarrega un arxiu comprimit de nom “psw-ce-3.13.0.0-25.zip”. A continuació, es descomprimeix i la carpeta “psw-ce-3.13.0.0-25” s’ubica al directori arrel “c:\”:

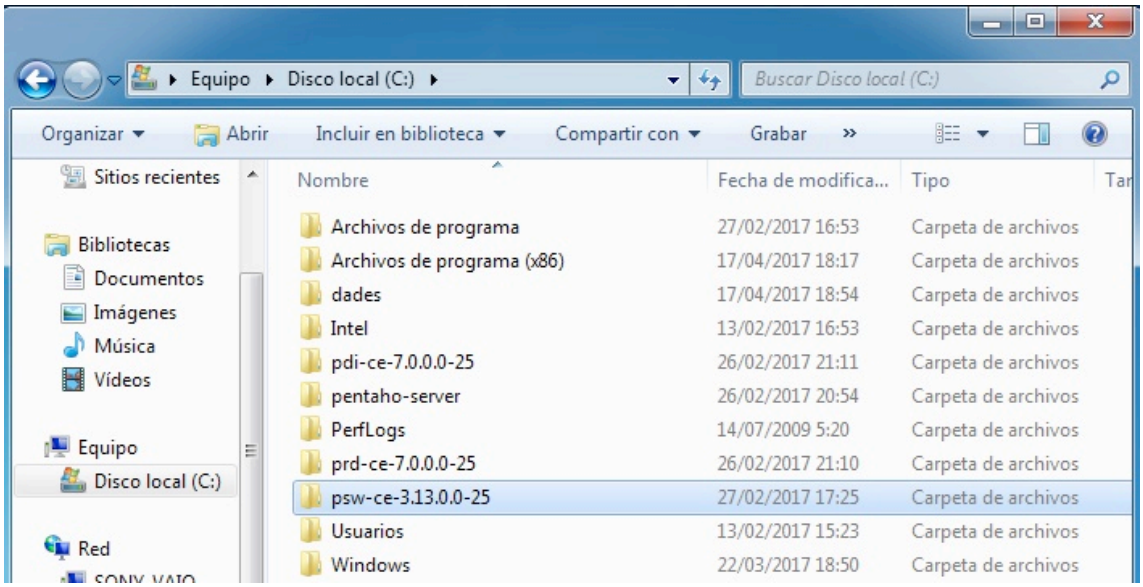


Figura 16 Ubicació dels arxius de la utilitat *Schema Workbench(Mondrian)*

Finalment, a l'interior de la carpeta "schema-workbench" s'executa l'arxiu *Workbench.bat* i apareix l'IDE de la utilitat.

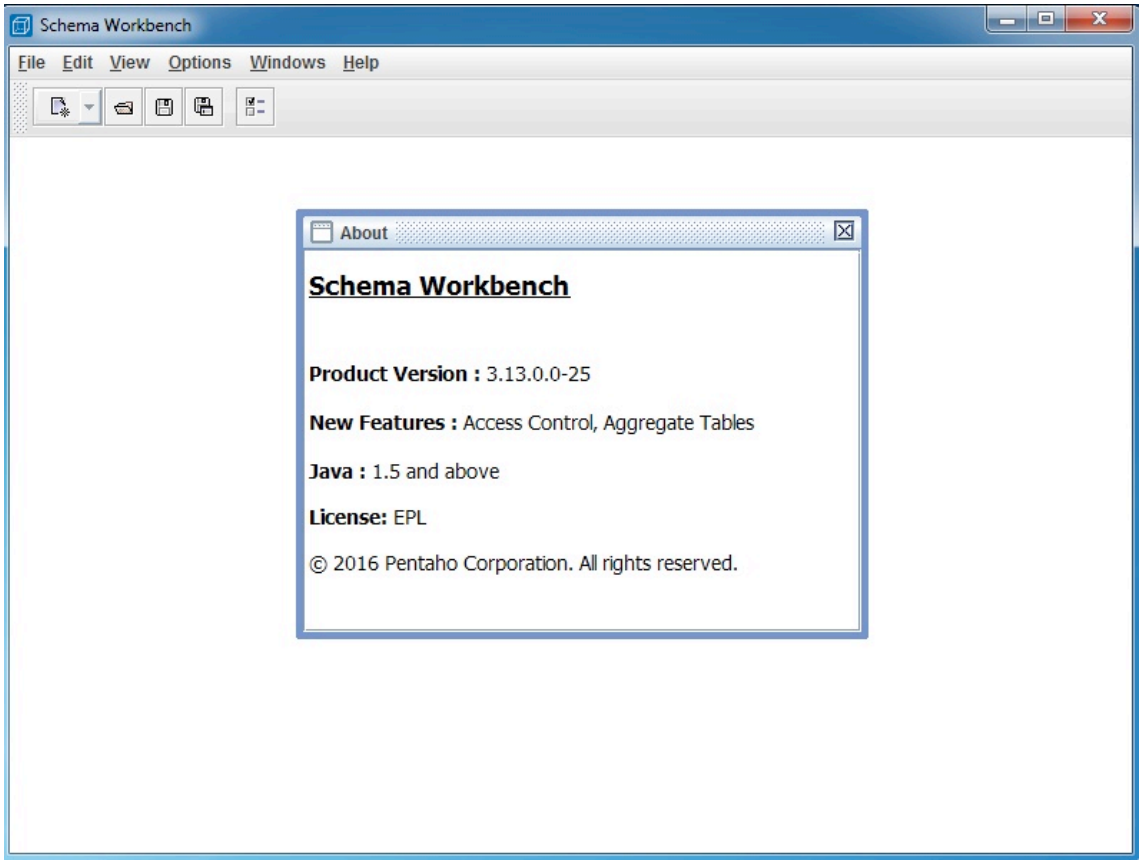


Figura 17 Entorn de disseny de *Schema Workbench*

3. Disseny del *data warehouse*

D'altra banda, el *data warehouse* o magatzem de dades s'implementarà a través del SGBD o sistema de gestió de base de dades *MySQL Workbench 6.3*[3].

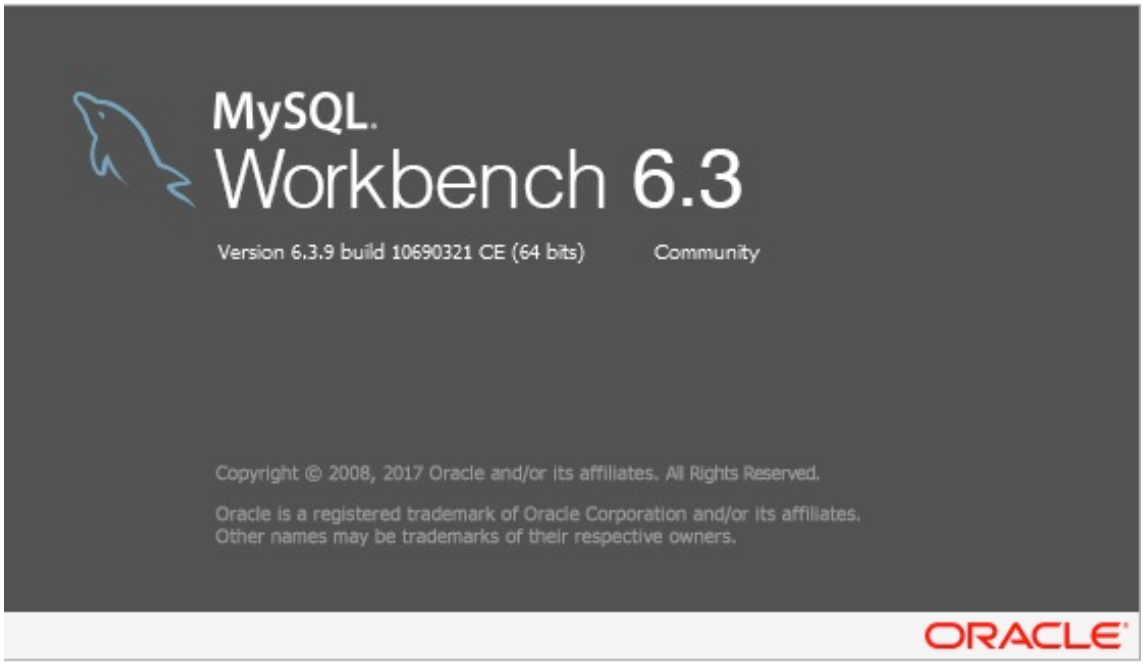


Figura 18 *Data warehouse* implementat en *MySQL*

MySQL Workbench proporciona als administradors de base de dades i desenvolupadors un entorn d'eines integrat per a:

- Disseny i modelat de bases de dades.
- Desenvolupament de llenguatge SQL.
- Administració de la base de dades.
- Migració de la base de dades.

Així doncs, es disposarà del repositori de dades on a traves de les diferents eines de la suite de *Pentaho* no tan sols s'emmagatzemaran les dades de diferents orígens de dades sinó que serà el proveïdor de dades del quals s'extraurà coneixement.

[3] MySQL$_{TM}$ (2017). *Download MySQL Workbench* [en línia] https://dev.mysql.com/downloads/workbench/ [data de consulta: 16 de març de 2017]

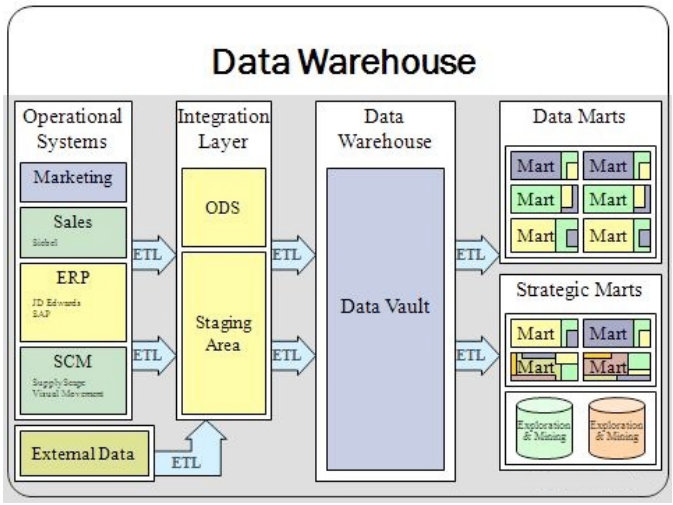


Figura 19 Representació gràfica d'un *data warehouse*[4]

Una vegada instal·lada el SGBD de *MySQL Workbench 6.3*, s'executa l'IDE d'administració i es mostra la següent pantalla:

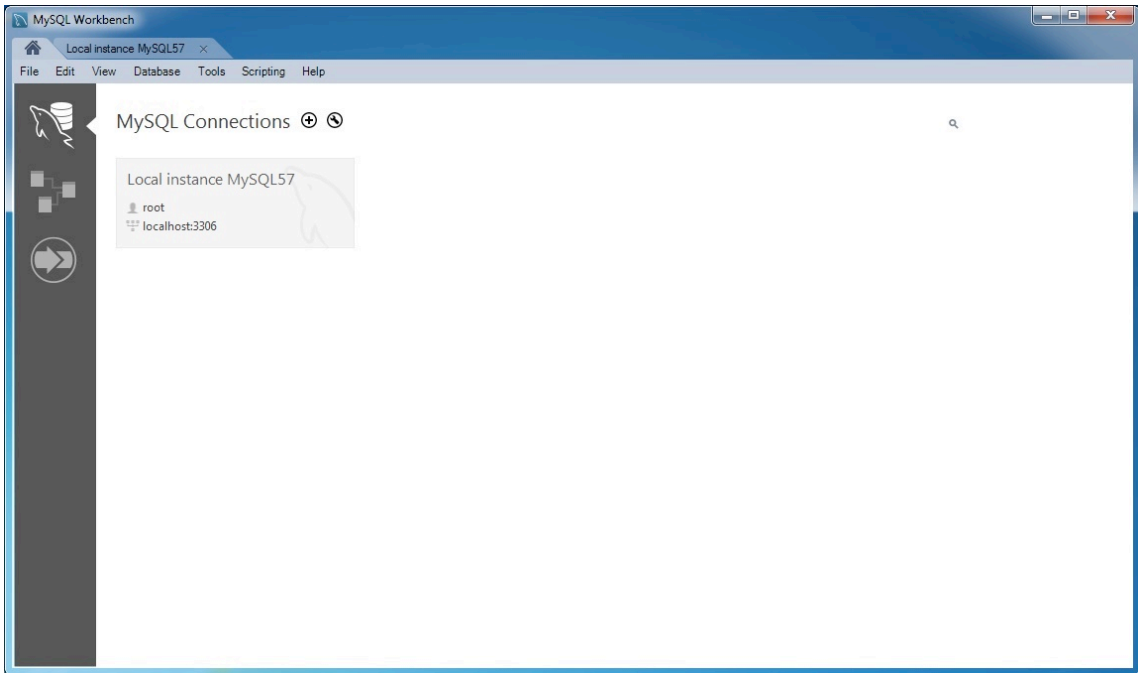


Figura 20 IDE de la base de dades *MySQL Workbench 6.3*

S'observa com s'ha creat una instància de nom *MySQL57* amb l'esquema per defecte que és *sys*. En aquest sentit, es crea un nou esquema de nom *pentaho* que és on s'albergaran totes les taules que farà ús la suite de *Pentaho*.

[4] WIKIPEDIA La enciclopèdia libre (2017). *Almacén de datos* [en línia] https://es.wikipedia.org/wiki/Almacén_de_datos [data de consulta: 17 de març de 2017]

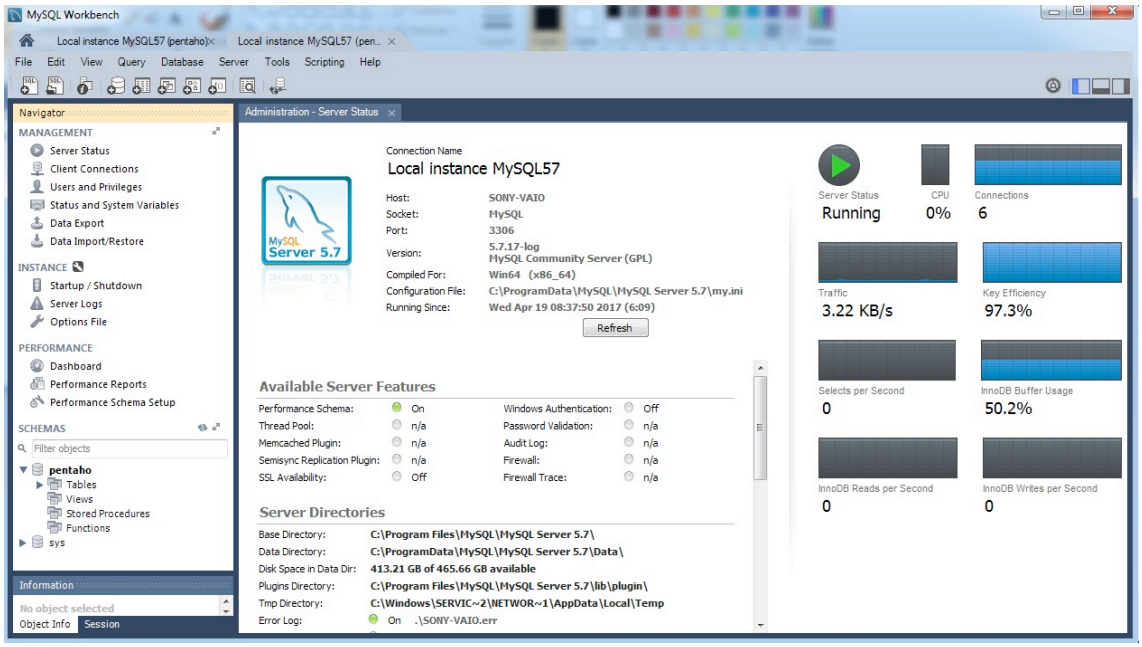


Figura 21 IDE de la base de dades *MySQL Workbench 6.3*

4. Disseny dels processos ETL (*Entry - Transformation - Load*)

Com ja es va apuntar a la justificació del treball, l'escenari inicial es el d'una organització que disposa d'un gran volum de dades a través d'un sistema no integrat donat que moltes d'aquestes dades es recopilen per mitjans diferents i en repositoris de dades diferents (*ORACLE, SLQ SERVER, MySQL, Excel, Microsoft Access, S390*, etc.). Per extreure informació d'aquestes dades es fa un disseny individual i un tractament arbitrari de les dades fent ús d'eines ofimàtiques com l'*Excel* que, amb la funcionalitat de les taules dinàmiques, permet la creació de diverses gràfiques, si bé el seu manteniment és costós i no està exempt de males praxis quan al seu maneig i interpretació.

Tota vegada que ja es disposa del software necessari per extreure coneixement de les dades, es defineix l'objectiu pràctic d'aquest treball. Es tracta d'analitzar un servei d'atenció al ciutadà concret com és una oficina de denúncies d'una administració pública local on es vol analitzar el següent:

- Dies amb més demandes del ciutadà.
- Hores amb major afluència del ciutadà.
- Tipus de delictes més denunciats.

Per a la obtenció d'aquesta informació, es disposa d'un arxiu d'*Excel* exportat d'un sistema de base de dades *Oracle* on hi ha la informació corresponent a cada una de les denúncies del ciutadà.

Per a tal fi, s'haurà d'emprar l'eina *Data Integration* per tractar la informació de l'arxiu d'*Excel* i emmagatzemar-la al *data warehouse* o magatzem de dades, en aquest cas, *MySQL Workbench 6.3*.

D'aquesta manera, s'inicia l'aplicació *Data Integration* i es selecciona crear una nova transformació:

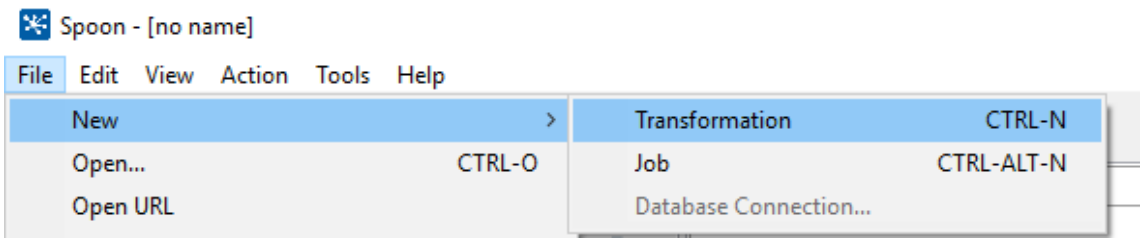


Figura 22 Creació d'una transformació

Pentaho permet definir tasques d'operacions seqüencials així com transformacions per transformar les dades. Per provar la connexió a la de base de dades es dissenya una transformació molt simple.

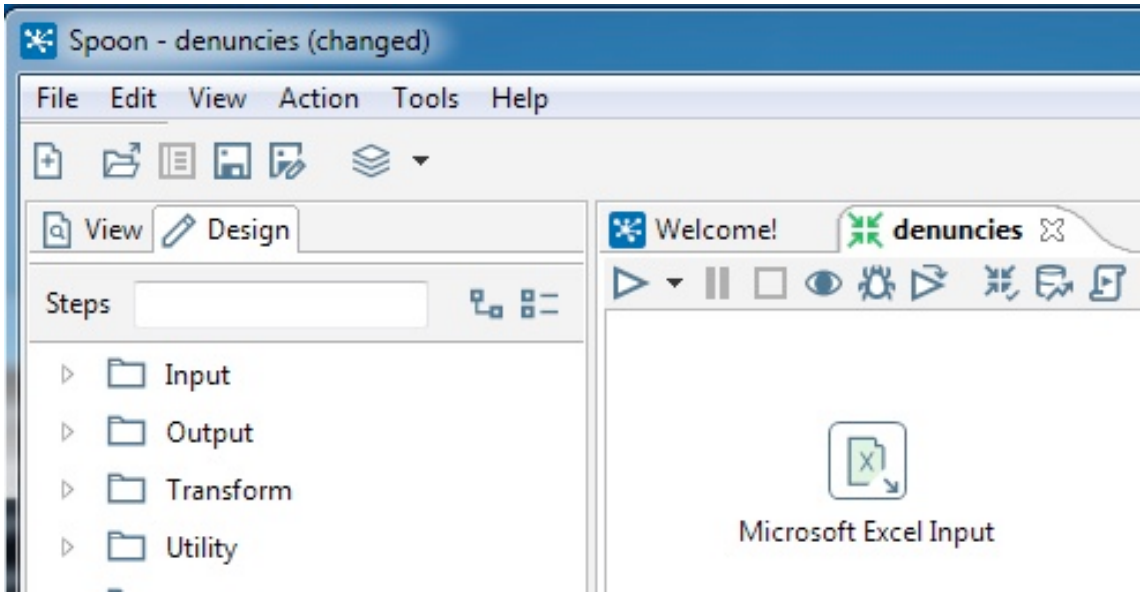


Figura 23 Operació d'entrada de dades

Microsoft Excel Input permet recopilar dades d'un arxiu d'*Excel*.

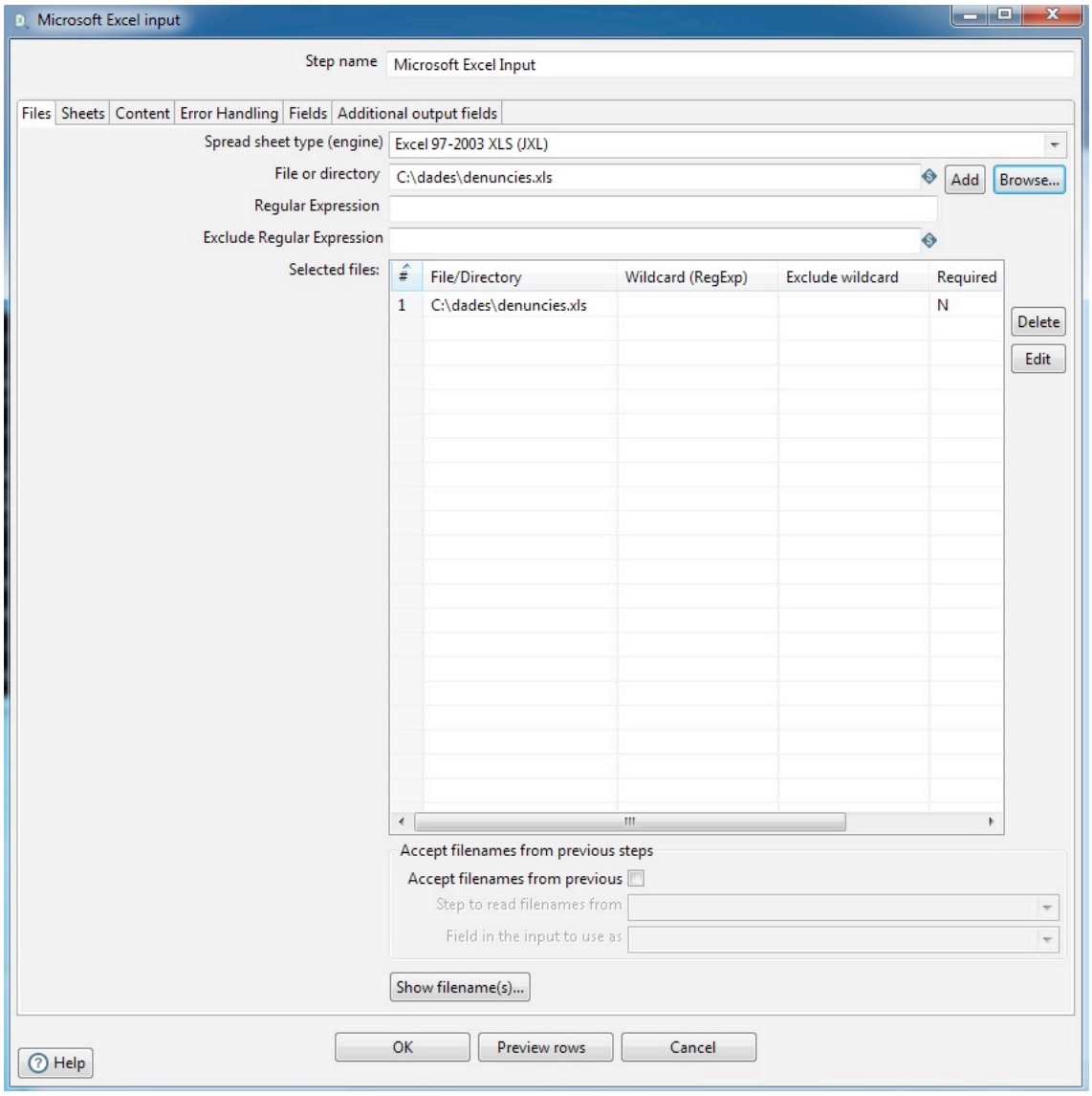


Figura 24 Selecció de l'arxiu d'*Excel*

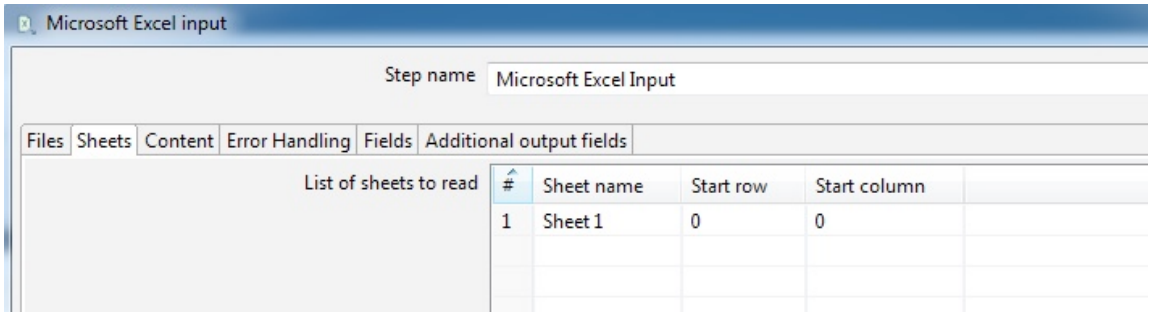
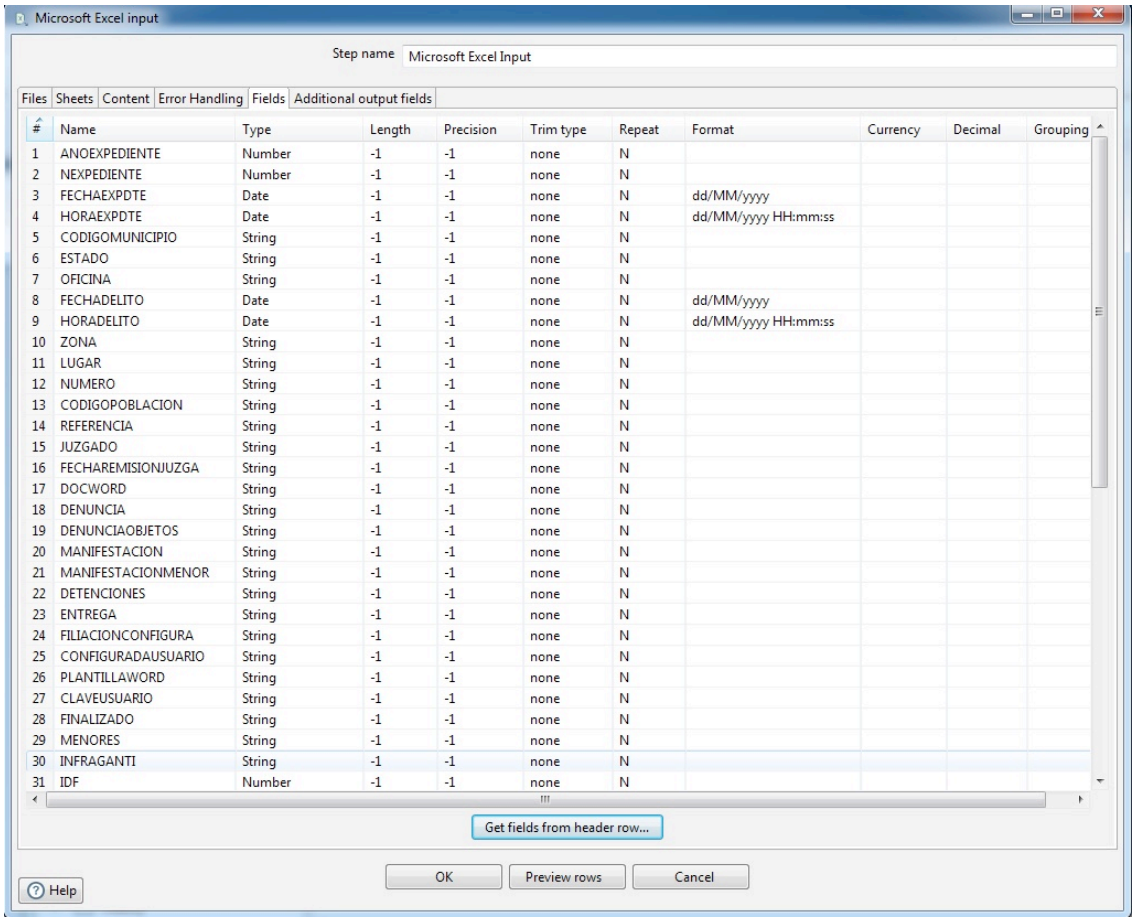


Figura 25 Selecció de la fulla de l'arxiu d'*Excel*

A la pestanya *Fields* es relacionen les diferents columnes que detecta *Pentaho Data Integration* i la possibilitat d'efectuar certes modificacions com són definir el tipus de dades, longitud, format, etc.



#	Name	Type	Length	Precision	Trim type	Repeat	Format	Currency	Decimal	Grouping
1	ANOEXPEDIENTE	Number	-1	-1	none	N				
2	NEXPEDIENTE	Number	-1	-1	none	N				
3	FECHAEXPDTE	Date	-1	-1	none	N	dd/MM/yyyy			
4	HORAEXPDTE	Date	-1	-1	none	N	dd/MM/yyyy HH:mm:ss			
5	CODIGOMUNICIPIO	String	-1	-1	none	N				
6	ESTADO	String	-1	-1	none	N				
7	OFICINA	String	-1	-1	none	N				
8	FECHADELITO	Date	-1	-1	none	N	dd/MM/yyyy			
9	HORADELITO	Date	-1	-1	none	N	dd/MM/yyyy HH:mm:ss			
10	ZONA	String	-1	-1	none	N				
11	LUGAR	String	-1	-1	none	N				
12	NUMERO	String	-1	-1	none	N				
13	CODIGOPOBLACION	String	-1	-1	none	N				
14	REFERENCIA	String	-1	-1	none	N				
15	JUZGADO	String	-1	-1	none	N				
16	FECHAREMISIONJUZGA	String	-1	-1	none	N				
17	DOCWORD	String	-1	-1	none	N				
18	DENUNCIA	String	-1	-1	none	N				
19	DENUNCIAOBJETOS	String	-1	-1	none	N				
20	MANIFESTACION	String	-1	-1	none	N				
21	MANIFESTACIONMENOR	String	-1	-1	none	N				
22	DETENCIONES	String	-1	-1	none	N				
23	ENTREGA	String	-1	-1	none	N				
24	FILIACIONCONFIGURA	String	-1	-1	none	N				
25	CONFIGURADAUSUARIO	String	-1	-1	none	N				
26	PLANTILLAWORD	String	-1	-1	none	N				
27	CLAVEUSUARIO	String	-1	-1	none	N				
28	FINALIZADO	String	-1	-1	none	N				
29	MENORES	String	-1	-1	none	N				
30	INFRAGANTI	String	-1	-1	none	N				
31	IDF	Number	-1	-1	none	N				

Figura 26 Columnes detectades al full de l'arxiu d'*Excel*

A continuació, de la part esquerre de l'IDE es desplega l'apartat *Output* per a acte seguit seleccionar l'objecte *Insert/Update* i col·locar-lo després de l'objecte *Microsoft Excel Input*.

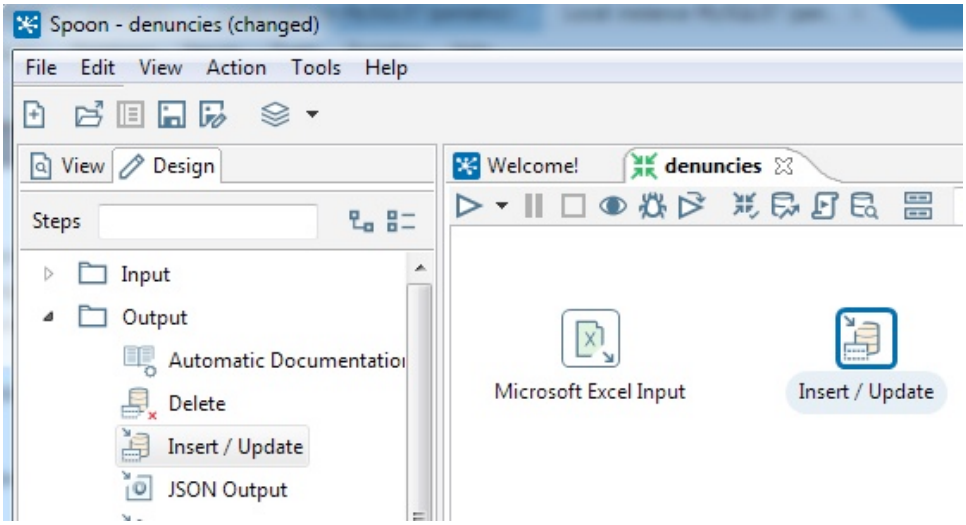


Figura 27 Operació de sortida de dades

A l'objecte *Insert/Update* es defineixen els paràmetres per a emmagatzemar les dades de l'objecte d'entrada a la base de dades.

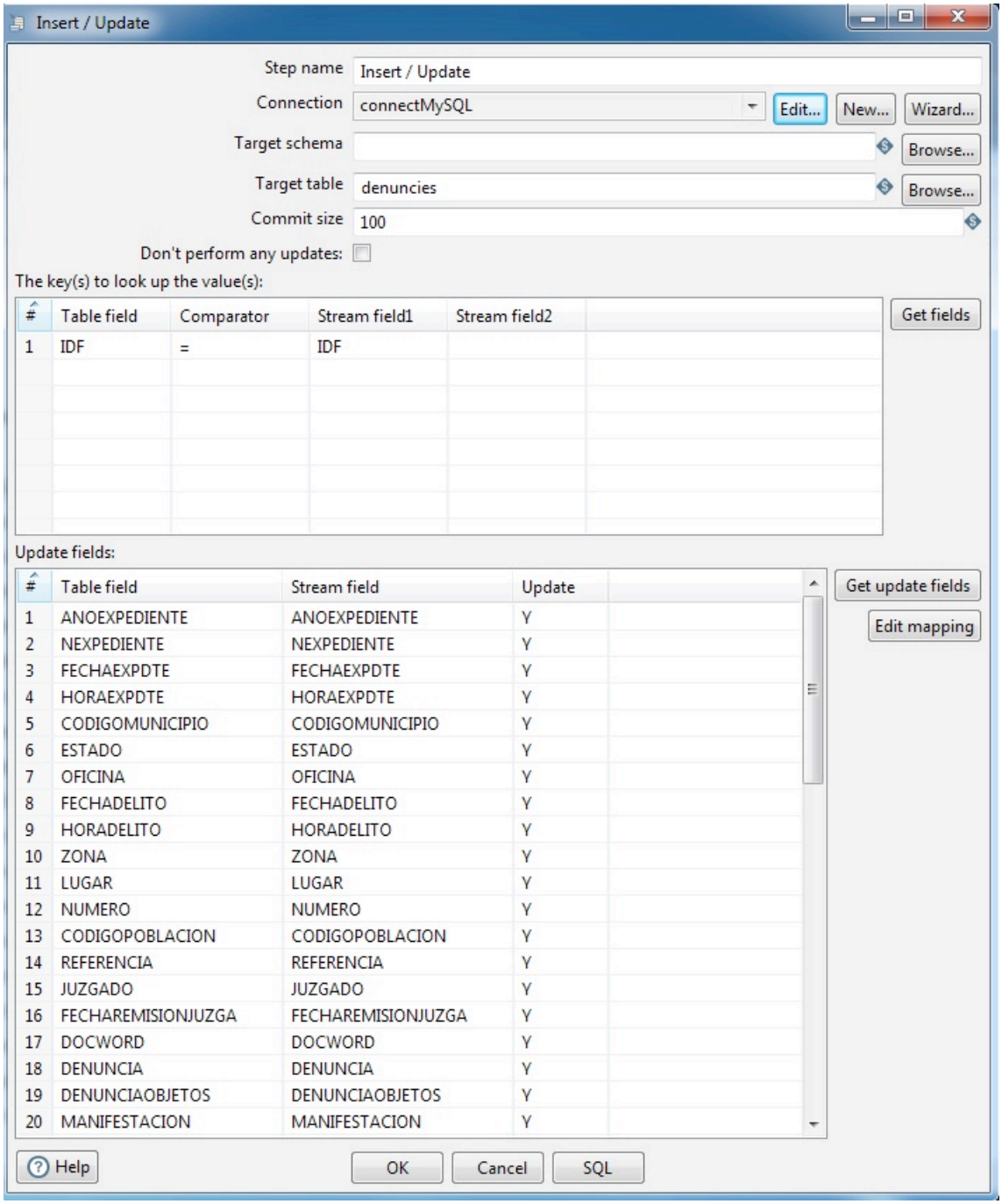


Figura 28 *Insert / Update*

Per això es defineix la connexió a la base de dades a través de l'opció *Connection*.

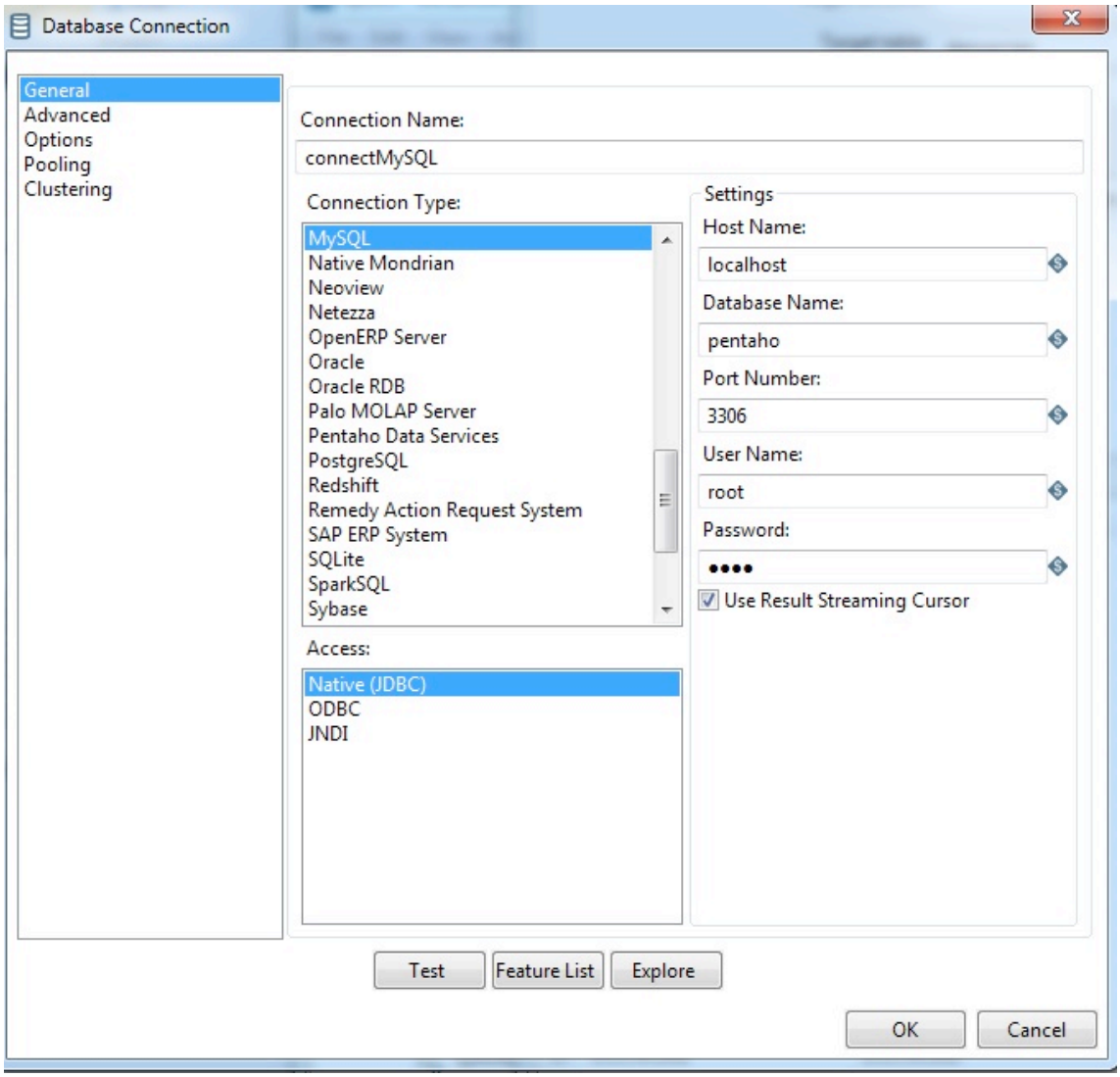


Figura 29 Connexió a *MySQL*

A continuació es defineix el paràmetre *Target Table* corresponent al nom de la taula on s'ubicaran les dades, en aquest cas "denuncies". És important en cas que sigui la primera vegada que s'executa l'operació pitjar l'opció SQL per crear la taula a l'esquema *pentaho*.

Simple SQL editor

SQL statements, separated by semicolon ';'

```
CREATE TABLE pentaho.denuncies
(
  IDF DOUBLE
, ANOEXPEDIENTE DOUBLE
, NEXPEDIENTE DOUBLE
, FECHAEXPDTE DATETIME
, HORAEXPDTE DATETIME
, CODIGOMUNICIPIO TINYTEXT
, ESTADO TINYTEXT
, OFICINA TINYTEXT
, FECHADELITO DATETIME
, HORADELITO DATETIME
, ZONA TINYTEXT
, LUGAR TINYTEXT
, NUMERO TINYTEXT
, CODIGOPOBLACION TINYTEXT
, REFERENCIA TINYTEXT
, JUZGADO TINYTEXT
, FECHAREMISIONJUZGA TINYTEXT
, DOCWORD TINYTEXT
, DENUNCIA TINYTEXT
, DENUNCIAOBJETOS TINYTEXT
, MANIFESTACION TINYTEXT
```

Line 1 column 0

Execute | Clear cache | Close

Figura 30 Pantalla de creació de la taula denuncies

Es pitja el botó *Execute* i es mostra el resultat.

Results of the SQL statements

The SQL statements had the following results

```
SQL executed: CREATE TABLE pentaho.denuncies
(
  IDF DOUBLE
, ANOEXPEDIENTE DOUBLE
, NEXPEDIENTE DOUBLE
, FECHAEXPDTE DATETIME
, HORAEXPDTE DATETIME
, CODIGOMUNICIPIO TINYTEXT
, ESTADO TINYTEXT
, OFICINA TINYTEXT
, FECHADELITO DATETIME
, HORADELITO DATETIME
, ZONA TINYTEXT
, LUGAR TINYTEXT
, NUMERO TINYTEXT
, CODIGOPOBLACION TINYTEXT
, REFERENCIA TINYTEXT
, JUZGADO TINYTEXT
, FECHAREMISIONJUZGA TINYTEXT
, DOCWORD TINYTEXT
, DENUNCIA TINYTEXT
, DENUNCIAOBJETOS TINYTEXT
, MANIFESTACION TINYTEXT
, MANIFESTACIONMENOR TINYTEXT
, DETENCIONES TINYTEXT
, ENTREGA TINYTEXT
, FILIACIONCONFIGURA TINYTEXT
, CONFIGURADAUSUARIO TINYTEXT
, PLANTILLAWORD TINYTEXT
, CLAVEUSUARIO TINYTEXT
, FINALIZADO TINYTEXT
, MENORES TINYTEXT
, INFRAGANTI TINYTEXT
, ROBODEVEHICULO TINYTEXT
, ROBOENVEHICULO TINYTEXT
, DANOVEHICULO TINYTEXT
, NEXPEDIENTEREGENT DOUBLE
, ANOEXPEDIENTEREGEN TINYTEXT
, SANCIONESANIMALES TINYTEXT
, NOMBREESTABLECIMIE TINYTEXT
, DIASEMANA TINYTEXT
, LEVANTAMIENTOCAD TINYTEXT
, VIOLENCIADOMESTICA TINYTEXT
, DETENCIONMENOR TINYTEXT
, ALCOHOLEMIA TINYTEXT
, SCLUGAR TINYTEXT
, INDICATIVO TINYTEXT
, GRUPO TINYTEXT
, INSTRUCTOR TINYTEXT
, SECRETARIO TINYTEXT
, EXTENSION TINYTEXT
, TIPODELITO TINYTEXT
, ROBOENINMUEBLE TINYTEXT
, IMPUTADOS TINYTEXT
, NUM_CEA TINYTEXT
```

OK | Cancel

Figura 31 Resultat d'execució

S'ha creat la taula "denuncies" a l'esquema *pentaho* tal com es mostra a la següent il·lustració.

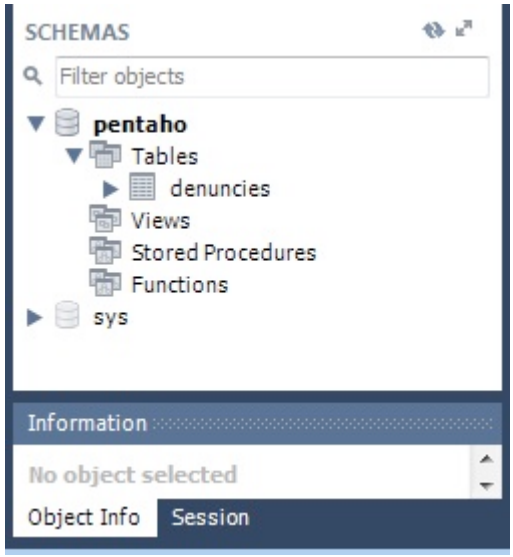


Figura 32 Taula creada

A continuació s'uneixen els dos objectes del dissenyador de *Pentaho Data Integration* per a que les operacions es duguin a terme de forma seqüencial.

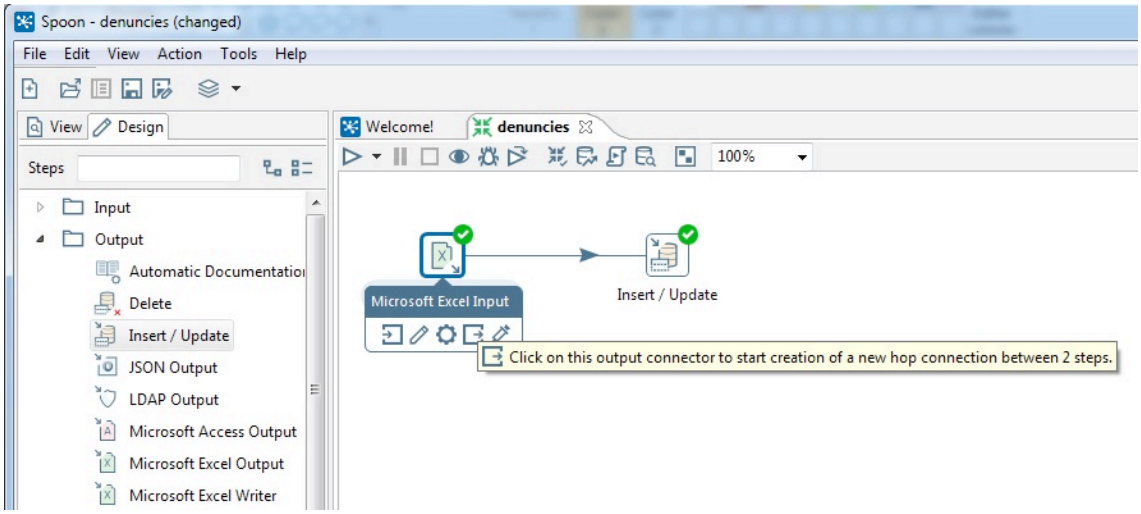


Figura 33 Transformació de nom "*denuncies*" creada

Finalment s'executa el procés de transformació que en aquest traslladarà totes les dades de l'arxiu d'*Excel* al repositori de dades de MySQL. Aquesta operació es durà a terme a través de l'opció *Run* corresponent a la primera icona de la barra d'eines de la transformació que es correspon amb un triangle dirigit cap a la dreta. El resultat d'execució és el que es detalla a continuació:

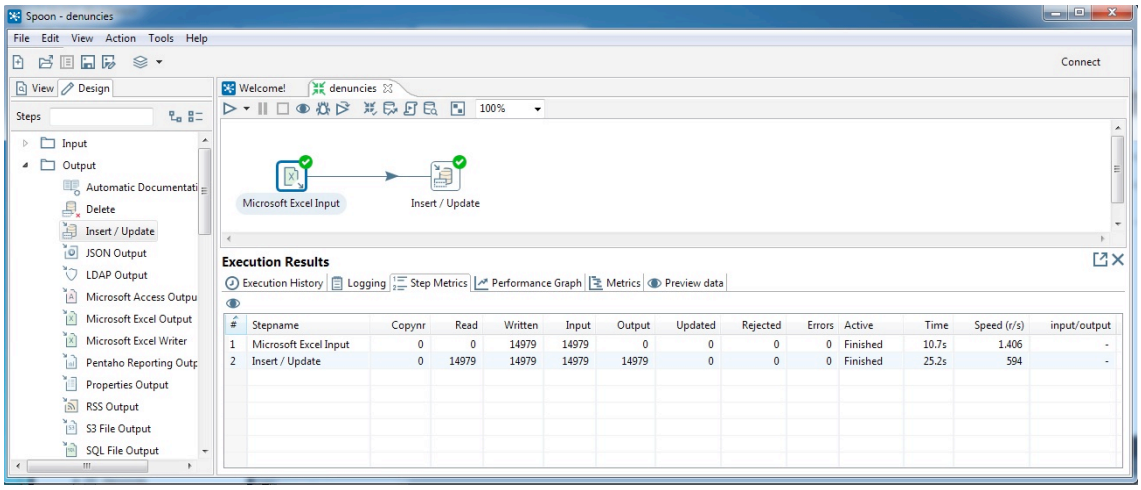


Figura 34 Resultat de l'execució del procés de transformació

A l'IDE d'administració de la base de dades es pot comprovar com la taula "denuncies" de l'esquema *pentaho* té les dades que es trobaven a l'arxiu d'Excel.

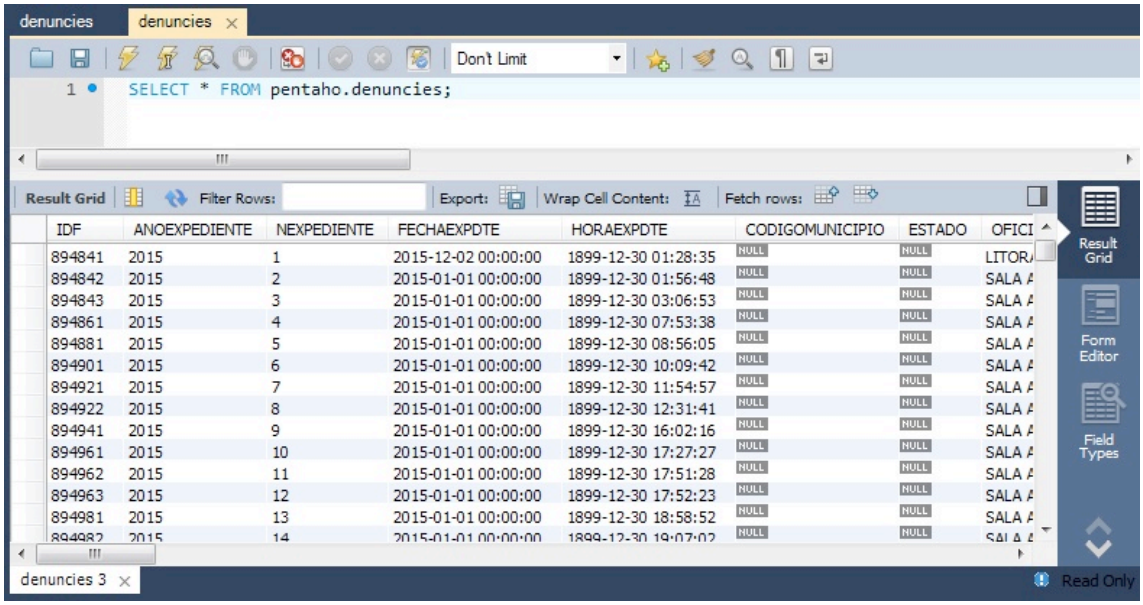


Figura 35 Execució del procés de transformació a la base de dades

5. Disseny d'anàlisi *OLAP* (*On-Line Analytical Processing*)

Com ja es va apuntar anteriorment, l'eina per a la creació de cubs OLAP en Pentaho és *Schema Workbench*. En primer lloc s'ha de definir la connexió amb la base de dades que es fa servir com a *Data Warehouse*, es a dir, en aquest cas My*SQL Workbench 6.3*. Des de l'opció del menú "Options" es configuren les dades de la connexió.

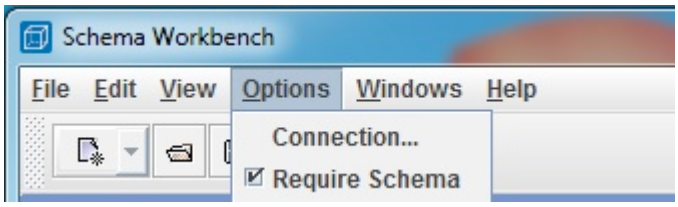


Figura 36 Opció per definir la connexió

A continuació el sistema mostra la pantalla per poder establir els paràmetres de connexió i es comprova la connexió a través de l'opció "Test".

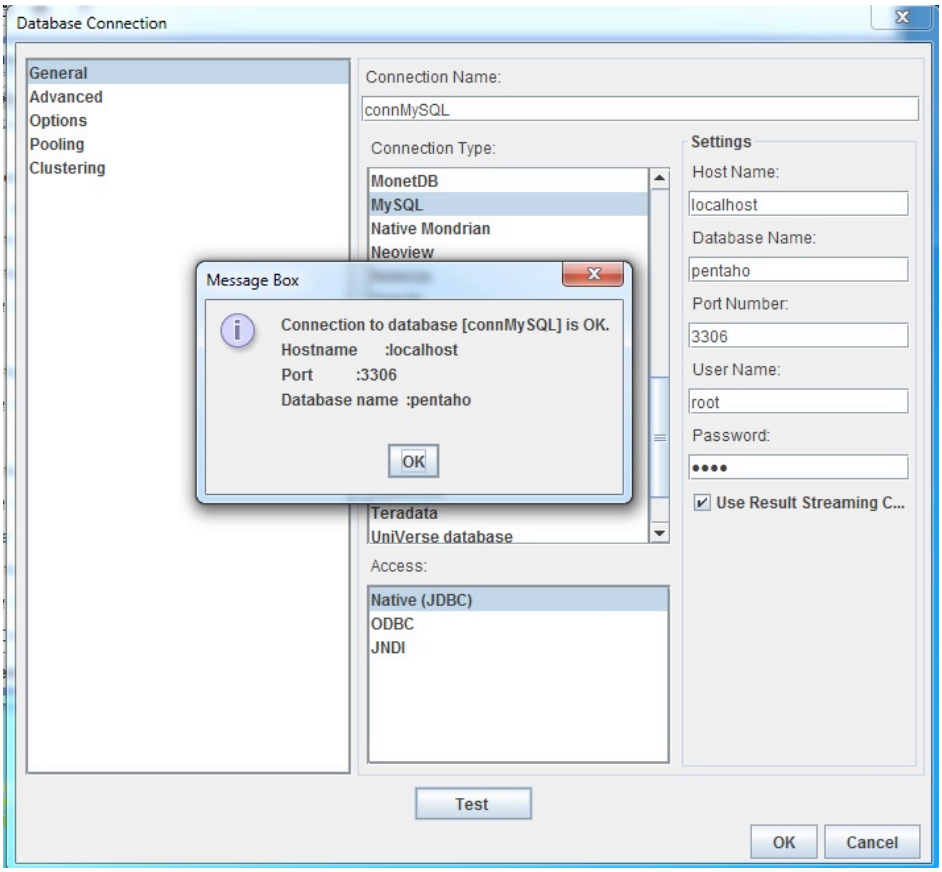


Figura 37 Paràmetres de la connexió a *MySQL* i comprovació

A partir d'aquí es comença amb la creació de l'esquema (*Schema*) que serà el contenidor dels cubs necessaris.

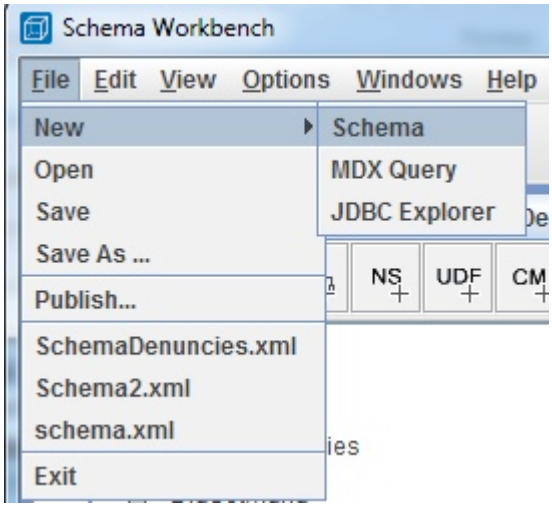


Figura 38 Creació nou esquema (*schema*)

L'esquema permet definir una sèrie de propietats com són el nom, la descripció, un nom per a la dimensió que agruparà a les mesures i un rol por defecte per a utilitzar en les connexions de bases de dades. Al crear el cub, s'indica un nom i una descripció del mateix i és possible marcar l'opció "cache", per a que el motor *Mondrian* treballi amb memòria cache amb aquest cub. Destacar també l'opció "enabled", que s'encarrega de fer visible el cub, en cas contrari, si no es marca, el cub no apareixerà. En aquest cas s'ha definit un esquema de nom "Schema" i un cub de nom "Denuncies".

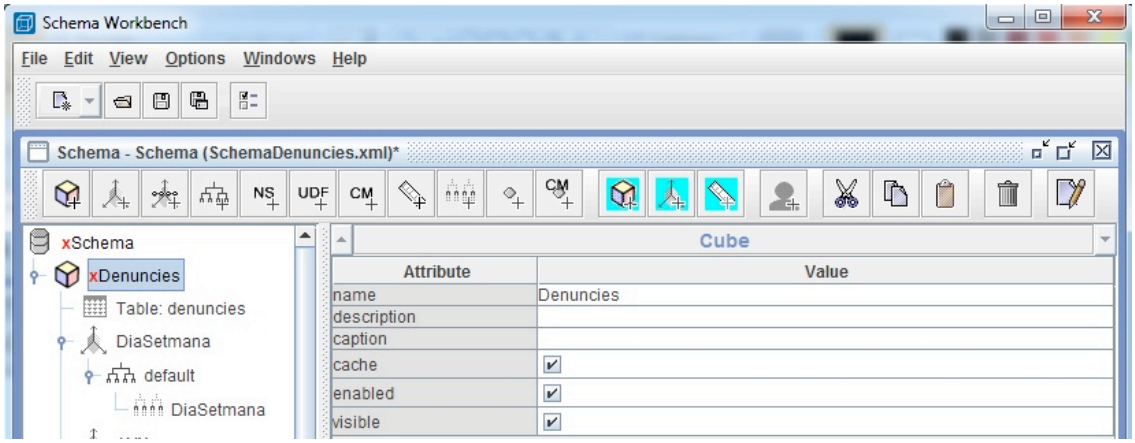


Figura 39 Creació del cub "Denuncies"

Al cub definit anteriorment afegim la taula de fets "denuncies" a partir de la qual es calculen tant les mesures com els indicadors. Aquests darrers s'obtindran

d'afegir les dimensions necessàries per a obtenir la informació que s'ha proposat com a objectiu. En aquest cas les dimensions necessàries són el dia de la setmana, l'hora i l'any que s'extreu de la data corresponent a la interposició de la denúncia i d'altra banda, el tipus de delicte que correspon als fets denunciats. La mesura serà el nombre de denuncies que s'ha interposat en base a les anteriors dimensions.

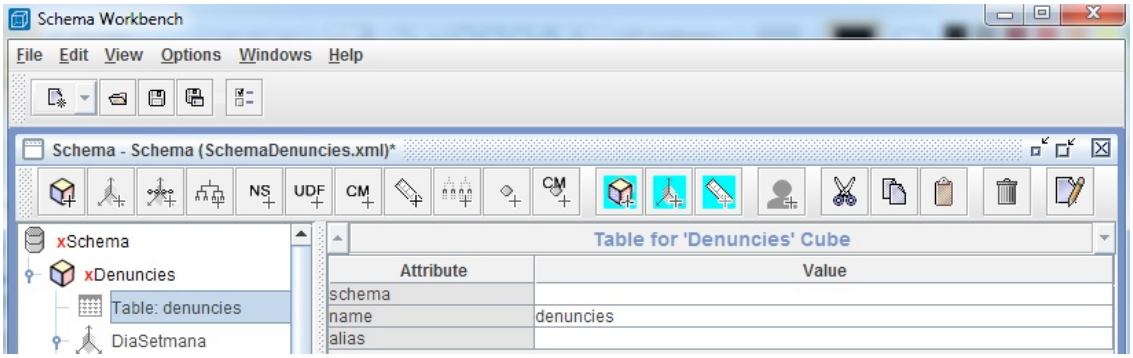


Figura 40 Taula de fets "denuncies"

La propietat "name" de la taula de fets es correspon amb el nom de la taula que existeix a la base de dades *MySQL* on s'han emmagatzemat les dades en el procés ETL(*ENTRY - TRANSFORMATION - LOAD*) amb l'eina *Pentaho Data Integration*. Amb l'opció "Add Dimension Usage" s'afegeixen totes les dimensiones necessàries. El cub hereta totes les característiques que contengui la dimensió, abastant totes les jerarquies i els seus corresponents atributs. S'identifica amb un nom significatiu la dimensió i s'indica el tipus. En el cas de la dimensió "DiaSetmana", es defineixen les propietats "name", foreignKey que es correspon amb l'atribut de la taula que conté el dia de la setmana i el tipus.

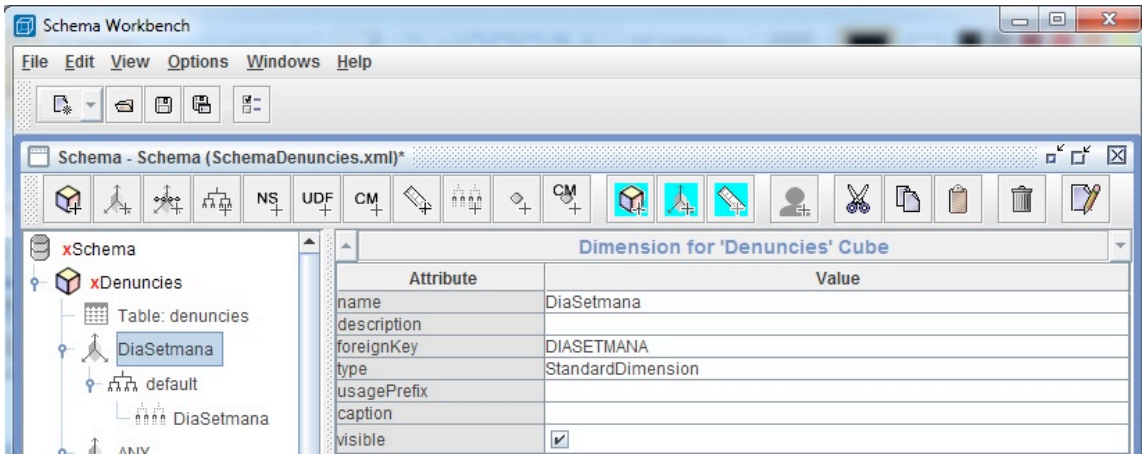


Figura 41 Dimensió “DiaSetmana”

A continuació, es creen les diferents jerarquies que contengui la dimensió i, en aquest cas, com que només es fa ús d’una taula, es a dir, no hi ha un model relacional de taules, es fa ús de la jerarquia per defecte definint l’atribut “allMemberName”.

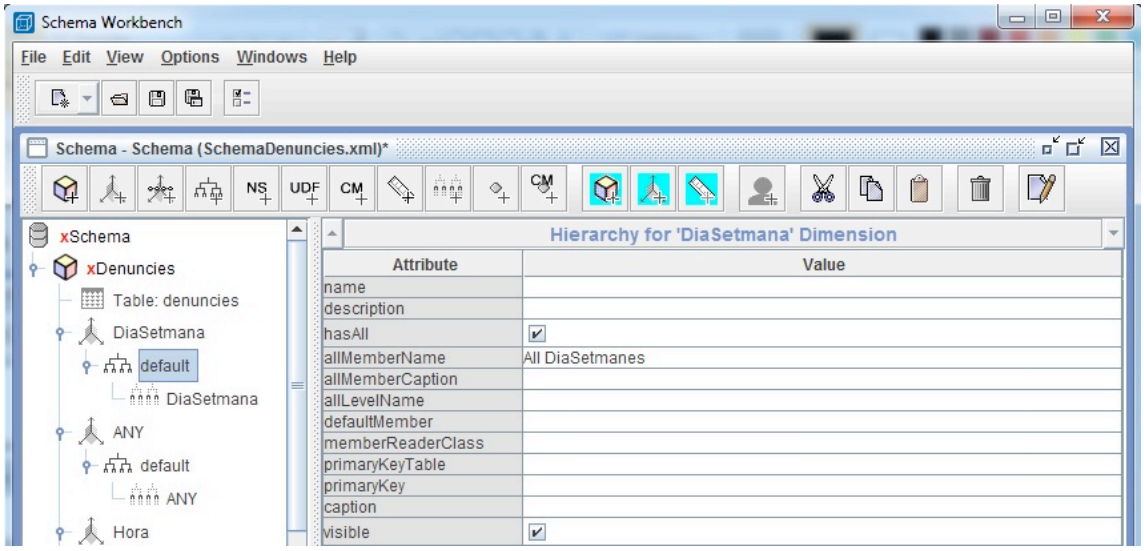


Figura 42 Jerarquia per defecte

Per acabar de definir la dimensió “DiaSetmana”, s’afegeix un nivell “level” o més segons convingui i s’anomena com a “DiaSetmana” definint les propietats “name”, “table” corresponent al nom de la taula que conté la informació, “column” corresponent a l’atribut de la taula anterior que conté la informació del dia de la setmana, “type” segons el tipus de dada, “internalType”, “levelType” i “hideMemeberif”.

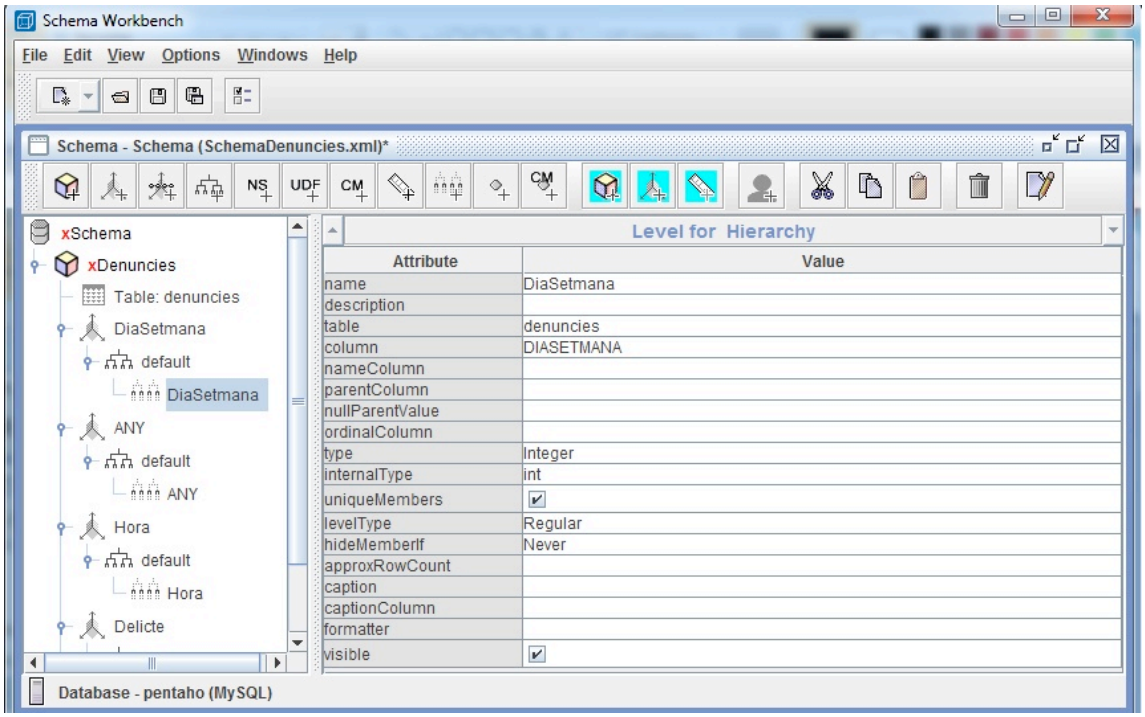


Figura 43 Nivell “DiaSetmana”

Quan s’han creades les dimensions considerades oportunes per a dur terme l’anàlisi, i como a darrer pas del disseny del cub, es defineixen les mètriques seran els valors que mostraran de l’anàlisi de la informació. En el aquest cas concret, es defineix una mètrica per a veure l’evolució de les demandes ciutadanes en relació als dies de la setmana així com a les diferents hores dels dies i aquesta evolució es fa en base a les dades corresponents als anys 2015 i 2016.

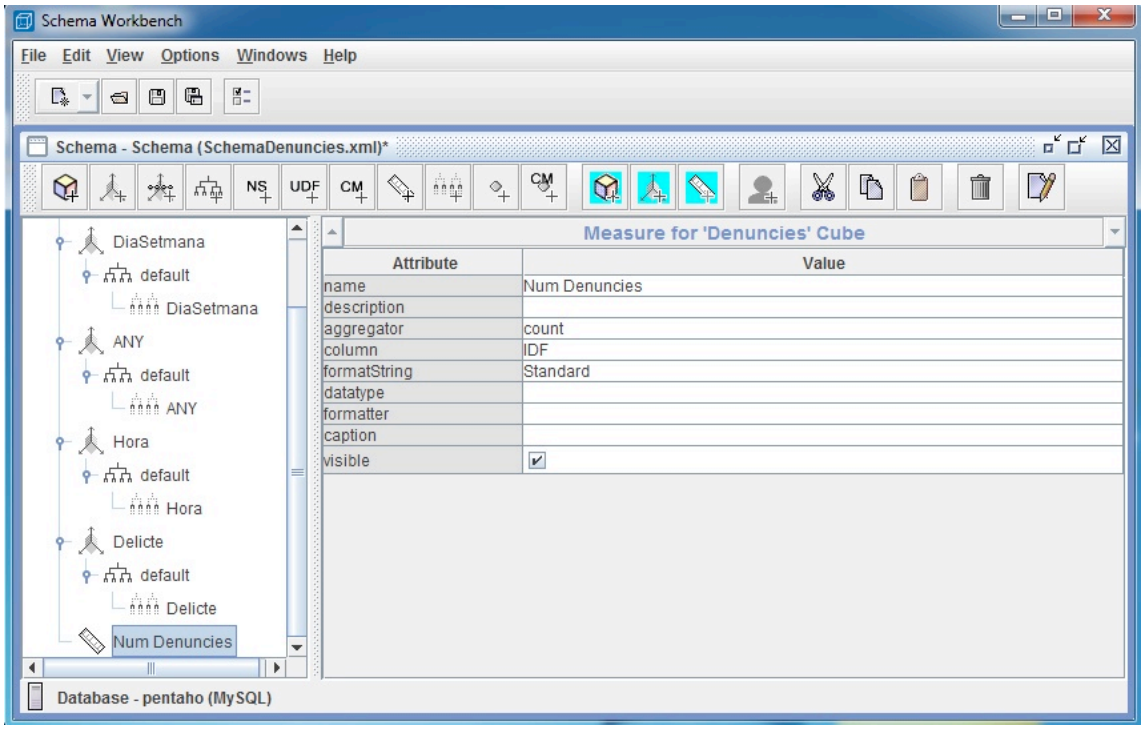


Figura 44 Mètrica “Num Denuncies”

En aquests cas els atributs que es defineixen en la mètrica són “name”, “aggregator” que ens permet diferents operacions (contar, sumar, mínim, màxim, mitjana, etc.), “column” corresponent a l’atribut de la taula que conté la dada a mesurar i “formatString”.

Una vegada definit el cub, s’ha de publicar a Pentaho Server a través de l’opció “Publish” del menú principal.

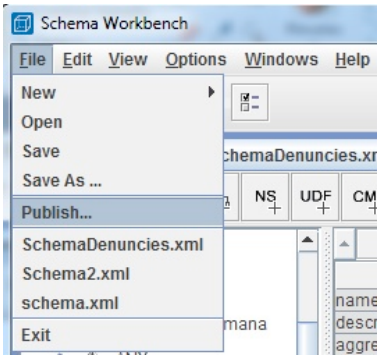


Figura 45 Publicació del cub

S’introdueixen les dades de connexió a Pentaho i, si no hi ha errors, es publicarà el cub.

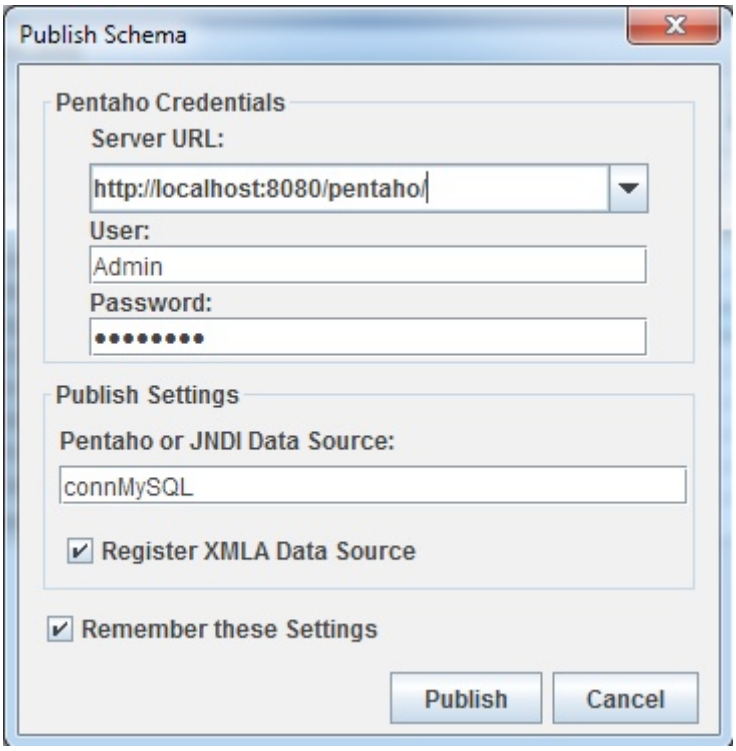


Figura 46 Dades de connexió per publicar el cub

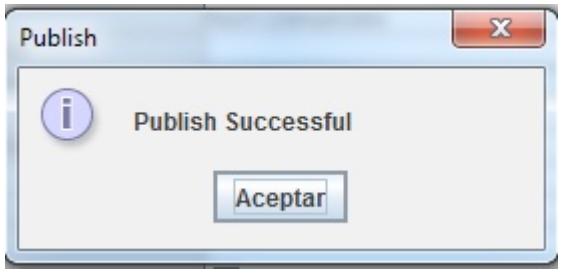


Figura 47 Publicació amb èxit

6. Visualització del cub *OLAP* (*On-Line Analytical Processing*)

Per verificar si el cub s'ha dissenyat de forma correcta, Pentaho disposa de diferents eines entre les que es troba *Sayku Analitycs* que si bé no es troba instal·lada per defecte, és possible afegir-la a través de l'opció *Pentaho Marketplace*. Es tracta d'un únic lloc que permet l'accés, gestió i actualització de tots els *plugins* del servidor.

S'ha optat per aquest plugin perquè permet efectuar consultes d'una forma molt més intuïtiva i gràfica que l'eina per defecte *JPivot View.*

Per a accedir a *Saiku Analitycs* en el servidor *Pentaho*, es pitja sobre el botó amb el nom "Saiku Analitycs" que està inclòs dins de l'opció de menú "Create New".

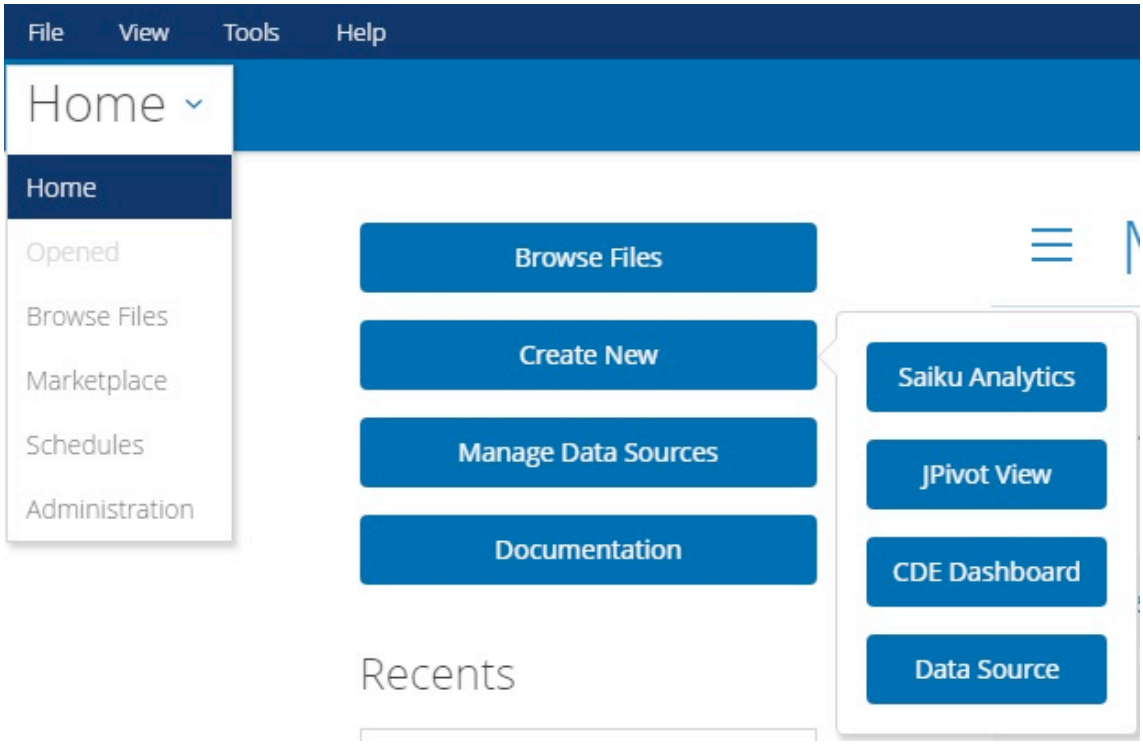


Figura 48 Accés a *Saiku Analytics*

El sistema mostra la pantalla inicial amb diferents opcions, si bé per comprovar si el disseny del cub *OLAP* és correcte, es seleccionarà l'opció "Create a new query".

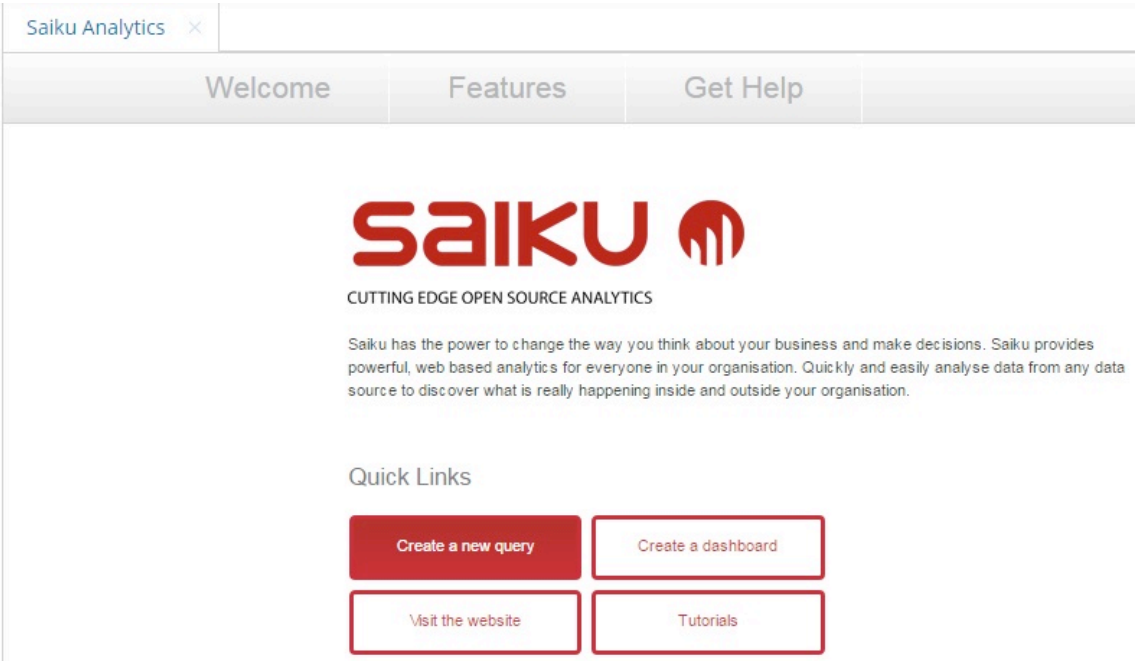


Figura 49 Pantalla inicial de *Saiku Analytics*

A continuació el sistema mostra els esquemes que s'ha publicat a *Pentaho Server* d'entre els quals es troba el que s'ha dissenyat per a l'apartat *DISSENY D'ANÀLISI OLAP*.

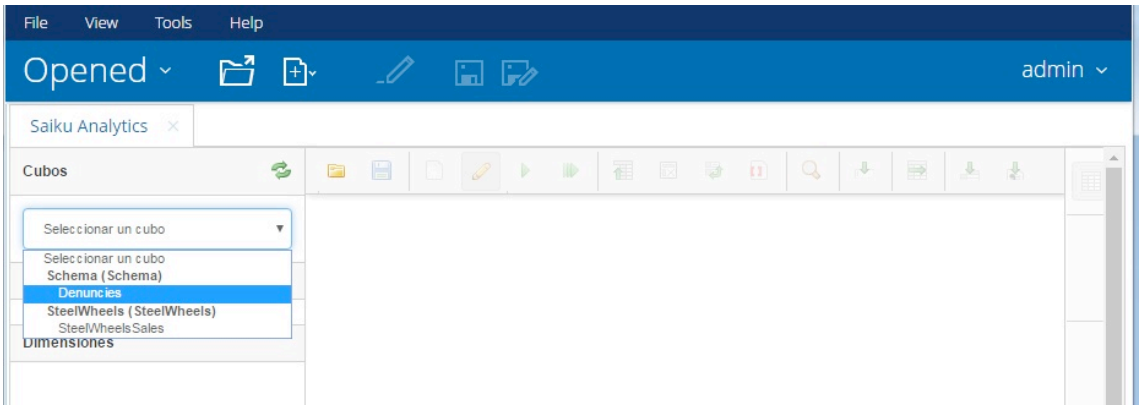


Figura 50 Pantalla inicial de *Saiku Analytics*

Es selecciona l'esquema i cub que es correspon amb les dades que desitjant visualitzar i acte seguit es mostra que el procés d'elaboració del cub s'ha dut a terme correctament.

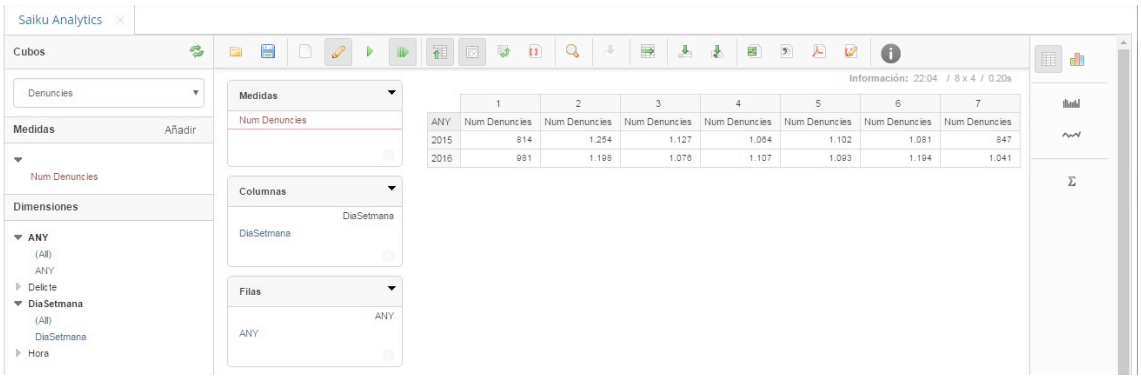


Figura 51 Número de denuncies per dia de la setmana (2015-2016)

Amb l'opció d'emmagatzemar del menú de *Saiku Analytics*, es guarda la consulta que proporciona aquesta informació que posteriorment s'emprarà per dissenyar el quadre de comandament o *Dashboard*.

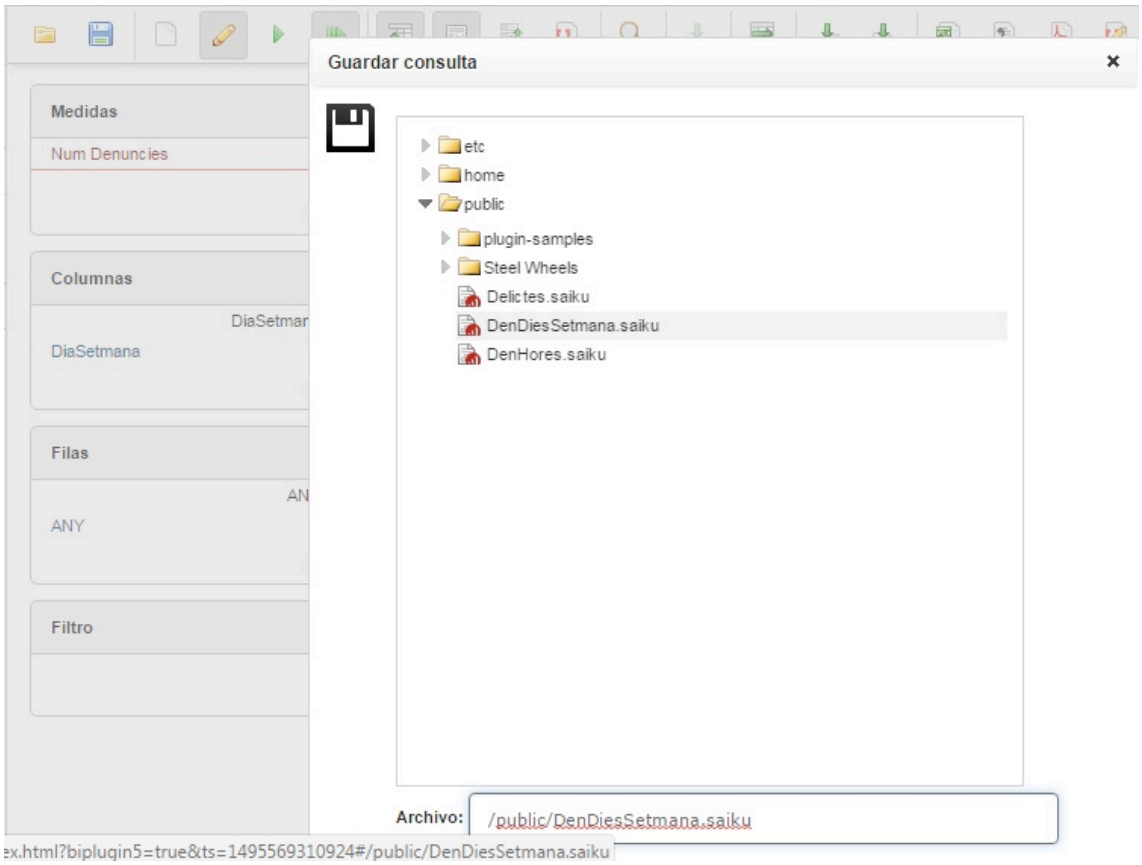


Figura 52 Emmagatzemat de la consulta.

A partir dels camps seleccionats que formen el cub *OLAP*, existeix la possibilitat de representar gràficament la informació en taules. Per exemple en el cas del nombre de denúncies per trams horaris, la visualització de la informació a través d'un gràfic de barres serà més fàcil d'entendre que no pas en taula numèrica.

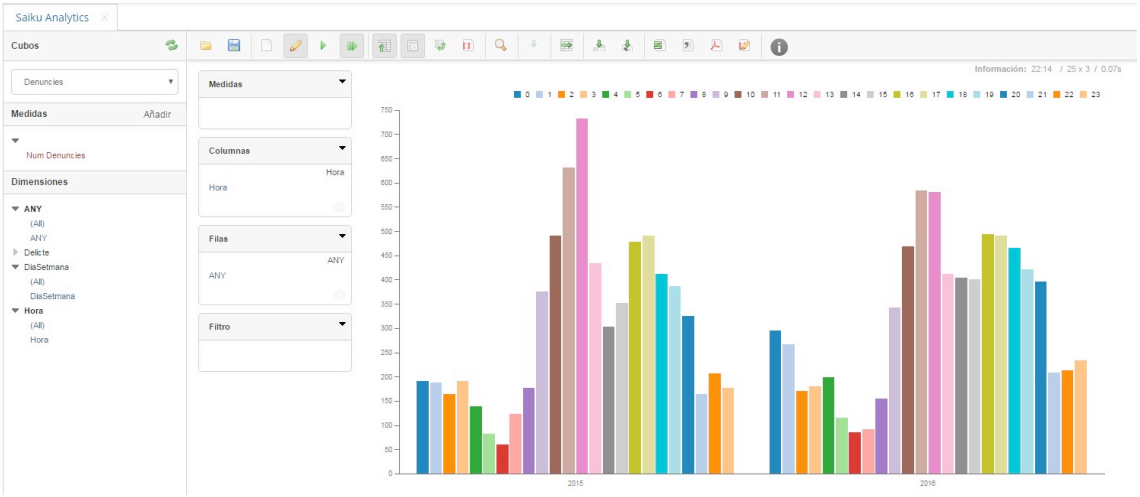
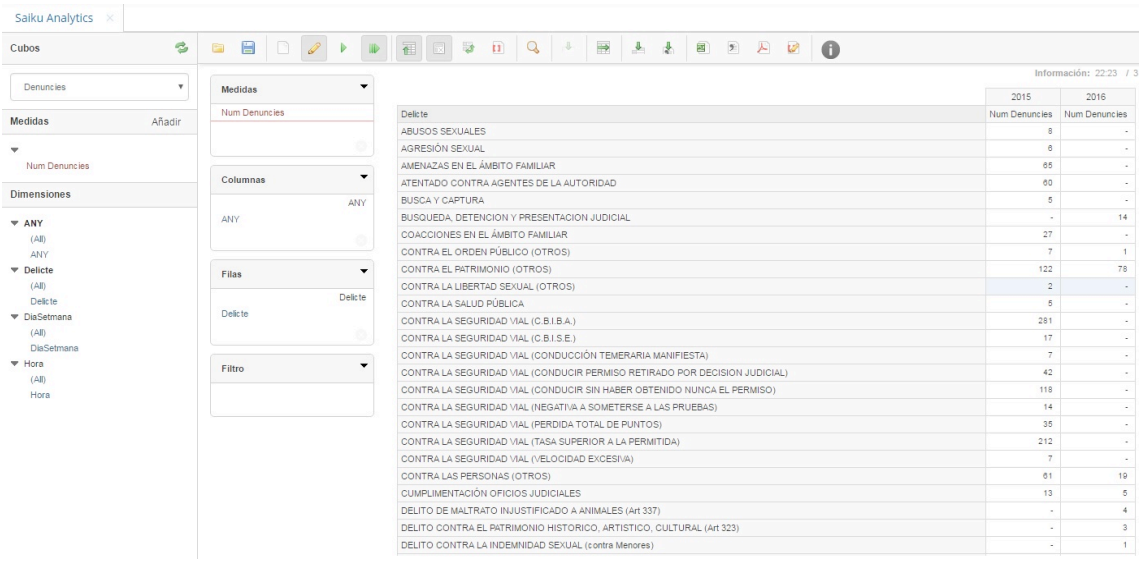


Figura 53 Número de denuncies per hores del dia dels anys 2015 i 2016



Delicte	2015 Num Denuncies	2016 Num Denuncies
ABUSOS SEXUALES	8	-
AGRESIÓN SEXUAL	6	-
AMENAZAS EN EL ÁMBITO FAMILIAR	65	-
ATENTADO CONTRA AGENTES DE LA AUTORIDAD	60	-
BUSCA Y CAPTURA	5	-
BUSQUEDA, DETENCION Y PRESENTACION JUDICIAL	-	14
COACCIONES EN EL ÁMBITO FAMILIAR	27	-
CONTRA EL ORDEN PÚBLICO (OTROS)	7	1
CONTRA EL PATRIMONIO (OTROS)	122	78
CONTRA LA LIBERTAD SEXUAL (OTROS)	2	-
CONTRA LA SALUD PÚBLICA	5	-
CONTRA LA SEGURIDAD VIAL (C.B.I.B.A.)	281	-
CONTRA LA SEGURIDAD VIAL (C.B.I.S.E.)	17	-
CONTRA LA SEGURIDAD VIAL (CONDUCCIÓN TEMERARIA MANIFIESTA)	7	-
CONTRA LA SEGURIDAD VIAL (CONDUCIR PERMISO RETIRADO POR DECISION JUDICIAL)	42	-
CONTRA LA SEGURIDAD VIAL (CONDUCIR SIN HABER OBTENIDO NUNCA EL PERMISO)	118	-
CONTRA LA SEGURIDAD VIAL (NEGATIVA A SOMETERSE A LAS PRUEBAS)	14	-
CONTRA LA SEGURIDAD VIAL (PERDIDA TOTAL DE PUNTOS)	35	-
CONTRA LA SEGURIDAD VIAL (TASA SUPERIOR A LA PERMITIDA)	212	-
CONTRA LA SEGURIDAD VIAL (VELOCIDAD EXCESIVA)	7	-
CONTRA LAS PERSONAS (OTROS)	61	19
CUMPLIMENTACIÓN OFICIOS JUDICIALES	13	5
DELITO DE MALTRATO INJUSTIFICADO A ANIMALES (Art 337)	-	4
DELITO CONTRA EL PATRIMONIO HISTORICO, ARTISTICO, CULTURAL (Art 323)	-	3
DELITO CONTRA LA INDEMNIDAD SEXUAL (contra Menores)	-	1

Figura 54 Número de denuncies per delicte dels anys 2015 i 2016

7. Disseny del quadre de comandament (*CDE Dashboard*)

En aquest apartat es passa a la fase de creació d'eines d'intel·ligència de negoci que proporcionen a realment a la direcció d'una organització informació valuosa per a prendre decisions. Aquesta informació és representada a través dels anomenats quadres de comandaments o *Dashboard*. Mitjançant la seva creació, es pretén presentar les dades d'una manera intuïtiva i fàcilment comprensible.

D'aquesta forma , *Pentaho Community Edition 7.0* disposa de l'opció "CDE Dashboard" per al disseny dels quadres de comandaments.

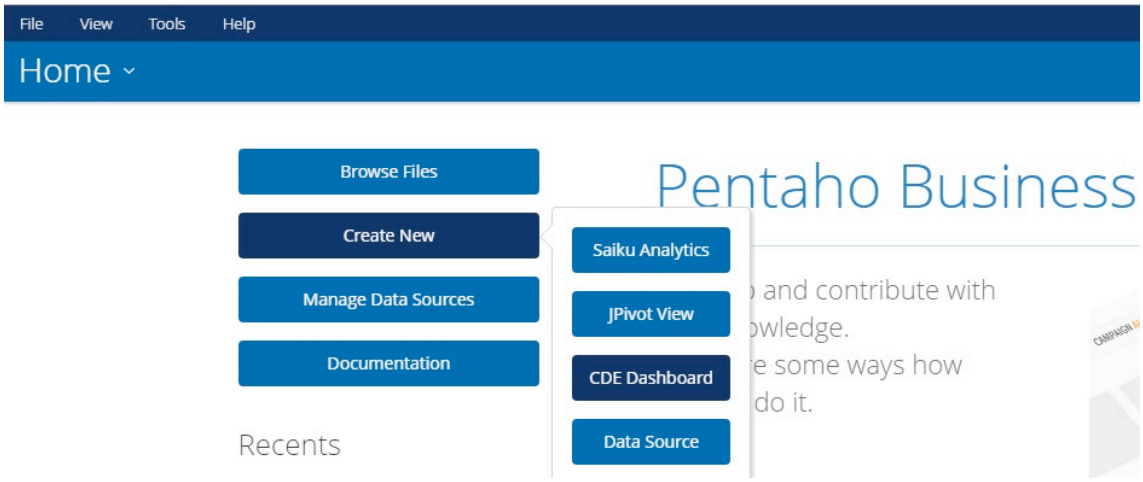


Figura 55 Opció de creació d'un *Dashboard*

L'entorn de la interfície inclou tres tipus de panells, els quals han de ser degudament combinats per obtenir el resultat final. En aquest sentit, i amb la intenció d'oferir una descripció de cada panell que composa la interfície de treball del *CDE Dashboard*, es detalla cada un d'ells:

- *Layout Panel*: Es tracta del panell dedicat al disseny. En aquest panell, es poden realitzar diferents modificacions sobre l'aspecte final amb el que es presentarà el *Dashboard*.

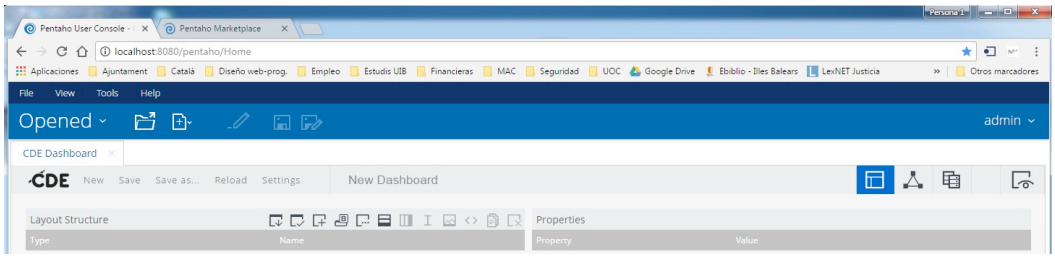


Figura 56 *Layout Panell* del *Dashboard*

- *Components Panel*: És el panell que conté els diferents tipus de gràfics que es poden elegir per a representar la informació.

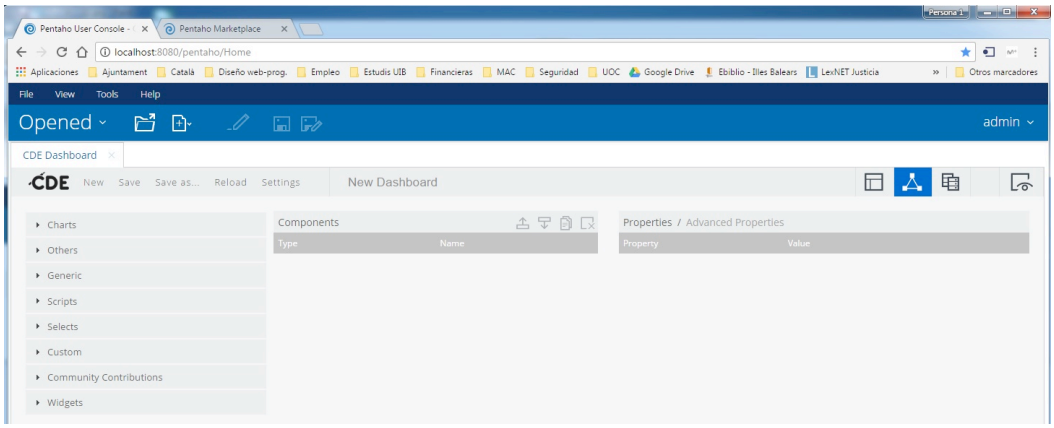


Figura 57 *Components Panell* del *Dashboard*

- *Datasources Panel*: Com el seu nom indica, es tracta del component on es defineix l'origen de dades, es a dir, inclou les fonts de les quals s'obtenen les dades per a la creació del quadre de comandament.

A més dels panells explicats anteriorment, existeix l'opció "Preview your Dashboard", que permet veure el resultat final del *Dahboard* i, per tant, efectuar les modificacions que es considerin oportunes per obtenir el resultat desitjat.

7.1 *Layout Panel*

Es selecciona “Create New” i a continuació es tria l’opció “CDE Dashboard”:

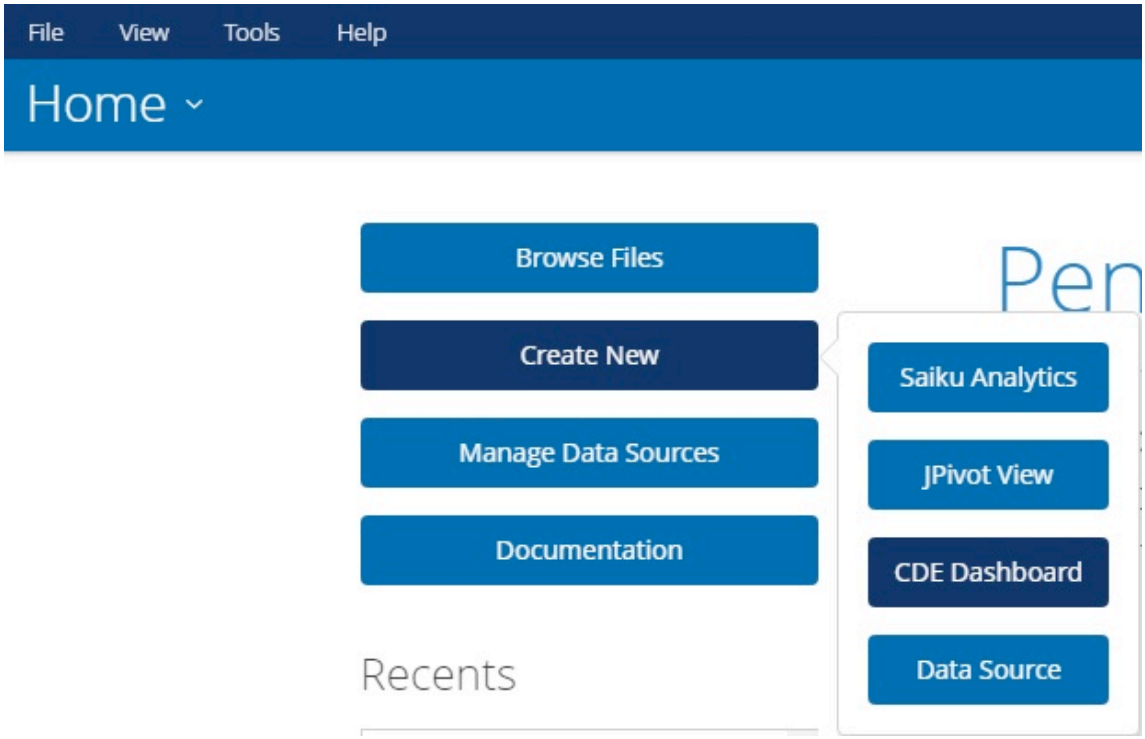


Figura 58 Opció de creació d’un *Dashboard*

El sistema mostra la pantalla inicial per dissenyar el quadre de comandament que es correspon amb l’opció de *Layout Panel*. Abans de continuar s’emmagatzema el document amb l’opció “Save as” emplenant els diferents paràmetres a la carpeta pública:

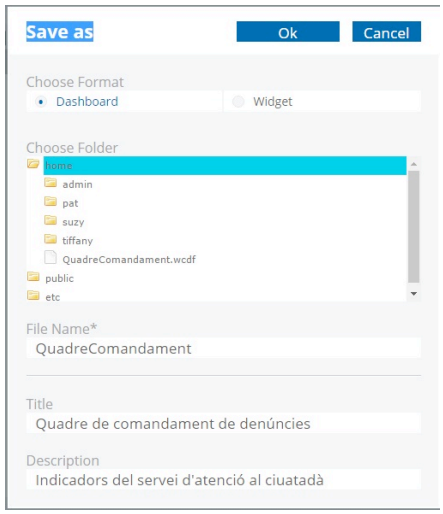


Figura 59 Opció de emmagatzemat d’un *Dashboard*

A continuació es crea l'estructura física on s'ubicaran els diferents elements del *dashboard* . En aquest cas, s'ha optat per una plantilla de pantalla formada per un encapçalament (*Header*), un cos (*Body*), un altres cos (*Body1*) i un peu de pàgina (*Footer*). A la part esquerra de la pantalla, es mostra cada un dels elements i a la part dreta, les propietats de l'element seleccionat.

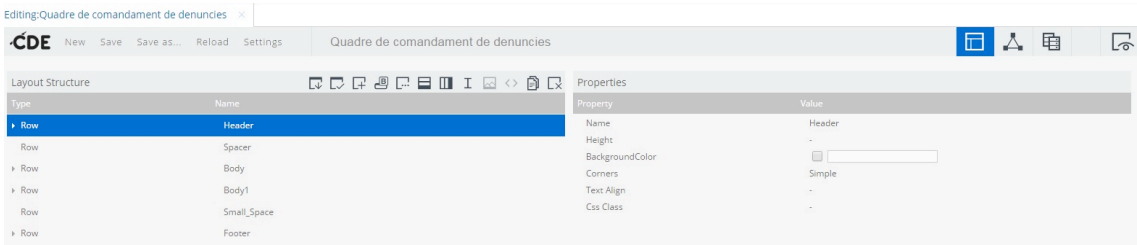


Figura 60 Elements *html* del *Layout Panell*

Si es desplega cada una dels elements *Row* de la plantilla, a la vegada es mostra els elements *Column* que conté ja que es segueix l'entorn de treball o *framework* de *Bootstrap* que és un model basat en caixes (files i columnes) on la summa total de columnes ha de ser 12.

<table>
<tr><td colspan="2">Header</td></tr>
<tr><td colspan="2">Spacer</td></tr>
<tr><td>Panel1</td><td>Panel2</td></tr>
<tr><td colspan="2">Panel3</td></tr>
<tr><td colspan="2">Small_Space</td></tr>
<tr><td colspan="2">Footer_Image</td></tr>
</table>

Figura 61 Format del *Layout Panell*

Indagant en cada uns dels elements que conformen la plantilla, es pot comprova que el títol de la pàgina es pot modificar definit l'etiqueta *html* corresponent.

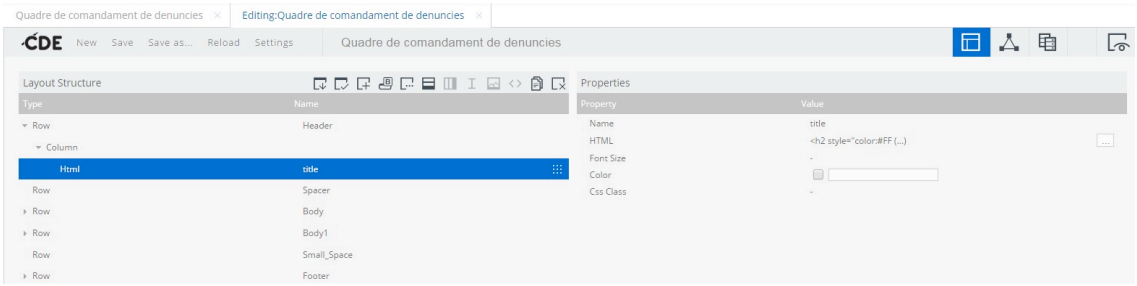


Figura 62 Etiqueta *html* del títol de la pàgina

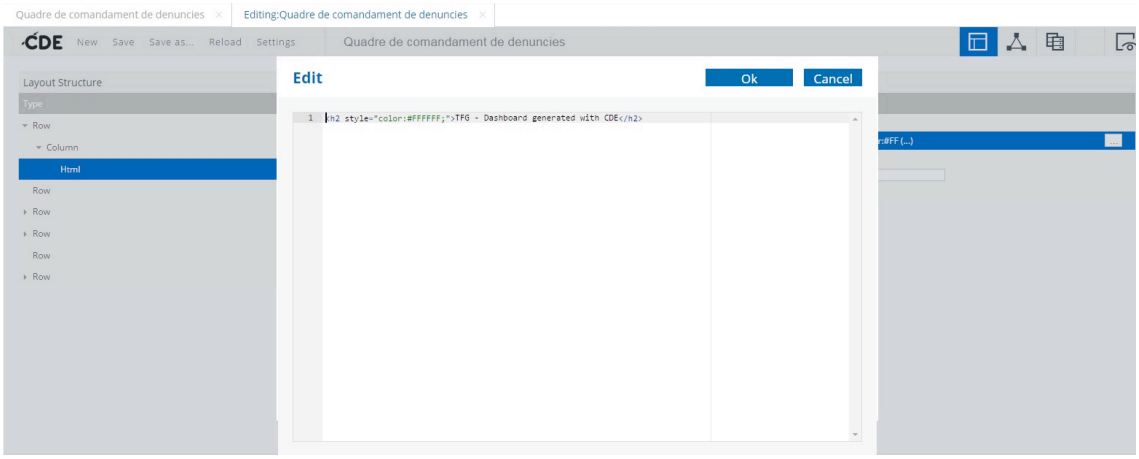


Figura 63 Modificació de l'etiqueta *html* del títol de la pàgina

En definitiva, en aquest apartat es dissenya l'aspecte final de la pàgina segons els elements que es vagin a mostrar. En l'exemple concret, es poden observar que hi han tres panells que contindran els elements d'informació que es definiran a l'apartat *Component Panel*: "Panel1", "Panel2" i "Panel3".

7.2 Component Panel

En aquest apartat es defineixen els elements que es fan servir per a la representació de la informació. Es disposa de una gran col·lecció de diferents objectes organitzats per categories (gràfics, selectors, etc.).

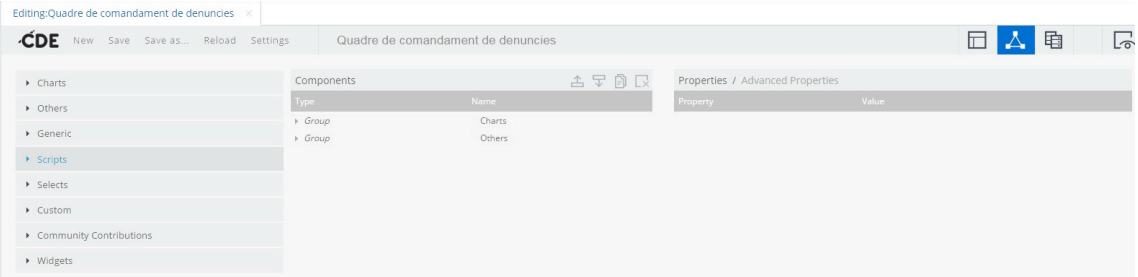


Figura 64 Pantalla de disseny del *Component Panell*

A la part esquerre de la pantalla es troben les diferents categories d'elements disponibles per mostrar la informació dels quals s'optarà per un element dels diferents gràfics (*Charts*) així com un element de la categoria Others, concretament una taula (*Table Component*).

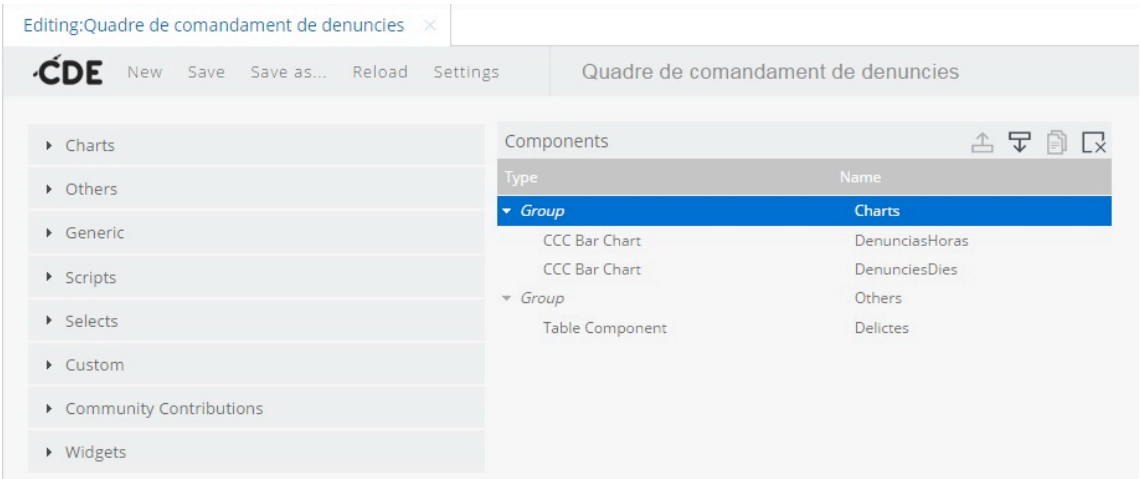


Figura 65 Elements on es mostrarà la informació.

Així doncs per a la representació de la informació corresponent al nombre de denuncies per hores així com el nombre de denúncies per dies de la setmana, en ambdós casos es farà servir una gràfic de barres (*CCC Bar Chart*) i en el cas del nombre de delites denunciats, es fa servir una taula (*Table Component*).

Cada un d'aquest elements disposa d'una sèrie de propietats que s'ha de definir per tal que es mostri la informació adient.

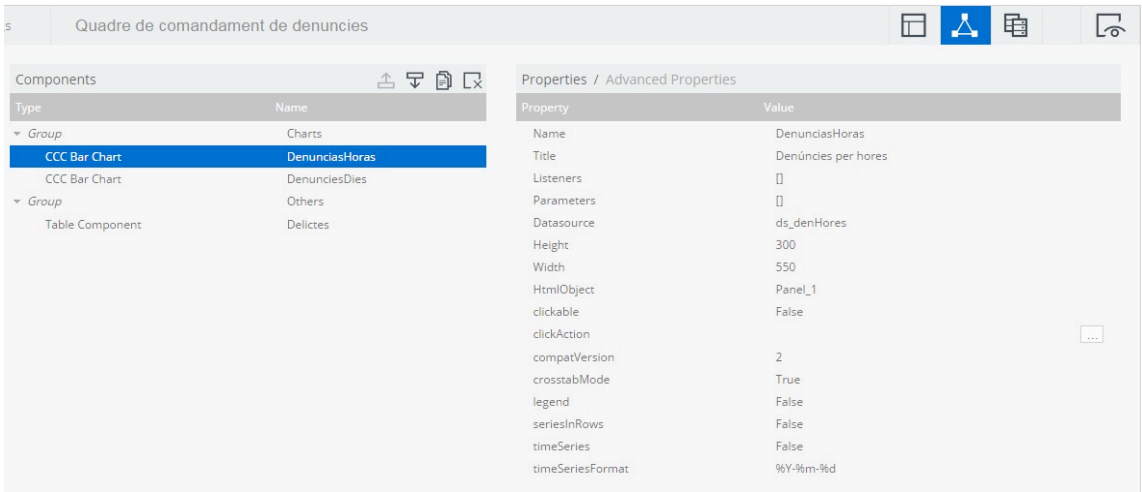


Figura 66 Propietats de l'element *CCC Bar Chart DenunciasHora*.

Les propietats que en aquest cas s'han definit són *Name* per a identificar l'element, *Title* per a definir el títol del gràfic que es mostrarà, *Datasource* per a definir l'origen de les dades que es defineix a *Datasources Panel*, *HtmlObject* per a indicar a quin element del *Layout Panel* s'ubicarà el gràfic. La resta s'han deixat com estaven per defecte si bé es poden modificar segons el resultat final. A l'igual s'ha procedit amb la resta de elements del *Component Panel*.

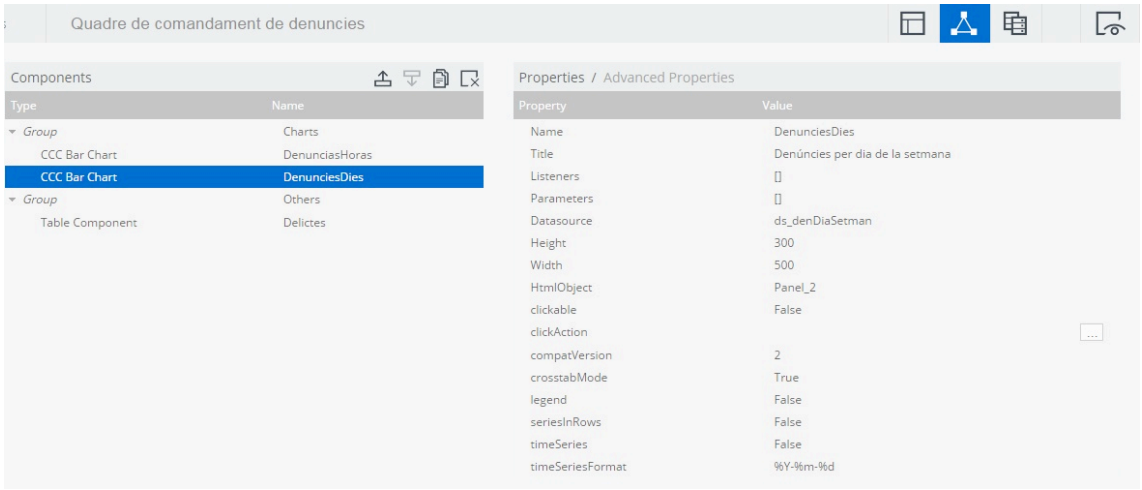


Figura 67 Propietats de l'element *CCC Bar Chart DenunciasDias*.

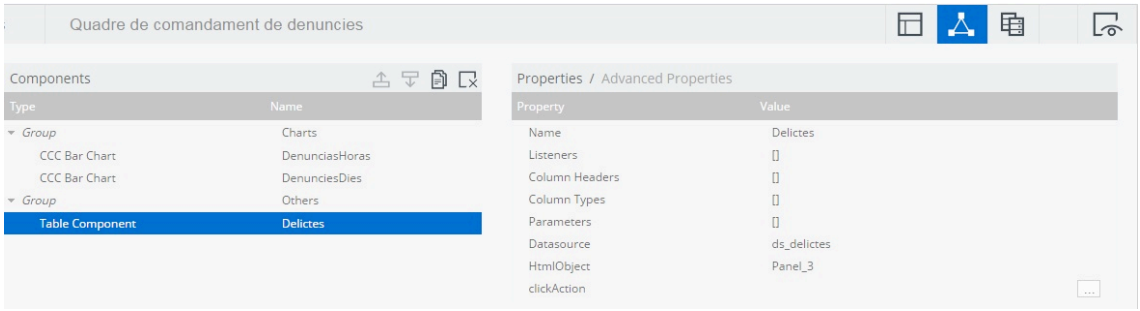


Figura 68 Propietats de l'element *Table Component Delictes*.

7.3 *Datasources Panel*

En aquest apartat es defineix l'origen de les dades que es mostraran en els components del panells de l'apartat anterior. *CDE Dashboard* disposa de diferents formes d'obtenir les dades de diferents fonts d'informació. En la part esquerre de la pantalla es mostren organitzades per categories les diferents opcions per definir l'origen de dades.

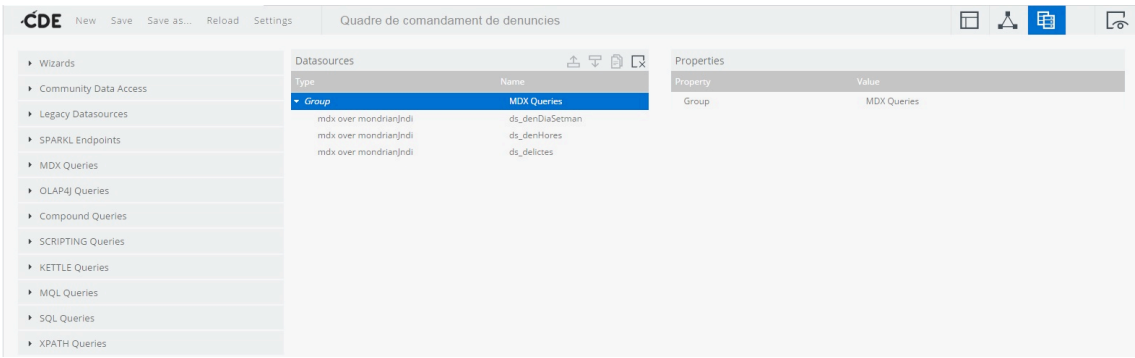


Figura 69 Pantalla de disseny del *Datasources Panell*

Com a l'apartat de la visualització del cub d'OLAP s'ha fet ús de l'eina *Saiku Analytics*, en aquest cas el més adient, és definir l'origen de dades d'acord a la consulta MDX (*mdx over mondrianjndi*) que resulta de la visualització del cub OLAP. D'aquesta manera es pot observar que cada element Datasources disposa d'una sèrie de propietats que s'han de definir.

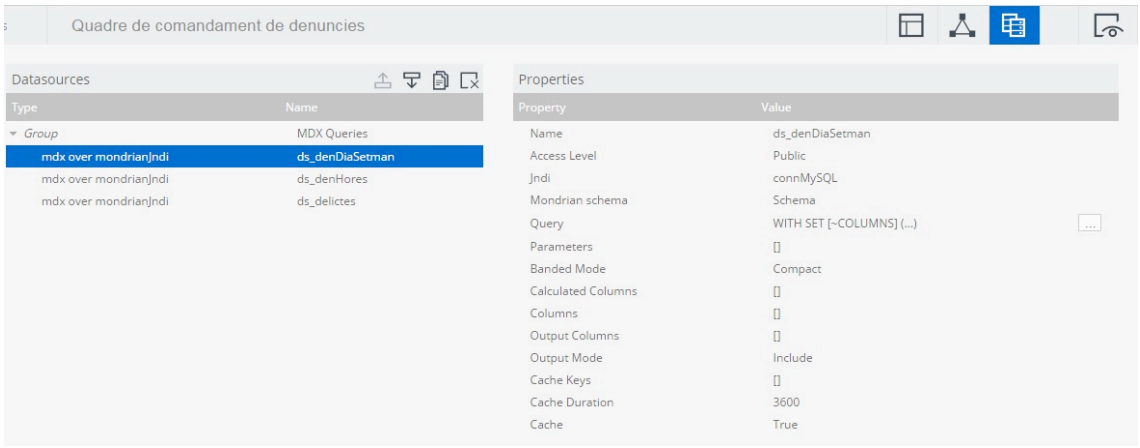


Figura 70 Propietats de l'element *mdx over mondrianjndi ds_denDiaSetman*.

Les propietats que en aquest cas s'han definit són *Name* per a identificar l'element, *Jndi* per a definir la connexió a la base de dades, *Mondrian schema* per a fer referència a l'esquema del cub OLAP, *Query* per a definir la consulta a l'origen de les dades. Aquest darrer paràmetre s'obté de la definició del cub OLAP efectuada per a través de *Saiku Analytics* on al menú existeix l'opció de mostrar la consulta (MDX query).

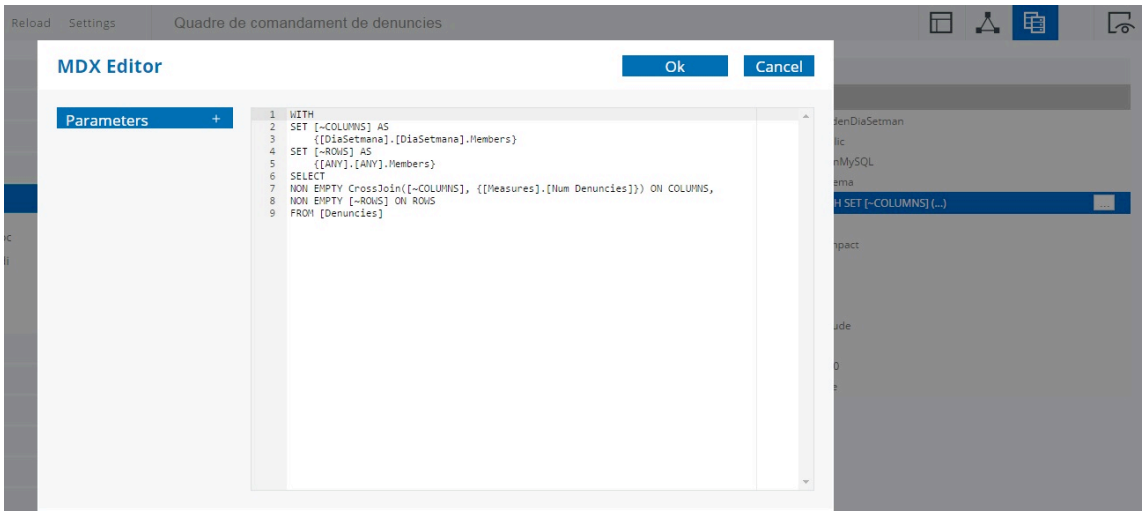


Figura 71 Definició de la consulta MDX.

Seguint les mateixes indicacions anteriors, es defineixen les propietats per a cada un dels origen de dades que proporcionaran les dades als elements del *Component Panel*.

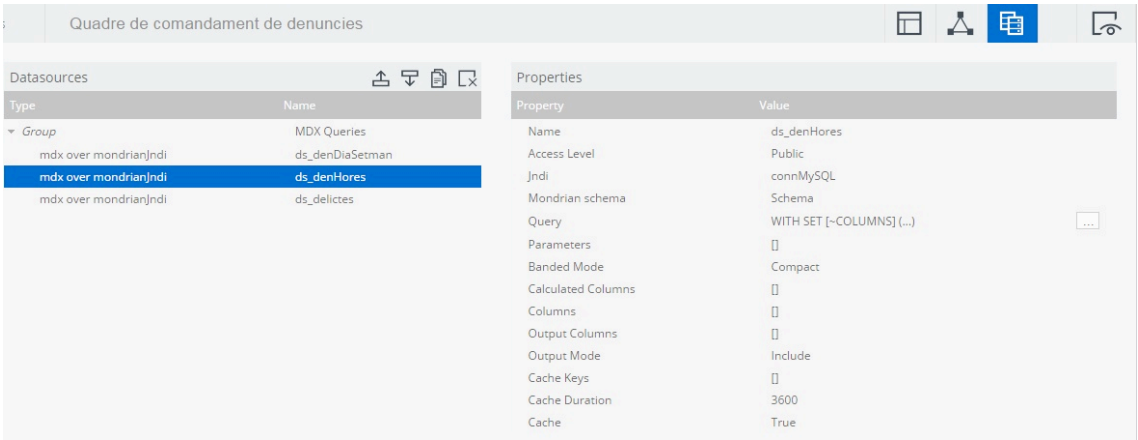


Figura 72 Propietats de l'element *mdx over mondrianjndi ds_denHores*.

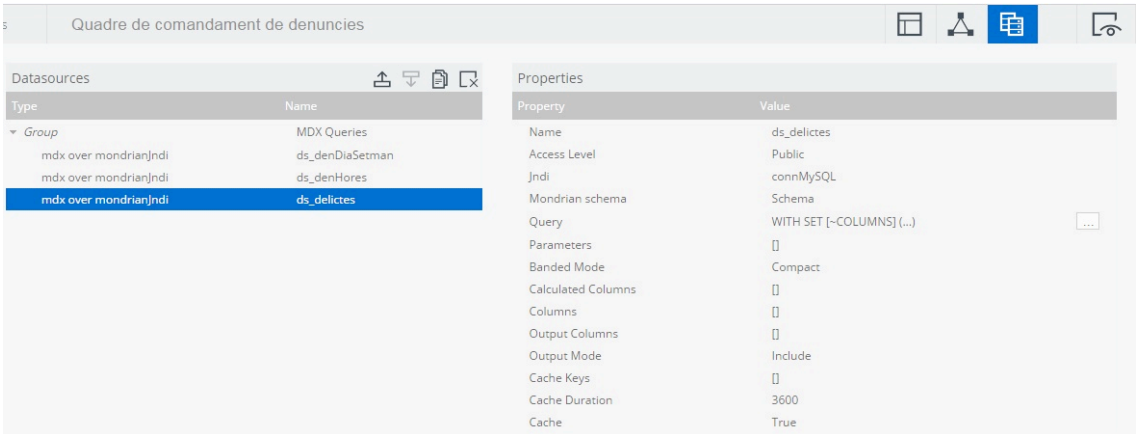


Figura 73 Propietats de l'element *mdx over mondrianjndi ds_delictes*.

Una vegada arribat a aquest punt es pot visualitzar el resultat final amb l'opció del menú *Preview your Dashboard*.

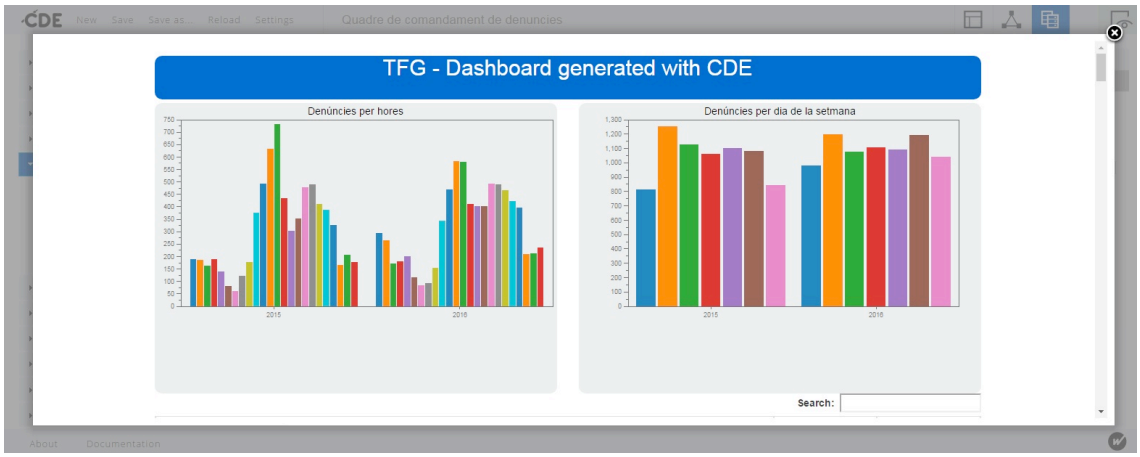


Figura 74 Vista preliminar del *Dashboard* (part superior)

Si el resultat preliminar és satisfactori, a través de l'opció *Save* del menú s'emmagatzema el resultat final. D'aquesta manera a través de l'opció principal del menú de Pentaho, *Browse Files*, es podrà accedir al resultat final del quadre de comandament.

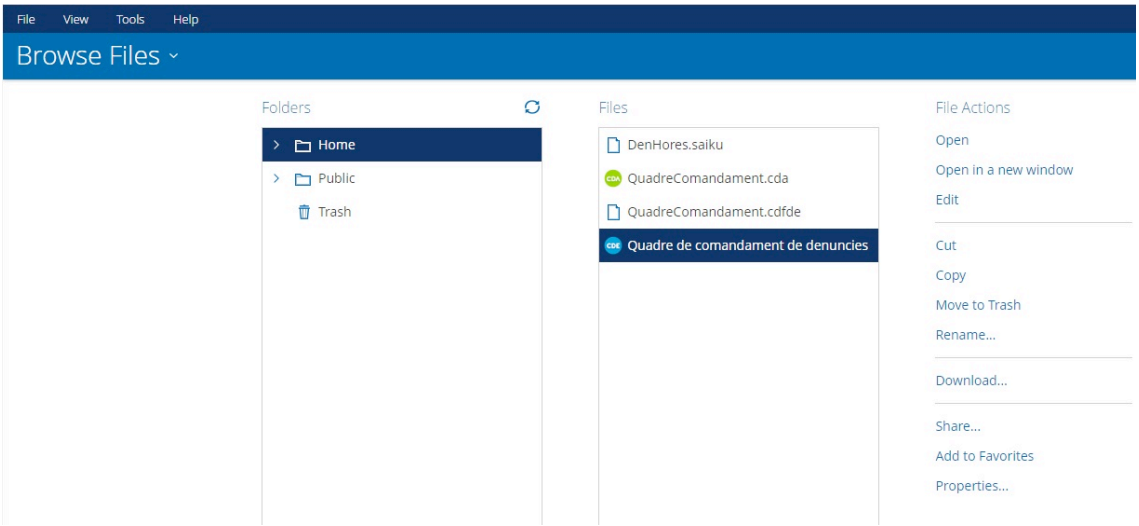


Figura 75 Localització del quadre de comandament.

Es selecciona l'opció del menú de la part dreta *Open* i la pantalla mostrarà el quadre de comandament final. Com es pot observa, es disposa d'altres opcions per a entre d'altres eliminar, editar i copiar.

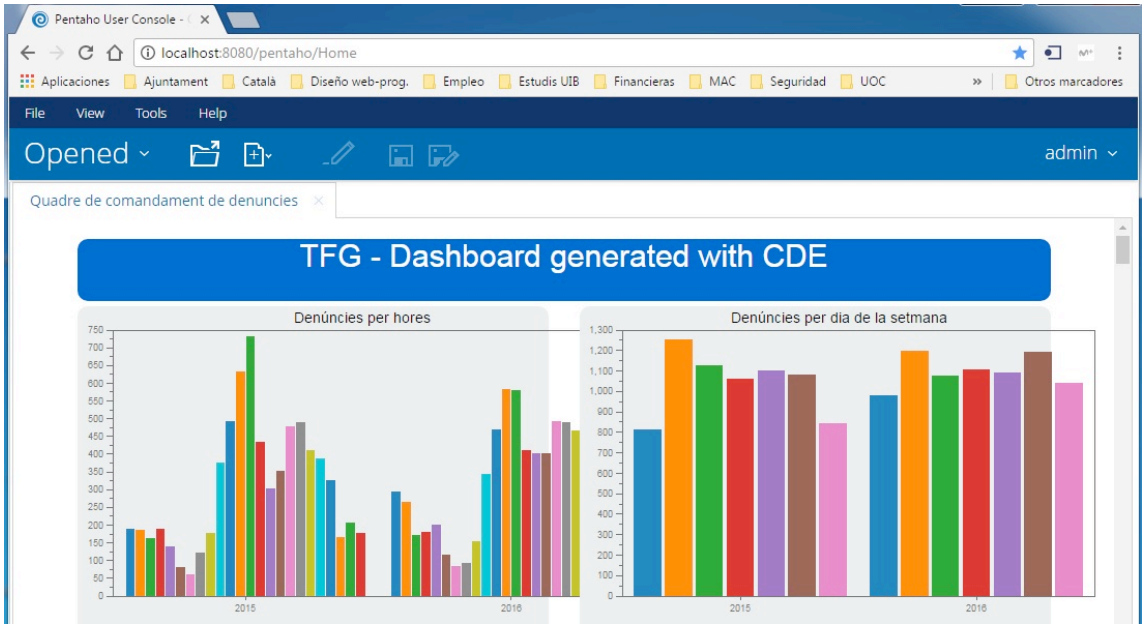


Figura 76 Resultat final del *Dashboard* (part superior)

Quadre de comandament de denuncies

Delicte	2015/Num Denuncies	2016/Num Denuncies
ABUSOS SEXUALES	8	
AGRESIÓN SEXUAL	6	
AMENAZAS EN EL ÁMBITO FAMILIAR	65	
ATENTADO CONTRA AGENTES DE LA AUTORIDAD	60	
BUSCA Y CAPTURA	5	
BUSQUEDA, DETENCION Y PRESENTACION JUDICIAL		14
COACCIONES EN EL ÁMBITO FAMILIAR	27	
CONTRA EL ORDEN PÚBLICO (OTROS)	7	1
CONTRA EL PATRIMONIO (OTROS)	122	78
CONTRA LA LIBERTAD SEXUAL (OTROS)	2	
CONTRA LA SALUD PÚBLICA	5	
CONTRA LA SEGURIDAD VIAL (C.B.I.B.A.)	281	
CONTRA LA SEGURIDAD VIAL (C.B.I.S.E.)	17	
CONTRA LA SEGURIDAD VIAL (CONDUCCIÓN TEMERARIA MANIFIESTA)	7	
CONTRA LA SEGURIDAD VIAL (CONDUCIR PERMISO RETIRADO POR DECISION JUDICIAL)	42	

Figura 77 Resultat final del *Dashboard* (part inferior)

8. Conclusions

El desenvolupament d'aquest treball en el marc del Business Intelligence ha permès aplicar en gran mesura els coneixements assolits al llarg del grau. En concret s'han aplicat metodologies de gestió de projectes (PMBOK) que han fet possible arribar a bon terme amb els objectius establerts. D'altra banda, l'elecció d'eines de codi obert front a les de solucions baix llicència ha evidenciat la possibilitat d'implantació a qualsevol organització si bé s'ha de disposar de personal amb un mínim de qualificació per al seu ús ja que presenta certes dificultats de maneig que usuari bàsic seria incapaç de sortir-ne de l'entrebanc.

Encara que la planificació ha hagut d'ésser modificada en part pel desconeixement de l'eina seleccionada, dur a terme la implantació de *Pentaho Community 7.0* i el seu posterior ús per a una aplicació pràctica per al disseny d'un quadre de comandament, ha estat possible arribar a l'objectiu establert si bé al llarg del decurs del present treball, s'ha pogut comprovar que encara existeix un gran potencial d'ús. La intenció principal ha estat demostrar l'abast d'aquesta eina pel que fa a l'extracció de coneixement de les dades tot seguint les metodologies del Big Data quant a la combinació de dades de múltiples origen de dades, la seva transformació i el seu emmagatzematge en un magatzem de dades (*Data warehouse*).

Quant a les diferents eines que conformen la suite de *Pentaho Community Edition 7.0*., dir que ha estat relativament senzill i intuïtiu realitzar gran part del projecte, a causa de les eines gràfiques que posseeix, com poden ser el *Pentaho Data Integration* o *Schema Workbench*. En el cas concret del primer, es tracta d'una eina molt potent que ha sorprès enormement a l'autor quant a la gran quantitat de possibilitats que ofereix per a la transformació de les dades, fins i tot, una possibilitat que encara no s'ha abordat i que podria ésser una de les línies de futur a treballar com és la implantació de processos d'execució programats (*jobs*) que permeten una gran autonomia al sistema de actualització de la informació de forma periòdica.

Malgrat ser un eina Open Source, els Dashboard que s'han pogut realitzar, posseeixen una gran varietat de gràfics, selectores, taules…, la qual cosa permet facilitar la presa de decisions a les persones de Creu Vermella que vagin a utilitzar el sistema d'informació.

Finalment, l'aplicació de tècniques de *Business Intelligence* amb l'ús d'un programari totalment desconegut ha estat un repte prou complex i a la vegada que ha despertat una enorme curiositat. El repte d'explorar un nou camp d'estudi totalment desconegut ha suposat un al·licient extra quant a les expectatives creades des d'un inici. D'altra banda, han quedat molts d'aspectes pendents de ser explorats en un futur proper que podria suposar un camp d'anàlisi més profund i fins i tot d'especialització en aquesta àrea de les tecnologies.

9. Glossari

BI

De l'anglès business intelligence, es denomina intel·ligència empresarial, intel·ligència de negocis o BI al conjunt d'estratègies i aspectes rellevants enfocades a l'administració i creació de coneixement sobre el mitjà, a través de l'anàlisi de les dades existents en una organització o empresa. .ii, 5, 6, 8, 12, 14

Bootstrap

és un framework o conjunt d'eines de Codi obert per a disseny de llocs i aplicacions web. Conté plantilles de disseny amb tipografia, formularis, botons, quadres, menús de navegació i altres elements de disseny basat en HTML i CSS, així com, extensions de Javascript opcionals addicionals. 46

data warehouse

És una col·lecció de dades orientada a un determinat àmbit (empresa, organització, etc.), integrat, no volàtil i variable en el temps, que ajuda a la presa de decisions en l'entitat en la qual s'utilitza. 3

Datasources

és un nom donat a la configuració de connexió a una base de dades d'un servidor. 44

diagrama de *Gantt*

és una eina visual per a la planificació i programació d'activitats o tasques sobre una línia del temps. Permet a l'usuari establir la durada i el començament de cada activitat. A través d'una gràfica, fàcil d'interpretar que permet a l'usuari el control de la planificació del seu treball. 7

ETL

(extreure, transformar i carregar), freqüentment abreujat ETL) és el procés que permet a les organitzacions moure dades des de múltiples fonts, reformatearlos i netejar-los, i carregar-los en una altra base de dades, data mart, o data warehouse per analitzar, o en un altre sistema operacional per recolzar un procés de negoci. .....iv, 3, 23, 34

framework

entorn de treball o marc de treball, és un conjunt estandarditzat de conceptes, pràctiques i criteris per enfocar un tipus de problemàtica particular que serveix com a referència, per enfrontar i resoldre nous problemes d'índole similar. 46

IDE

Un entorn de desenvolupament integrat o entorn de desenvolupament interactiu, en anglès Integrated Development Environment (IDE), és una aplicació informàtica que proporciona serveis integrals per facilitar-li al desenvolupador o programador el desenvolupament de programari. 16

OLAP

acrònim en anglès de processament analític en línia (On-line Analytical Processing). És una solució utilitzada en el camp de l'anomenada Intel·ligència de negocis (o Business Intelligence) que el seu objectiu és agilitar la consulta de grans quantitats de dades. S'usa en informes de negocis de vendes, màrqueting, informes d'adreça, mineria de dades i àrees similars.....iv, 3, 6, 9, 18, 32, 39, 40, 42

open source

o codi font és l'expressió amb la que es denomina al programari distribuït i desenvolupat lliurement. Es caracteritza pel lliure accés al codi font. 7

Pentaho Community 7.0

Com a plataforma de BI dinàmic i cada vegada més popular, Pentaho transforma la anàlisi de negocis en informació valuosa i comprensible per a l'alta direcció. La plataforma ajudarà a organitzar els continguts desenvolupats al servidor a través de connectors, així com els publicats al servidor a través de les aplicacions d'escriptori.....1, ii

Pentaho Marketplace

loc que permet accedir, gestionar i mantenir tots els plugins del servidor. Un únic lloc amb una interfície clara simple i senzilla per poder llistar els plugins disponibles, provar tots els que es desitgin instal·lar i mantenir actualitzats els que es vagin finalment a utilitzar en el servidor de producció. 39

plugins

En informàtica, un complement o plug-in és una aplicació (o programa informàtic) que es relaciona amb una altra per agregar-li una funció nova i generalment molt específica. Aquesta aplicació addicional és executada per l'aplicació principal i interactuen per mitjà de la interfície de programació d'aplicacions. 39

PMBOK

guia o llibre en el qual es presenten estàndards, pautes i normes per a la gestió de projectes. 7

SGBD

és un conjunt de programes que permeten l'emmagatzematge, modificació i extracció de la informació en una base de dades, a més de proporcionar eines per afegir, esborrar, modificar i analitzar les dades. Els usuaris poden

accedir a la informació usant eines específiques de consulta i de generació d'informes, o bé mitjançant aplicacions a aquest efecte..... 20, 21

10. Bibliografia

- ADVENTURES WITH OPEN SOURCE BI (2010) *A Simple Date Dimension for Mondrian Cubes* [en línia] http://type-exit.org/adventures-with-open-source-bi/2010/07/a-simple-date-dimension-for-mondrian-cubes/ [data de consulta: 25 d'abril de 2017]

- Business Intelligence (2009). *Qué es el Business Intelligence* [en línia] http://www.buyto.es/general-business-intelligence/que-es-y-para-que-sirve-el-business-intelligence [data de consulta: 14 de març de 2017]

- Business Intelligence (2016) *Install the Saiku Analytics plugin in Pentaho BIServer CE* [en línia] http://edpflager.com/?p=3322#more-3322 [data de consulta: 16 de maig de 2017]

- El Rincón del BI (2010). *Business Intelligence Open Source. Proyecto EnoBI usando Pentaho.* [en línia] https://churriwifi.wordpress.com/2010/04/15/15-business-intelligence-open-source-pentaho/ [data de consulta: 14 de març de 2017]

- Jordi Conesa Caralt (coord.) Josep Curto Díaz, Introducción al Business Intelligence, Edició maig 2010, Editorial UOC, Barcelona.

- María Jesús Marco Galindo, Josep Maria Marco Simó, Josep Prieto Blázquez, Ramon Segret Sala, Escanejant la informàtica, Edició maig 2010, Editorial UOC, Barcelona.

- MySQL$_{TM}$ (2017). *Download MySQL Workbench* [en línia] https://dev.mysql.com/downloads/workbench/ [data de consulta: 16 de març de 2017]

- Pentaho.com (2017). *Components Reference* [en línia] https://help.pentaho.com/Documentation/7.0/0D0/160/000 [data de consulta: 16 de març de 2017]

- Pragmatic software applied (2011) *Pragmatic Pentaho Community Edition @AWS Platform* [en línia] http://www.pragtech.co.in/products/other-products/pragmatic-pentaho-community-edition-aws-platform.html [data de consulta: 16 de maig de 2017]

- SINNAPS (2015). *¿Qué es un diagrama de gantt y para qué sirve?* [en línia] https://www.sinnaps.com/blog-gestion-proyectos/diagrama-gantt-sirve [data de consulta: 16 de març de 2017]

- Sistemas, informàtica & Electrónica (2016) *Kettle -Pentaho Schema Workbench – Saiku Reporting – Community Dashboard Editor* [en línia] https://jossjack.wordpress.com/2016/05/08/kettle-pentaho-schema-workbench-saiku-reporting-community-dashboard-editor/ [data de consulta: 14 de març de 2017]

- WIKIPEDIA La enciclopèdia libre (2017). *Almacén de datos* [en línia] https://es.wikipedia.org/wiki/Almacén_de_datos [data de consulta: 17 de març de 2017]

- WIKIPEDIA La enciclopèdia libre (2017). *Bootstrap (framework)* [en línia] https://es.wikipedia.org/wiki/Bootstrap_(framework) [data de consulta: 11 de juny de 2017]

- WIKIPEDIA La enciclopèdia libre (2017). *Framework* [en línia] https://es.wikipedia.org/wiki/Framework [data de consulta: 11 de juny de 2017]

- WIKIPEDIA La enciclopèdia libre (2017). *OLAP* [en línia] https://es.wikipedia.org/wiki/OLAP [data de consulta: 25 d'abril de 2017]