

Diseño de informes

Josep Curto Díaz

PID_00243065

Índice

Introducción.....	5
1. Informes e inteligencia de negocio.....	7
1.1. Tipos de informes	7
1.2. Elementos de un informe	7
1.3. Tipos de métricas	8
1.4. Tipos de gráficos	9
1.5. Ciclo de vida de un informe	12
2. Informes en el contexto de Pentaho.....	14
2.1. Pentaho Reporting Engine	16
2.2. Pentaho Report Designer	16
2.3. Pentaho Metadata Editor	19
3. Caso práctico.....	21
3.1. Diseño de la capa de metadatos en Pentaho	21
3.2. Diseño de un informe mediante el <i>wizard</i> en Pentaho	28
3.3. Diseño de un informe mediante Pentaho Report Designer	36
3.4. Publicación de informes en el servidor	41
Abreviaturas.....	45
Bibliografía.....	46

Introducción

El punto de entrada tradicional para una herramienta de inteligencia de negocio en el contexto de una organización es la necesidad de informes operacionales.

A lo largo de la vida de una empresa, la cantidad de datos que se generan por su actividad de negocio crece de modo exponencial y esa información se guarda tanto en las bases de datos de las aplicaciones de negocio como en ficheros en múltiples formatos.

Es necesario generar y distribuir informes para conocer el estado del negocio y poder tomar decisiones a todos los niveles: operativo, táctico y estratégico.

El primer enfoque es modificar las aplicaciones de negocio para que estas puedan generar los informes. Frecuentemente, el impacto en las aplicaciones es considerable, y afecta tanto al rendimiento de los informes, como a las operaciones que soporta la aplicación.

Es en ese momento cuando se busca una solución que permita generar informes sin impactar en el rendimiento de las aplicaciones de negocio.

Es necesario comentar que:

- Las herramientas de informes existen desde hace mucho tiempo y, por ello, son soluciones maduras que permiten cubrir las necesidades de los usuarios finales respecto a los informes.
- Cada fabricante soporta la creación de todo tipo de informes; en función del enfoque, la dependencia de los usuarios finales respecto al departamento IT puede ser diferente.
- Las fuentes de origen de los informes son varias, desde el propio *data warehouse*, OLAP, metadatos u ODS.
- Las últimas tendencias en informes son incorporar mayores capacidades de visualización, proporcionar mayor libertad a los usuarios finales –para la creación y consumo– y funcionalidades para embeber informes dinámicos en PDF, PPT o dispositivos móviles.

El objetivo de este módulo es presentar los elementos de informe, criterios de realización y un ejemplo por medio de un caso práctico.

1. Informes e inteligencia de negocio

Las herramientas de informes (o también llamadas de *reporting*) permiten responder principalmente a la pregunta de ¿qué pasó? Dado que esa es la primera pregunta que se formulan los usuarios de negocio, la gran mayoría de las soluciones de *business intelligence* del mercado incluyen un motor de generación de informes.

Definamos primero qué es un **informe**: es un documento por medio del cual se presentan los resultados de uno o varios procesos de negocio. Suele contener texto acompañado de elementos como tablas o gráficos para agilizar la comprensión de la información presentada.

Los informes están destinados a usuarios de negocio que tienen la necesidad de conocer la información consolidada y agregada para la toma de decisiones.

Ahora podemos definir formalmente las herramientas de *reporting*: se entiende por **plataforma de reporting** aquellas soluciones que permiten diseñar y gestionar (distribuir, planificar y administrar) informes en el contexto de una organización o en una de sus áreas.

1.1. Tipos de informes

Existen diferentes tipos de informes en función de la interacción ofrecida al usuario final y la independencia respecto al departamento TI:

- **Estáticos**: tienen un formato preestablecido inamovible.
- **Paramétricos**: presentan parámetros de entrada y permiten múltiples consultas.
- **Ad hoc**: son creados por el usuario final a partir de la capa de metadatos que permite usar el lenguaje de negocio propio.

Informes *ad hoc*

Este tipo de informes están orientados al autoconsumo y frecuentemente a la exploración.

1.2. Elementos de un informe

Principalmente, un informe puede estar formado por:

- **Texto:** que describe el estado del proceso de negocio, proporciona las descripciones necesarias para entender el resto de los elementos del informe, así como etiquetas (título) o metadatos (fecha de ejecución).
- **Tablas:** que tiene forma de matriz y permite presentar una gran cantidad de información.
- **Gráficos:** que persiguen el objetivo de mostrar información con un alto impacto visual que sirva para obtener información agregada con mucha más rapidez que a través de tablas.
- **Mapas:** que permiten mostrar información geolocalizada.
- **Métricas:** que permiten conocer cuantitativamente el estado de un proceso de negocio.
- **Alertas visuales y automáticas:** que permiten definir avisos automáticos de los cambios de estado de un proceso de negocio. Estas alertas están formadas por elementos gráficos, como fechas, iconos o colores resultados y deben estar automatizadas en función de reglas de negocio encapsuladas en el cuadro de mando.

1.3. Tipos de métricas

Los informes incluyen métricas de negocio. Por ello es necesario definir los diferentes tipos de medidas existentes basadas en el tipo de información que recopilan, así como la funcionalidad asociada:

1) **Métricas:** valores que recogen el proceso de una actividad o sus resultados. Estas medidas proceden del resultado de la actividad de negocio.

- Métricas de realización de actividad (*leading*): miden la realización de una actividad. Por ejemplo, la participación de una persona en un evento.
- Métricas de resultado de una actividad (*lagging*): recogen los resultados de una actividad. Por ejemplo, la cantidad de puntos de un jugador en un partido.

2) **Indicadores clave:** se trata de los valores que hay que alcanzar, y que suponen el grado de asunción de los objetivos. Estas medidas proporcionan información sobre el rendimiento de una actividad o sobre la consecución de una meta.

- *Key performance indicator* (KPI): indicadores clave de rendimiento. Más allá de la eficacia, se definen unos valores que nos explican en qué rango óptimo de rendimiento nos deberíamos situar al alcanzar los objetivos. Son

métricas del proceso. Por ejemplo, la ratio de crecimiento de altas en un servicio.

- *Key goal indicator* (KGI): indicadores de metas. Definen mediciones para informar a la dirección general si un proceso TIC ha alcanzado sus requisitos de negocio, y se expresan por lo general en términos de criterios de información. Si consideramos el KPI anterior, sería marcar un valor objetivo de crecimiento del servicio que se pretende alcanzar, por ejemplo, un 2 %.

Debemos distinguir que:

- Existen también indicadores de desempeño. Los indicadores clave de desempeño (en definitiva, son KPI) definen mediciones que determinan cómo de bien se está desempeñando el proceso de TI para alcanzar la meta. Son los indicadores principales que indican si será factible lograr una meta o no, y son buenos indicadores de las capacidades, prácticas y habilidades.
- Los indicadores de metas de bajo nivel se convierten en indicadores de desempeño para los niveles altos.

1.4. Tipos de gráficos

En el proceso de confección de un informe, uno de los puntos más complicados es la selección del tipo de gráfico. Debemos empezar primero por la definición formal de este concepto.

Se entiende por **gráfico** la representación visual de una serie de datos.

El gráfico puede ser una herramienta eficaz, ya que:

- Permite presentar la información de manera clara, sencilla y precisa.
- Facilita la comparación de datos y habilita destacar tendencias y diferencias.

El uso del gráfico va a depender del tipo de dato, que podemos clasificar en:

a) Cualitativos: se refieren a cualidades o modalidades que no pueden expresarse numéricamente. Pueden ser ordinales (siguen un orden) o categóricos (sin orden).

b) Cuantitativos: se refieren a cantidades o valores numéricos. Pueden ser discretos (toman valores enteros) o continuos (toman cualquier valor en un intervalo).

Revisemos ahora algunos de los tipos de gráficos más relevantes:

1) **Gráficos de barras:** es una representación gráfica en un eje cartesiano de las frecuencias de una variable cualitativa o discreta. La orientación puede ser vertical u horizontal. Se pueden clasificar en sencillo (representa una única serie de datos), agrupado (contiene varias series de datos) o apilado (se divide en segmentos de diferentes colores o texturas y cada uno de ellos representa una serie).

2) **Histograma:** se usa para representar las frecuencias de una variable cuantitativa continua. En uno de los ejes se posicionan las clases de la variable continua (los intervalos o las marcas de clase que son los puntos medios de cada intervalo) y en el otro eje las frecuencias. Existen también los histogramas bidireccionales, que contienen dos series de datos cuyas barras de frecuencias crecen en sentidos opuestos.

3) **Gráfico de líneas:** es una representación gráfica en un eje cartesiano de la relación que existe entre dos variables. Se suelen usar para presentar tendencias temporales.

4) **Gráfico de Pareto:** es un tipo de gráfico de barras vertical ordenado por frecuencias de forma descendente que identifica y da un orden de prioridad a los datos.

5) **Gráfico de sectores:** es una representación circular de las frecuencias relativas de una variable cualitativa o discreta que permite, de una manera sencilla y rápida, su comparación.

6) **Pictograma:** es un gráfico que representa mediante figuras o símbolos las frecuencias de una variable cualitativa o discreta.

7) **Gráfico de dispersión:** muestra en un eje cartesiano la relación que existe entre dos variables e informa del grado de correlación entre ellas. El tipo de correlación se puede deducir según la forma de la nube de puntos, a saber: nula, lineal o no lineal.

8) **Gráfico de burbujas:** es una variante del gráfico de dispersión al que se añade una tercera dimensión vinculada al tamaño de los puntos (que se convierte en burbujas) e incluso puede añadirse una cuarta vinculada con el color de cada burbuja. Por lo tanto, permite estudiar la relación de tres variables.

9) **Cartograma:** es un mapa en el que se presentan datos por regiones bien poniendo el número o bien coloreando las distintas zonas en función del dato que representan.

Histogramas bidireccionales

La pirámide de población es un tipo de histograma bidireccional que muestra la estructura demográfica de una población, por sexo y edad, en un momento determinado.

Gráfico de líneas

Si se unen los puntos medios de las bases superiores de las barras en los gráficos de barra, se obtiene el polígono de frecuencias.

10) **Gráficos en cascada:** es un tipo de gráfico normalmente usado para comprender cómo un valor inicial se ve afectado por una serie de cambios intermedios positivos y negativos.

11) **Diagrama de caja:** es un tipo de gráfico que utiliza los cuartiles para representar un conjunto de datos. Permite observar de un vistazo la distribución de los datos y sus principales características: centralidad, dispersión, simetría y tamaño de las colas.

12) **Mapa de calor:** es una representación gráfica de los datos donde los valores individuales contenidos en una matriz se representan como colores

13) **Treemap:** es un método para la visualización de datos jerárquicos mediante el uso de rectángulos anidados.

14) **Diagramas ternarios:** son usados para representar el porcentaje relativo de tres componentes donde el único requisito es que los tres componentes tienen que sumar un 100 %.

Aun conociendo algunos de los gráficos más conocidos, puede ser complicado elegir entre ellos. Existen algunos criterios para seleccionar el tipo de gráfico adecuado:

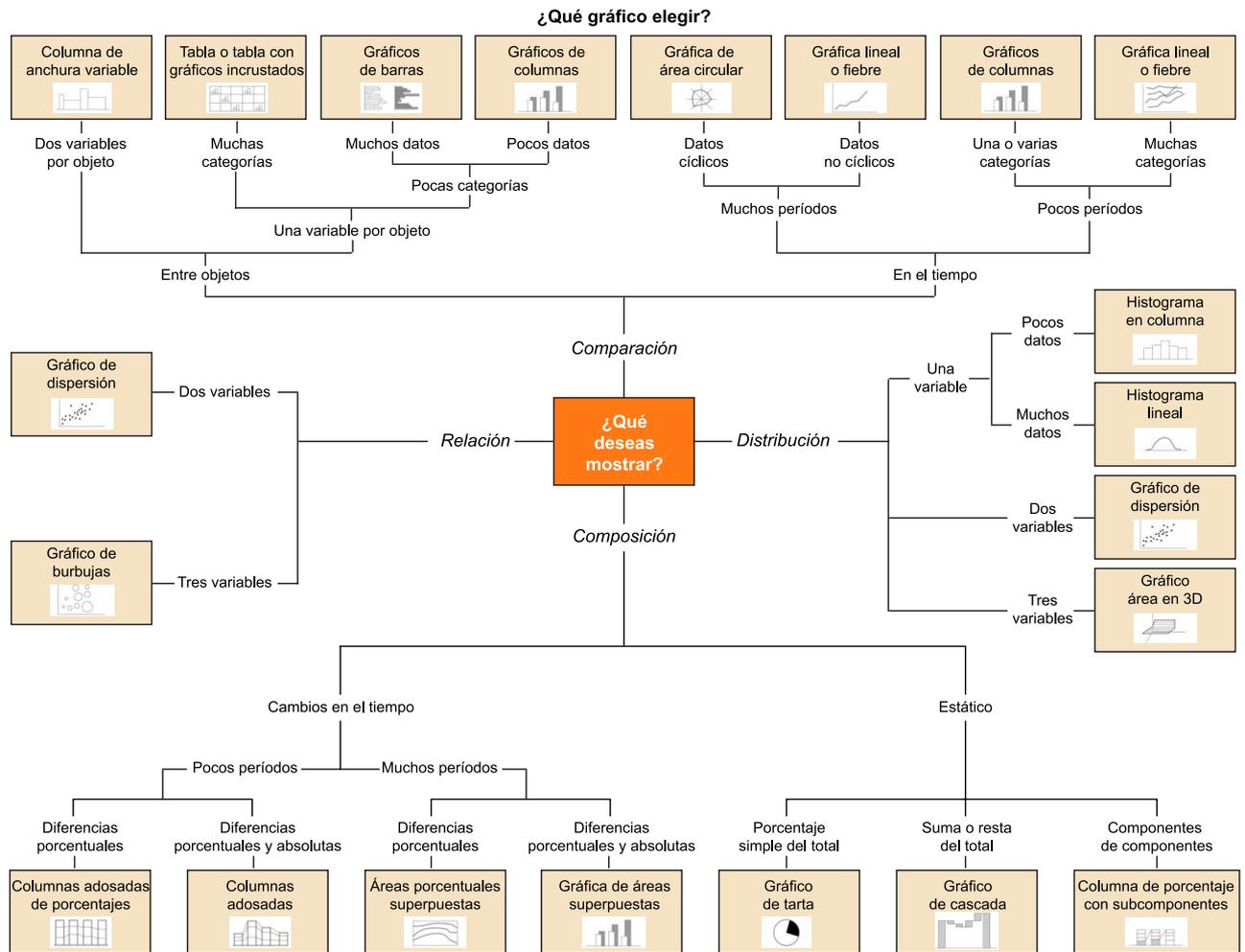
- Preguntar qué se desea mostrar. Existen varias opciones: comparación, distribución, relación o composición.
- Identificar el tipo de dato: cuantitativo o cualitativo.
- Necesidad de trabajar con una o más variables.
- Apalancarse en otras dimensiones: temporales o geográficas.

Para elegir un gráfico es necesario seguir un proceso sistemático a través de una serie de preguntas:

- ¿Qué se desea mostrar? Tenemos varias opciones: comparación, distribución, composición y relación.
- ¿Cuántas variables tenemos? Podemos tener una o más de una.
- ¿Es estático o cambia en el tiempo?

El siguiente diagrama de Abela permite identificar el gráfico más adecuado en función de las preguntas anteriores.

Figura 1. Selección de gráficos



Fuente: Andrew Abela.

1.5. Ciclo de vida de un informe

Como ya se ha definido, el objetivo de un informe es presentar los resultados de un área o proceso de negocio. En el momento de diseñar un informe no solo es necesario tener en cuenta la forma y el contenido que tendrá, sino su ciclo de vida para que pueda continuar generando valor para la organización. Por ello debemos introducir lo que se conoce como el ciclo de vida de un informe, que se compone por las siguientes etapas:

Ciclo de vida de un informe

Este ciclo de vida tiene similitudes con el de Demming, conocido como PDCA (*plan – do – check – act*), si bien algunas de sus fases están más detalladas.

1) Identificar: consiste en determinar los aspectos de negocio relevantes para su comprensión e identificar las métricas que representan dichos aspectos y que sean relevantes para la compañía y sus gestores.

2) Medir: consiste en desarrollar o revisar los sistemas de información que recopilan la información necesaria para las métricas. Inicialmente la compañía debería tener ya estos sistemas implementados, pero no es extraño encontrarse con la necesidad de habilitar este tipo de sistemas.

3) Revisar: consiste en comprobar que el dato de los sistemas anteriores representa de manera efectiva, válida, completa y con calidad los procesos de negocio, por lo que el sistema de *reporting* posterior tendrá dichas características. En esencia, estamos hablando de gobernanza del dato.

4) Crear: consiste en crear el informe y en habilitar su distribución a las partes interesadas.

5) Recopilar: consiste en recopilar de manera continua el *feedback* por parte de los usuarios, así como futuras necesidades.

6) Mejorar: consiste en implementar las mejoras recopiladas en el punto anterior en el sistema de *reporting*. Estas mejoras pueden ser en forma, contenido, distribución, calidad del dato, etc.

2. Informes en el contexto de Pentaho

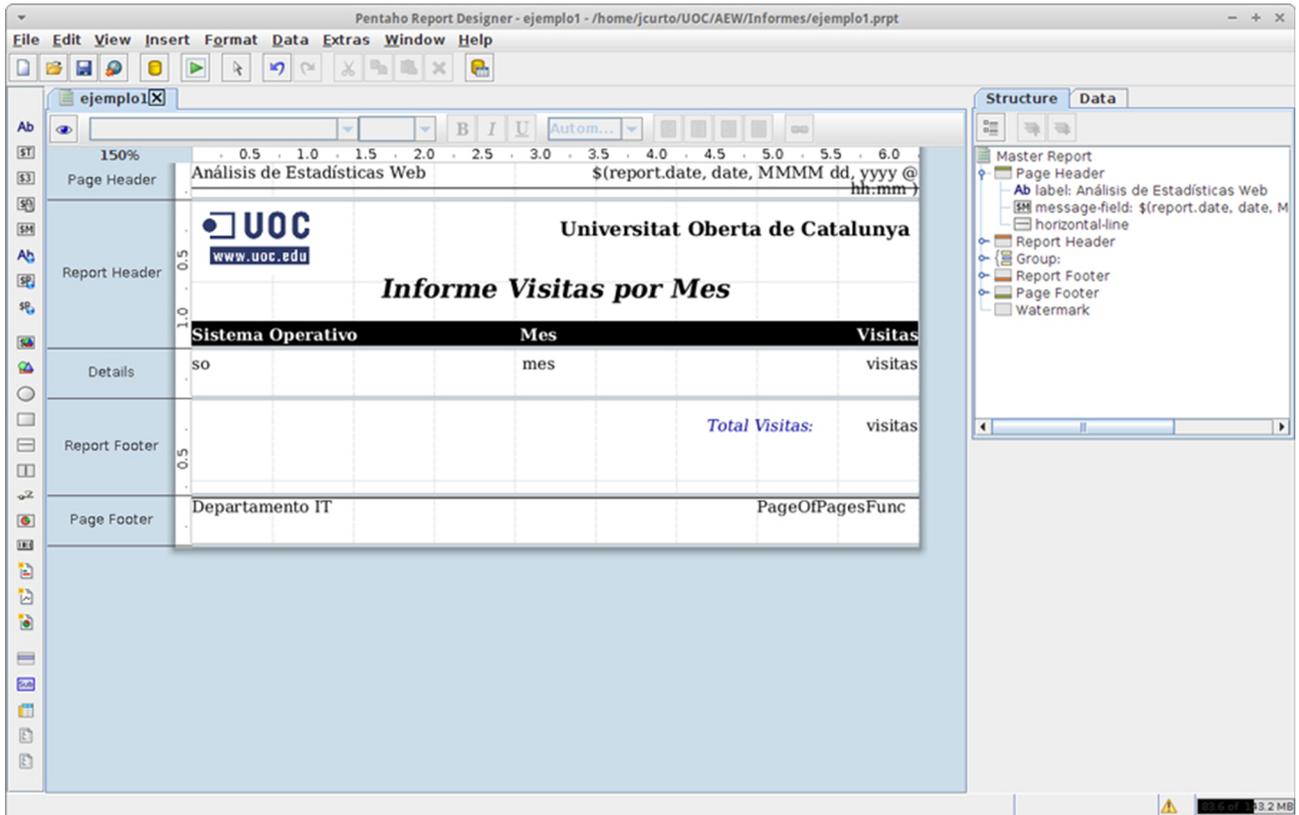
Pentaho Reporting es el motor de informes de Pentaho que está integrado en la *suite*. El proyecto se inició bajo el nombre JFreeReports en 2002. Su origen se remonta a un desarrollo de David Gilbert, creador de JFreeChart, para cubrir necesidades de generación de informes. Pronto, se unió Thomas Morgner al proyecto y se erigió en el principal desarrollador.

En el 2006, Pentaho adquirió el proyecto y Thomas entró a formar parte de la compañía. El proyecto sigue siendo *open source* y está bajo la licencia GNU Lesser General Public License (LGPL) Version 2.1.

La herramienta sigue la numeración de la *suite* de Pentaho y, en los últimos años, ha sido completamente rediseñada para incluir múltiples funcionalidades. Aunque anteriormente existían más soluciones (como *web ad-hoc query reporting*; WAQR), actualmente solo existe propiamente una herramienta de diseño: Pentaho Report Designer.

Esta herramienta es un editor basado en eclipse con prestaciones profesionales para la creación y personalización de informes destinado principalmente a desarrolladores.

Figura 2. Interfaz de Pentaho Report Designer



Fuente: Josep Curto.

Adicionalmente hay una herramienta que permite crear una capa de metadatos de negocio llamada Pentaho Metadata. Esta herramienta permite abstraer los datos consolidados en el *data warehouse* y transformarlos a lenguaje de negocio. Pentaho Metadata que está relacionada con las herramientas de *reporting* al ser una potencial fuente de origen.

Resumiendo:

Tabla 1

	Community	Enterprise
Motor generador de informes (ejecución de informes)	Pentaho Reporting Engine (JFreeReports)	
Herramientas de desarrollo (diseño de informes/capa de metadatos)	Pentaho Report Designer, Pentaho Metadata	

Tratamos, a continuación, estas herramientas.

2.1. Pentaho Reporting Engine

Pentaho Reporting Engine (PRE) es un motor de generación de informes basado en JFreeReports. Su única funcionalidad es la de ejecutar informes creados con la herramienta de diseño. Este motor está incluido en tres herramientas: Pentaho Server, Pentaho Report Designer y PDI.

Figura 3. Informe ejecutado en Pentaho Server

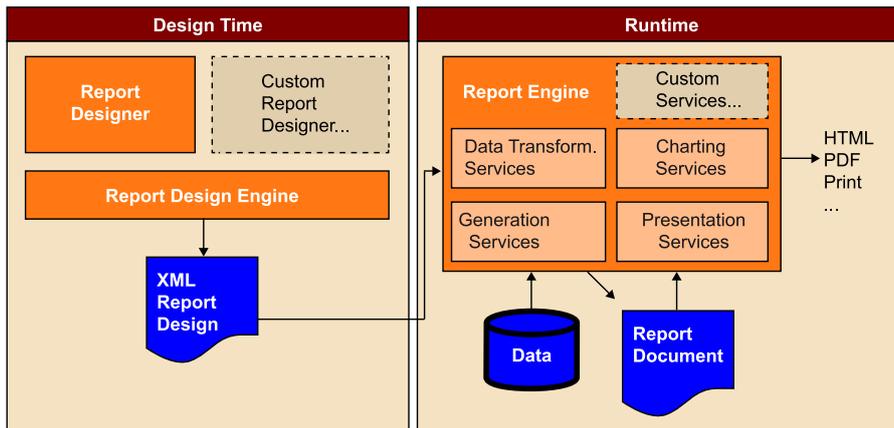
Revenue			
Direct Sales	\$ 400,000		
Channel Sales	\$ 150,000		
Total Revenue		\$ 550,000	
Beginning Inventory	\$ 40,000		
Net Purchases	\$ 325,000		
Ending Inventory	\$ 35,000		
Cost of Goods Sold		\$ 330,000	
Gross Margin			\$ 220,000
Expenses			
Selling expenses			
Sales Salaries	\$ 48,000		
Nonrecurring Item	\$ 12,000		
Other	\$ 13,000		
Total Selling Expenses		\$ 73,000	
General and administrative expenses			
Office Salaries	\$ 27,100		
Depreciation	\$ 5,500		

Fuente: Josep Curto.

2.2. Pentaho Report Designer

Pentaho Report Designer (PRD) es un editor programado en Java que encapsula la lógica de un informe en un fichero XML (con extensión). El flujo de creación de informes está representado por el siguiente esquema:

Figura 4. Flujo de creación de informes



Fuente: Pentaho.

El resultado de Report Designer es un fichero XML que no contiene datos, sino solo las instrucciones del informe. En el momento de ejecución, cuando un usuario pide la información o es distribuido, es cuando se recuperan los datos a través del motor de informes y se genera el informe en el formato deseado.

Las principales características de PRD son las siguientes:

- Generales: previsualizador, editor gráfico, creación basada en bandas, *drag & drop*.
- Elementos que se pueden integrar en los informes: informes, listas, tablas, tablas pivotantes, texto dinámico, imágenes, etiquetas, *subreports*, códigos de barras, *sparklines*, tabla de contenidos e índice
- Formato de salida de gráficos: PNG, JPG, SVG, EPS, PDF.
- Propiedades de informes: paginación, *templates*, *subreports*, *javascript scripting*, *hyperlinks*, etc.
- Tipos de informes: estáticos y paramétricos.
- Formato de salida: PDF, HTML, Excel, Excel 2007, texto, RTF y CSV.
- Distribución: *bursting*, *email*, *web service*.
- Fuentes de datos: JDBC, ODBC, Reflection, Hibernate, Kettle, Mondrian, OLAP4J, Pentaho Metadata, Scripted data y MongoDB.
- Existencia de un API de programación.

Bursting

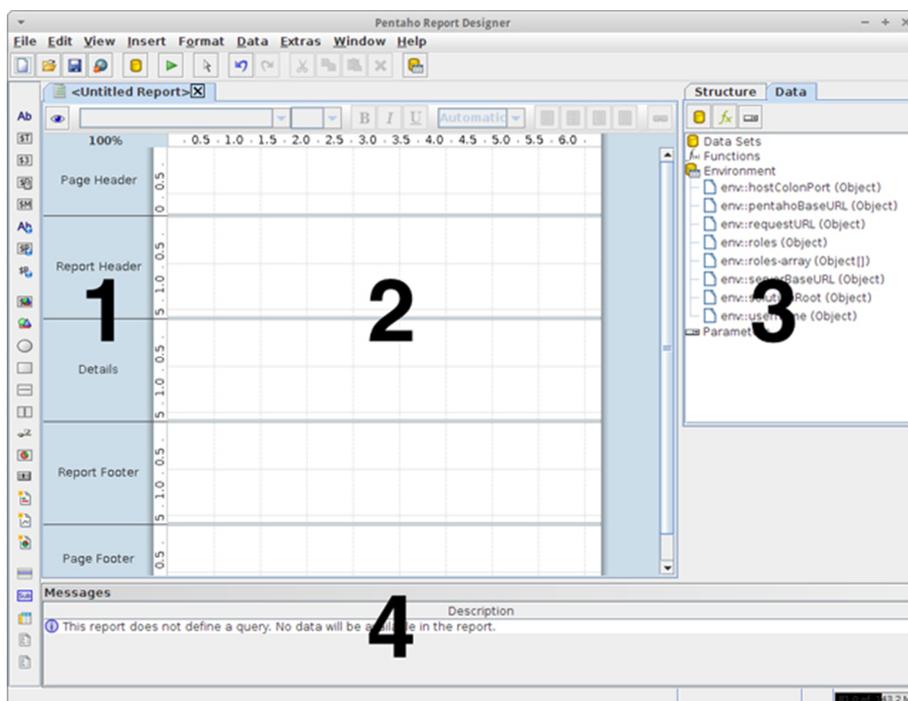
El término hace referencia a la generación de un informe y a su distribución automática a múltiples usuarios mediante la plataforma de *reporting*, siguiendo reglas de negocio como necesidades, fecha, tipo de usuario y privilegios de acceso de información.

- Gráficos: de barras, de líneas, de área, de sectores, combinación de varios de los anteriores, combinación de barras y líneas, en anillo, de burbujas, de dispersión, de radar y en cascada.
- Funcionalidades avanzadas de creación de mensajes de correo electrónico.
- Integración de un *wizard* para la creación de informes, conocido como Pentaho Report Design Wizard (PRDW), que facilita la creación de informes basado en plantillas.

Esta herramienta sigue los despliegues por área tradicionales de las herramientas de diseño:

1. Paleta de elementos.
2. Área de trabajo (cada informe está formado por diferentes bandas, cabecera, agrupación, cuerpo, pie de página, etc.).
3. Propiedades de los elementos.
4. Zona de errores y *logs*.

Figura 5. Secciones interfaz de PRD



Fuente: Josep Curto.

2.3. Pentaho Metadata Editor

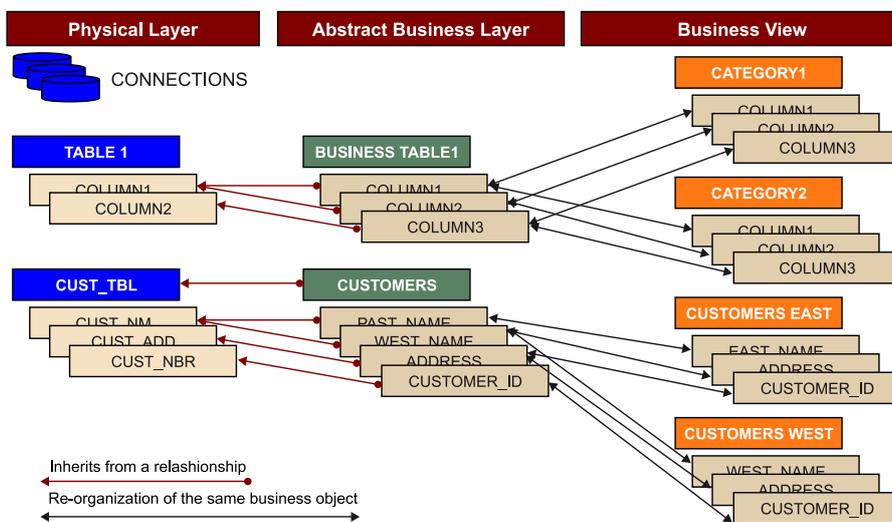
Pentaho incluye en su *suite* una capa de metadatos basada en la especificación del Object Management Group llamada Common Warehouse Metamodel (CWM). Este proyecto está auspiciado por Matt Casters, creador de PDI. En el contexto de la inteligencia de negocio, la existencia de este elemento ofrece diversos beneficios:

- Independencia de los elementos de negocio del *data warehouse*.
- Ofrecer la información del *data warehouse* a los usuarios finales mediante su lenguaje de negocio.
- Encapsular funcionalidades propias de diferentes departamentos en la capa de negocio: color, fuentes, descripción diferentes, etc.
- Soporte para la internacionalización de elementos.
- Definir la seguridad.

En el caso concreto de Pentaho, tanto las herramientas de informes como el servidor (publicando el esquema de metadatos como fuente de datos) están explotando actualmente las capacidades de la capa de metadatos.

Esta herramienta, conocida como Pentaho Metadata Editor (PME), permite crear vistas de negocio. Estas vistas de negocio deben referenciar a una tabla del *data warehouse*. A partir de dicha referencia se crean las relaciones que estarán disponibles al usuario final en función de las necesidades que han sido identificadas.

Figura 6. Funcionamiento Pentaho Metadata

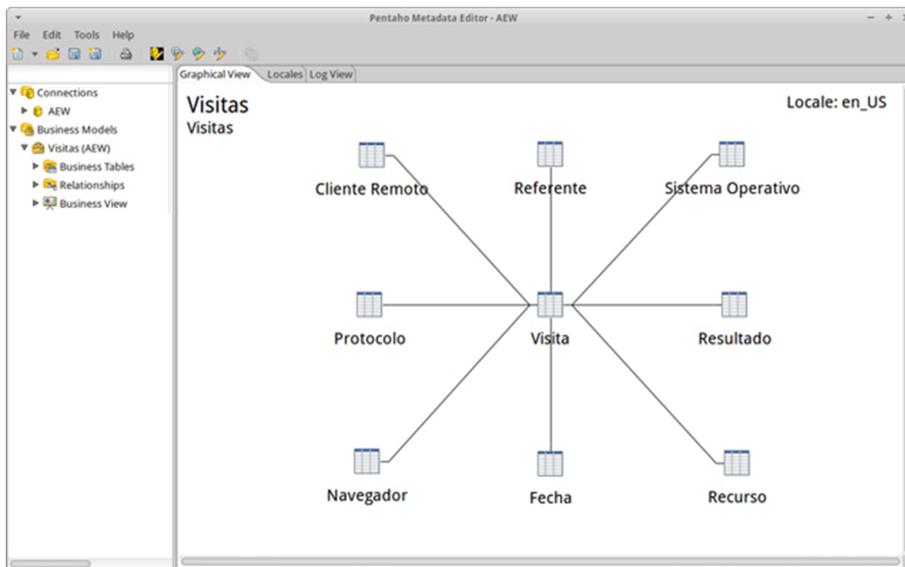


Fuente: Pentaho.

El área de trabajo se divide en dos partes:

- Esquema jerárquico de elementos, que incluye las conexiones y las vistas de negocio.
- Esquema gráfico de la vista de negocio.

Figura 7. Interfaz Pentaho Metadata Editor



Fuente: Josep Curto.

Cabe comentar que esta herramienta en los últimos años no ha evolucionado de forma tan significativa como el resto de las herramientas de la *suite* de Pentaho.

3. Caso práctico

En el contexto de Pentaho existen diferentes opciones para publicar informes o datos:

- Mediante la capa de metadatos (que recordemos que desempeña el papel de nueva fuente de datos).
- Mediante el *wizard* incluido en Pentaho Report Designer.
- Creando un informe desde cero con Pentaho Report Designer.

3.1. Diseño de la capa de metadatos en Pentaho

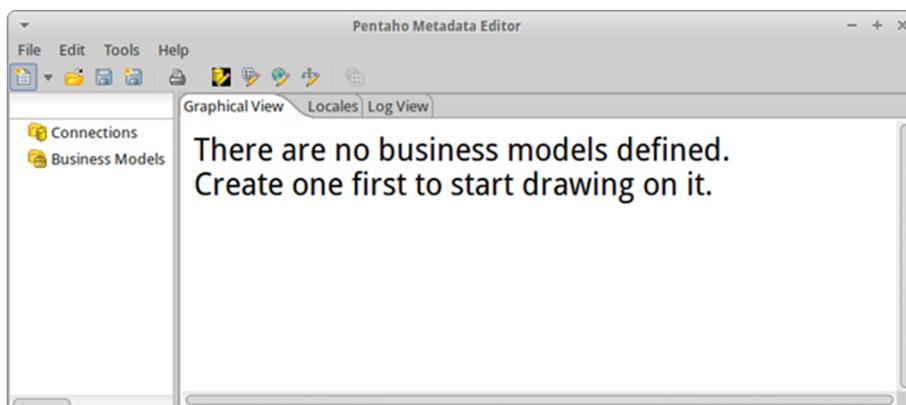
Para poder crear informes basados en la capa de metadatos, es necesario primero crear una capa de metadatos. Esta capa es una abstracción entre el *data warehouse* y los informes cuyo beneficio principal es proporcionar al usuario final la capacidad de crearse sus propios informes sin depender del departamento IT.

La herramienta PME es la que nos permite crear esta capa y se inicia a partir del fichero metadata-editor.sh (o metadata-editor.bat en el caso de Windows).

Cuando se inicia la herramienta no existe ningún elemento creado. Para crear una capa de metadatos, es necesario:

- Crear una conexión al *data warehouse*.
- Crear un modelo de negocio.
- Crear las relaciones del modelo de negocio.
- Crear las categorías.

Figura 8. Primer contacto con PME

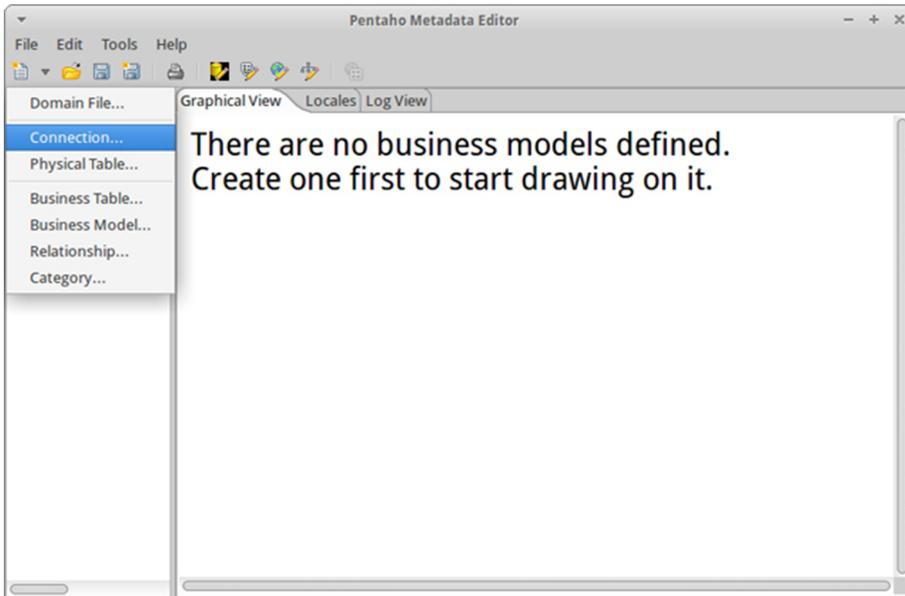


Fuente: Josep Curto.

Procedemos a explicar el proceso.

Creamos una nueva conexión.

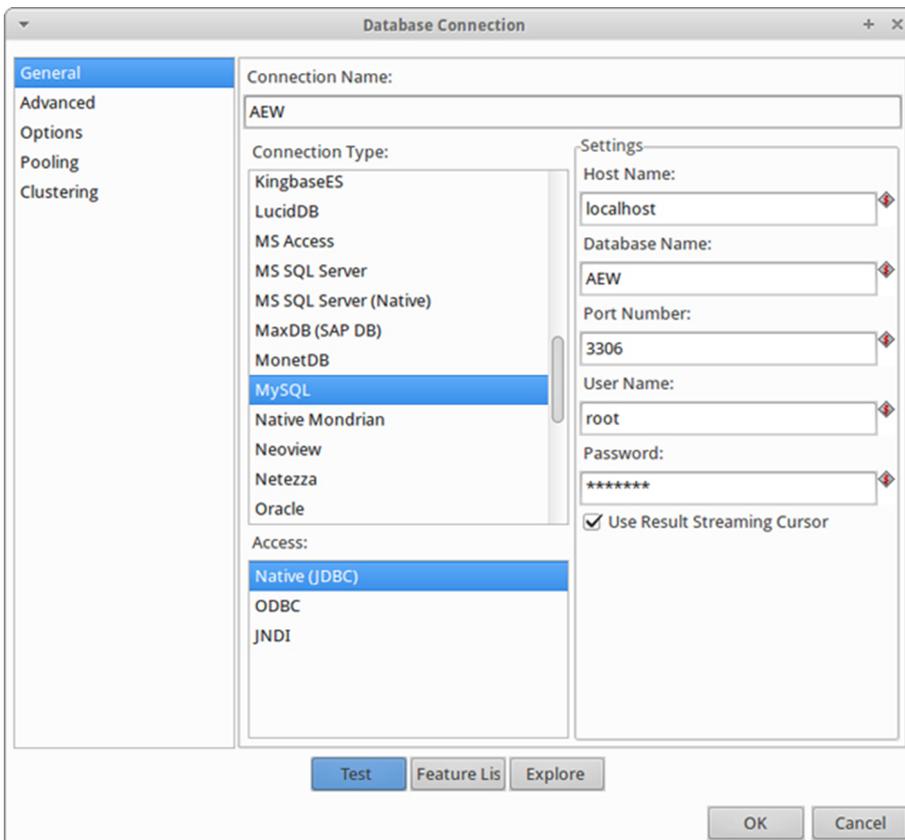
Figura 9. Crear conexión en PME



Fuente: Josep Curto.

Completamos los parámetros de conexión (con los datos que hemos estado usando a lo largo de los diferentes módulos).

Figura 10. Completar datos de conexión en PME



Fuente: Josep Curto.

Una vez conectados correctamente, seleccionamos las tablas que deben formar parte de nuestro modelo. Se han definido por lo tanto las tablas físicas de la capa de metadatos.

Figura 11. Selección tablas para el modelo

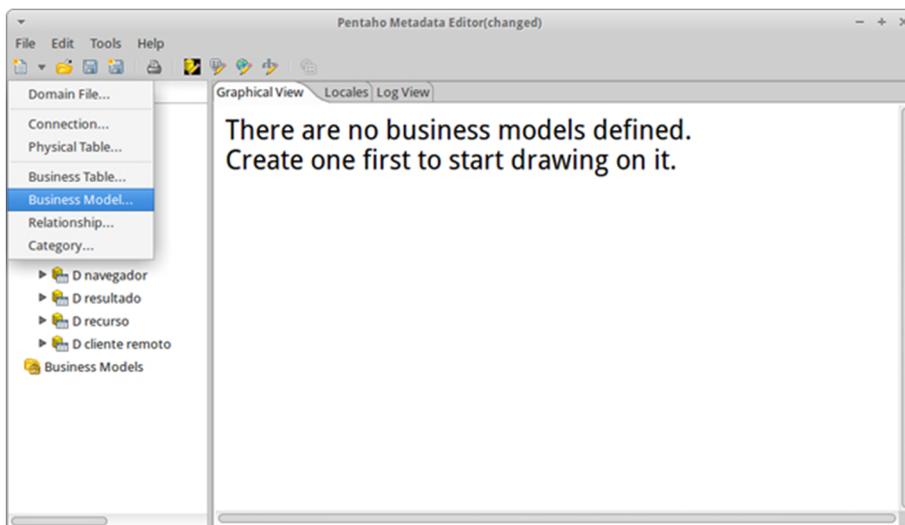


Fuente: Josep Curto.

El hecho de tener las tablas a nivel de conexión no supone que sean consideradas realmente en la capa de datos. Simplemente que están disponibles para el uso, por lo que es necesario mapear las tablas físicas con su equivalente en el modelo de negocio usando lenguaje de negocio para el usuario final.

Para ello necesitamos crear un modelo de negocio.

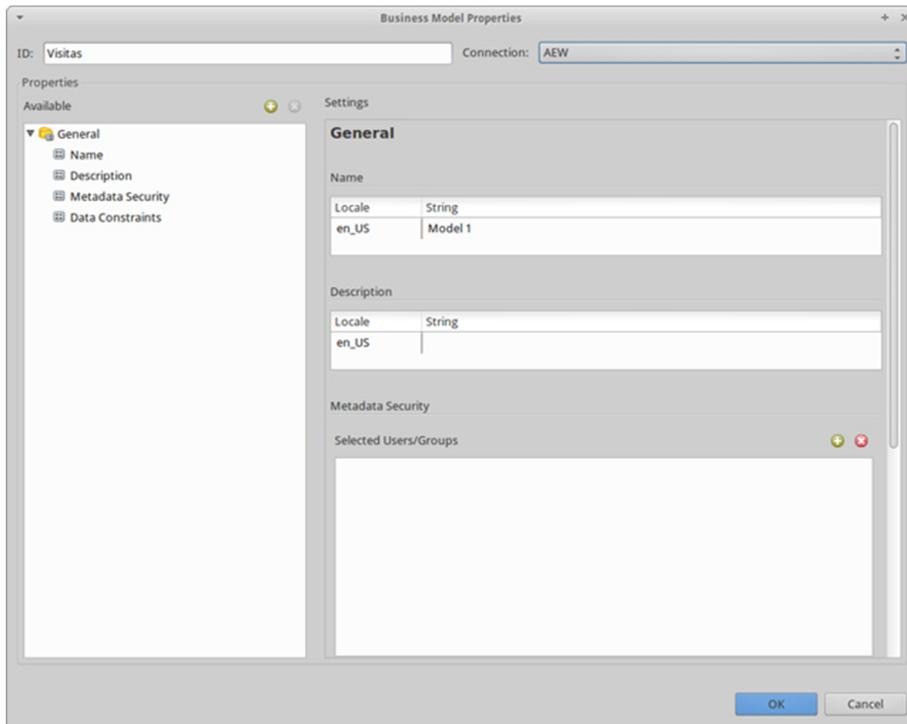
Figura 12. Creación de modelo de negocio



Fuente: Josep Curto.

Configuramos el modelo. Como mínimo, la conexión y el nombre. Aunque podríamos definir usuarios, idioma y restricciones de acceso.

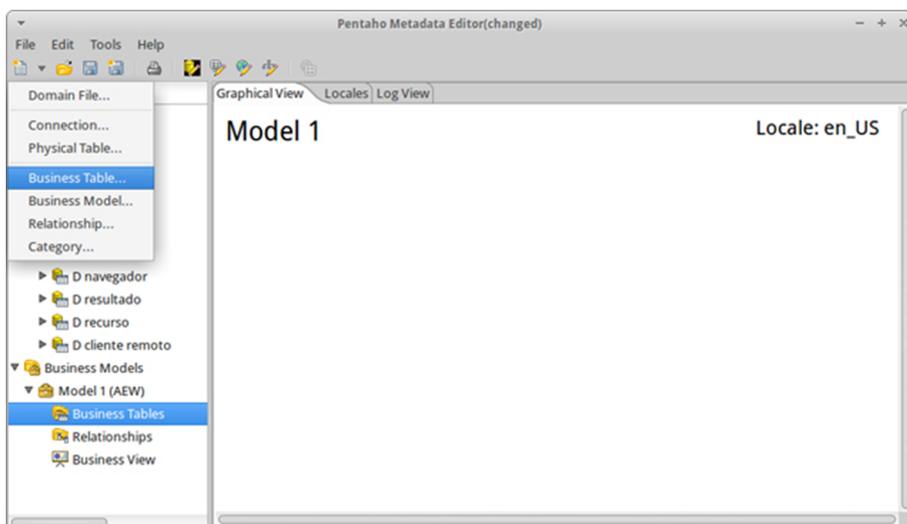
Figura 13. Propiedades del modelo



Fuente: Josep Curto.

Ahora es necesario crear las tablas de negocio, que consiste en el proceso de mapear las tablas físicas de la conexión con los nombres de negocio que usar tanto en el nombre de la tabla, como en los atributos de esta.

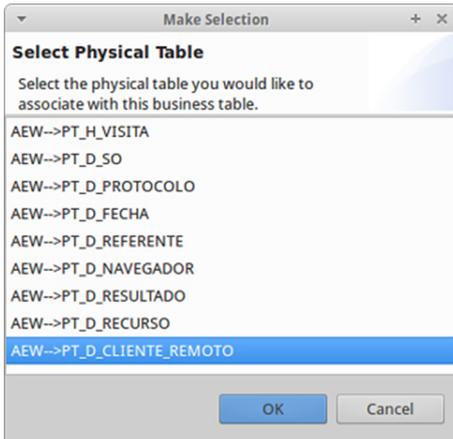
Figura 14. Creación de tablas del modelo de negocio



Fuente: Josep Curto.

Seleccionamos la tabla de cliente remoto.

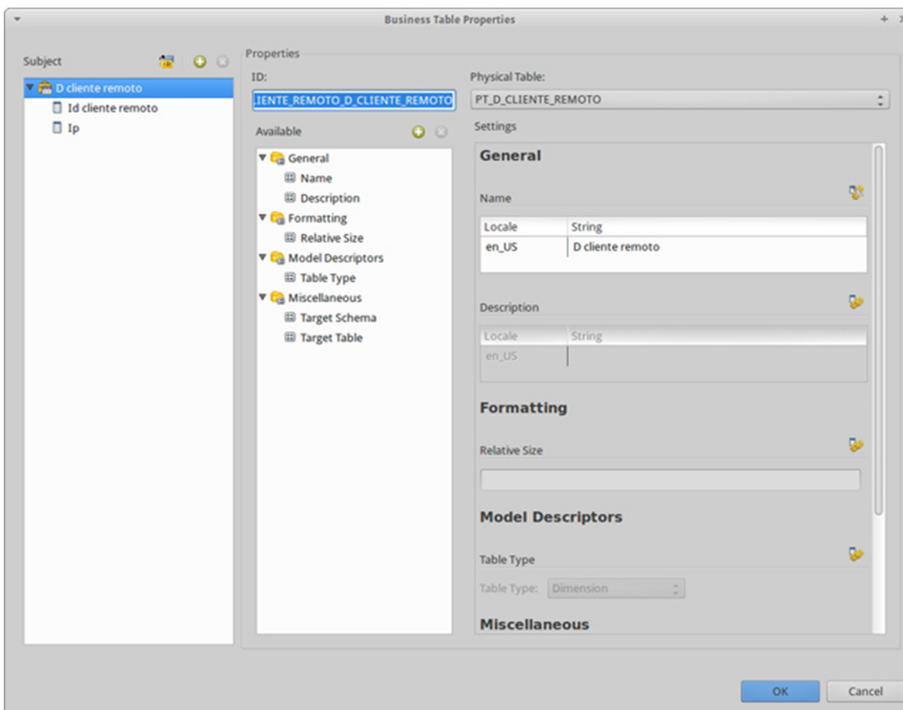
Figura 15. Selección de tablas



Fuente: Josep Curto.

Damos el nombre adecuado tanto a la tabla como a sus campos. Podemos definir otras características pero para nuestro ejemplo nos centramos solo en el nombre.

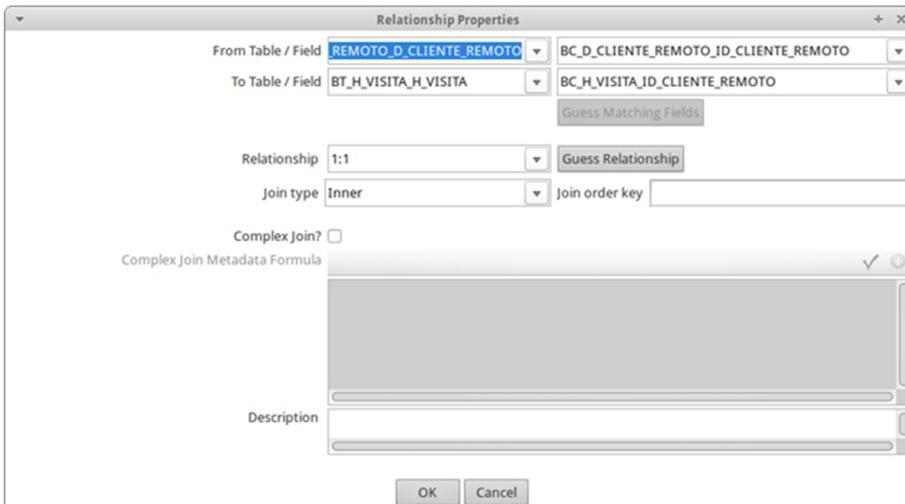
Figura 16. Propiedades de la tabla



Fuente: Josep Curto.

Este proceso se debe realizar para todas las tablas que han sido exportadas del modelo y posteriormente crear las relaciones de negocio entre ellas para que el modelo esté bien definido.

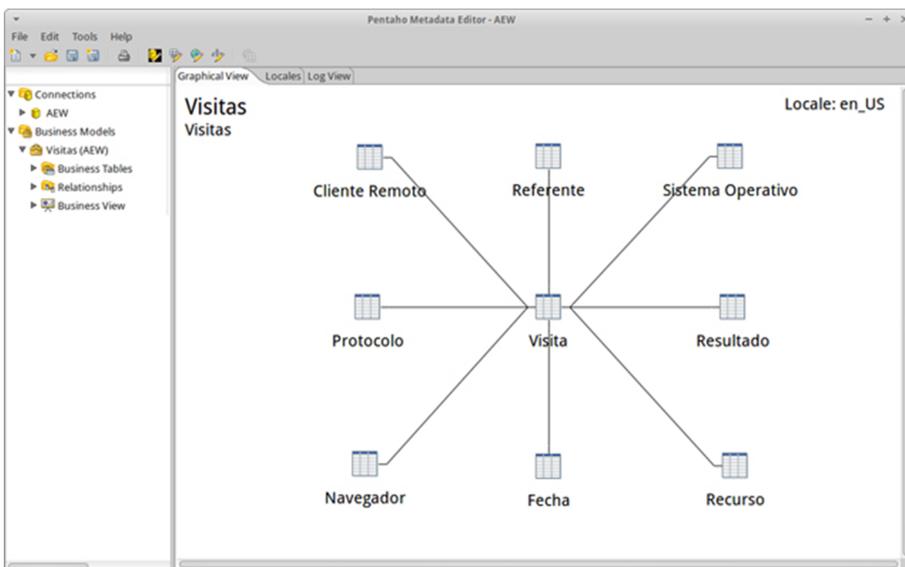
Figura 17. Vinculación de dimensión y tabla de hecho



Fuente: Josep Curto.

El resultado final de crear todas las tablas en la vista de negocio y sus relaciones es el siguiente.

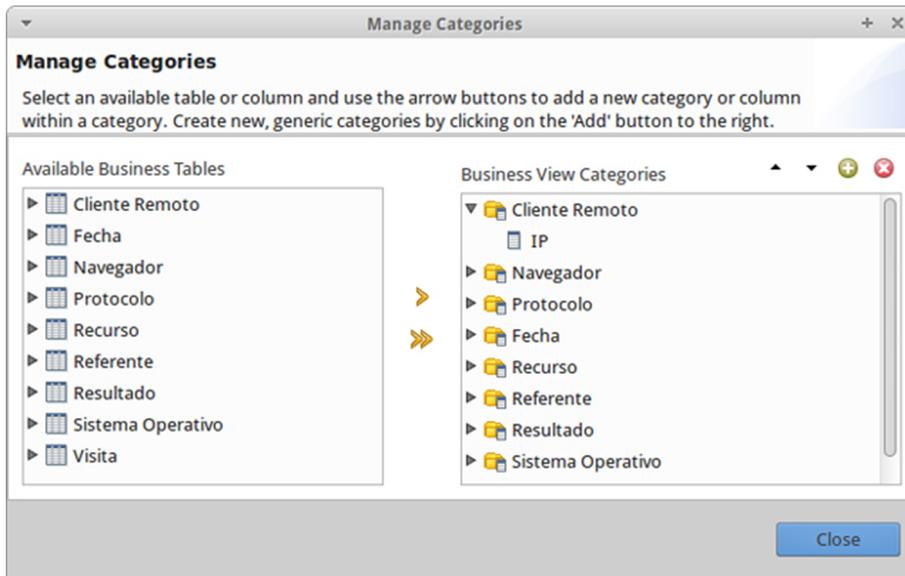
Figura 18. Modelo de negocio



Fuente: Josep Curto.

Sin embargo, no es suficiente con tener el modelo de negocio y sus relaciones. Es necesario definir las categorías disponibles en la capa de metadatos al usuario final. En el ejemplo, el usuario solo podrá acceder y usar el atributo IP de cliente remoto.

Figura 19. Gestión de categorías

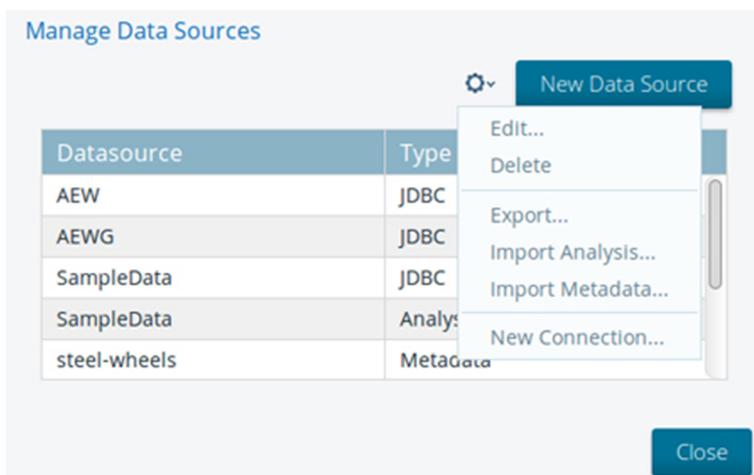


Fuente: Josep Curto.

Una vez terminado el proceso, guardamos el fichero y publicamos la capa de metadatos. Es necesario que el servidor de Pentaho esté en funcionamiento, puesto que el proceso de publicación se realiza a través del servidor.

Una vez encendido el servidor, vamos a la sección *Manage Data Sources* y seleccionamos *Import Metadata*.

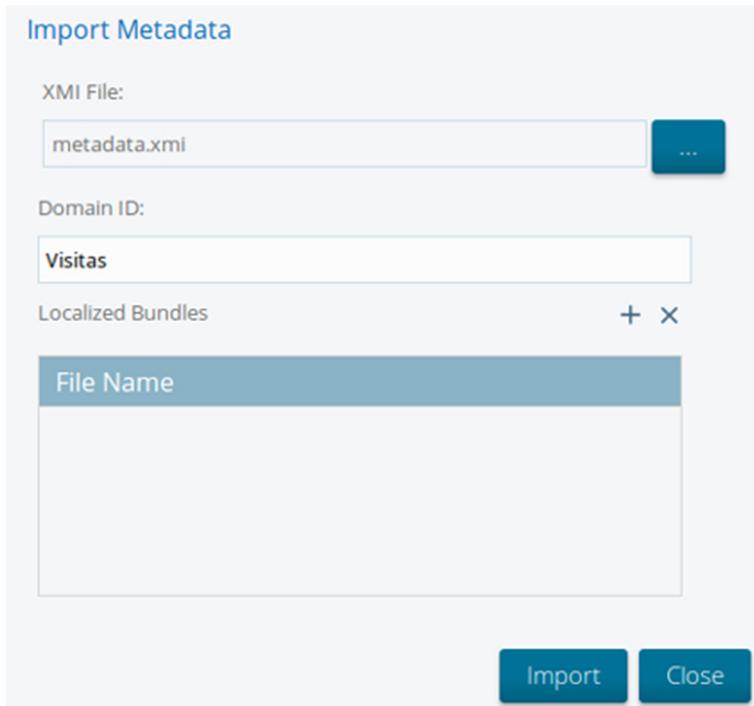
Figura 20. Importación del esquema de metadatos



Fuente: Josep Curto.

Seleccionamos el fichero y le damos un nombre a la capa de metadatos.

Figura 21. Publicación del esquema de metadatos



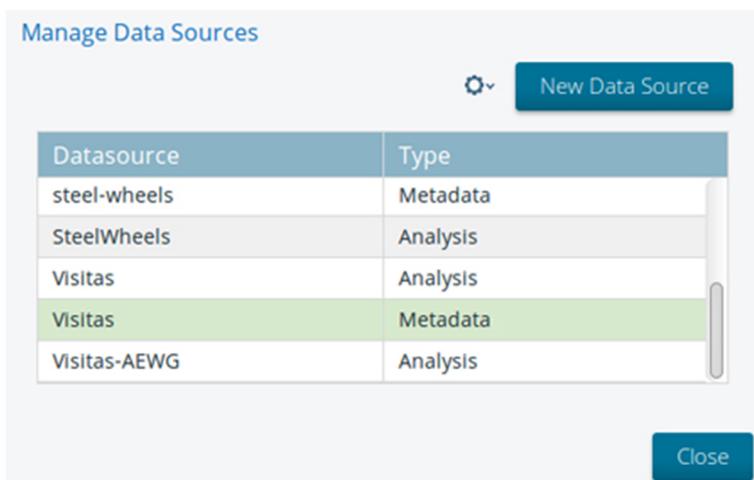
The screenshot shows a dialog box titled "Import Metadata". It contains the following elements:

- XMI File:** A text input field containing "metadata.xml" and a blue button with three dots to its right.
- Domain ID:** A text input field containing "Visitas".
- Localized Bundles:** A section with a "+ X" icon and an empty table with a header "File Name".
- Buttons:** "Import" and "Close" buttons at the bottom right.

Fuente: Josep Curto.

Una vez importado, tendremos acceso a la capa de metadatos.

Figura 22. Esquema de metadatos publicado



The screenshot shows a dialog box titled "Manage Data Sources". It contains the following elements:

- Buttons:** A gear icon and a "New Data Source" button at the top right.
- Table:** A table with two columns: "Datasource" and "Type".
- Table Content:**

Datasource	Type
steel-wheels	Metadata
SteelWheels	Analysis
Visitas	Analysis
Visitas	Metadata
Visitas-AEWG	Analysis

- Buttons:** A "Close" button at the bottom right.

Fuente: Josep Curto.

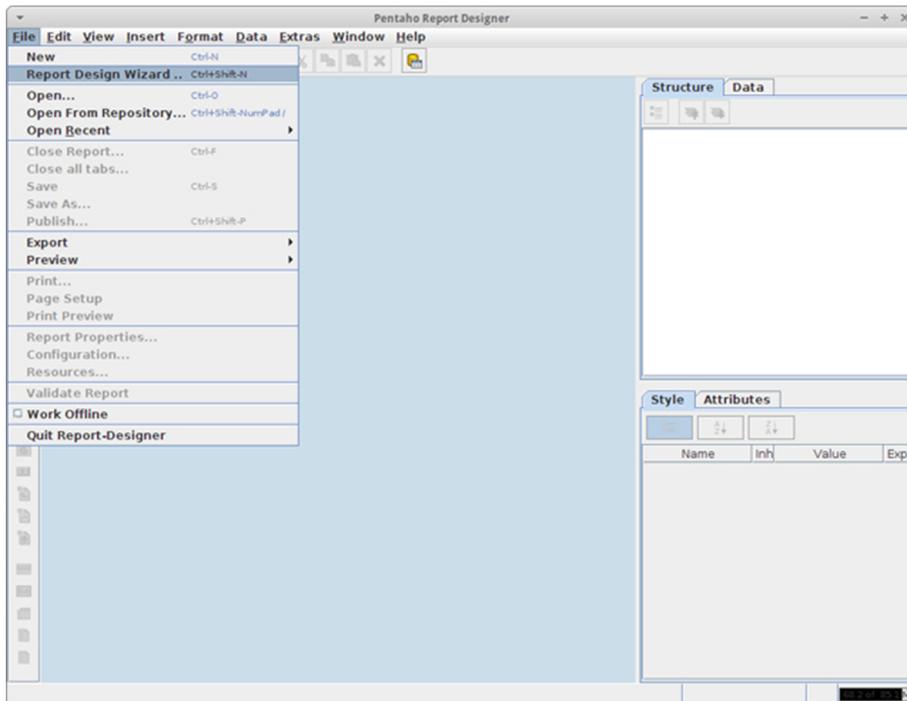
3.2. Diseño de un informe mediante el *wizard* en Pentaho

Pentaho Report Designer (PRD) ofrece dos posibles formas para crear informes:

- Mediante el *wizard*.
- En formato libre.

Iniciamos el PRD mediante el fichero report-designer.sh. Elegimos crear un informe mediante el *wizard*.

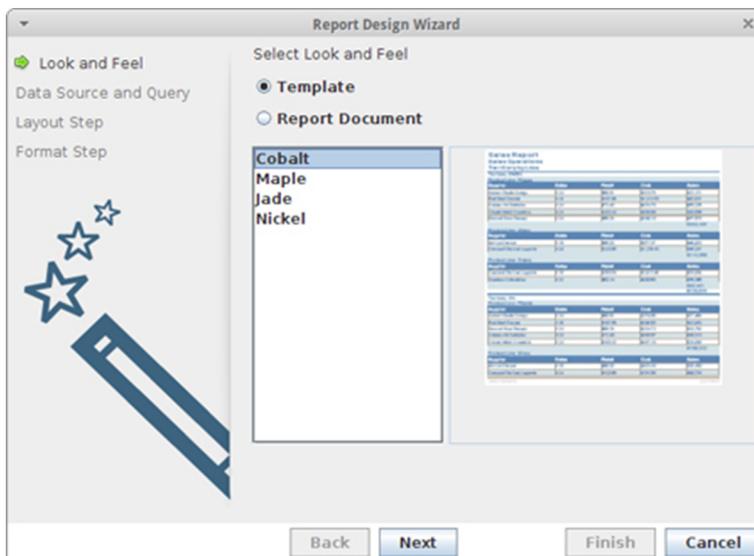
Figura 23. Iniciar PRDW



Fuente: Josep Curto.

La herramienta nos guía por el proceso de creación del informe. El primer paso es elegir una de las plantillas disponibles. Elegimos, por ejemplo, *cobalt*.

Figura 24. Elección de plantilla



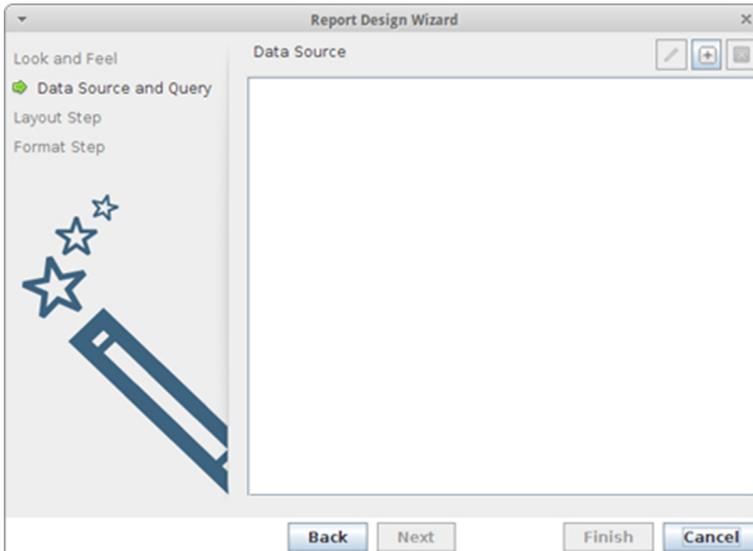
Fuente: Josep Curto.

El siguiente paso es definir la fuente de datos y la consulta. En el caso de haber creado otros informes con anterioridad, habría un listado de consultas.

Plantillas

El conjunto de plantillas puede extenderse, por ejemplo, para tener una plantilla corporativa.

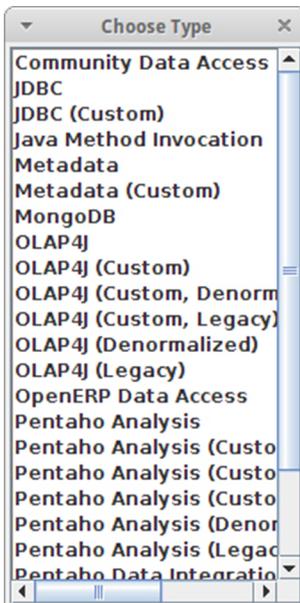
Figura 25. Elección de consulta y fuente de datos



Fuente: Josep Curto.

Pulsamos el botón + en la esquina superior y se abrirá un nuevo menú que mostrará los tipos de fuentes de datos disponibles. Están soportadas múltiples entradas de datos (desde la capa de metadatos hasta base de datos pasando por OLAP). Tenemos a nuestra disposición tres: base de datos, OLAP y metadatos. Elegimos la primera y nos conectaremos mediante JDBC.

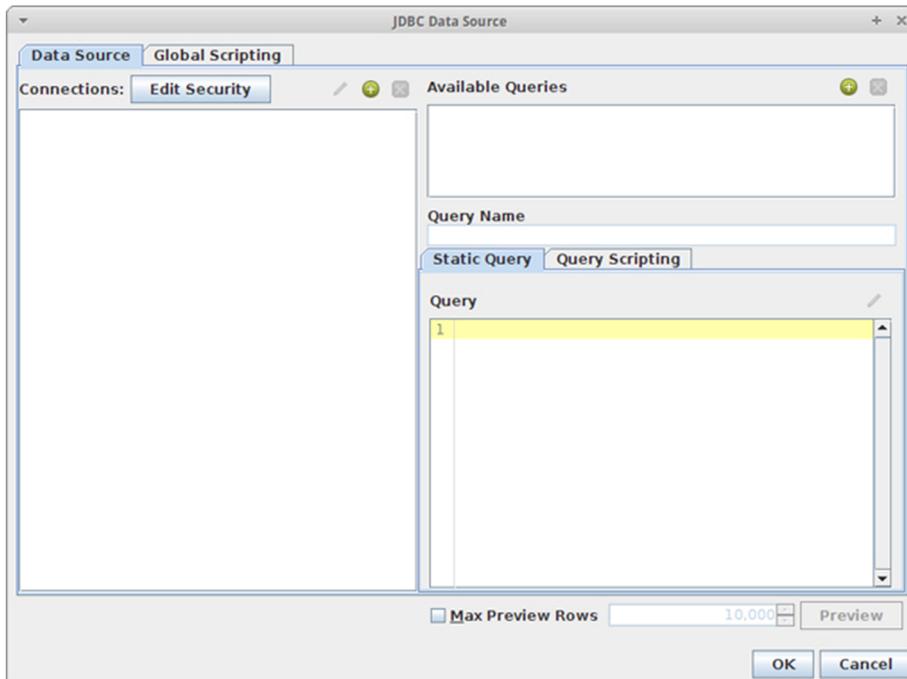
Figura 26. Fuente de datos disponibles



Fuente: Josep Curto.

Inicialmente no hay disponible ninguna conexión. A medida que trabajamos con la herramienta, van guardándose las conexiones disponibles y las consultas.

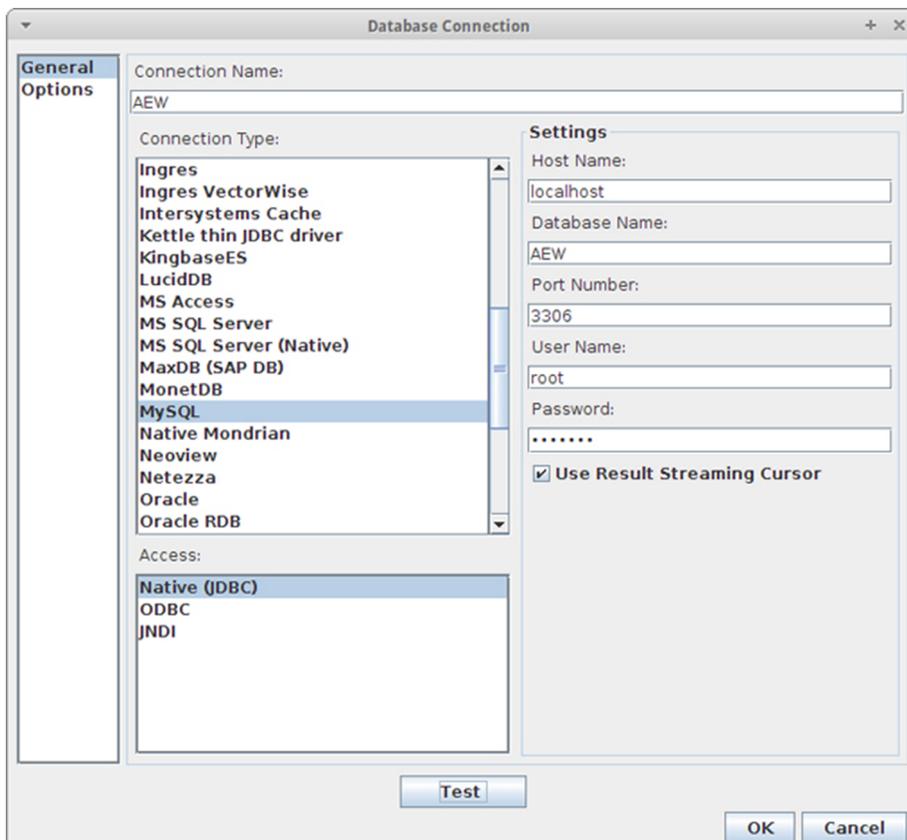
Figura 27. Interfaz de consultas y conexiones



Fuente: Josep Curto.

Necesitamos definir la conexión y el conjunto de datos que será usado en nuestro informe. De nuevo completamos un menú similar para definir la conexión de base de datos.

Figura 28. Conexión con AEW



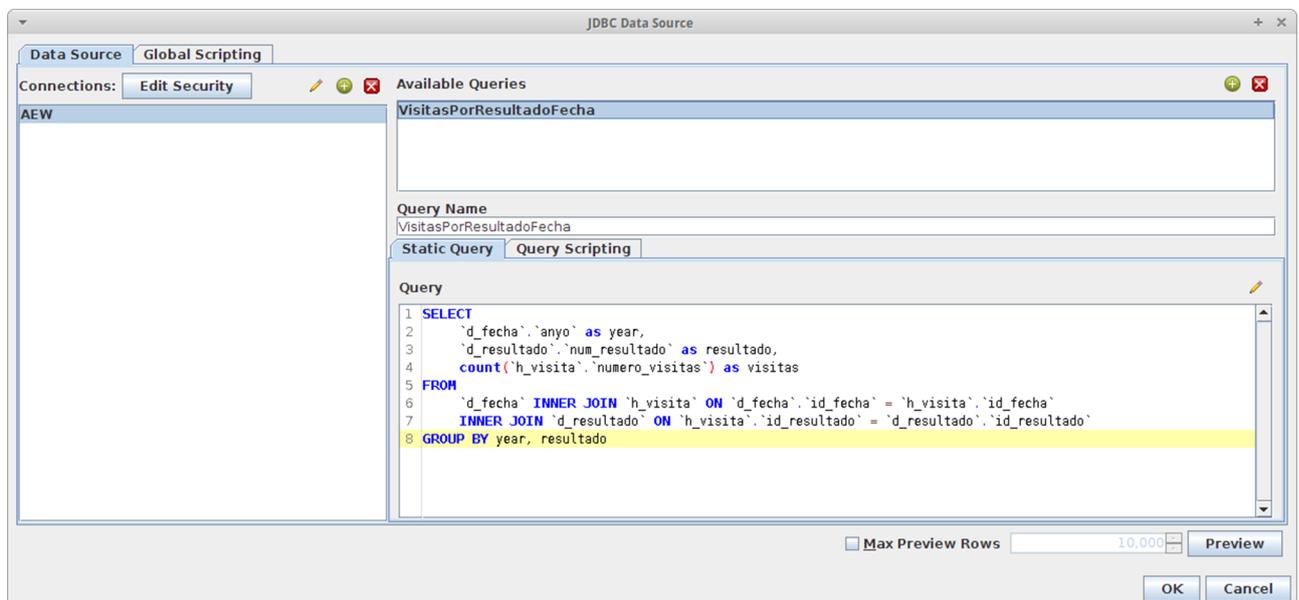
Fuente: Josep Curto.

En este caso, creamos una consulta para recuperar las visitas agrupadas por resultado y año. La consulta es:

```
SELECT
    'd_fecha`.`anyo` as year,
    'd_resultado`.`num_resultado` as resultado,
    count('h_visita`.`numero_visitas`) as visitas
FROM
    'd_fecha` INNER JOIN 'h_visita` ON 'd_fecha`.`id_fecha` = 'h_visita`.`id_fecha`
    INNER JOIN 'd_resultado` ON 'h_visita`.`id_resultado` = 'd_resultado`.`id_resultado`
GROUP BY year, resultado
```

Esta consulta la podemos construir mediante la aplicación o simplemente escribiéndola en espacio disponible tal y como ilustra la siguiente imagen:

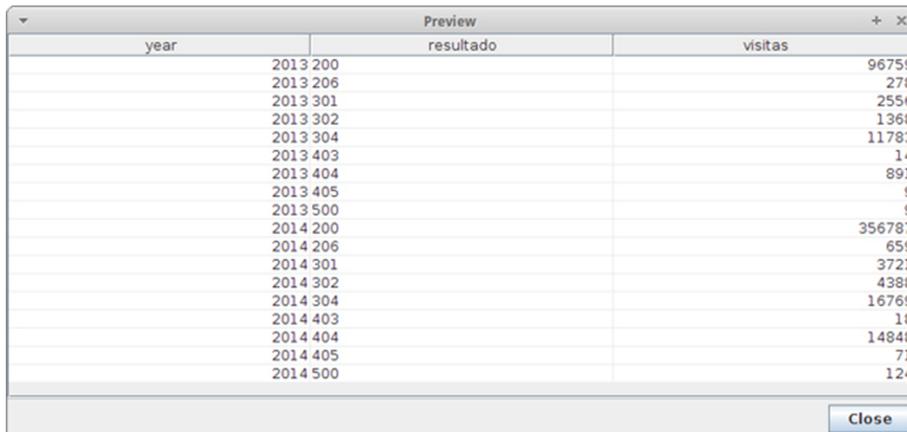
Figura 29. Consulta SQL



Fuente: Josep Curto.

Podemos visualizar la consulta para comprobar que los resultados son correctos.

Figura 30. Visualización de la consulta

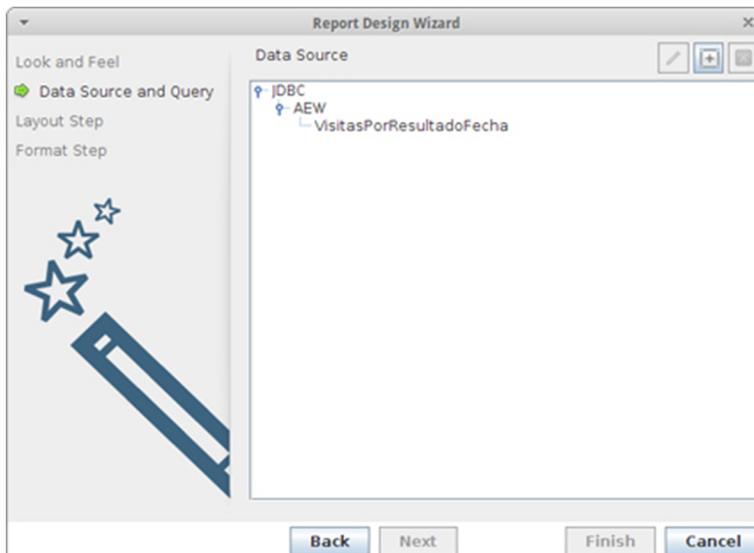


year	resultado	visitas
2013 200		96759
2013 206		278
2013 301		2556
2013 302		1368
2013 304		11783
2013 403		14
2013 404		891
2013 405		9
2013 500		9
2014 200		356787
2014 206		659
2014 301		3721
2014 302		4388
2014 304		16769
2014 403		18
2014 404		14848
2014 405		71
2014 500		124

Fuente: Josep Curto.

Una vez definida la conexión y la consulta, las tenemos disponibles para nuestro informe.

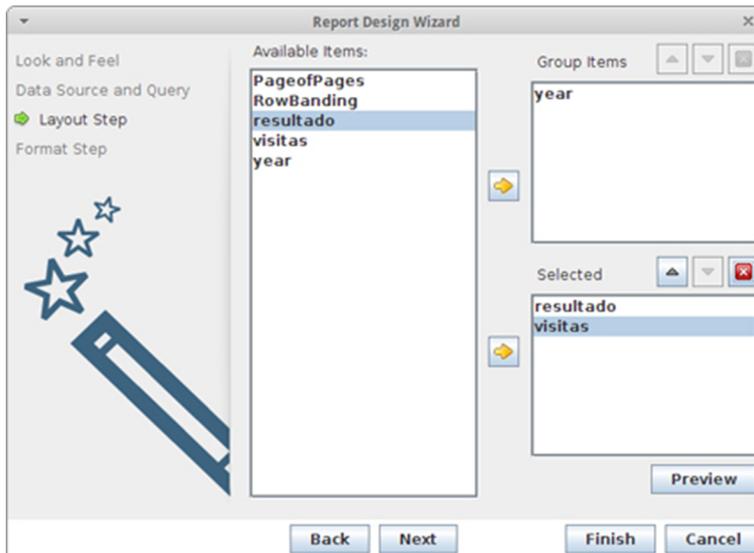
Figura 31. Selección de la consulta



Fuente: Josep Curto.

El siguiente paso es definir el nivel de agregación. Por ejemplo, en nuestro caso decidimos que vamos a mostrar agregar la información por año y que el detalle será las visitas por resultados.

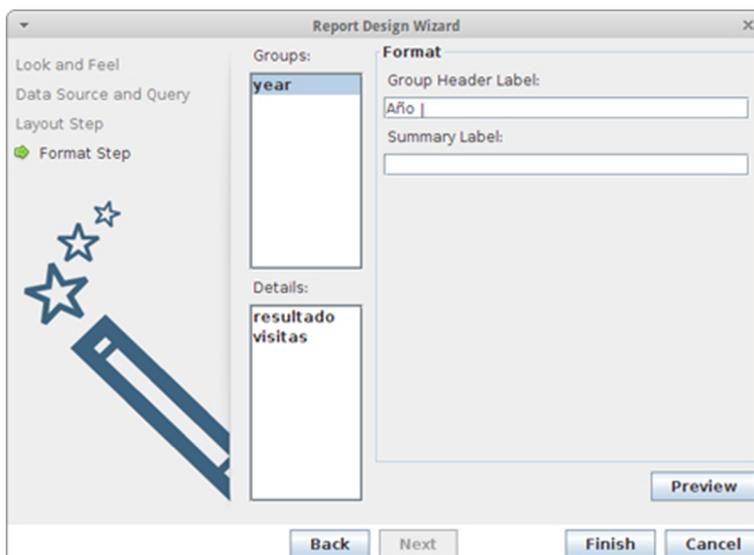
Figura 32. Definición nivel agregación



Fuente: Josep Curto.

El paso final es ajustar detalles estéticos por cada campo.

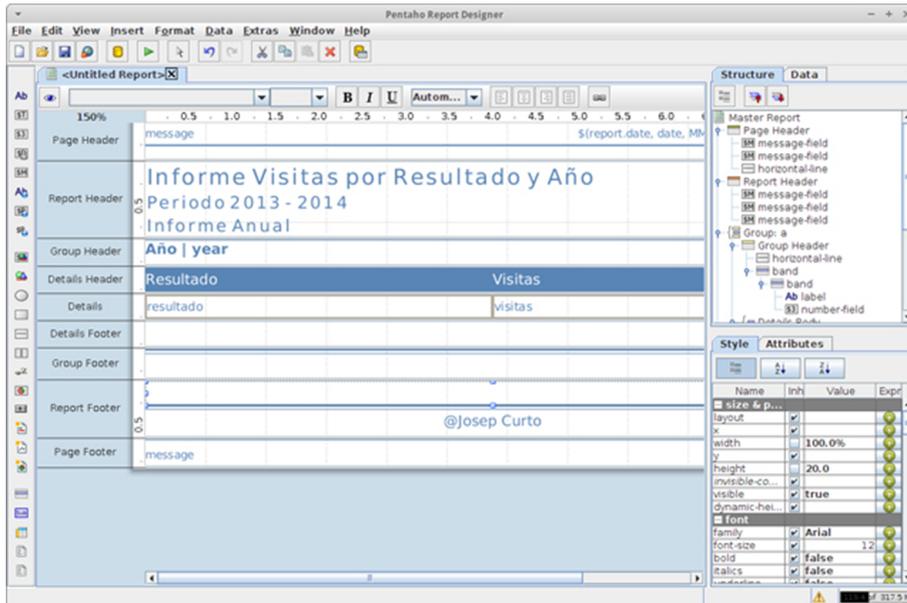
Figura 33. Ajustar detalles estéticos



Fuente: Josep Curto.

Y finalmente, ajustamos los títulos, subtítulos y el pie de página.

Figura 34. Ajustar finales



Fuente: Josep Curto.

Podemos visualizar el informe y lo guardamos para posteriores ejecuciones.

Figura 35. Informe visitas por resultado y año

August 09, 2015 @ 10:30

Informe Visitas por Resultado y Año
Periodo 2013 - 2014
Informe Anual

Año 2,013	
Resultado	Visitas
200	96,759
206	278
301	2,556
302	1,368
304	11,783
403	14
404	891
405	9
500	9

Año 2,014	
Resultado	Visitas
200	356,787
206	659
301	3,721
302	4,388
304	16,769
403	18
404	14,848
405	71
500	124

©Josep Curto

Sun Aug 09 10:30:09 EST 2015

1/1

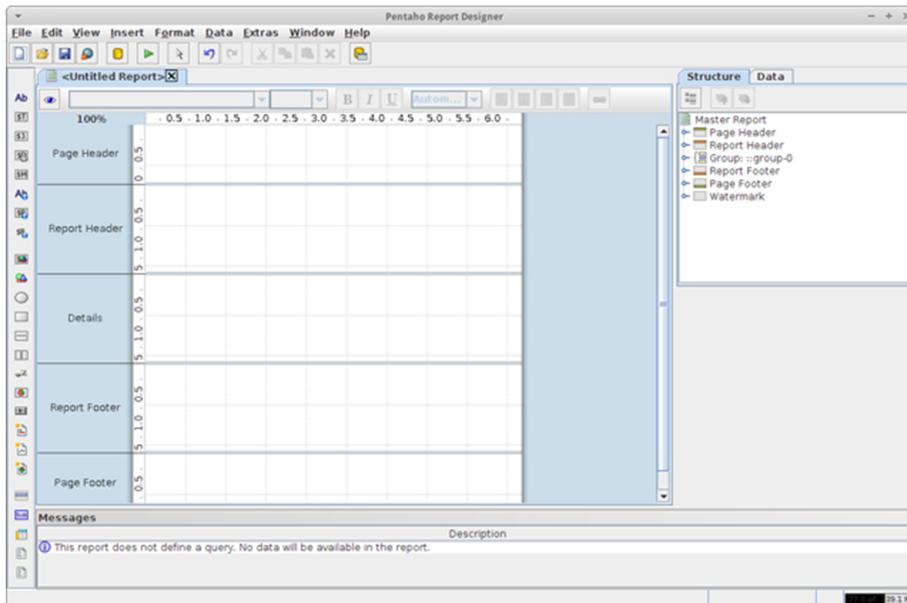
Fuente: Josep Curto.

3.3. Diseño de un informe mediante Pentaho Report Designer

Aunque el *wizard* permite crear informes, básicamente estos son listados. Para crear informes más sofisticados (con gráfico, *subreports*, etc.), es aconsejable usar Pentaho Report Designer. Vamos por lo tanto a crear un informe desde cero con esta herramienta.

Un está divididos en bandas que debemos ir completando progresivamente.

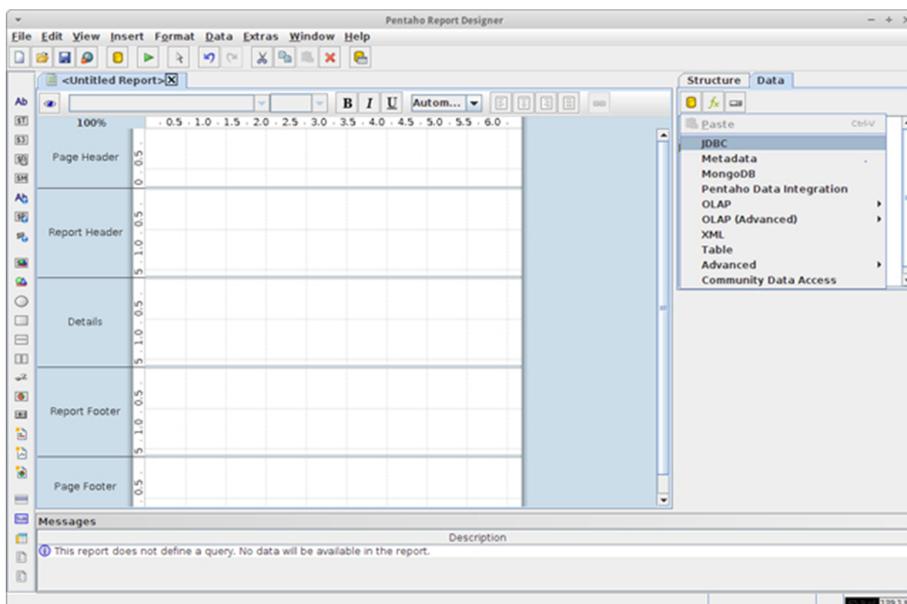
Figura 36. Estructura de bandas



Fuente: Josep Curto.

El primer paso es crear la consulta vinculada al informe. Debemos, como en el caso anterior, añadir una conexión (AEW) y crear la consulta.

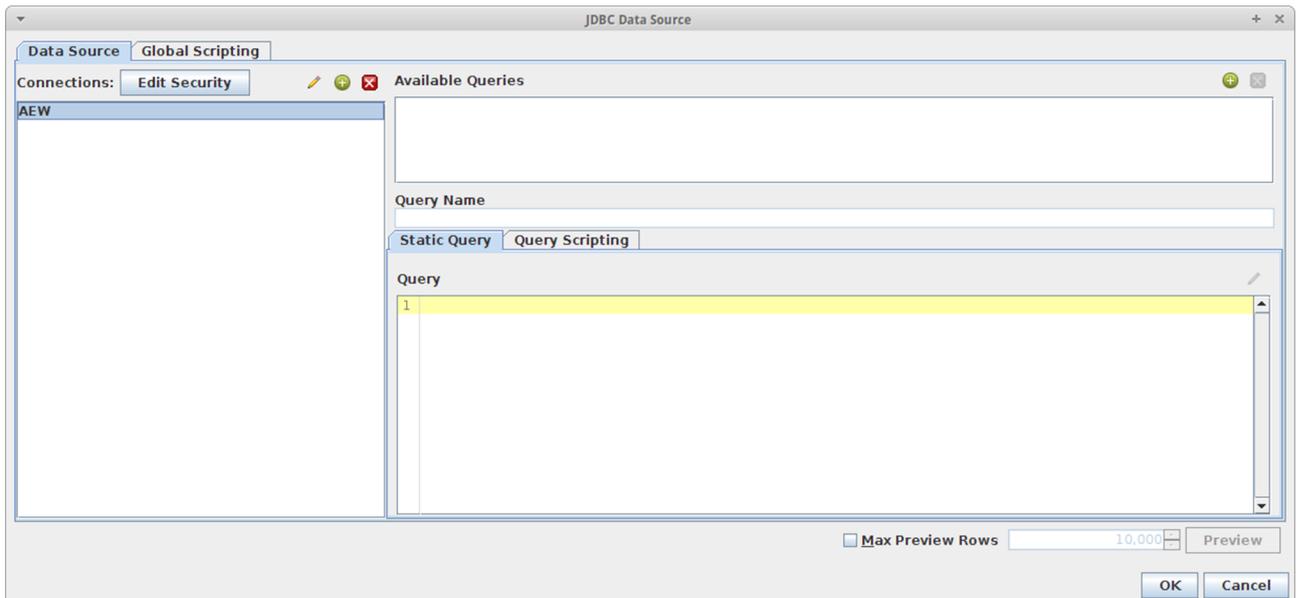
Figura 37. Creación de nueva fuente de datos



Fuente: Josep Curto.

Para ello, creamos una nueva fuente de datos (*data source*) mediante JDBC. Puesto que ya hemos creado anteriormente la conexión AEW, la tendremos a nuestra disposición.

Figura 38. Espacio de creación de consultas

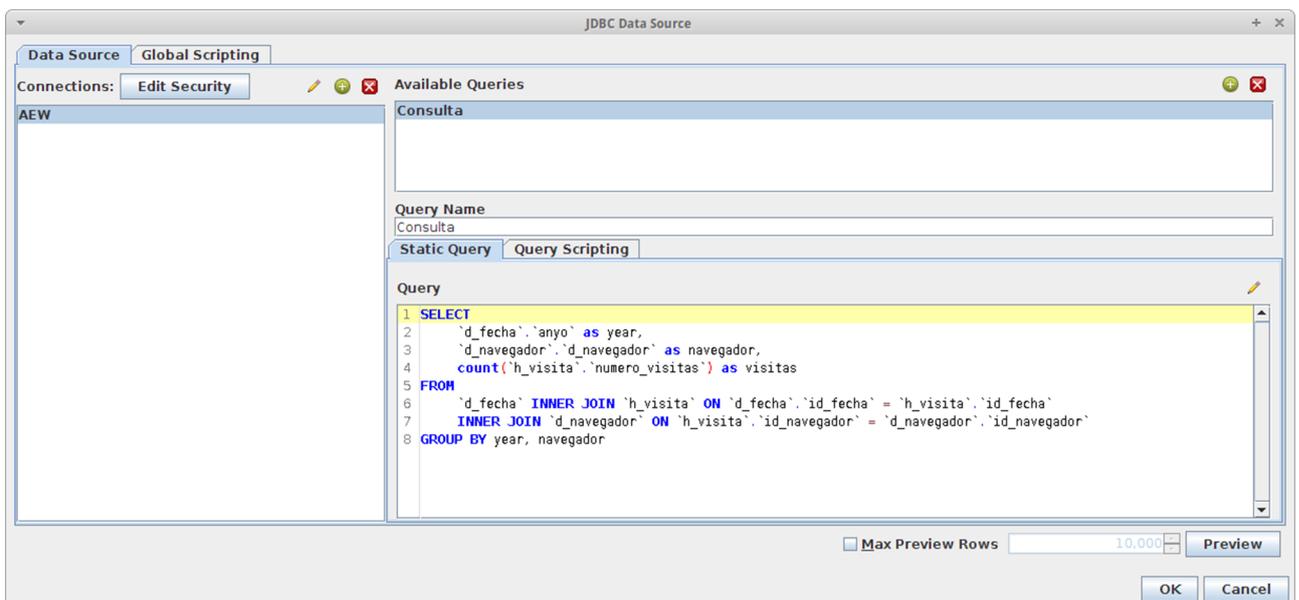


Fuente: Josep Curto.

En este caso, creamos una consulta para recuperar las visitas agrupadas por navegador y año. La consulta es:

```
SELECT
    'd_fecha'.'anyo' as year,
    'd_navegador'.'d_navegador' as navegador,
    count('h_visita'.'numero_visitas') as visitas
FROM
    'd_fecha' INNER JOIN 'h_visita' ON 'd_fecha'.'id_fecha' = 'h_visita'.'id_fecha'
    INNER JOIN 'd_navegador' ON 'h_visita'.'id_navegador' = 'd_navegador'.'id_navegador'
GROUP BY year, navegador
```

Figura 39. Consulta de visitas por navegador y año



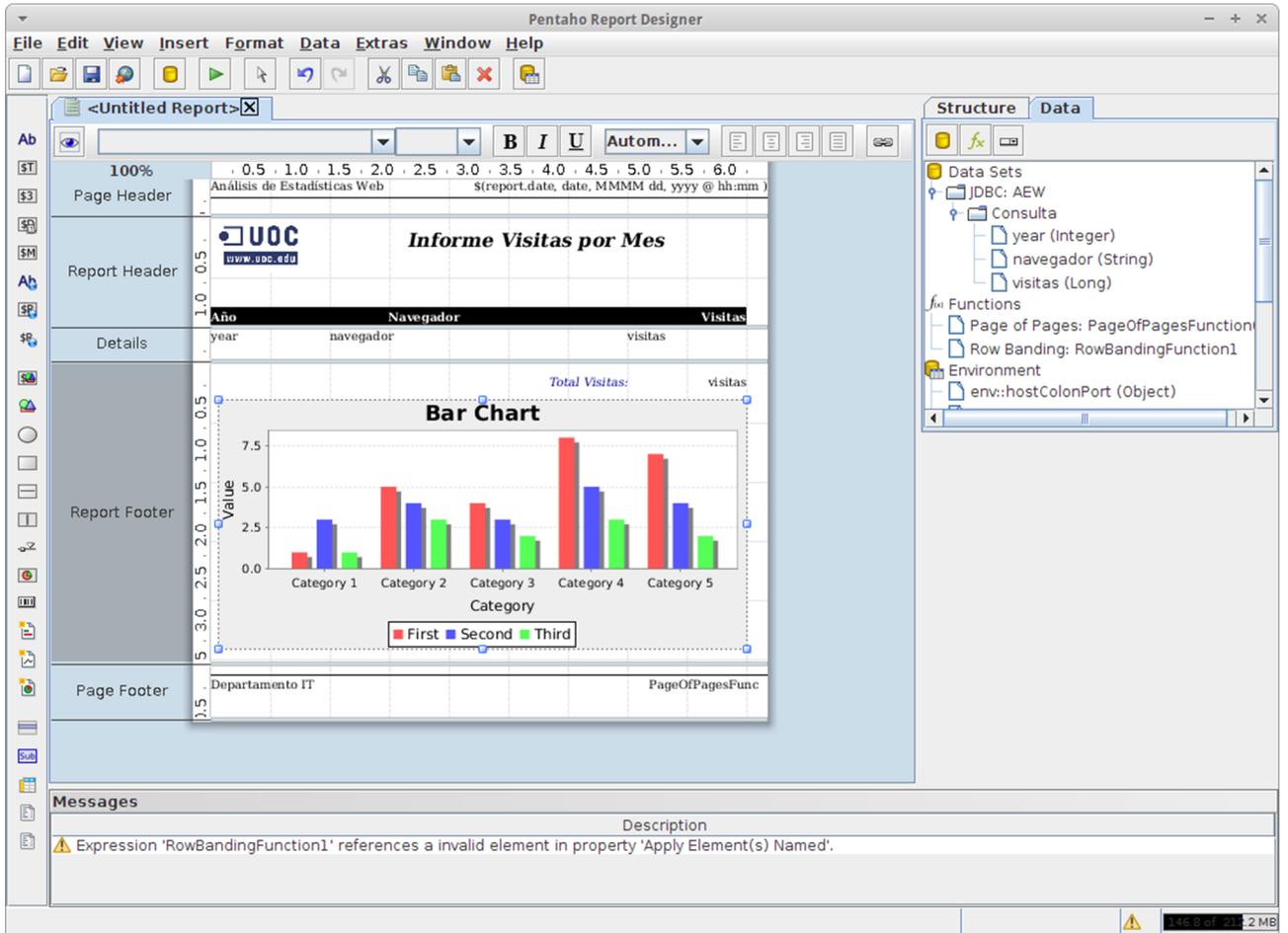
Fuente: Josep Curto.

Ahora podemos incluir todos los elementos que queremos en el informe. Estos elementos se añaden a partir del menú lateral de elementos:

- Un *label* con una referencia a la temática del informe (análisis de estadísticas web) en *page header*.
- En la misma banda añadimos fecha de ejecución del informe y una línea horizontal. Además, reducimos el tamaño de la banda.
- En la banda de *report header* añadimos el título del informe y el logo de la UOC.
- En el *page footer* añadimos quién ha hecho el informe (por ejemplo, el departamento IT), el número de páginas (que es una función que se debe crear primero en el lateral, en la pestaña de datos) y una línea horizontal.
- En la banda *details*, vamos a considerar un ejemplo fácil de crear un listado, lo que hacemos es arrastrar los tres campos de la consulta y alinearlos horizontalmente.
- Creamos también unas etiquetas para cada uno de los campos en la banda *report header*.
- Para enriquecer este informe respecto el anteriormente creado, incluimos un gráfico de barras (completamos *chart-title*, *category-column*, *value-columns* y *series-by-field* como mínimo) para representar las visitas por años.

De esta manera, tendremos el siguiente informe.

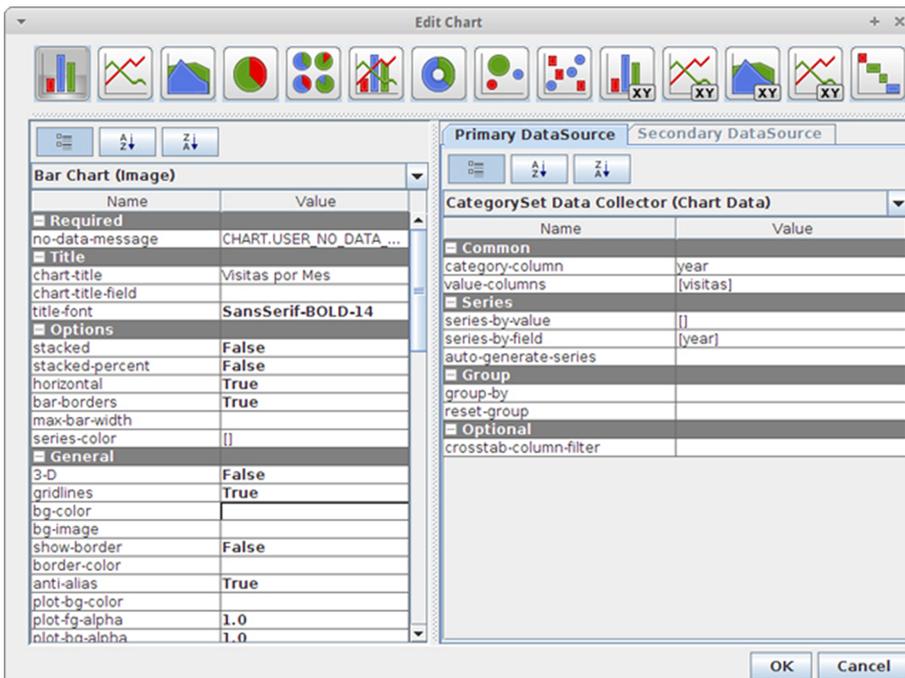
Figura 40. Informe final



Fuente: Josep Curto.

En la siguiente imagen adjuntamos el detalle de configuración del gráfico.

Figura 41. Parámetros gráfico



Fuente: Josep Curto.

El resultado final es un informe de tres páginas con un gráfico al final.

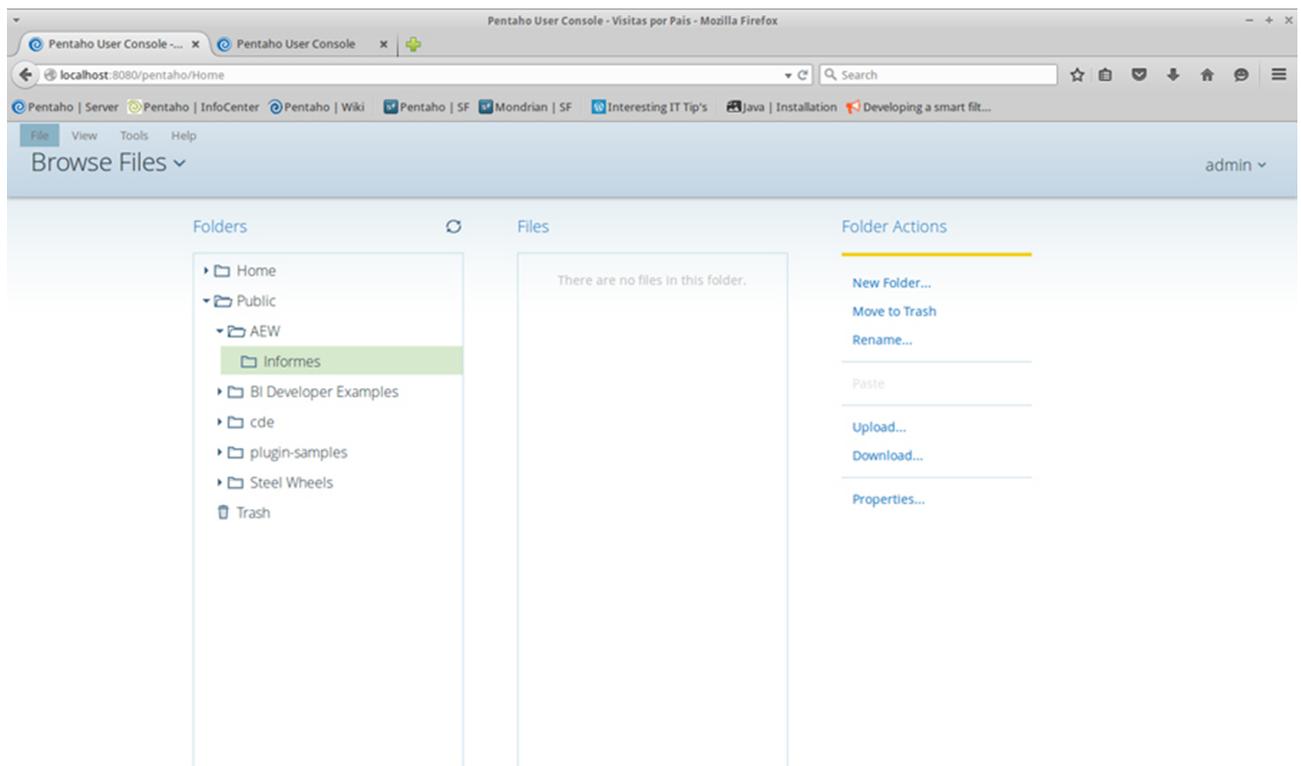
3.4. Publicación de informes en el servidor

En los anteriores apartados hemos revisado cómo crear un informe. Consideremos el primer informe que hemos creado (y al que hemos nombrado como ejemplo1.prpt) y vamos ahora a publicarlo para que esté disponible para todos los usuarios.

Este proceso se realiza de nuevo teniendo en cuenta el servidor. Vamos a la sección Browse. Accedemos a una estructura de soluciones (carpetas) para gestionar proyectos en Pentaho Server. Es posible apreciar que hay diversas soluciones ya publicadas (la propia de ejemplo, Steel Wheels), así como otras que buscan ayudar al neófito en la solución de desarrollar nuevos proyectos.

Aprovechamos para crear una nueva carpeta llamada AEW en la parte pública (lo que significa que es accesible a todos los usuarios) y dentro de ella otra carpeta llamada informes.

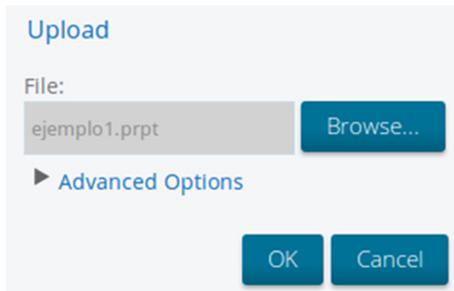
Figura 42. Estructura de soluciones en Pentaho



Fuente: Josep Curto.

Pulsamos Upload y seleccionamos el informe que hemos creado previamente.

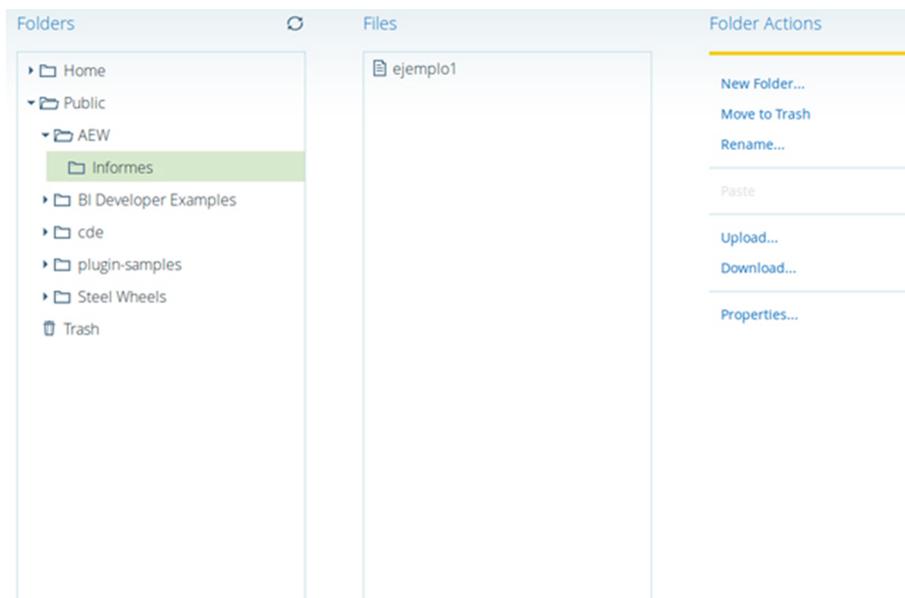
Figura 43. Proceso de publicación de un objeto



Fuente: Josep Curto.

Una vez pulsemos OK, tendremos el informe disponible.

Figura 44. Informe disponible



Fuente: Josep Curto.

Si pulsamos en el informe, tendremos la opción de abrirlo, lo que hará que se ejecute el informe en el servidor.

Figura 45. Informe ejecutado en el servidor

The screenshot shows a web browser window displaying a report from the Pentaho User Console. The report is titled "Informe Visitas por Resultado y Año" and covers the period "Período 2013 - 2014". It is an annual report ("Informe Anual") for the year 2013. The data is presented in a table with two columns: "Resultado" (Result) and "Visitas" (Visits). The results are as follows:

Resultado	Visitas
200	96,759
206	278
301	2,556
302	1,368
304	11,783
403	14
404	891
405	9
500	9

The interface also shows a "View Report" button and a checked "Auto-Submit" option. The report is displayed in a paginated HTML format.

Fuente: Josep Curto.

Abreviaturas

CSV *Comma separated value.*

CWM *Common warehouse metamodel.*

EJB *Enterprise javabeans.*

EPS *Encapsulated postscript.*

HTML *Hypertext markup language.*

JDBC *Java database connection.*

MDX *Multidimensional expressions.*

ODBC *Open database connectivity.*

ODS *Operational data store.*

OLAP *Online analytical processing.*

PDF *Portable document format.*

PNG *Portable network graphics.*

POJO *Plain old java object.*

PSV *Pipe separated value.*

RTF *Rich format text.*

SQL *Structured query language.*

SSV *Semi-colon separated value.*

SVG *Scalable vector graphics.*

TSV *Tabular separated value.*

XML *Extensible markup language.*

WAQR *Web ad-hoc query reporting.*

Bibliografía

Bouman, R.; Van Dongen, J. (2009). *Pentaho® Solutions: Business Intelligence and Data Warehousing with Pentaho® and MySQL*. Indianápolis: Wiley Publishing.

García Mattío, M.; Bernabeu, D. R. (2013/2009). *Pentaho Reporting 5.0 by Example: a begginer's guide*. Birmingham: Packt Publishing.

Patil, M. R.; Ramazzina, S. (2015). *Pentaho Reporting 5 for Java Developers* (2.^a ed.). Birmingham: Packt Publishing.