

Guía Básica de Instalación y Uso del Servidor de Inteligencia de Negocio Pentaho Business Analytics Server 8/9.X

Autores: Alejandro Maté Morgia
Juan Carlos Trujillo Mondéjar

Fecha: Febrero 2021



Universitat d'Alacant
Universidad de Alicante





1. Introducción

El servidor de Inteligencia de Negocio de Pentaho BA (Business Analytics) Community¹ es una aplicación de código abierto, aunque desde hace algún tiempo existe también una versión comercial denominada Pentaho EE (Enterprise Edition)². Pentaho incluye integrado en su instalación el servidor de cubos OLAP Mondrian y se ejecuta sobre un servidor de aplicaciones Tomcat. En esta guía veremos cómo realizar la instalación más sencilla de Pentaho, utilizando la versión preconfigurada que se puede descargar de la web. Esta versión incluye ya configurados un Tomcat y una base de datos Hypersonic³ para los distintos repositorios de datos que existen en Pentaho. Tales repositorios de datos son:

- **Jackrabbit:** almacena los archivos de las aplicaciones de Inteligencia de Negocio que desarrollemos. Hasta la llegada de la versión 5 este repositorio no existía y los archivos se gestionaban usando el sistema de archivos propio del sistema operativo.
- **Quartz:** almacena los datos relacionados con generación programada de informes (reporting) y la ejecución programada de procesos ETL (Extracción, Transformación y Carga de datos)
- **Hibernate:** almacena los datos relacionados con la auditoria (logging)

Entre las tareas que se pueden realizar con el servidor de Inteligencia de Negocio se incluye el diseño de cuadros de mando, la generación y publicación de informes, la ejecución programada de procesos ETL, la inserción de anotaciones y el análisis ad-hoc de cubos. Durante el curso nos centraremos únicamente en esta última tarea para validar el almacén de datos que diseñemos.

2. Descarga e instalación de Pentaho

Para comenzar es necesario tener instalado y configurado Java 1.8 o superior. A continuación, puede descargarse el software Pentaho Business Analytic Server 8.X/9.X (Community Edition) en la siguiente dirección:

<https://sourceforge.net/projects/pentaho/files/Business%20Intelligence%20Server/7.1/pentaho-server-ce-7.1.0.0-12.zip/download>

Versión X.2:

<https://sourceforge.net/projects/pentaho/files/Pentaho%208.2/server/pentaho-server-ce-8.2.0.0-342.zip/download>

<https://sourceforge.net/projects/pentaho/>

¹ Versión de código abierto de Pentaho BA Server y el resto de herramientas de Pentaho <https://community.hds.com/docs/DOC-1009931-downloads>

² Versión comercial de la suite Pentaho <http://www.pentaho.com/>

³ Es posible configurar una base de datos distinta para la gestión de los repositorios (ej. MySQL), siendo recomendado en entornos de producción.



Pentaho Community Edition 7.1

Business Analytics Platform

Pentaho's simplified, and interactive approach empowers business users to access, discover and blend any data types regardless of their size.

[All OS](#)

Data Integration

Pentaho's Data Integration, also known as Kettle, delivers powerful extraction, transformation, and loading (ETL) capabilities.

[All OS](#)

Report Designer

The Report Designer is a graphical tool that generates reports from data streamed through the Data Integration engine.

[Windows / Linux](#)

[Mac OS X](#)

[Other artifacts](#)

Design Tools

Aggregation Designer

The Aggregation Designer provides a simple interface that allows you to create and deploy aggregate tables.

[All OS](#)

Schema Workbench

The Schema Workbench is a visual design interface that allows you to create and test Mondrian OLAP cube schemas.

[All OS](#)

Metadata Editor

Metadata Editor is a tool that simplifies your experience when creating reports, and allows you to build metadata domains and relational data models.

[All OS](#)

Figura 1. Repositorio online para la descarga de Pentaho 7.1

Al extraer el archivo, obtendremos el directorio **/pentaho-server** cuyo contenido podemos ver a continuación:


















Nombre	Fecha de modifica...	Tipo	Tamaño
 data	16/05/2017 21:25	Carpeta de archivos	
 licenses	16/05/2017 21:25	Carpeta de archivos	
 pentaho-solutions	16/05/2017 21:25	Carpeta de archivos	
 third-party-tools	16/05/2017 21:26	Carpeta de archivos	
 tomcat	16/05/2017 21:19	Carpeta de archivos	
 import-export.bat	16/05/2017 21:19	Archivo por lotes ...	1 KB
 import-export.sh	16/05/2017 21:19	Shell Script	1 KB
 promptuser.js	16/05/2017 21:19	Archivo JavaScript	2 KB
 promptuser.sh	16/05/2017 21:19	Shell Script	1 KB
 set-pentaho-env.bat	16/05/2017 21:19	Archivo por lotes ...	5 KB
 set-pentaho-env.sh	16/05/2017 21:19	Shell Script	4 KB
 start-pentaho.bat	16/05/2017 21:19	Archivo por lotes ...	2 KB
 start-pentaho.sh	16/05/2017 21:19	Shell Script	2 KB
 start-pentaho-debug.bat	16/05/2017 21:19	Archivo por lotes ...	2 KB
 start-pentaho-debug.sh	16/05/2017 21:19	Shell Script	2 KB
 stop-pentaho.bat	16/05/2017 21:19	Archivo por lotes ...	1 KB
 stop-pentaho.sh	16/05/2017 21:19	Shell Script	1 KB

Figura 2. Directorio pentaho-server



En esta carpeta se encuentran todos los archivos del servidor de Inteligencia de Negocio. Anteriormente a las versiones 5.X, la consola de administración se encontraba separada de la consola de usuario, pero ahora se encuentran juntas en una única consola (aplicación web). Para arrancar dicha consola hemos de ejecutar el script **start-pentaho.bat** si nuestro SO es Windows (**start-pentaho.sh** en el caso de SO's basados en Linux). Al ejecutar este script arrancará el servidor Tomcat y podremos acceder al servidor Pentaho en la dirección **http://localhost:8080/pentaho/**. Cuando deseemos detener el servidor (asegurarnos de haber guardado cualquier archivo que estuviésemos editando en Pentaho), hemos de ejecutar el script **stop-pentaho**. Más adelante detallaremos el acceso y las distintas opciones disponibles en la consola del servidor Pentaho.

Es importante tener en cuenta que para poder ejecutar Pentaho ha de estar definida la variable de entorno **JAVA_HOME** la cual debe indicar la ruta de la instalación de Java. Pentaho ejecuta por defecto un script que busca la ruta y define dicha variable. No obstante, en algún caso podría no funcionar e impedir de esta forma el correcto arranque del servidor. En ese caso, definir dicha variable de forma manual o usando un script como "for /d %%i in ("%Program Files%\Java\jdk*") do set JAVA_HOME=%%i" para asignar el **JAVA_HOME** a la versión más reciente de java.

La subcarpetas más relevantes dentro la carpeta **/pentaho-server** son i) la carpeta **tomcat**, que contiene el servidor de aplicaciones sobre el que se ejecutará Pentaho, y ii) la carpeta **pentaho-solutions**, que previamente a las versiones 5.x se usaba para almacenar los proyectos de Inteligencia de Negocio que se desarrollaban y ahora contiene los archivos de sistema.

En la carpeta **tomcat/lib** se encuentran los conectores (de tipo JDBC) de las bases de datos que usaremos como fuente de nuestras vistas de análisis multidimensionales (análisis ad-hoc de cubos), cuadros de mando o informes. Por defecto incluye algunos como el de MySQL. Pero si el driver de nuestra base de datos no se encuentra en esta carpeta o corresponde a una versión distinta, bastará con añadir o sustituir el correspondiente driver JDBC en dicha carpeta.

La otra carpeta relevante es **pentaho-solutions**. En las versiones anteriores a la versión 5.X de Pentaho esta carpeta almacenaba todos los archivos de nuestras aplicaciones creadas para el servidor de Inteligencia de Negocio. Sin embargo, ahora tales archivos se almacenan y gestionan en el repositorio Jackrabbit por lo que no es posible tener acceso directamente desde el SO, si no que hemos de usar una API. No obstante, es posible instalar una aplicación llamada **Repository Synchronizer** desde el Marketplace de Pentaho (la "tienda de aplicaciones", también de código abierto) que permite sincronizar los archivos del repositorio con los de la carpeta **pentaho-solutions**, ahora inicialmente vacía. No vamos a entrar en detalle en el uso del repositorio Jackrabbit y la aplicación Repository Synchronizer, pues en principio no vamos a necesitar estos conocimientos para el desarrollo de las prácticas.

3. Consola del servidor de Inteligencia de Negocio

A continuación se explicarán algunos de los aspectos básicos para comenzar a utilizar Pentaho BA Server. Antes de comenzar el primer paso es arrancar el servidor con el ya comentado script **start-pentaho.bat** (start-pentaho.sh en Linux). Haciendo clic sobre él, se abrirá una nueva consola del sistema donde podremos ver el log de la ejecución del servidor.

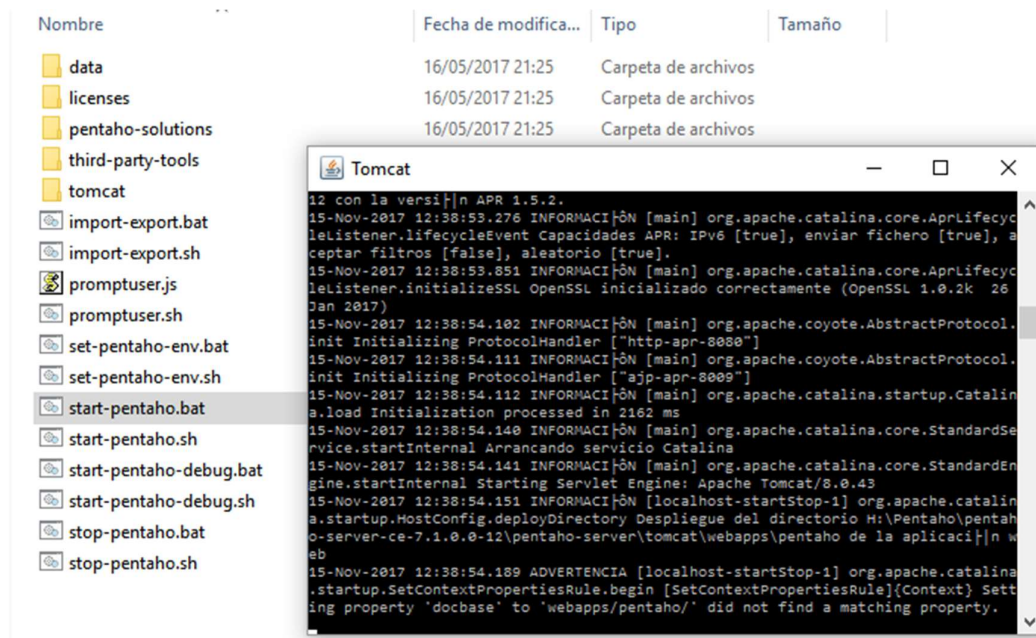


Figura 3. Arranque de Pentaho Server

Esperamos algunos segundos y ya podremos acceder a la consola desde cualquiera de los navegadores web instalados en nuestro equipo.

a. Acceso a la consola y secciones.

Podemos acceder a la consola en la dirección <http://localhost:8080/pentaho>. El primer paso es hacer login. Pentaho trae por defecto algunos usuarios, como es el caso del usuario **admin** y clave **password**. Usaremos este usuario para conectar la primera vez.

Una vez hayamos hecho login correctamente, tendremos ante nosotros la página inicial de la consola del servidor.

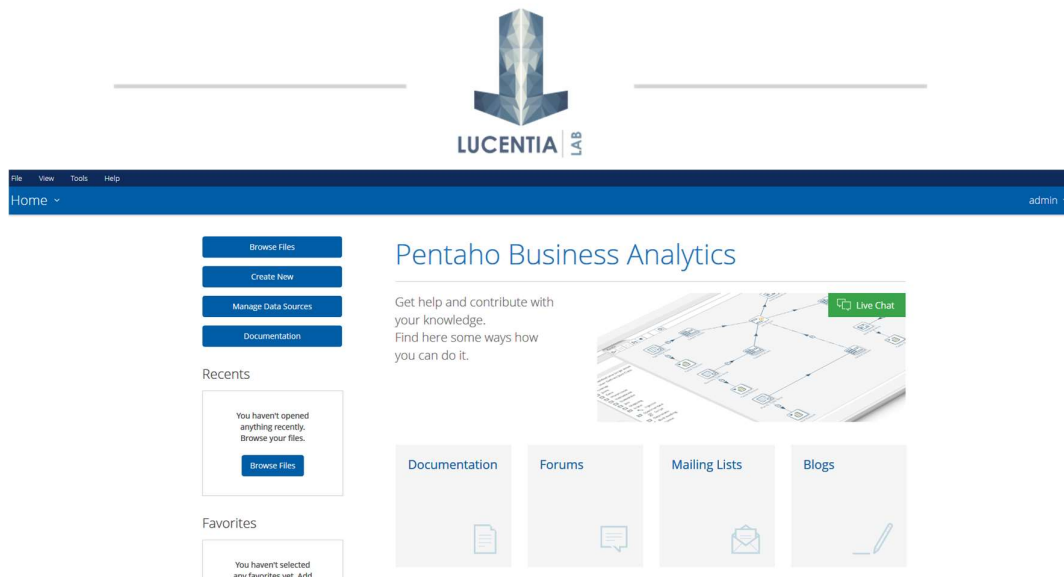



Figura 4. Página de inicio de la consola de Pentaho Server 7

En la parte superior tenemos un menú desplegable (donde está seleccionado inicialmente Home) donde podemos acceder las distintas secciones de la aplicación. Estas son:

- **Home:** Página principal de la aplicación. Accesos rápidos a otras secciones de la aplicación, archivos recientes y enlaces a páginas relevantes del mundo de Pentaho como tutoriales, documentación,...
- **Browse Files:** Permite gestionar las carpetas que contienen los archivos de nuestros proyectos de Inteligencia de Negocio. Es posible subir y descargar archivos desde y hasta nuestro PC respectivamente. El botón “refresh”  permite actualizar los directorios mostrados cuando se ejecuta alguna operación sobre los archivos de los mismos.
- **Recent:** Archivos de la aplicación abiertos recientemente vistas de análisis (análisis de cubos ad-hoc), visualización y edición de cuadros de mando o informes. Podemos tener más de un archivo abierto al mismo tiempo y movernos entre las distintas secciones de la aplicación sin necesidad de cerrarlos previamente.
- **Marketplace (dentro de Home desde la versión 6.X):** Permite instalar y actualizar plugins desarrollados por el equipo de Pentaho u otros usuarios de la comunidad de Pentaho (Pentaho Community). Saiku Analytics y Community Dashboard Editor (CDE), son interesantes aplicaciones para la creación de vistas de análisis avanzadas y cuadros de mando respectivamente.
- **Administration (dentro de Home desde la versión 6.X):** Gestión de los usuarios del servidor (ej. admin) y roles que es posible asignar a los mismos (ej. business analyst)

b. Ejemplo de creación de una de vista de análisis multidimensional (análisis ad-hoc de cubos)

Una vez conocemos los aspectos fundamentales del servidor de Inteligencia de Negocio Pentaho, podemos llevar a cabo un sencillo ejemplo que consiste en la creación de una vista de análisis multidimensional para cubo OLAP en un Almacén de Datos (Data Warehouse, DW) que incluye Pentaho como ejemplo.

Una vista de análisis nos permite llevar a cabo el análisis o explotación de un cubo OLAP sobre Almacén de Datos previamente definido e implementado, de forma que podamos extraer conocimiento de tal análisis para dar soporte a la toma de decisiones estratégicas en una organización. Este es el objetivo final de las aplicaciones de Inteligencia de Negocio.

En Pentaho BA Server la implementación de los cubos OLAP se hace mediante la herramienta Mondrian, la cual veremos más adelante en la asignatura y para la que disponemos de otro manual. De momento solo es necesario entender que esta aplicación es la que permite la definición de un cubo OLAP sobre un Almacén de Datos en Pentaho, así como la ejecución y gestión de las consultas que se realicen sobre el mismo.

El Almacén de Datos de ejemplo que incluye Pentaho y que usaremos recibe el nombre de “Steel Wheels” y contiene datos históricos de ventas de productos de modelismo: maquetas de coches, motos, trenes...

Para crear una vista de análisis sobre la base de datos de ejemplo SteelWheels la opción más sencilla es seleccionar la opción **Archivo→New→JPivot View** en el menú superior de la consola de la aplicación.

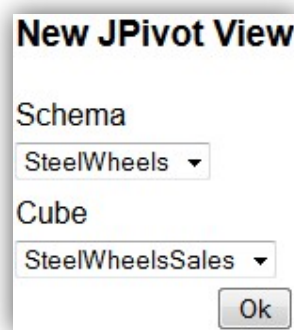


Figura 5. Selección del esquema para realizar análisis ad-hoc

Basta con seleccionar el esquema que deseamos analizar y hacer clic en el botón "OK". En nuestro caso seleccionamos el esquema del DW SteelWheel y seleccionamos el cubo OLAP SteelWheelSales.



En el caso de que acabásemos de definir y publicar un nuevo esquema en Pentaho con la aplicación Mondrian, sería necesario hacer clic en la opción **Herramientas→Actualizar→Vaciar el caché de esquema de Mondrian**. Para el ejemplo de SteelWheel no es necesario ya que este cubo viene por defecto en Pentaho.

Tras seleccionar un esquema y cubo, pulsamos Ok y aparecerá el cubo de análisis. Para navegar por el cubo basta con hacer clic en los signos "+" y "-" (Roll-Up y Drill-Down) para agrupar la información o entrar en detalle. Si hacemos clic en el botón del cubo en la parte superior izquierda también aparecerán otras opciones, tales como el movimiento de elementos entre filas y columnas (Pivoting) o la aplicación de filtros específicos (Slice&Dice).

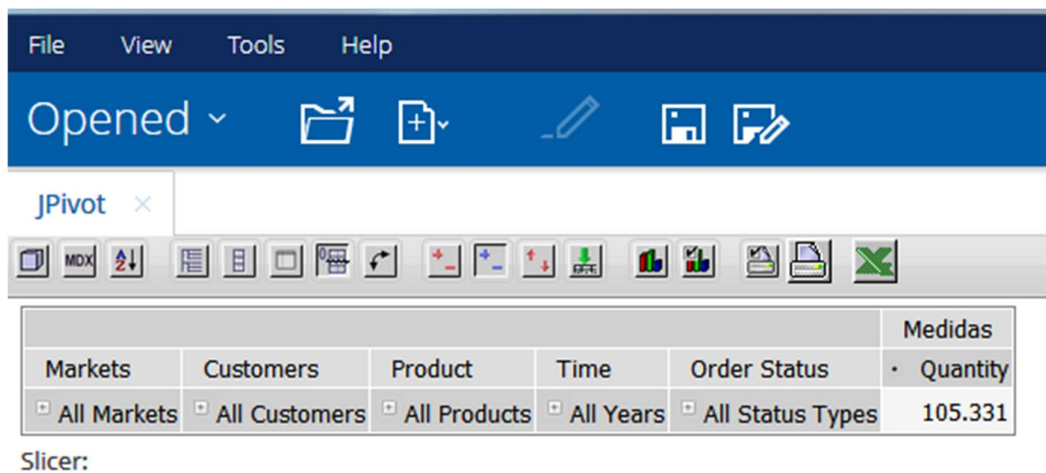


Figura 6. Pantalla de análisis ad-hoc mediante JPivot

Dado que JPivot se incluye únicamente como herramienta de muestra con funcionalidad limitada, se recomienda conectarse al Marketplace e instalar:

- Pivot4J.** La herramienta más sencilla de instalar pero también la más limitada en funcionalidad. El desarrollo se encuentra parado pero tiene mucha mayor funcionalidad que JPivot
- Saiku Analytics.** La versión community está dotada de una gran funcionalidad, pero resulta ligeramente complicada de instalar por la necesidad de instalar una licencia community que debe generarse en una página propia de Saiku.
- STPivot.** Herramienta similar a Saiku en funcionalidad que podríamos considerar el sucesor de Pivot4J. Ofrece gran funcionalidad y no requiere licencia, pero debe descargarse desde Github, compilarse e instalarse manualmente dentro de pentaho-solutions, siguiendo las instrucciones en: <https://github.com/stratebi/STPivot4>