

Pentaho BI

Introducción

1.- ¿Qué es un BI?

El Business Intelligence o Inteligencia Empresarial engloba los métodos y técnicas a través de las cuales se pueden transformar los datos en conocimiento. Éste, extrae los datos de diferentes fuentes para unificarlos y analizarlos para su posterior interpretación en la toma de decisiones.

2.- Fases de la implementación del BI

- 1) Definición de objetivos y prioridades
- 2) Recolección de la información mediante análisis exhaustivo
- 3) Análisis de datos y medición de resultados
- 4) Colaboración mediante la difusión de resultados
- 5) Revisión y aprovechamiento de las herramientas disponibles
- 6) Identificación e implementación de las herramientas para BI

3- Características del BI

- Open source
- Compatibilidad con múltiples bases de datos
- Muestreo de datos en tiempo real
- Autoservicio
- Fácil importación

4.- Software libre Pentaho

Pentaho BI Suite es una plataforma de BI que incluye un conjunto de programas libres útiles para la inteligencia empresarial, que permite realizar informes, minería de datos, etcétera.

Está orientada a la solución centrada en procesos, ofreciendo los componentes necesarios para su implementación. Estas soluciones se componen de una infraestructura de herramientas de análisis e informes junto a un motor de workflow de procesos de negocio.

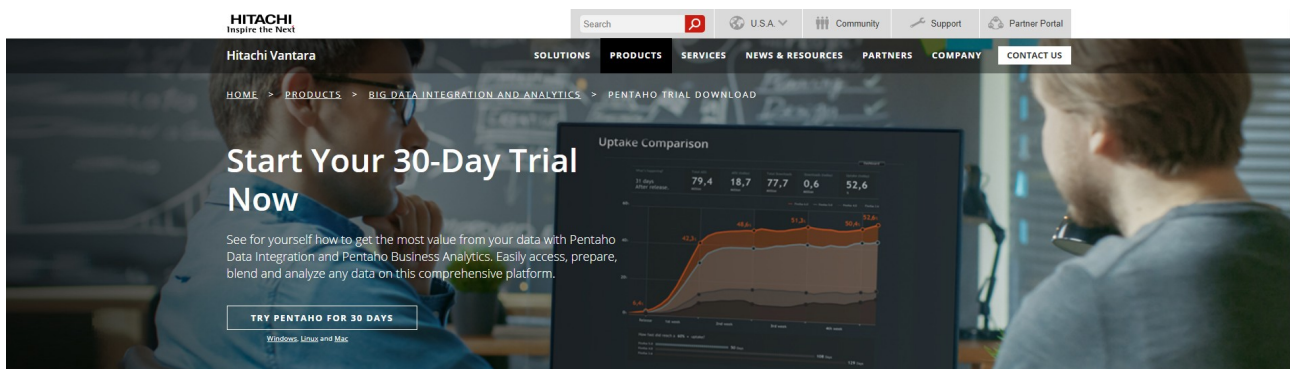
Configuración

1.- Descarga

Los requisitos son la disponibilidad de un entorno de ejecución de Java (JDK) y una base de datos MySQL.

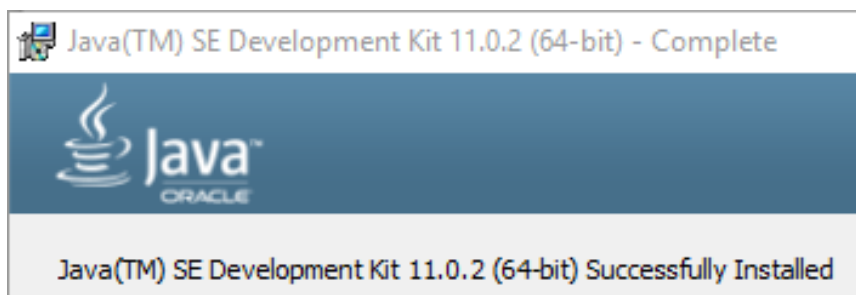
Java SE Development Kit 11.0.2		
You must accept the Oracle Technology Network License Agreement for Oracle Java SE to download this software.		
Thank you for accepting the Oracle Technology Network License Agreement for Oracle Java SE; you may now download this software.		
Product / File Description	File Size	Download
Linux	147.28 MB	jdk-11.0.2_linux-x64_bin.deb
Linux	154.01 MB	jdk-11.0.2_linux-x64_bin.rpm
Linux	171.32 MB	jdk-11.0.2_linux-x64_bin.tar.gz
macOS	166.13 MB	jdk-11.0.2_osx-x64_bin.dmg
macOS	166.49 MB	jdk-11.0.2_osx-x64_bin.tar.gz
Solaris SPARC	186.78 MB	jdk-11.0.2_solaris-sparcv9_bin.tar.gz
Windows	150.94 MB	jdk-11.0.2_windows-x64_bin.exe
Windows	170.96 MB	jdk-11.0.2_windows-x64_bin.zip

El instalador de Pentaho se puede conseguir en la web oficial

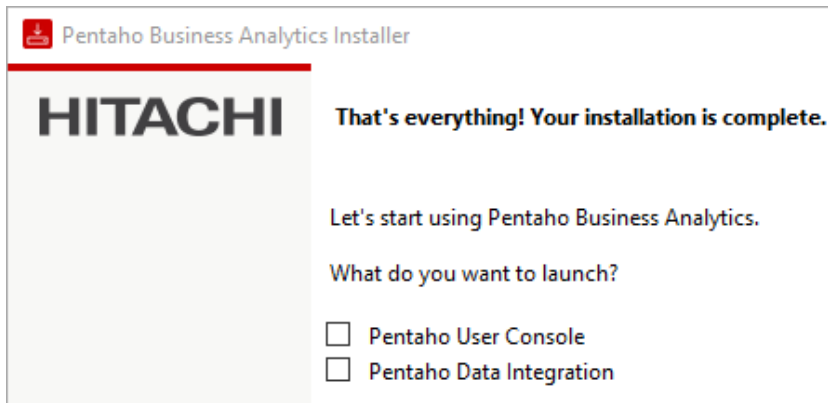


2.- Instalación

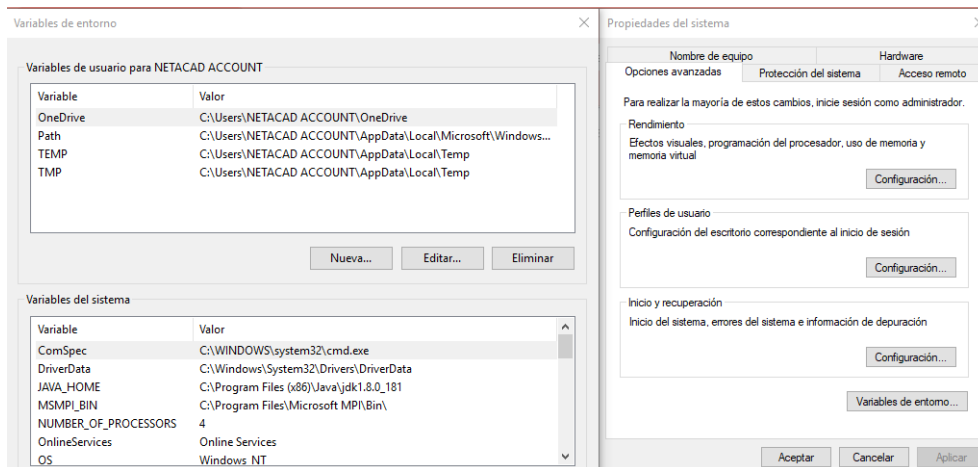
La instalación de ambos componentes es muy sencilla, para el entorno de Java primero se ejecuta el fichero descargado como administrador y tras aceptar la instalación, se encuentra instalado



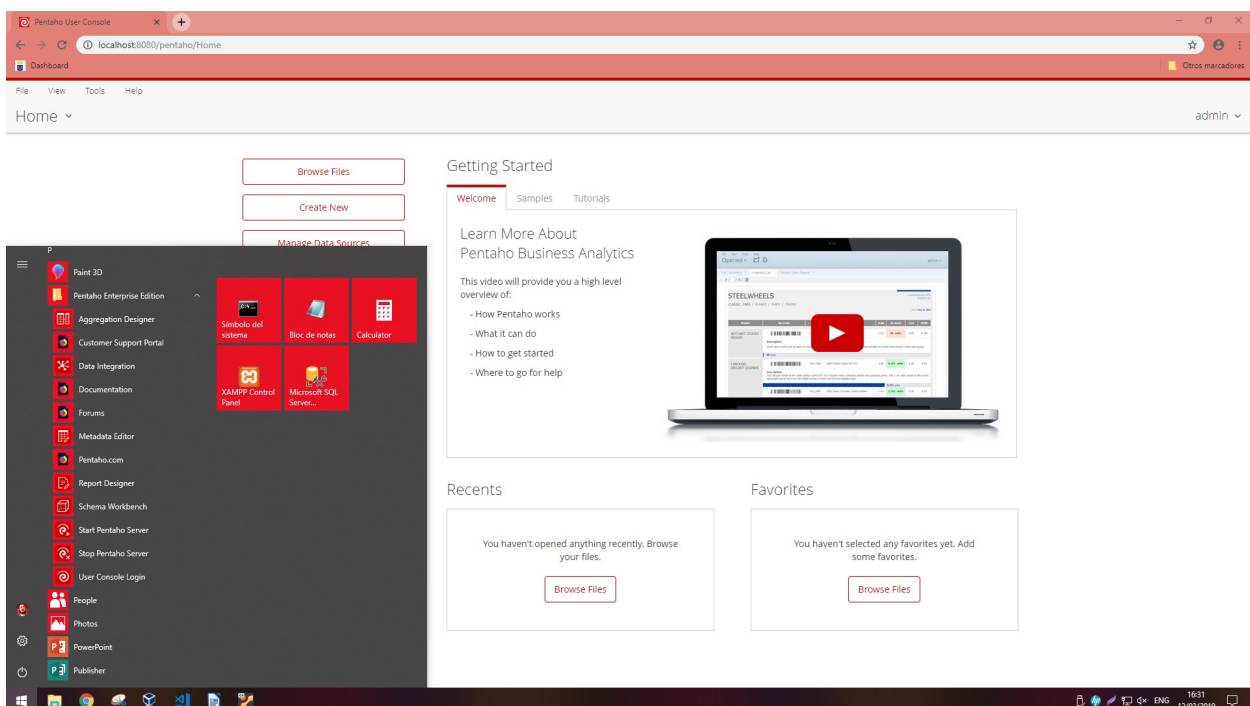
De la misma manera, se instala Pentaho



En las variables de entorno del sistema, se cambia el directorio de Java y se añade al final de la variable *Path*



Y tras arrancar el servicio, se puede acceder a la interfaz web como evaluador

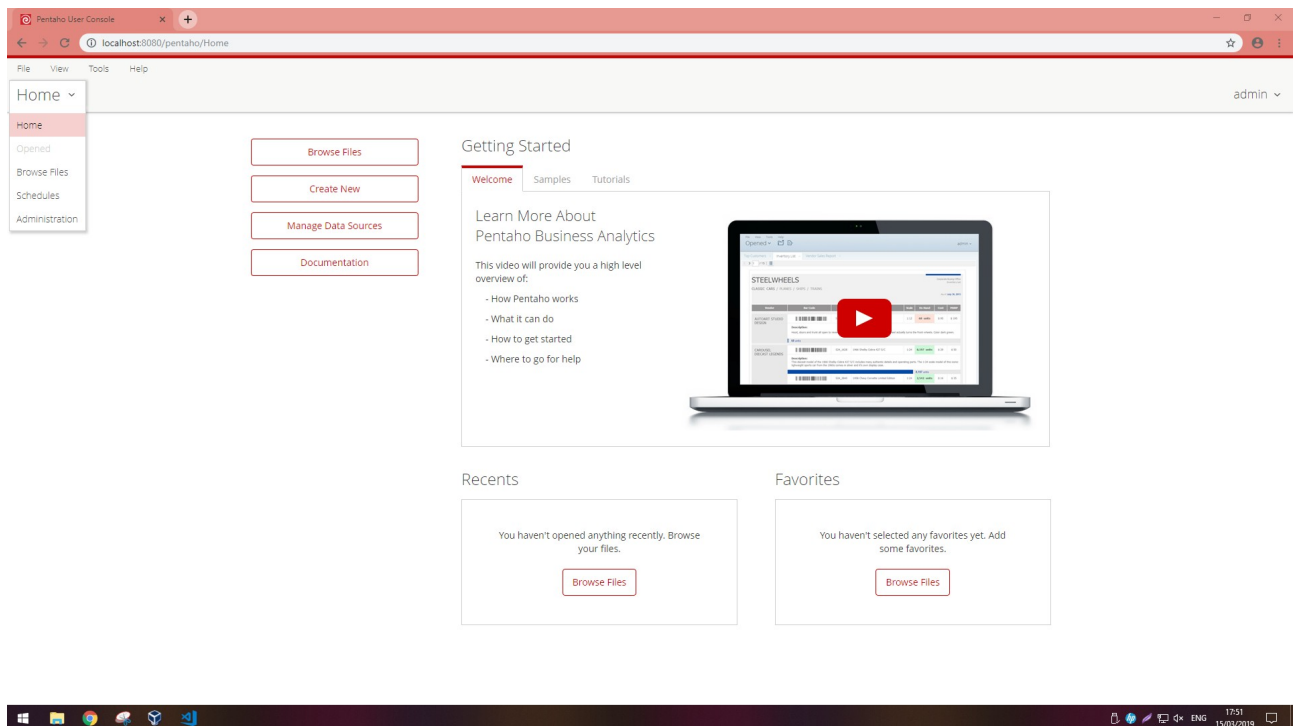


Interfaz de Pentaho

1.- Introducción a la interfaz de Pentaho

La consola de usuario de Pentaho (PUC) se compone por diferentes páginas, siendo el Home la principal

2.- Arquitectura de interfaz



La consola se compone de un menú superior con 4 elementos para manipular informes, cambiar las vistas, las herramientas y la documentación de Pentaho

Abajo a la izquierda, aparece también una lista con otras 4 funciones: búsqueda de archivos, creación de archivos, administración de fuentes de datos y documentación

También, se muestra la ventana de bienvenida con 3 elementos que se componen de un vídeo-introducción, ejemplos y tutoriales

Y en la parte inferior se ofrece la opción de abrir archivos recientes y ver los archivos favoritos

En el menú desplegable del Home hay 4 opciones

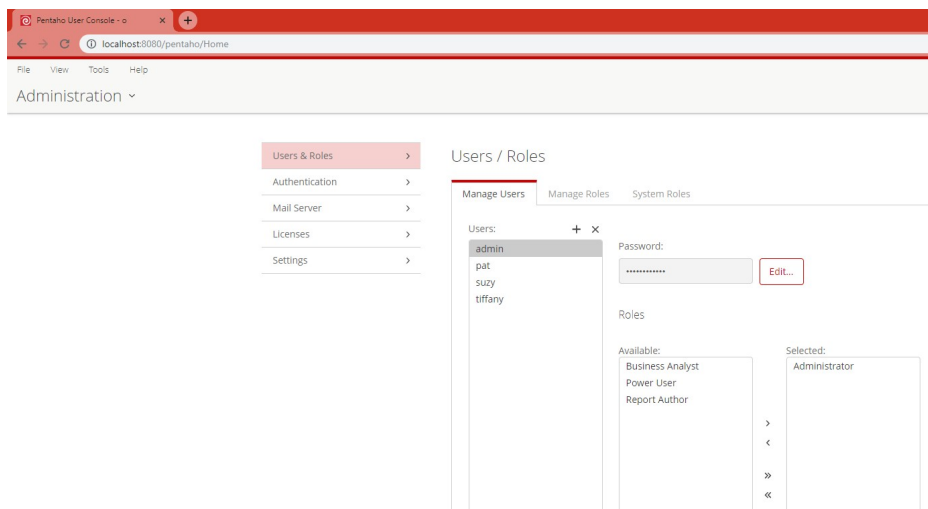
Visualizar el fichero abierto, que se activa al abrir un fichero

Buscar un fichero, para la mejor organización de los ficheros

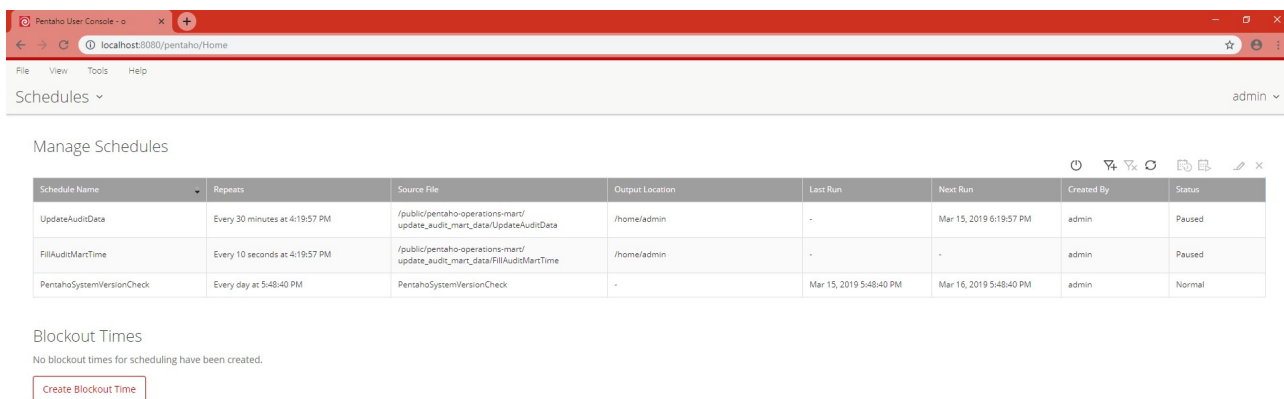
Los horarios, para programar ejecuciones de informes y ver sus estados

Y el panel de administración, donde los usuarios con permisos realizan las tareas de mantenimiento y configuración de Pentaho. En este panel, se pueden administrar usuarios y roles, configurar el proveedor de seguridad para el servidor, configurar el servidor de correo electrónico y su cuenta, gestionar las licencias y administrar la configuración.

Administración de usuarios



Horarios



3.- Principales herramientas de Pentaho

Pentaho no es más que una suite que aloja una serie de componentes que contienen diferentes herramientas para cada una de las tareas a realizar

Pentaho Business Analytics Platform (BA): Es el núcleo de Pentaho. Consiste en una aplicación web que contiene lo que se ha creado tanto en la propia web como en las aplicaciones de escritorio.

Pentaho Data Integration (PDI): Herramienta ETL para extraer datos de las diversas fuentes. Estos datos serán transformados y enviados a diferentes destinos.

Pentaho Report Designer: Para la sencilla configuración de la fuente de datos que será utilizada en la generación de informes.

Estructura de Pentaho

1.- Plataforma Pentaho BI

Como se explicó anteriormente, Pentaho es una suite que se compone de diferentes módulos que facilitan la realización de ciertas tareas. Estos módulos son:

- **Servidor Pentaho BI:** Donde se ubican la infraestructura y arquitectura necesarias para la construcción de las aplicaciones de BI a utilizar.
- **Repositorio de soluciones:** Donde se almacenarán la configuración y las soluciones de Pentaho. Todos los informes, análisis, etcétera de una misma aplicación son almacenados en una misma solución.
- **Consola de usuario:** Donde se accede al servidor. El acceso está restringido para los usuarios de las aplicaciones.
- **Consola de administración:** Aplicación web a la que acceden los administradores para gestionar usuarios, conexiones con la BDD y demás herramientas administrativas.
- **Aplicaciones y portales web:** Las aplicaciones se componen por varias aplicaciones o portales web que acceden a informes y datos importantes de su solución a través de la plataforma BI.
- **Fuentes de datos:** La plataforma accede a las fuentes de datos que usen las diversas aplicaciones según las definiciones de sus soluciones, ya sea un fichero o una base de datos.

2.- Pentaho Data Integrator

La integración de datos de Pentaho (también conocida como Kettle), se trata de una de las herramientas ETL (Extract, Transform & Load) más potentes y versátiles para el diseño del proceso de integración según las necesidades de una empresa. Ésta, permite preparar, limpiar y manejar los datos de cualquier fuente, eliminando la codificación y complejidad que puedan presentar mediante herramientas visuales, mostrando los datos más relevantes al usuario al cargarlos en un Data Warehouse. En esta herramienta, se crea una nueva transformación y se añade una conexión a la base de datos para realizar cualquier tipo de tarea a continuación.

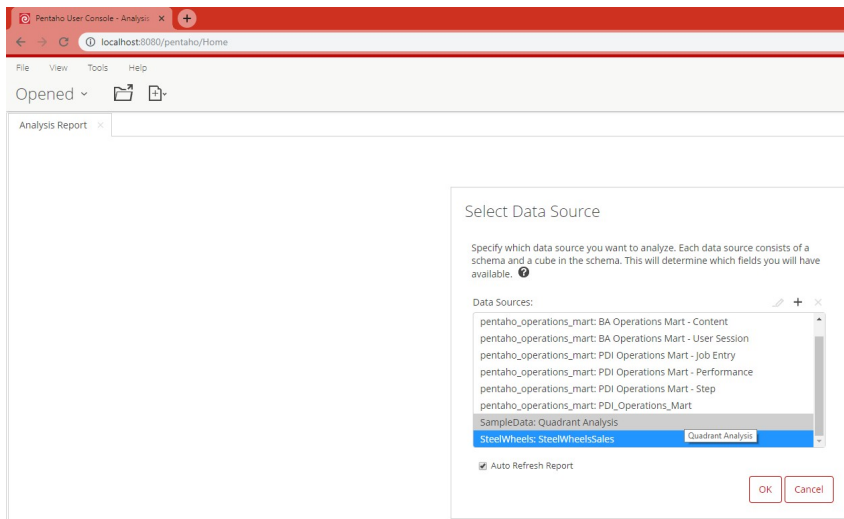
Creación de un informe a partir de una fuente de datos existente

1.- Creación de un nuevo informe

En el menú de Archivos, se crea un nuevo Informe de Análisis

2.- Fuente de datos

A continuación, se debe seleccionar una base de datos, que en este caso viene con la instalación de Pentaho para trabajar con sus datos



3.- Paneles del informe

En el diseñador del informe se muestran 3 paneles: el de las fuentes de datos (con medidas, posicionamiento y elementos a añadir), el de diseño (con las líneas, columnas y medidas añadidas en la fuente de datos) y el de la tabla (que muestra los datos del panel de diseño de forma gráfica)

4.- Adición de elementos

En el panel de fuente de datos, con click derecho en el elemento a añadir se muestra la opción de añadir al informe.

Una vez realizado, se añade al panel de diseño donde se pueden cambiar las posiciones en la que se muestran los datos.

Tras haber seleccionado una fuente de datos (y dado una medida) y haber cambiado el diseño, se mostrará la tabla resultante en el panel de la misma.

Generación y análisis de informes

1.- Acceso al informe

Desde la barra de herramientas, se abre un nuevo archivo y se selecciona del directorio público, el informe que se desea abrir.

2.- Informes interactivos

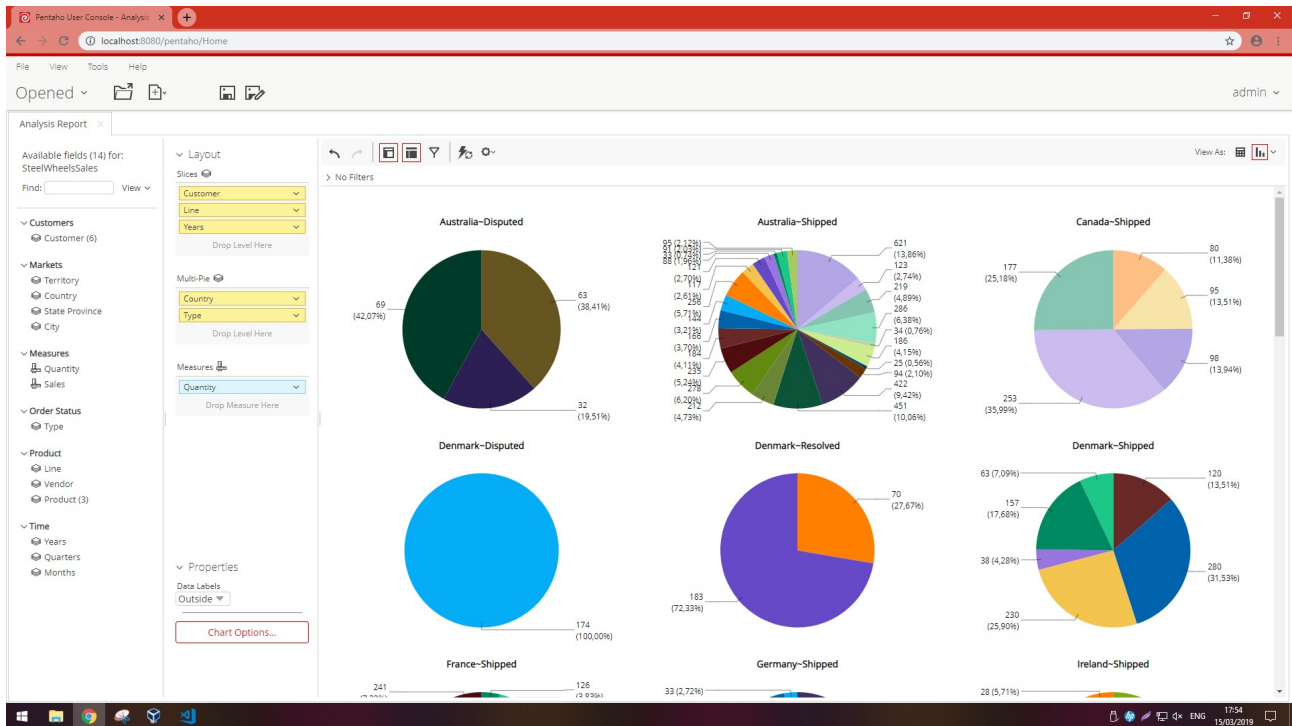
En este lugar se puede crear informes operativos sencillos sin necesidad de conocimientos en desarrollo, solo se precisa de conocimientos básico en el uso del equipo y su sistema operativo. Son usados para la creación de informes destinados a la respuesta de preguntas comerciales de forma inmediata para su posterior impresión o diseño de los elementos gráficos de éstos.

The screenshot displays the Pentaho User Console - Analysis interface. The main window shows an 'Analysis Report' for the 'SteelWheelsSales' dataset. The interface includes a sidebar with a 'Layout' section containing 'Rows' and 'Columns' lists, and a 'Measures' section. The 'Rows' list includes 'Customer', 'Country', 'Type', and 'Line'. The 'Columns' list includes 'Years' and 'Quantity'. The 'Measures' section shows 'Quantity' and 'Sales'. The main data table displays the following data:

Customer	Country	Type	Line	Years	Quantity
All Stores, Co.	UK	Shipped	Classic Cars	2003	297
			Ships	2004	331
			Trains	2004	257
			Vintage Cars	2003	120
			Vintage Cars	2004	273
Alpha Cognac	France	Shipped	Classic Cars	2003	126
			Planes	2003	218
			Ships	2005	75
Amica Models & Co.	Italy	Shipped	Vintage Cars	2003	172
			Classic Cars	2003	96
			Ships	2004	149
			Trains	2004	82
			Trucks and Bus...	2004	22
Anna's Decorations, Ltd	Australia	Shipped	Vintage Cars	2004	24
			Classic Cars	2004	566
			Classic Cars	2003	621
			Motorcycles	2005	123
			Trucks and Bus...	2003	219
Atelier graphique	France	Shipped	Vintage Cars	2005	286
			Vintage Cars	2003	34
			Classic Cars	2003	166
			Motorcycles	2004	156
			Vintage Cars	2004	71
Austrian Collectables, Ltd	Australia	Disputed	Planes	2005	43
			Ships	2005	63
			Vintage Cars	2005	32
			Vintage Cars	2005	69
			Classic Cars	2003	25
		Shipped	Classic Cars	2004	94
			Vintage Cars	2003	422
			Classic Cars	2004	451

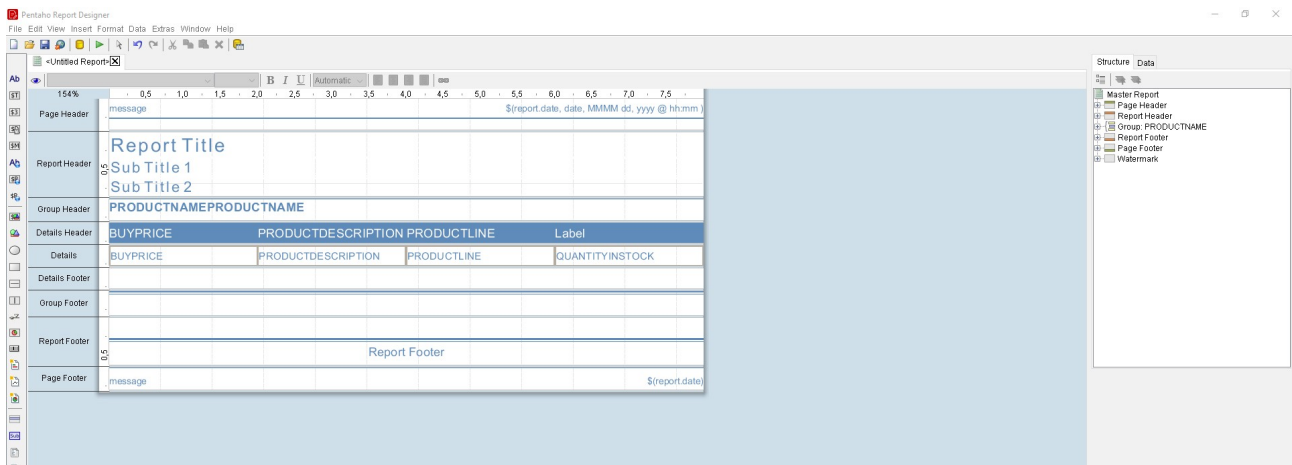
3.- Informes del analizador

Es una herramienta visual para el análisis que se centra en la información empresarial contenida en las fuentes de datos de Pentaho. Su uso requiere los mismos conocimientos que los informes interactivos y tiene como objetivo, normalmente, la compilación de datos de forma rápida mientras son explorados visualmente, además de su filtrado y clasificación en gráficos.



4.- Diseñador de tablero

Permite la creación de paneles de forma sencilla y su uso es recomendable en casos donde se desea crear una sola interfaz en la que ver muchos informes al mismo tiempo y tener acceso a páginas web visitadas con frecuencia o ver gráficos dentro de un espacio mientras se crean informes en otro.



Creación de informes con una fuente de datos propia

1.- Creación de la base de datos del repositorio

Primero, se debe instalar la base de datos. Para este propósito se puede usar Xampp (para manejar la base de datos desde phpMyAdmin) o cualquier otro software como puede ser MySQL Workbench con la configuración correcta.

Una vez instalados los servicios necesarios (ya sea Xampp o MySQL Workbench, por ejemplo), se crea la base de datos deseada. En ella se crean las tablas e introducen los datos mediante consultas SQL o la interfaz visual

En este caso se ha usado una máquina virtual de Ubuntu con el paquete mysql-server

```
Ubuntu Server 18.04 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
root@danielserv:~# sudo mysql
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 14
Server version: 5.7.25-0ubuntu0.18.04.2 (Ubuntu)

Copyright (c) 2000, 2019, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> USE Hospital;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> SHOW TABLES;
+-----+
| Tables_in_Hospital |
+-----+
| Appointment         |
| Doctor              |
| Patient             |
| Room                |
+-----+
4 rows in set (0.00 sec)

mysql>
```

2.- Creación del repositorio

Desde la página oficial de MySQL, se descarga el conector Java en la versión 5.1.47

Generally Available (GA) Releases

Connector/J 5.1.47

Select Operating System:

Platform Independent


[Looking for the latest GA version?](#)

Recommended Windows Download:

MySQL Installer
for Windows

All MySQL Products. For All Windows Platforms.
In One Package.

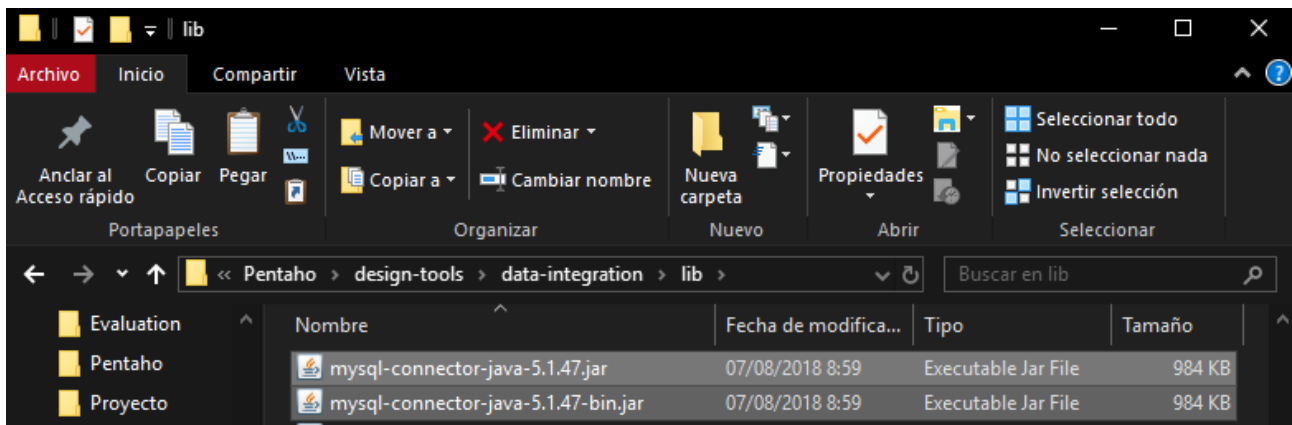
Starting with MySQL 5.6 the MySQL Installer package replaces the standalone MSI packages.



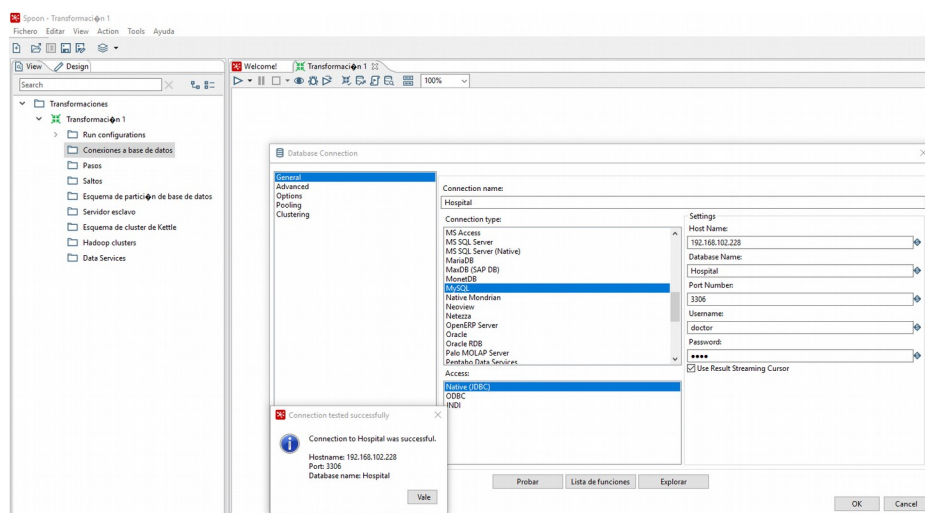
Windows (x86, 32 & 64-bit), MySQL Installer MSI [Go to Download Page >](#)

Platform Independent (Architecture Independent), Compressed TAR Archive <small>(mysql-connector-java-5.1.47.tar.gz)</small>	5.1.47	4.2M	Download <small>MD5: dd1481a25a74fee9cf9cc23e61614a73 Signature</small>
Platform Independent (Architecture Independent), ZIP Archive <small>(mysql-connector-java-5.1.47.zip)</small>	5.1.47	4.6M	Download <small>MD5: c83017d13655f71e192389c70bb6bc93 Signature</small>

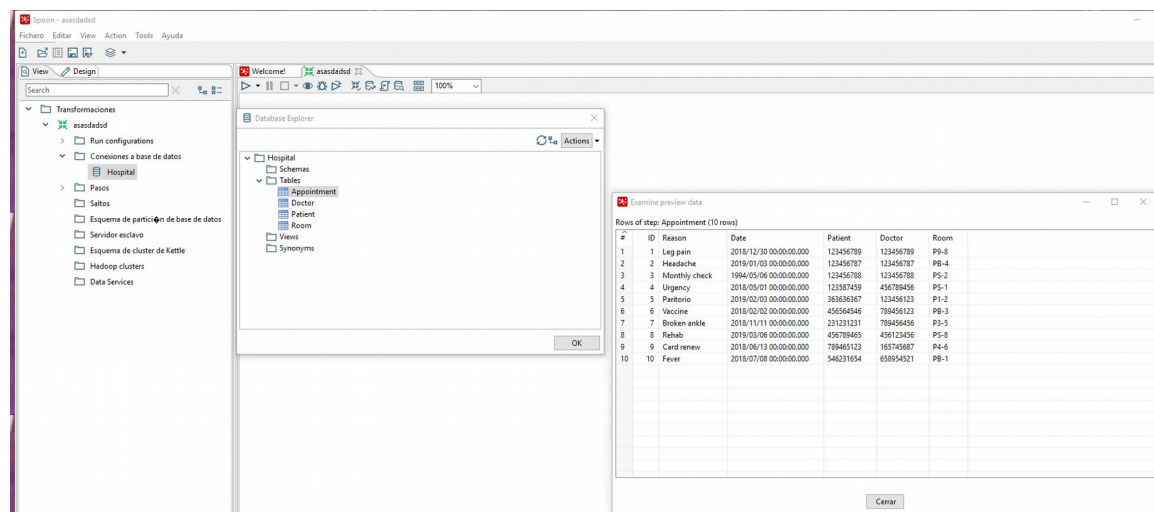
Se extrae el .jar en la carpeta de Pentaho Data Integration, donde se encuentran las librerías de Pentaho



Arrancando el bat de spoon, se abrirá el PDI. Una vez dentro de Spoon, se crea una nueva transformación y en el panel de vistas, se crea una nueva conexión a una base de datos (donde se introduce la información de la base de datos creada anteriormente). Tras confirmar, se comprueba que se añadió la conexión entrando en el explorador de la BDD.



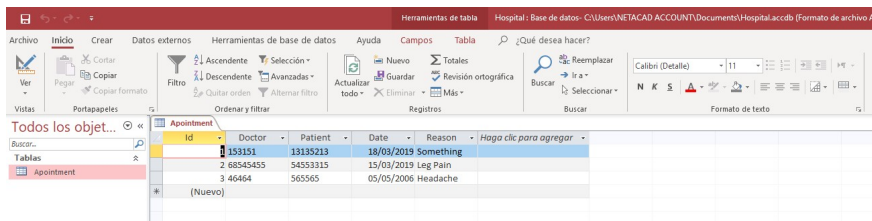
En el mismo explorador, se pueden visualizar los datos contenidos en MySQL



3.- Extracción de datos desde Microsoft Access

Pentaho ofrece la posibilidad, mediante un INPUT, de acceder a datos almacenados en bases de datos de Access, el gestor de bases de datos de Office que funciona mediante consultas SQL.

Una vez en Access, se crea una base de datos y se añaden las tablas y datos

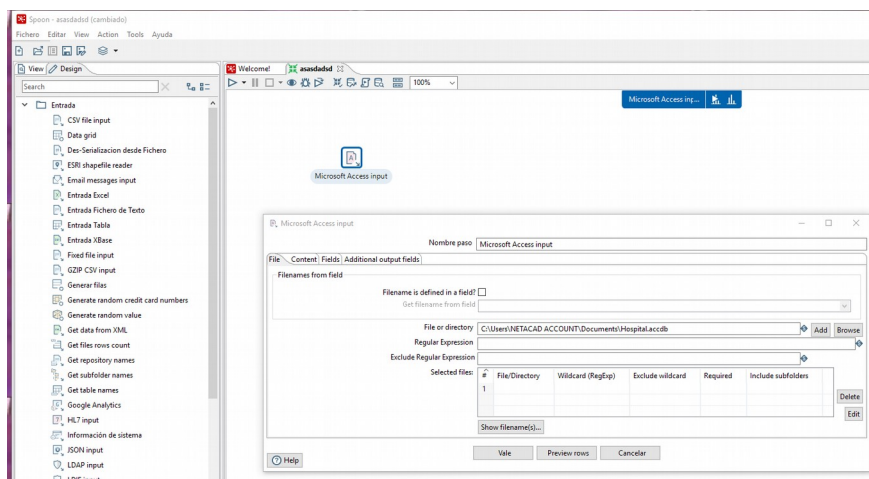


4.- Creación de una tabla INPUT

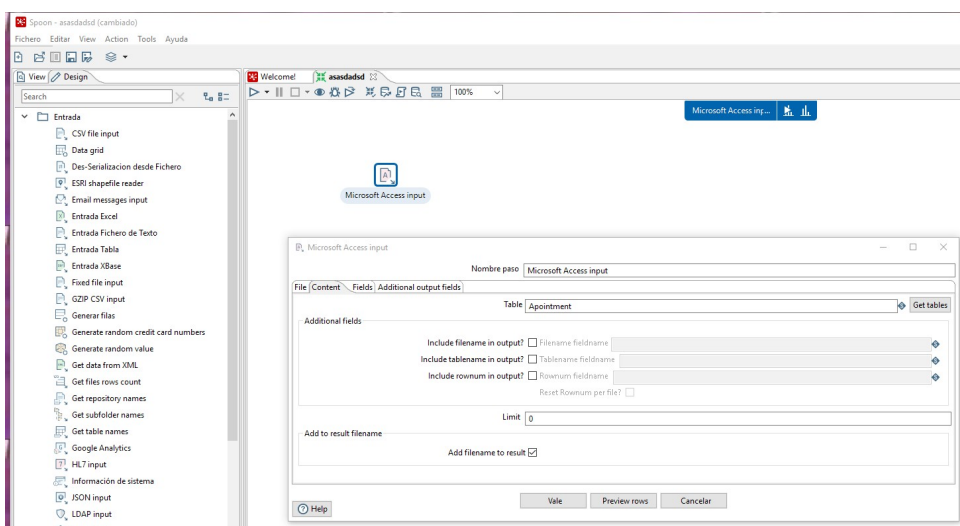
En el PDI, se busca la opción de Microsoft Access input y se añade

Una vez en el panel, se puede proceder con su configuración haciendo doble click sobre el input

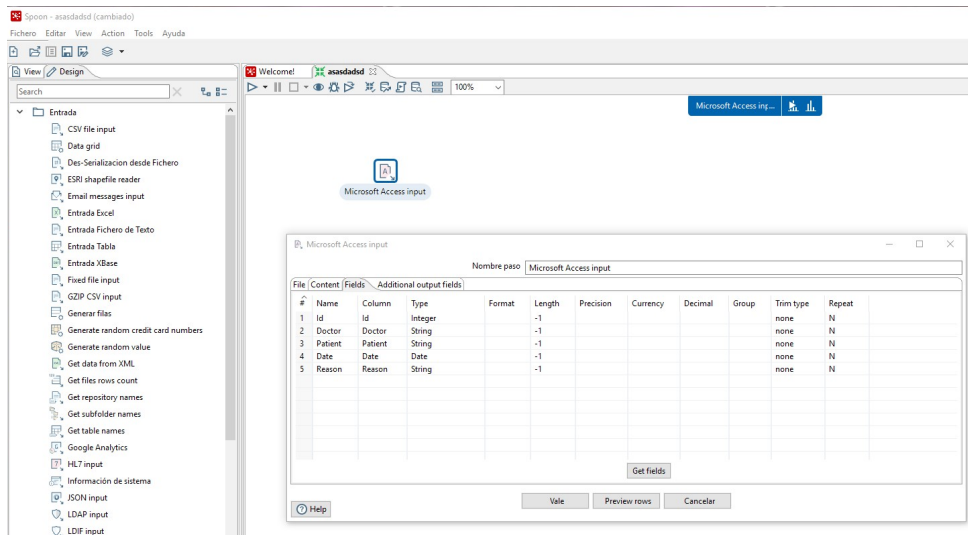
Se ofrecen varias opciones y se escoge la de fichero, donde se seleccionará la base de datos creada en Access. Tras especificar la ubicación, se añade



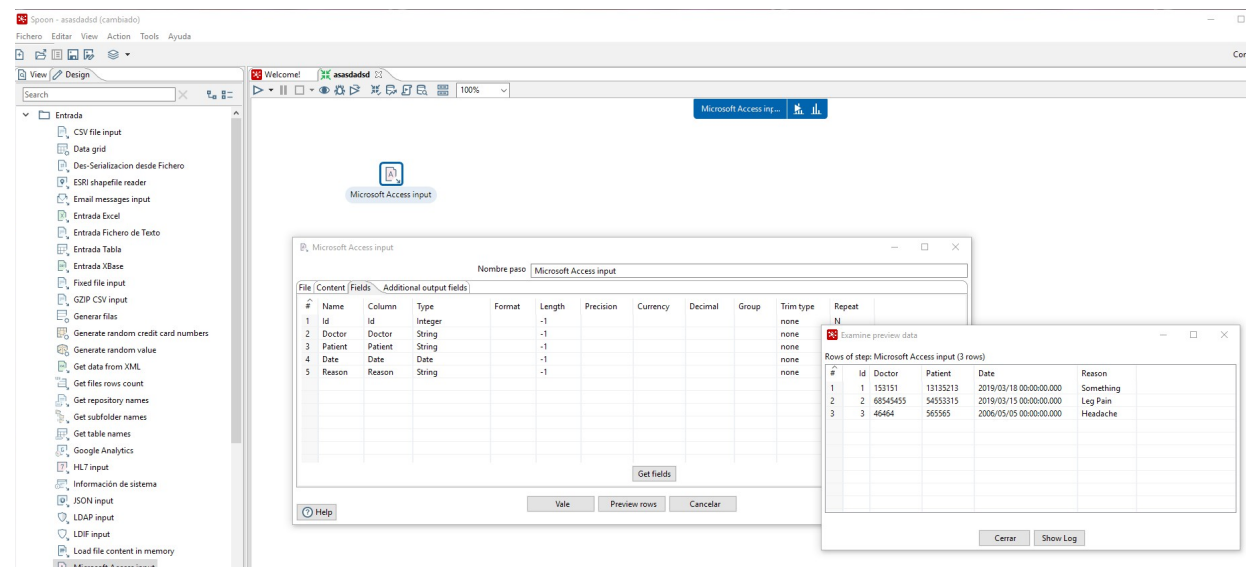
En la pestaña de contenido se acciona la opción de obtener tablas y se selecciona la deseada



Una vez seleccionada, en la pestaña de campos, se obtienen con el botón en la parte inferior



Finalmente, se previsualizan las filas y se muestran los datos de la tabla



Tablas de hechos y dimensiones

1.- Funcionalidad

Las tablas de hechos y dimensiones son utilizadas en esquemas para el soporte de aplicaciones de análisis de datos, cumpliendo cada una de ellas con su rol conteniendo distintos tipos de datos. Estos tipos de tablas son utilizados en esquemas en estrella para estructurar los datos de forma correcta, lo que los convierten en esquemas muy utilizados en la inteligencia empresarial.

2- Tablas de dimensiones

Las tablas de dimensiones contienen los datos organizados en columnas, manteniendo los datos asociados a una clave externa de la tabla de hechos para el acceso a los datos de las tablas de dimensiones a través de la tabla de hechos central

3.- Tablas de hecho

La tabla de hecho es el corazón del esquema de la estrella, conteniendo la relación que la une con el resto de tablas de dimensiones. Es la encargada de la documentación de los datos de la entidad, mientras que las tablas de dimensiones almacenan los detalles relacionados con el elemento de la tabla de hecho.

4.- Carga o exportación de la tabla de hecho

Gracias a la compatibilidad con diferentes bases de datos, el muestreo y la relación entre ellos resulta beneficioso en el caso de las tablas de hecho ya que facilita la comprensión a simple vista.

La tabla de hecho contendría sólo la información de las claves asociadas a las demás tablas mientras que cada una de esas claves se relaciona con la clave de una tabla de dimensiones que contiene la información detallada.

Para cargar este tipo de tablas en el PDI, se siguen los mismos pasos que anteriormente, ya que las tablas están alojadas en la base de datos de Access.