Pentaho: software líder de Inteligencia de Negocio de código abierto.

Introducción

La Inteligencia de Negocios, del inglés Business Intelligence (BI), se refiere a tecnologías, aplicaciones y prácticas para la recolección, integración, análisis y presentación de información para soportar la toma de mejores decisiones de negocios. Un tipo emergente de aplicaciones son diseñadas para generar información correcta y disponible para los empleados, clientes, suministradores, socios y consumidores en el momento exacto que se necesite.

Estas aplicaciones son construidas sobre nuevas tecnologías siendo ya Business Intelligence (BI) una categoría de software bien establecida y bien conocida que se extiende por una amplia gama de capacidades funcionales. Sin embargo sigue siendo costoso en tiempo y recursos para cualquier organización el proceso de evaluación, selección y mantenimiento de la tecnología BI adecuada para su entorno y sus necesidades.

Las herramientas BI profesionales de código abierto (Open Sourse BI) ayudan a ser más económicos estos procesos de manera que reduce los costos y riesgos para posibles usuarios de BI, y disminuyen el costo total de propiedad (TCO). Por supuesto, no siempre es posible enfrentar un proyecto BI con herramientas de este tipo en todas las organizaciones. Sin embargo, existe una tendencia a considerar la posibilidad de BI de código abierto en primer lugar, y sólo iniciar una evaluación tradicional de software BI propietario, cuando se determine que el requisito del negocio no puede ser satisfecho con una alternativa profesional de código abierto.

Por otra parte las herramientas profesionales de código abierto ofrecen ventajas significativas para la fase de evaluación del producto, minimizando el tiempo y el margen de error en este proceso. Comparado con el modelo cerrado del software propietario, basado en el pago por adelantado de las tasas de licencia, el modelo profesional de código abierto se distingue por un mayor nivel de acceso a la información así como al producto con todas las funcionalidades.

Las organizaciones pequeñas generalmente no cuentan, debido a su estructura, con un personal de programadores para evaluar y desplegar soluciones de BI. El personal de las TIC en estas organizaciones está dedicado mayormente al mantenimiento de los servicios de correo electrónico, sistemas de recuperación, gestión de la seguridad, el soporte técnico de usuarios, entre otros.

Las aplicaciones de código abierto están disponibles sin restricciones, lo cual permite a las organizaciones crear rápidamente prototipos de soluciones de negocio, con pequeños equipos de trabajo. Mientras que las soluciones con software propietarios, necesitan mayor tiempo y estructura para formalizar un retorno de la inversión, los equipos ágiles pueden ir resolviendo paulatinamente problemas discretos de negocio, entregando valores a lo largo del camino, construyendo así capacidades de conocimiento.

INCONVENIENTES DEL SOFTWARE PROPIETARIO PARA SOLUCIONES BI

A menudo la solución a un problema de negocio es un proceso que incluye inteligencia de negocio. Las herramientas BI necesitan entender los procesos, o cómo ser parte de ellos, para poder integrarse en una solución completa.

Las implementaciones BI típicamente tienen varias etapas:

1. Presentación de informes e información necesaria
2. Mostrar KPI y tableros (Dashboards)
3. Análisis avanzados de las probables causas de excepciones
4. La integración de informes con gestión de procesos y flujo de trabajo
5. Análisis y ajuste de esos procesos

Sin embargo, muchos clientes solo llegan hasta la implementación de reportes. La deficiente gestión de procesos y flujos de trabajo emparejado con la ausencia de metodologías les impide llegar más lejos. Las herramientas BI tradicionales no garantizan el cumplimiento de estos objetivos, sus principales deficiencias radican en:

- Las licencias de software, el soporte técnico y los servicios son muy costosos
- Son difíciles de usar para la mayoría de los usuarios
- Se hace difícil transferir las habilidades de los proveedores a los clientes
- Es difícil para los clientes crear soluciones e integrar reglas de negocio
- Las “soluciones” son una conjunto de herramientas y no una solución completa
- Las soluciones son difíciles de ampliar y dirigir para los clientes. Los clientes no compran el software, pagan por adelantado el derecho a usarlo
- Las soluciones se centran en los reportes y análisis de indicadores claves de desempeño (KPI) e ignoran el desempeño de los procesos que afectan las métricas
- Inversiones significativas y acuerdos contractuales deben ser firmados antes que una evaluación completa y la creación de prototipos puedan ser culminados.

LA SUITE BI DE CÓDIGO ABIERTO PENTAHO

Pentaho proporciona un espectro completo de herramientas de inteligencia de negocio, reportes, análisis, dashboards, minería de datos e integración de datos. Ofrece además, una serie de servicios críticos entre los que están la autentificación, programación de tareas, seguridad y servicios Web. Este conjunto de herramientas y servicios forman una plataforma integral de inteligencia de negocio, convirtiendo a Pentaho en el proveedor líder de soluciones BI de código abierto.

A diferencia de otras ofertas de BI, Pentaho es una plataforma centrada en procesos y orientada a la solución, con componentes que posibilita a las organizaciones desarrollar soluciones integrales a los problemas de inteligencia de negocio.

La plataforma BI Pentaho, es centrada en procesos porque el controlador central es un motor de flujo de trabajo (Workflow), que utiliza definiciones de procesos para definir los procesos de inteligencia de negocios que se ejecutan dentro de la plataforma BI. Los procesos pueden ser fácilmente personalizados y otros nuevos se pueden añadir. La plataforma de BI incluye componentes e informes para analizar el rendimiento de estos procesos.

Es orientada a la solución porque las operaciones de la plataforma están especificadas en definiciones de proceso y en documentos de acción que especifican cada actividad. Estos procesos y operaciones conjuntamente definen la solución a un problema de inteligencia de negocio, la cual puede ser fácilmente integrada dentro de los procesos de negocio que son externos a la plataforma.

Pentaho mucho más que un simple producto, es una Suite de inteligencia de negocio que está constituida por una colección de programas de computadora que trabajan juntos para crear y entregar soluciones de inteligencia de negocio. Algunos de estos componentes suministran funcionalidades que son muy básicas, tales como la autentificación o las conexiones a las bases de datos. Otras suministran funcionalidades que operan a alto nivel como es la visualización de información utilizando gráficos y mapas.

En la figura 1 se muestra una representación gráfica de los componentes que conforman la Suite Pentaho. Las principales capas son claramente identificadas, con la capa de presentación en lo más alto y la capa de datos e integración de aplicaciones en lo más bajo. Las áreas funcionales más importantes reportes, análisis, dashboards y administración de procesos- constituyen la capa intermedia, en tanto que la plataforma BI en sí misma entrega características básicas para seguridad y administración.

La mayoría de los usuarios finales interactúan con la capa de presentación, la cual puede tomar varias formas. La integración de datos es necesaria para llevar datos desde varios sistemas orígenes hacia el interior de un entorno compartido de un almacén de datos.

Las áreas funcionales cubiertas por la suite son:

- **Análisis:** permite a los usuarios ver datos dimensionalmente. El motor de análisis es proporcionado por el servidor OLAP Mondrian y la librería JPivot para la navegación y el análisis.

- **Reportes:** el motor de reportes permite diseñar, crear y distribuir reportes en varios formatos conocidos (HTML, PDF, entre otros formatos estándares) desde diferentes tipos de fuentes.

- **Minería de datos:** permite a usuarios avanzados descubrir patrones ocultos en los datos, realizar análisis predictivos y de tendencias.

• **Dashboards:** son usados para monitorear y analizar indicadores clave de desempeño (KPI). Interesantes dashboards pueden crearse incluyendo gráficos, reportes, vistas de análisis entre otros contenidos de Pentaho, con relativa facilidad.

• **Integración de datos:** es utilizado para integrar información dispersa que provienen de diferentes orígenes (aplicaciones, ficheros, bases de datos, hojas de cálculo) para crear un almacén de datos central, eslabón fundamental para una solución BI. Pentaho Data Integration es el motor que provee esta funcionalidad.

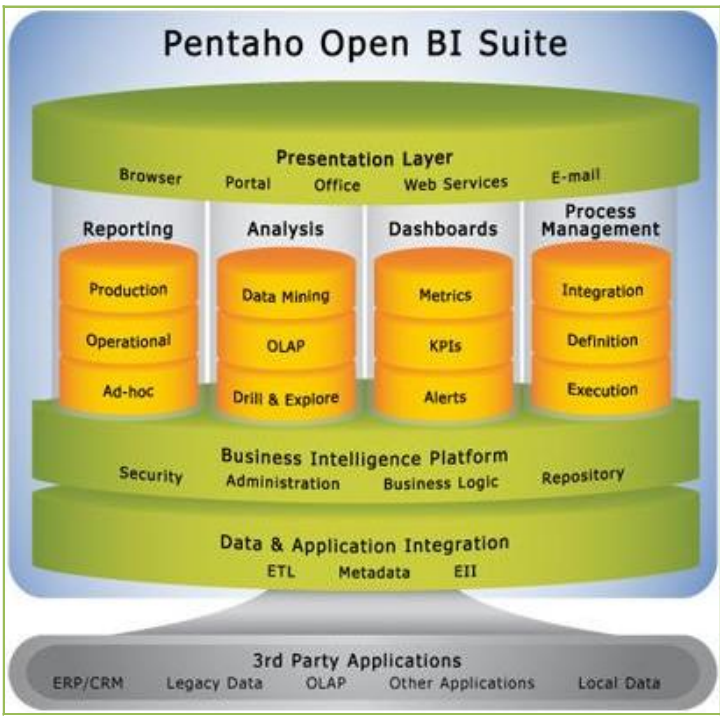

Figura 1. Suite BI de código abierto Pentaho

SERVIDOR PENTAHO

El Servidor Pentaho es una colección de programas que trabajan en conjunto para proveer un número de funciones de la Suite BI Pentaho. Estos programas son implementados como servlets de Java.

En un nivel funcional el servidor Pentaho puede ser divido en tres capas:

- **La plataforma**
- **Los componentes BI**
- **La capa de presentación**

La plataforma

La colección de componentes conocidos colectivamente como la plataforma, ofrece los siguientes servicios:

- Repositorio de solución y motor de solución

- Gestión de conexión de bases de datos
- Autentificación de usuarios y servicios de autorización
- Registro y servicio de auditoría
- Programación de tareas
- Servicios de correo electrónico

La funcionalidad ofrecida por estos servicios es relativamente de bajo nivel y constituye la infraestructura básica de la plataforma BI.

Repositorio de solución y motor de solución

La plataforma Pentaho organiza el contenido BI dentro de las llamadas "soluciones". Una solución Pentaho puede ser pensada como una carpeta de sistema de archivo con todo el contenido BI para solucionar algún problema de negocio. Las carpetas sirven simplemente para facilitar una organización general al contenido BI.

Una solución Pentaho puede contener carpetas e ítems llamados secuencias de acción (conjunto predefinido de acciones que son ejecutadas en el servidor Pentaho). Las secuencias de acción son servicios que pueden ser invocados para entregar algún contenido BI. Pueden ser invocados directamente por interacción del usuario o dirigidos como un servicio web desde otra aplicación. Esta última propiedad permite la integración de Pentaho con otras aplicaciones.

Las secuencias de acción son representadas utilizando XML y son almacenadas en ficheros de texto plano con extensión .xaction, es por ello que son también llamadas xactions. Las más sencillas pudieran ser creadas con un editor de texto simple, pero las secuencias de acción más complejas son creadas utilizando Pentaho Design Studio o usando Eclipse con un plugin de secuencia de acción de Pentaho.

Las secuencias de acción son ejecutadas por el componente de la plataforma conocido como el motor de solución (solution engine). Cuando quiera que un cliente invoque una secuencia de acción, el motor lee la definición de la secuencia de acción y ejecuta sus pasos

Lógicamente, las soluciones Pentaho son almacenadas y mantenidas en el repositorio de solución. Físicamente, el repositorio de solución puede ser almacenado como ficheros en un sistema de fichero o en una base de datos relacional. Básicamente ambos métodos son suficientes, sin embargo el repositorio de solución basado en fichero no soporta autorización. Entonces para precisar el control de usuario y el control de acceso al contenido, el repositorio de solución necesita ser almacenado en una base de datos.

Gestión de conexión de bases de datos

En la mayoría de los casos, los datos presentados en las aplicaciones de inteligencia de negocio son almacenados en una base de datos. Para acceder a estos datos la aplicación necesita establecer una conexión a la base de datos. Establecer esta conexión puede ser una tarea relativamente costosa. Algunas veces se requiere buscar el lugar donde está hospedada la base

de datos y emplear algún tiempo en negociación de protocolos, autentificación de usuario y establecimiento de una sesión.

Para evitar el establecer una nueva conexión para cada consulta o lote de consultas, las conexiones a las bases de datos pueden ser abiertas una vez y almacenadas en una "reserva". Cuando quiera que un cliente necesite una conexión a la base de datos, una conexión libre puede ser escogida desde la reserva, usada para hacer algún trabajo y ser liberada nuevamente dentro de la reserva. Esto es una vía también de limitar el número de conexiones a bases de datos abiertas simultáneamente y protegerlas de ser desbordadas con peticiones de conexión.

Autentificación de usuarios y servicios de autorización

La plataforma Pentaho emplea Seguridad Spring (formalmente conocida como Acegi Security System for Spring) para tratar la autentificación y autorización de usuarios. Esto es una solución de seguridad estándar del marco de trabajo Java Spring. Proporciona la lógica a seguir si un usuario necesita ser autenticado y puede delegar solicitudes de autenticación a un mecanismo de autenticación externo, tales como un servidor de base de datos, un directorio LDAP, o autenticación NTLM sobre una red Windows.

Programación de tareas

La plataforma Pentaho usa Quartz como un componente de programación de tarea. Quartz es creado y mantenido por el proyecto OpenSymphony.

La programación de tareas es utilizada para acciones como: ejecución periódica de tareas de mantenimiento, ejecución en segundo plano de reportes o programar trabajos de ETL.

Servicios de correo electrónico

La plataforma BI incluye la capacidad de enviar correos electrónicos empleando un servidor SMTP estándar. También es incluido un fichero de configuración para el uso de una cuenta de Gmail.

Los componentes BI

La plataforma forma el cimiento para un número de componentes que ofrecen funcionalidad típica de inteligencia de negocio. En esta capa podemos encontrar los siguientes componentes:

- Capa de metadatos
- Servicios de reportes Ad hoc
- Motor de ETL
- Motor de reportes
- Motor OLAP
- Motor de minería de datos

Capa de metadatos

La función de la capa de metadatos Pentaho (PML del inglés Pentaho Metadata Layer) es proteger a los usuarios finales de la complejidad del lenguaje SQL y las bases de datos. PML está basado en la especificación Common Warehouse Metamodel y es capaz de generar SQL desde una consulta escrita en el lenguaje de consulta de metadatos (MQL, del inglés Metadata Query Language). La consulta MQL a su vez es creada por un usuario final, construyendo la selección deseada desde un conjunto de objetos en un modelo de metadatos.

Servicios de reportes Ad hoc

El servicio Web de reporte y consulta ad hoc (WAQR del inglés Web Ad Hoc Query and Reporting service) ofrece a los usuarios finales una vía fácil para crear reportes utilizando la capa de metadatos. La herramienta trabaja en una vía intuitiva guiando al usuario a través de un asistente.

Motor de ETL

EL motor ETL de Pentaho, es el soporte para las tareas de integración de datos y ejecuciones de tareas y trasformaciones creadas con la herramienta Pentaho Data Integration. Este motor ETL es parte de los componentes que conforman la Suite Pentaho, pero puede también funcionar en un servidor diferente o incluso en varios servidores en un modo de clúster.

Motores de reportes

La plataforma Pentaho hospeda múltiples motores de reportes. Los motores nativos son el ya mencionado WAQR y JFreeReport. Adicionalmente Pentaho soporta JasperReports y BIRT, lo cual significa que tiene capacidad para manipular todos los reportes creados por las tres herramientas de reportes de código abierto más populares.

Motor OLAP

Mondrian es el motor OLAP de Pentaho. Traduce las consultas MDX en SQL sobre la base de un modelo multidimensional. Se encarga de esconder y reducir los resultados precedentes e intermedios para optimizar el rendimiento. Esto significa que la primera vez que un análisis es ejecutado sobre un modelo multidimensional, toma más tiempo que los subsiguientes análisis, porque Mondrian intenta mantener los resultados previos, las jerarquías y los cálculos en memoria. Otra importante característica es su modelo de seguridad que soporta roles, los cuales pueden utilizarse para restringir los datos que son accesibles por un usuario. Esto permite limitar el número de vistas OLAP diferentes y los reportes que necesitan ser desarrollados.

Motor de minería de datos

El motor de Pentaho de minería de datos es posiblemente una de las partes más potentes y menos utilizadas de la plataforma. El motor de minería de datos Weka, que ha sido adoptado por Pentaho, es el que se ocupa de estas tareas. El mismo consiste en una colección de algoritmos de minería de datos tales como los necesitados para clúster, árboles de decisión, regresión y redes neuronales.

La capa de presentación

Pentaho tiene integrada una interface web llamada consola de usuario, la cual permite a los usuarios interactuar con el servidor. La capa de presentación puede ser usada para buscar y abrir el contenido existente (reportes, dashboards, análisis), pero de cierta forma puede usarse también para crear nuevo contenido BI. La figura 2 muestra la consola de usuario con el árbol de carpetas en el panel izquierdo para organizar el contenido, y en la pantalla principal varios pestañas con documentos abiertos, permitiendo mantener abiertos varios reportes, dashboards o análisis al mismo tiempo.

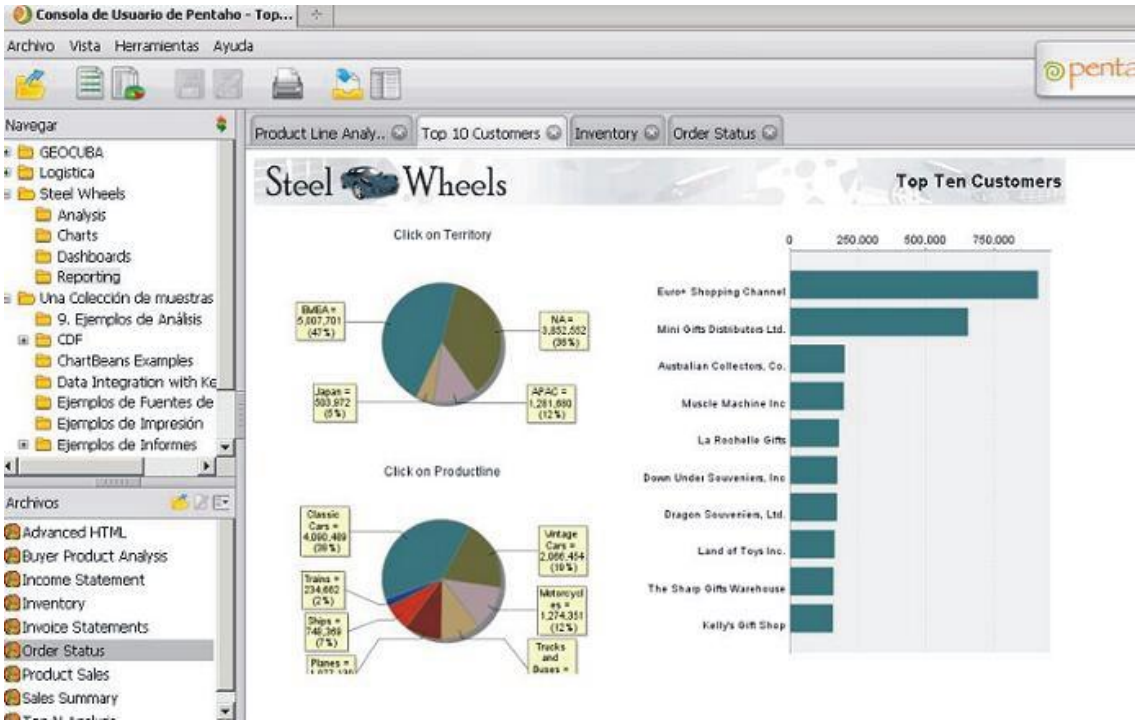

Figura 2. Consola de Usuario de Pentaho

PROGRAMAS DE ESCRITORIO

Los programas de Pentaho que no son servidores pueden ser clasificados como programas de escritorio. Algunos de ellos solo pueden actuar como clientes y necesitan interactuar con el servidor Pentaho, pero otros pueden ser usados de forma independiente también. Estas herramientas de escritorio son principalmente usadas por los desarrolladores como herramientas de diseño, aunque algunas pueden ser utilizadas por usuarios avanzados. A continuación se explica brevemente cuales son sus funcionalidades.

- Pentaho Metadata Editor (PME): con esta herramienta los diseñadores pueden construir capas de metadatos que sirven como una capa de abstracción entre una base de datos relacional y un usuario final.

- Pentaho Schema Workbench (PSW): con esta herramienta se construyen esquemas multidimensionales (cubos OLAP) para ser usados por el motor de Mondrian.

• Pentaho Aggregate Designer (PAD): herramienta para diseñar tablas agregadas que son usadas por Mondrian para incrementar el rendimiento de los cubos OLAP.

• Pentaho Report Designer (PRD): herramienta para la creación de informes para la plataforma de Pentaho. Provee un asistente que mediante un simple proceso, paso a paso, permite editar las características más comúnmente usadas para formatear reportes, permitiendo que usuarios avanzados puedan construir sus propios reportes.

• Pentaho Data Integration (PDI): Anteriormente conocida como Kettle, esta herramienta, con la cual se diseñan y despliegan transformaciones y tareas ETL., constituye un pilar importante en el éxito de Pentaho.

• Pentaho Data Mining: se basa en el proyecto Weka. Éste es un proyecto iniciado en la Universidad de Waikato en Nueva Zelanda, que ha sido adoptado por Pentaho como su herramienta estándar de minería de datos. Provee un ambiente moderno para la construcción de modelos analíticos.

• Pentaho Design Studio (PDS): se basa en el ambiente de desarrollo integrado Eclipse. Su propósito principal es la creación y mantenimiento de las secuencias de acción. Las secuencias de acción son los verdaderos “caballos de batalla” de una solución Pentaho porque vinculan todos los demás componentes juntos.

EDICIÓN ENTERPRISE (EMPRESARIAL) Y EDICIÓN COMMUNITY (COMUNIDAD) DE PENTAHO

Pentaho ofrece dos versiones de la Suite BI Pentaho. La principal distinción se hace entre la versión empresarial con licencia comercial y la edición de la comunidad de código abierto. Esta distinción tiene más que ver con el tipo de apoyo ofrecido que con las diferencias del software existente, aunque la versión empresarial ofrece algunos componentes que no están disponibles en la versión de la comunidad. No obstante, fuera de esto, prácticamente no existen límites entre lo que se puede construir y lograr entre ambas versiones.

CONCLUSIONES

La suite de código abierto Pentaho, proporciona soluciones de BI integrales que se integran a los procesos de negocio, permitiendo ajustar los mismos para optimizar las métricas de los indicadores claves de desempeño (KPI) de una organización y entregar soluciones completas a los problemas de negocio.

Permite desarrollar y desplegar poderosas aplicaciones BI iterativas con la participación de desarrolladores y usuarios finales, combinando el desarrollo de soluciones complejas en un solo proceso, con un ahorro considerable de tiempo.

Cuenta con miles de usuarios en el mundo, muchos de ellos organizaciones poderosas que pudieran permitirse el costo de una solución BI propietaria, sin embargo eligieron Pentaho. Y con una extensa comunidad de contribuidores que le permite ajustarse a los nuevos requerimientos técnicos con rapidez. Registró un crecimiento considerable en varios indicadores en el primer trimestre del presente año 2011, lo que lo ratifica como un suministrador líder y en ascenso, de soluciones BI de código abierto.

Su compatibilidad tanto con software de código abierto como propietario, la extensa gama de soluciones que oferta, la incorporación de estándares y su estabilidad en el mercado hacen de la Suite de código abierto Pentaho un software a considerar para crear capacidades de inteligencia de negocio en una organización.

REFERENCIAS

Casters, M. (Abril de 2010). Pentaho Data Integration 4 and MySQL. Pentaho Corporation.

Diaz, J. C. (2009). Adoption and Usage Survey: Open Source Bussiness Intelligence and Reporting. BeyeNETWORK.

Ferrando, S. (2008). Open Source on the Trading Desk. dbConcert, Inc.

Jaspersoft. (2008). Open Source Business Intelligence Costs and Benefits. Jaspersoft.

McKendrick, J. (2008). Open Source Business Intelligence A 2008 Progress Report. Unisphere Research.

Open source business intelligence. (s.f.). Obtenido de www.openbi.com

Pentaho Corporation. (2007). A New Business Model to Drive Business Intelligence Acceptance and Adoption. Pentaho Corporation.

Pentaho Corporation. (2006). Creating Pentaho Solutions. Pentaho Corporation.

Pentaho Corporation. (2008). Pentaho Open Source Business Intelligence Platform Technical White Paper . Pentaho Corporation.

Pentaho, open sourse business intelligence. (s.f.). Obtenido de sitio web de Pentaho: http://www.pentaho.com/

Roland Bouman, J. v. (2010). Pentaho Solutions Business Intelligence and Data Warehousing with Pentaho and MySQ. Indianapolis: Wiley Publishing, Inc.

Roldán, M. C. (2010). Pentaho 3.2 Data Integration. Packt Publishing Ltd. .

Ventana Research. (2006). Open Source BI A Ventana Research Primary Research Study. Ventana Research.