

Sistemas de Información BI: Estado Actual y Herramientas de Software Libre

Sabina Durá Subiela (sduras@uoc.edu)
Máster de Software Libre

Resumen:

Actualmente muchas empresas ya sean grandes o pequeñas, se plantean implantar en su organización un sistema de información que les ayude en la toma de decisiones. Una de las tendencias actuales es implantar un sistema de BI (Business Intelligence). Muchas organizaciones implantan este tipo de sistemas sin haber realizado un estudio previo de las necesidades y problemas de la empresa, de los beneficios que se pretenden obtener con la implantación del sistema, de la aceptación y formación de los usuarios, de las herramientas y proveedores de BI,... y así un largo etcétera de situaciones y condiciones que se deberían tener en cuenta para la implantación de esta clase de sistemas en cualquier tipo de organización, ya sea grande o pequeña.

En el presente artículo se realiza una síntesis de la amalgama de información existente en los sistemas de BI, se realiza una definición aclaratoria, así como también se expone su estado actual, beneficios, factores de riesgo, pasos para conseguir una implantación exitosa y los componentes y herramientas open source de los sistemas de BI. Primero se exponen las principales herramientas open source existentes en el mercado y se profundiza más especialmente en la Suite open source de BI Pentaho.

Palabras clave: Business Intelligence, Open Source, Pentaho, Factores de éxito, OLAP, ETL, Data warehouse, Data mining, Cuadros de mando, Reporting

1. Introducción

En la actualidad la mayoría de las organizaciones disponen de sistemas de información para realizar el tratamiento de sus datos. Estos sistemas pueden ser más o menos complejos o más o menos económicos dependiendo de las necesidades particulares de cada negocio. Con el paso del tiempo la información almacenada en estos sistemas de información constituyen la historia y el presente del negocio.

Ahora bien, ¿Por qué no utilizamos estos datos para tomar mejores decisiones que nos lleven a un aumento de beneficios del negocio? ¿Por qué no utilizar estos datos de forma predictiva para ayudarnos a tomar mejores decisiones? [11] ¿Y si utilizamos esta información como herramienta de marketing? ¿Cómo puedo tomar ventaja sobre mi competencia? Estas y otras cuestiones similares pueden ser respondidas mediante la implantación de sistemas de BI (Business Intelligence). [4]

Hay que tener en cuenta que el coste de implantación de un sistema BI puede llegar a ser bastante elevado, tanto a nivel económico como a nivel temporal y de organización.

Para conseguir una exitosa implantación de un sistema BI se debe tener en consideración diversos factores que se tratarán posteriormente y tener una gran capacidad de adaptabilidad de criterios, para que la implantación coincida con el tipo de negocio de la empresa.

Existen organizaciones que gracias al BI han conseguido un notable ahorro de costes y aumento de beneficios, pero en otras organizaciones debido a malas decisiones y un erróneo conocimiento de la empresa, los gastos de implantación han sido superiores a los beneficios obtenidos.[4]

Este panorama llevaba a que las empresas pequeñas y medianas (PYMES) [1] no adoptasen soluciones de BI, pero actualmente este paradigma está cambiando gracias a la aparición de herramientas BI Open Source [3], que les proporcionan costes de implantación más reducidos y procedimientos adaptados a estándares abiertos, lo que asegura la continuidad y validez de las soluciones aportadas.

El coste de la implantación de sistemas de BI ha disminuido en los últimos años debido a la caída de precios de los sistemas de almacenamiento de datos. Otros factores a tener en cuenta en la reducción de costes es que los sistemas de Data Warehouse son menos complejos de implantar en una organización; y que cada vez más se está implantando el BI en el cloud [10][11], ofreciendo el BI como un servicio, lo que implica un gran ahorro en infraestructuras hardware.

BI se define como la habilidad para tomar decisiones, esto se logra mediante procesos que hacen uso de metodologías, tecnologías y aplicaciones que permiten el tratamiento y la depuración de datos de diferentes fuentes y aplicar en ellos técnicas analíticas de extracción de conocimiento[1]. Gracias a los sistemas BI se pueden transformar los datos de la compañía en conocimiento para obtener una ventaja competitiva. En la siguiente figura se muestra el modelo integral de una solución BI.

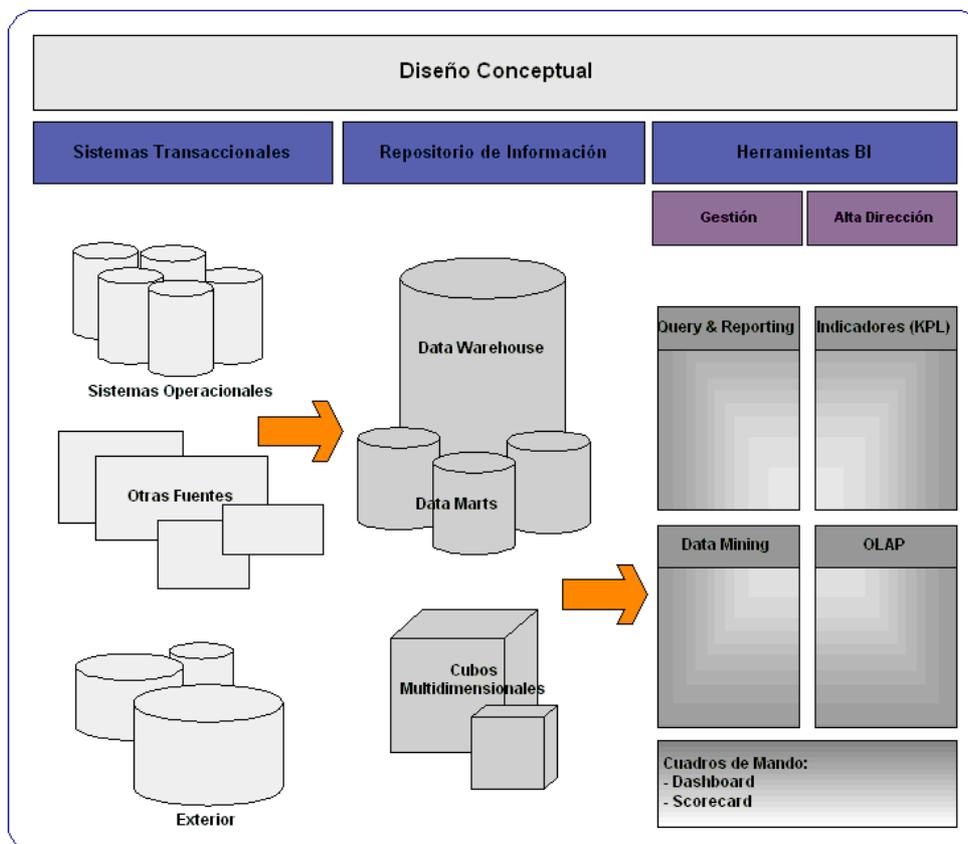


Figura 1: Modelo Integral de una solución BI [19]

El principal objetivo del presente artículo es realizar un estudio acerca del estado del arte de los sistemas de BI, el grado de aceptación y de éxito que tiene en las diferentes organizaciones y cuales son los beneficios que puede aportar la implantación de estos sistemas a una organización. Para ello se sintetiza y expone de forma ordenada los conceptos anteriormente nombrados y se organiza la multitud de herramientas y componentes que forman estos sistemas, centrándose especialmente en aquellos que son open source.

El artículo está estructurado de la siguiente forma, en el apartado 2 se exponen los beneficios que puede aportar una implantación exitosa de un sistema BI; en el apartado 3 se tratan los factores críticos y los pasos a seguir para conseguir una implantación exitosa; en el apartado 4 se estructuran los componentes de un sistema BI en los que posteriormente se clasificarán las herramientas BI de open source (BBDD, ETL, OLAP, Data Warehouse, Data Mining, Reporting, Cuadros de mando,...) del apartado 5; en el apartado 6 se introduce el concepto de BI en Internet; en el apartado 7 se expone el estado actual de los sistemas BI open source; en el apartado 8 se profundiza en las características de la principal herramienta open source del mercado, Pentaho; en el apartado 9 se exponen las tendencias de los sistemas BI y su futuro, y por último se realizan las conclusiones del presente estudio.

2. Beneficios de la implantación de BI

Los primeros sistemas de información a la dirección aportaban básicamente información económica y financiera, con lo que quedaban muchas lagunas por cubrir respecto a la organización de la empresa. Los nuevos sistemas de BI abarcan todas las áreas funcionales de una organización (recursos humanos, marketing, logística,...) y no sólo los económicos [17][19].

Los beneficios que pueden aportar la implantación de sistemas BI son: incremento de la eficiencia en la toma de decisiones, mejora de comunicación entre las diferentes áreas de la organización, mejora de rentabilidad, atracción de nuevos clientes, retención de los clientes actuales, predicción de ventas, reducción del fraude, identificación de riesgos, presentación de soluciones basadas en el conocimiento del negocio, ...

Las empresas que tienen una página web deben plantearse si su página presenta la suficiente inteligencia para que la oferta de sus productos y servicios se convierta en una venta. Para conseguir este hecho se pueden adoptar técnicas de BI que nos permitan obtener información acerca de las preferencias de los usuarios del sitio web, tanto para aumentar la rentabilidad del negocio como para focalizar los esfuerzos de mantenimiento y futuro crecimiento.[1]

Aprovechando el análisis predictivo de las plataformas BI las empresas pueden atraer a nuevos clientes, ofreciendo productos de una forma que resulte atractiva. Además también se pueden tomar decisiones proactivas para evitar la pérdida de clientes descontentos. Los sistemas BI también pueden valorar a los clientes para enfocar los mayores esfuerzos de retención, ya que generalmente cuesta menos mantener un cliente que captar uno nuevo.

Las herramientas BI ofrecen la posibilidad de analizar las transacciones con una mayor probabilidad de fraude, para, así si es posible, poder evitarlas. También son capaces de detectar patrones de riesgos, de esta forma se pueden mejorar los procesos y evitar riesgos.

La mayor parte de los beneficios de la implantación de un sistema de BI son intangibles, ya que derivan en la mejora de la gestión de la compañía. Esto dificulta la tarea de calcular su ROI (Return On Investment) para obtener beneficios cuantificables. Gracias a la implantación de un sistema de BI, se evidencia una reducción de costes al aumentar el rendimiento de la infraestructura TIC de la organización y un incremento de la productividad de los empleados debido a la disponibilidad de información y la

calidad de ésta. Pero como se ha comentado anteriormente este tipo de parámetros son difícilmente cuantificables desde una perspectiva económica-financiera [18]. En la siguiente figura se representa el espectro de los beneficios que se pueden obtener en una implantación de un sistema BI.

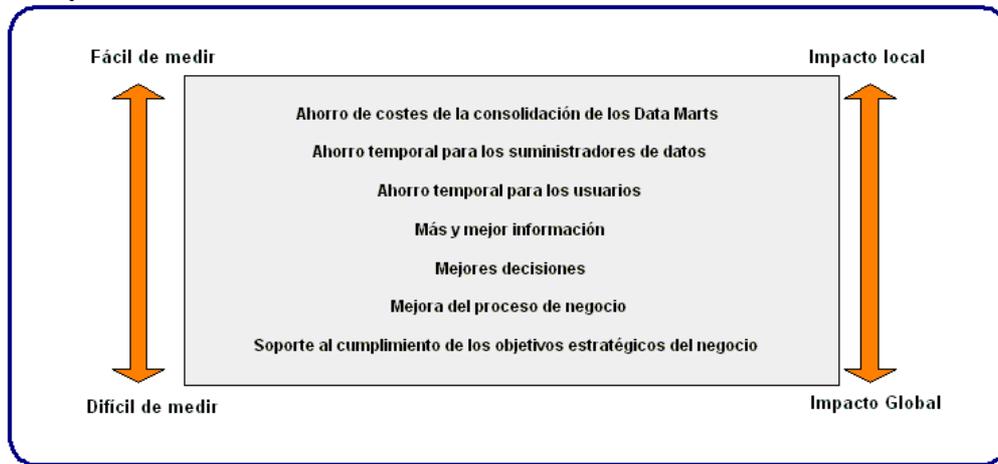


Figura 2: Espectro de beneficios BI [18]

3. Factores de Éxito

La implantación de un sistema de BI en una organización es un proceso muy complejo y requiere grandes cantidades de recursos. A pesar del gran esfuerzo de las organizaciones en la implantación de este tipo de sistemas, no se asegura que se vayan a obtener beneficios [2]. En recientes estudios se ha revelado que las compañías que identifican un juego de *factores críticos de éxito (CSFs)* tienen una mayor probabilidad de éxito y obtienen mejores resultados que las compañías que no identifican dichos factores [7].

Por lo tanto una de las metas de las compañías a la hora de enfrentarse a la implantación de un sistema BI, es la definición de unas métricas [5] que les ayuden en la toma de decisiones. Las compañías que afrontan este proceso con un mayor éxito desarrollan sistemas predictivos mediante la correcta identificación de KPIs (Key Performance Indicators) que se utilizarán en los cuadros de mando.

Las empresas que han tenido éxito en su implantación demuestran que el éxito no se basa fundamentalmente en la inversión tecnológica, sino que se basa más en conocer la infraestructura del negocio, el equipo humano y la política de negocio para poder establecer así los mejores CSFs posibles [7].

Algunos de los CSFs que influyen en el éxito de las soluciones BI son [5][7]:

- *Organización*: se debe alinear el BI con los objetivos estratégicos del negocio. El equipo de IT y el de negocio deben estar coordinados y trabajar bajo un fin común.
- *Procesos*: se deben definir procesos bien estructurados que afecten al equipo, a la metodología y a la dirección del negocio.
- *Tecnología*: es necesario remarcar que no todas las herramientas y aplicaciones BI valen para todo ni para todos los usuarios, se deben analizar los diferentes tipos de usuarios y asignar las herramientas y estrategias adecuadas para sus necesidades. Además la infraestructura de un sistema BI debe ser flexible y escalable, se debe poder tratar con diferentes fuentes de información y trabajar conjuntamente con distintos tipos de herramientas.

A continuación se identifican cinco pasos que deben seguir las organizaciones para obtener una implantación exitosa de un sistema de BI. Estos pasos cubren áreas que ayudan a identificar los CSFs antes de implementar una solución de BI [13].

1. *Identificar el problema:* Lo primero que debe realizar una organización es identificar el problema para saber como abordarlo. La solución de BI debe corresponderse con el problema identificado. No se debe implantar una solución de BI para alcanzar un objetivo de TI, el sistema de BI no debe perder de vista las necesidades reales de la organización y para ello los requisitos de los usuarios deben formar parte en la identificación del problema.
2. *Determinar expectativas respecto al uso:* Cuantificar correctamente el número de usuarios y consultas que va a tener un sistema de BI es un factor primordial en la solución implantada. Un número de usuarios y consultas estimados inferior al real, puede llevar a retrasos y fallos en el sistema que producirán una pérdida de confianza en el sistema de BI implantado. Además hay que elegir correctamente el tipo de herramienta dirigida a cada usuario para sacar el máximo provecho de la información disponible.
3. *Comprender la entrega de datos:* La correcta identificación de la información es la base del sistema de BI, por lo tanto como se va a proporcionar está información es algo que no puede tomarse a la ligera. Se debe identificar si los datos serán entregados por lotes o en tiempo real, qué tipo de sistemas de limpieza se aplicarán a los datos, por lo que las herramientas ETL (Extracción - Transformación y Carga) tienen un papel de vital importancia en la solución de BI.
4. *Implementar las iniciativas de capacitación:* La capacitación a los empleados debe realizarse durante o antes de la fase de implantación, no debe realizarse en una fase demasiado temprana, ya que la dilatación en el tiempo llevaría a la desmotivación y falta de interés de los futuros usuarios.
5. *Seleccionar un tipo de solución:* Las organizaciones que deban seguir unos reglamentos y normas pueden aprovechar las soluciones verticales de BI que cumplen con los requisitos específicos de su normativa. Las soluciones horizontales son más versátiles y se pueden adaptar más al funcionamiento específico de la empresa, pero esto implica una mayor inversión económica y temporal.

4. Componentes de una solución BI

Podemos clasificar los componentes de una solución de BI tal y como se representan en la siguiente figura [16][17][19].

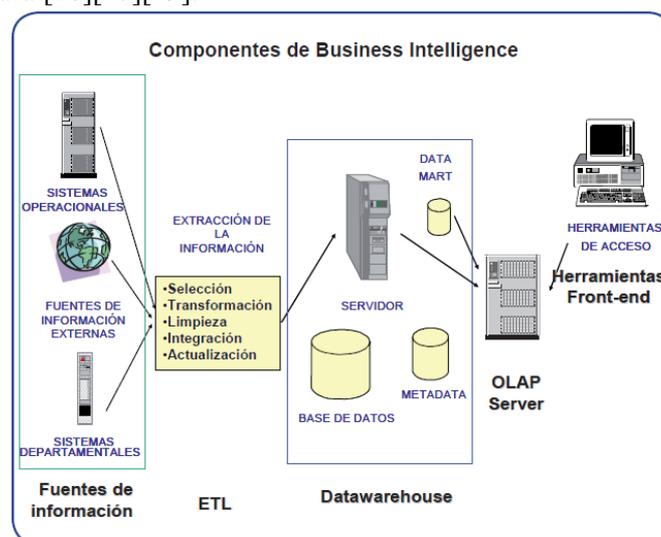


Figura 3: Componentes BI [16]

Las *fuentes de información* en un sistema de BI, sirven para alimentar el Data Warehouse. La información se suele obtener de: sistemas operacionales y transaccionales, que incluyen aplicaciones desarrolladas a medida para la organización

(ERP, CRM, SCM, ...); sistemas de información departamentales (presupuestos, hojas de cálculo, previsiones,...) y fuentes de información externa (estudios de mercado, estadísticas,...). Se debe remarcar que este tipo de información está estructurada, pero que cada vez más, gracias al avance tecnológico, se puede incorporar al sistema, información semiestructurada o desestructurada (imágenes, vídeos, correos electrónicos,...) ya que este tipo de información posee un gran valor para tenerla en cuenta en el modelo de negocio de la empresa. La calidad de los datos con la que se alimentará al Data Warehouse debe ser la máxima posible, ya que si en el Data Warehouse hay errores, éstos se propagarán por toda la organización y serán muy difíciles de localizar.

El *proceso de extracción, transformación y carga (ETL)* trata de recuperar los datos de las diferentes fuentes de información para alimentar el Data Warehouse. Este proceso consume la mayor parte del tiempo en un sistema de BI, por lo que se trata de un proceso de vital importancia dentro del sistema. El proceso ETL se divide en 5 subprocesos: Extracción de los datos en bruto de las diferentes fuentes de información; Limpieza de los datos al eliminar duplicados y valores erróneos para obtener los datos limpios y de alta calidad; Transformación de los datos limpios en datos consistentes para el análisis; Integración de los datos en las definiciones del Data Warehouse y Actualización de los datos periódicamente en el Data Warehouse.

El *Data Warehouse* o almacén de datos proporciona información consistente, integrada, preparada e histórica lista para ser analizada en un sistema BI y utilizarla en la toma de decisiones de una organización. Al almacenar información histórica el data warehouse puede proporcionar información relativa a la evolución de ventas en un determinado periodo. La construcción y el diseño de un data warehouse lleva tiempo y dinero, lo que conlleva que muchas organizaciones no acepten esta situación. Para asimilar con mayor facilidad el cambio, aparecen los Data Marts que almacenan información de una o un número limitado de áreas (marketing, producción, ventas,...). Los Data Marts son más pequeños que los Data Warehouse, almacenan menos información y dan soporte a un menor número de usuarios. Los Data Marts pueden ser independientes del Data Warehouse corporativo y obtener la información directamente de las fuentes de datos, o ser dependientes del Data Warehouse corporativo y así evitar posibles inconsistencias en la información.

En la siguiente figura se visualiza una representación de los dos tipos de Data Marts indicados.

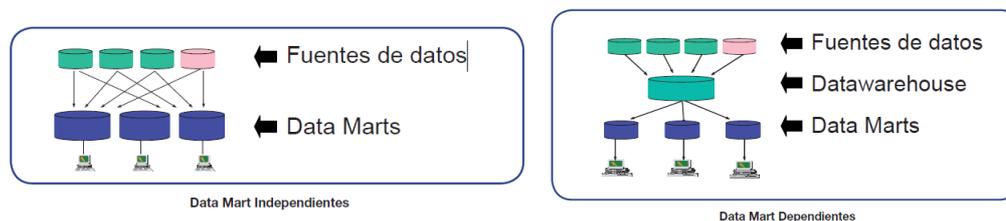


Figura 4: Tipos de Data Marts [16]

Las *Herramientas de BI para la explotación de la información* permiten tratar y visualizar la información que se almacena en el Data Warehouse. Estas herramientas son:

- OLAP (On-Line Analytical Procesing): Son herramientas que permiten realizar consultas complejas a las bases de datos, proporcionando un acceso multidimensional a los datos (cubos multidimensionales), indexación especializada y capacidades intensivas de cálculo. El servidor OLAP almacena los datos multidimensionales precalculados para que la explotación de los datos por parte del cliente sea más rápida.

- Query & Reporting: son herramientas para la realización de consultas y la elaboración de informes. Estas herramientas extraen la información del Data Warehouse o de los Data Marts de la organización. Pueden ser desarrolladas a medida o no.
- Cuadros de mando: existen dos tipos, los *Cuadros de mando analíticos o Dashboard* que permiten obtener, a partir de los Data Marts, la elaboración de informes e indicadores clave (KPI). Son operativos o tácticos y analizan áreas de negocio no relacionadas entre si. Los *Cuadros de mando integrales o Balanced Scorecard* se desarrollan a nivel estratégico de toda la organización, así los diferentes niveles de gestión y dirección de la organización disponen de una visión estratégica con un conjunto de objetivos e indicadores que abarcan a toda la organización.
- Data Mining: Son herramientas que a partir de la información disponible descubren patrones ocultos, tendencias,... y presentan esta información de forma sencilla a los usuarios. El data mining utiliza tecnología basada en redes neuronales, árboles de decisión,... y sus usos más comunes son la segmentación, ventas cruzadas, previsiones, optimizaciones,...

Se puede considerar a los *Usuarios* como componentes del sistema de BI. Básicamente existen dos tipos de usuarios: los usuarios productores de información y los usuarios consumidores de información. En la siguiente figura se representan ambos tipos de usuarios y su clasificación dentro de la organización.

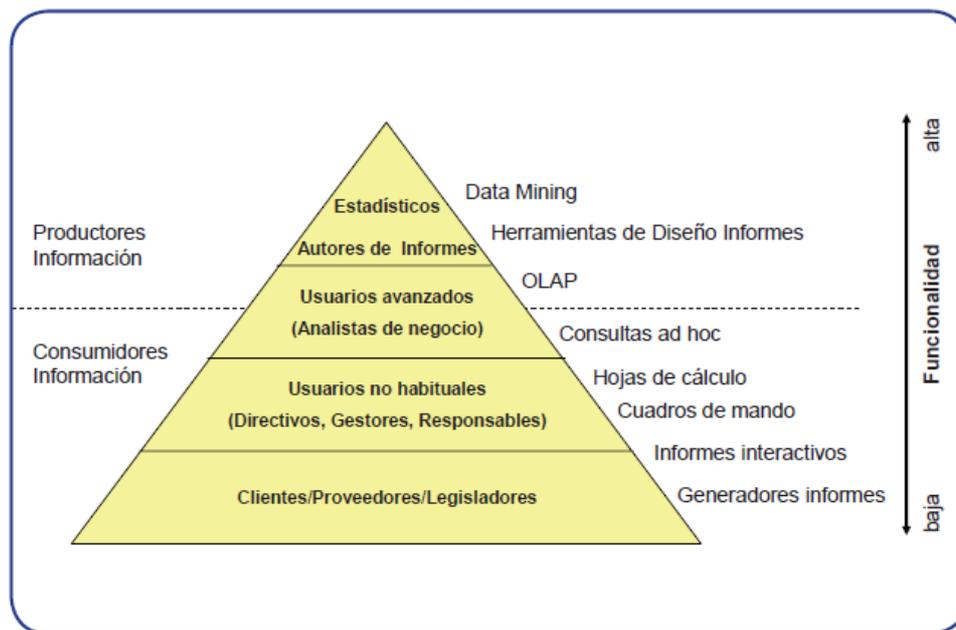


Figura 5: Usuarios BI [16]

5. Herramientas BI Open Source

Bases de Datos

A partir de la base de datos de trabajo diario se puede extraer conocimiento útil para el BI. A la hora de elegir una base de datos [6] para nuestro negocio debemos tener en cuenta una serie de criterios. En primer lugar la base de datos seleccionada debe ser capaz de trabajar con grandes cantidades de datos, así como otros factores como la velocidad de realización de consultas, el particionado y la replicación de datos.

Algunos motores de base de datos Open Source que cumplen estos requisitos son: PostgreSQL, MySQL, SAP DataBase, EnterpriseDB, FireBird, MonetDB y Max DB.

De las bases de datos anteriormente nombradas las más conocidas y utilizadas son MySQL y PostgreSQL¹. MySQL² se utiliza ampliamente en aplicaciones web, es rápida, posee herramientas gráficas para mantenimiento y administración y ofrece un buen soporte en sus versiones de pago (aunque desde que ha sido adquirida por Oracle el coste de estas licencias ha aumentado considerablemente). PostgreSQL es una base de datos robusta y dispone de todas las características necesarias para ser considerada una buena base de datos empresarial. Como desventaja podemos citar que su rapidez es menor que la versión no transaccional de MySQL.

ETL (Extraction, Transformation, Load)

Un sistema ETL realiza funciones de extracción de la fuente de datos (bases de datos transaccionales o externas), transformación de los datos y la carga de los mismos en un Data Warehouse. Estas herramientas se encargan de tomar la información de diferentes fuentes y realizar su carga en el almacén de datos.

En las herramientas ETL tenemos que considerar qué fuentes de datos y qué herramientas soportan. Se debe tener en cuenta si la herramienta se utiliza para realizar la carga en bases de datos relacionales (ROLAP – Relational OLAP), multidimensionales (MOLAP – Multidimensional OLAP) o ambas (HOLAP – Hybrid OLAP) [6]. También hay que tener en cuenta el soporte que ofrecen estas herramientas y si ofrecen una interfaz gráfica, o si por el contrario se debe aprender un lenguaje específico para poder manejarlas.

Algunos ejemplos de herramientas ETL Open Source son: Kettle (Pentaho Data Integration), JasperETL, Palo ETL, Bee y Octopus.

Kettle es la herramienta ETL que utiliza la Suite Pentaho, hace uso de la interfaz gráfica Spoon para diseñar los trabajos de transformación de datos. Kettle utiliza dos herramientas para realizar la transformación de los datos: PAN (motor de transformación, permite lectura y escritura sobre diferentes fuentes de datos) y KITCHEN (permite la ejecución de trabajos de forma planificada).

JasperETL (Talend ETL) es la herramienta ETL disponible en la Suite JasperSoft, es una herramienta independiente de JasperSoft y esta desarrollada en Java/Perl. Está orientada a usuarios con conocimientos de lenguajes de programación, lo que hace que sea una herramienta muy flexible aunque su curva de aprendizaje sea más larga.

Palo ETL es una herramienta completamente integrada en la Suite PALO, no existe un desarrollo independiente de esta herramienta. Es una herramienta muy intuitiva y con una baja curva de aprendizaje.

Bee es una suite que contiene una herramienta ETL, un servidor OLAP y un cliente OLAP con interfaz web. Su herramienta ETL trabaja con bases de datos relacionales (ROLAP). Posee una interfaz gráfica de usuario y el proceso de transformación se encuentra descrito en un fichero XML.

Octopus es una herramienta ETL basada en Java y que se puede conectar a las fuentes de datos relacionales (ROLAP) mediante JDBC. La transformación se encuentra en un archivo XML. Permite interoperar simultáneamente a varias bases de datos (MSSQL, Oracle, DB2, QED, Excel, Access, MySQL, CSV-files, XML-files). Se pueden incorporar valores por defecto en la carga de datos o incluso implementar transformaciones de datos mediante Java o JavaScript.

Data Warehouse

Data Warehouse es el proceso de extraer datos de diferentes aplicaciones, ya sean internas o externas, después depurarlos y estructurarlos para que sean guardados en un

¹ <http://www.postgresql.org/>

² <http://www.mysql.com/>

almacén de datos. Este almacén de datos será utilizado posteriormente para el análisis BI. Todo este proceso requiere de una gran infraestructura hardware y software para llevar a cabo la transformación de los datos, aunque últimamente han aparecido en el mercado algunos sistemas que permiten automatizar, y por tanto facilitar, gran parte del proceso. [11].

El Data Warehouse es una gran base de datos corporativa que almacena los datos extraídos a partir de diferentes fuentes, la información almacenada se utilizará posteriormente en las consultas e informes. La información se almacena en el Data Warehouse en diferentes periodos de tiempo, además cabe resaltar que esta información no se modifica como en los sistemas transaccionales, sino que se incrementa.

El metadata en el Data Warehouse es el lugar dónde se define la estructura que tendrá el propio Data Warehouse. En el metadata se definen las tablas y sus atributos, lo que implica que a nivel empresarial se deben definir el modelo de negocio y la información que se desea extraer del mismo [15].

OLAP (On-Line Analytical Processing)

El procesamiento analítico en línea [1] [6] permite obtener acceso a datos organizados, organiza subconjuntos de datos en una estructura multidimensional que pueda responder a las preguntas específicas de un negocio. Es decir, se basa en el análisis multidimensional (análisis del hipercono) de los datos y permite al usuario tener una visión más rápida e interactiva de los mismos.

Este tipo de sistemas son muy utilizados en las áreas de marketing de las empresas, ya que pueden responder a preguntas complejas en un corto espacio de tiempo. También se utilizan estos tipos de sistemas para generar índices que mejoran los tiempos de accesos a las consultas.

Según el tipo de base de datos en la que almacenan los datos podemos clasificar estos sistemas en: ROLAP, que almacena los datos en una base de datos de tipo relacional; MOLAP, que almacena los datos en una base de datos de tipo multidimensional y HOLAP, que es un sistema híbrido que combina los dos sistemas anteriores.

Existen tanto servidores como clientes OLAP, para elegir uno se tienen que tener en cuenta diferentes factores: En los servidores OLAP, lo primero que debemos tener en cuenta es si son ROLAP, MOLAP o HOLAP [6], también debemos tener en cuenta si son capaces de manejar grandes conjuntos de datos y si son independientes de la base de datos o deben trabajar específicamente con una. Otros factores que también debemos considerar en un servidor OLAP es la API que utiliza y el tipo de lenguaje de consulta que soporta. El cliente OLAP debe ser compatible con el servidor OLAP utilizado, la API utilizada, los informes predefinidos que soporta y en que formato pueden ser exportados.

Algunos de los servidores OLAP Open Source existentes son: Mondrian, Palo y Bee.

Mondrian es un servidor OLAP desarrollado en Java orientado a ROLAP. Puede ser utilizado por cualquier motor de base de datos que tenga un driver JDBC. Utiliza MDX¹ (MultiDimensional eXpressions) como lenguaje de consulta. Presenta los datos en un formato multidimensional a través de un API de Java.

Bee posee un servidor OLAP orientado a ROLAP dentro de su suite. Utiliza MySQL como base de datos y puede manejar hasta 50GB de forma eficiente.

Palo es un servidor OLAP orientado a MOLAP que soporta hasta 256 dimensiones con jerarquías dentro de cada dimensión.

¹Lenguaje de consulta para base de datos multidimensionales sobre cubos OLAP

Algunos de los clientes OLAP Open Source existentes son [6]: Bee y Jpivot.

Bee posee también un cliente OLAP, el cliente es un cliente web y se debe utilizar con el servidor OLAP Bee. Puede generar diferentes tipos de gráficos tanto en 2D como en 3D y los resultados se pueden exportar a Excel, PDF, PNG, PowerPoint, Text y XML.

Jpivot es un cliente web OLAP que trabaja con el servidor Mondrian. Como Modrian utiliza MDX como lenguaje de consulta. Genera gráficos tanto en 2D como en 3D pero solamente exporta los datos a PDF y Excel.

Reporting

Son los componentes de un sistema BI que se encargan de generar los informes, debemos tener en cuenta que estos informes sean presentados en formatos estándar para mantener al máximo la compatibilidad con otras aplicaciones.

Algunos generadores de informes BI Open Source son: Pentaho Reporting (JfreeReports) , Jasper Report, y Eclipse Birt.

Pentaho Reporting es el generador de informes incluido en la suite Pentaho. Se pueden generar informes tanto a nivel de cliente utilizando el asistente o el diseñador de informes o también se pueden generar los informes a través de la web .

Jasper Reports es un generador de informes desarrollado en Java, los informes se pueden entregar en los formatos PDF, XML, HTML y CSV .

Eclipse Birt es un generador de informes basado en Eclipse, genera informes en formato XML. Puede ser embebida en proyectos J2EE.

Cuadros de Mando (Dashboards/Scorecard)

Un cuadro de mando es una herramienta que permite alinear los objetivos de las diferentes áreas con la estrategia de la organización y seguir su evolución. El uso que se le puede dar a un cuadro de mando es muy variado, puede ir desde evaluaciones de personal, acciones de formación, plan de negocio,...

En un cuadro de mando se transcribe la estrategia de la empresa en indicadores de rendimiento (KPI – Key Performance Indicators) [3]. Se utilizan para hacer evolucionar la estrategia de la empresa, gestionar el rendimiento y medir el impacto de las iniciativas relacionadas con la optimización de los métodos.

Algunos de los programas para diseño de cuadros de mando que podemos encontrar son: Pentaho Dashboard Designer, Jasper, Birt y Palo.

Como se puede observar la mayoría de los cuadros de mando están integrados en suites de BI Open Source. Otras herramientas de cuadros de mando independientes de las suites open source son: JetSpeed, Jboss Portal y Marvelit.

Data Mining

La minería de datos consiste en seleccionar, explorar, modificar y modelar grandes cantidades de datos para poder analizarlos, este proceso debe ser automático o semi-automático. Se utiliza en todo tipo de sistemas desde en la emisión de tarjetas de crédito hasta en biología pasando por la web semántica.

Las herramientas de análisis en BI son un punto crítico dentro del proceso, ya que un mal análisis puede llevar al fracaso de un negocio.

Weka es una herramienta Open Source de Data Mining realizada en Java. Tiene una gran cantidad de algoritmos implementados, pero su punto fuerte es la posibilidad de ser adaptado a cualquier entorno. Es un sistema multiplataforma y está integrado dentro de la suite ofimática Pentaho.

Suites BI

Muchas de las herramientas anteriormente comentadas se encuentran incluidas dentro de plataformas de BI.

Tradicionalmente las plataformas de BI no eran de Open Source y estaban enfocadas principalmente a grandes empresas, ya que estas eran las únicas que tenían la capacidad, infraestructura y presupuesto necesario para llevar a cabo una implantación de BI. Algunas de las plataformas de BI comerciales más conocidas son [11]: MicroStrategy, TIBCO, IBM's Business Intelligent y Microsoft Business Intelligence Platform.

Con el tiempo han ido apareciendo soluciones integradas de BI Open Source cada vez más competitivas. Algunas de las soluciones de BI Open Source existentes en el mercado son: Pentaho, Jasper, Birt y Palo.

Pentaho ofrece una plataforma completa de herramientas BI Open Source. Estas herramientas son: reportes, análisis OLAP, dashboards, integración de datos y minería de datos. Actualmente es la plataforma BI Open Source más utilizada [1][2][8][9][10].

Pentaho elimina el coste de la licencia de software y obtiene beneficios a través del pago de una licencia anual que incluye soporte, servicios y una optimización del servicio. Esta plataforma la utilizan desde las pequeñas empresas hasta las multinacionales. Su tecnología soporta análisis de rentabilidad, análisis de clientes, reportes financieros, herramientas KPI, reportes operacionales, ...

Los proyectos Open Source incluidos en la plataforma Pentaho son: Kettle (ETL Open Source), Mondrian (Servidor OLAP Open Source), Jpivot (Cliente OLAP Open Source), Weka (Data Mining Open Source), Pentaho Reporting Engine (Generador de Informes Open Source) y Pentaho (Suite BI completa).

Jasper es la principal competidora de software BI Open Source de Pentaho, en su suite incluye también una gama de herramientas que constituyen una plataforma BI completa [1][2][8]. Estas herramientas son: JasperETL (ETL Open Source), Mondrian (Servidor OLAP Open Source), Jpivot (Cliente OLAP Open Source), Jasper Reports (Generador de Informes Open Source) y Jasper Server (Servidor web).

Birt (Business Intelligence Reporting Tools) [2][8] es un plugin de Eclipse que permite una integración completa en proyectos J2EE. No posee herramientas ETL ni OLAP específicas (aunque sí posee un motor interno para tablas cruzadas) ya que se focaliza principalmente en la generación de informes.

Palo Suite se centra en su motor MOLAP y todo lo que ello conlleva, escenarios, análisis multidimensional,...[8] Las herramientas desarrolladas son: Palo ETL (ETL Open Source), Palo Web (Aplicación Web), PALO OLAP (Motor OLAP Open Source), Pivot (Cliente OLAP Open Source) y Palo for Excel (Extrae Informes del motor MOLAP).

A continuación se muestra una tabla con las herramientas anteriormente expuestas y sus urls correspondientes:

HERRAMIENTAS BI OPEN SOURCE	
BASES DE DATOS	
MySQL	http://www.mysql.com
PostgreSQL	http://www.postgresql.org
ETL	
Kettle	http://kettle.pentaho.com/

JasperETL	http://www.jaspersoft.com/jasperetl
Palo ETL	http://www.jedox.com/es/producto/palo-suite/etl-server.html
Bee	http://sourceforge.net/projects/bee
Octopus Enhydra (ETL)	http://octopus.objectweb.org
OLAP	
Mondrian	http://mondrian.pentaho.com
Jpivot	http://jpivot.sourceforge.net
Palo	http://www.palo.net
Bee	http://sourceforge.net/projects/bee
REPORTING	
Pentaho Reporting (JfreeReport)	http://sourceforge.net/projects/jfreereport
Jasper Report	http://jasperforge.org/projects/jasperreports
Eclipse Birt	http://www.eclipse.org/birt/phoenix
CUADROS DE MANDO	
JetSpeed	http://portals.apache.org/jetspeed-2
Jboss Portal	http://www.jboss.com/products/jbossportal
Marvelit	http://www.marvelit.com/index.html
DATA MINING	
Weka	http://www.cs.waikato.ac.nz/ml/weka/
SUITES BI	
Pentaho	http://www.pentaho.org
Eclipse Birt	http://www.eclipse.org/birt
JasperReports	http://jasperforge.org/projects/jasperreports
Palo	http://www.jedox.com/en/products/Palo-Suite.html

6. BI en Internet

Cada vez son más las empresas que necesitan que sus sistemas de información estén distribuidos y descentralizados, esto conlleva a que un mayor número de empresas utilicen los sistemas de información en la nube [14].

En el marco económico actual, las empresas deben ser más innovadoras y tener una mayor visión estratégica del negocio. A través de una solución de BI se pueden alcanzar dichos objetivos, pero como se ha comentado anteriormente la implantación de un sistema de BI tradicional requiere de gran capital humano, temporal y económico. Las pequeñas y medianas empresas tienen ahora la oportunidad de maximizar su rendimiento a través de soluciones Cloud Computing, también llamadas Cloud BI, en las que el software es ofrecido como servicio, reduciendo así los costes en la infraestructura de información y explotando al máximo los beneficios del Cloud Computing [20].

La seguridad es uno de los grandes retos a afrontar con el Cloud BI, ya que debido a la sensibilidad de los datos a tratar es el punto de mayor reticencia para decidirse por un

sistema Cloud BI frente a un sistema BI tradicional. Otros aspectos a tener en cuenta son la velocidad de acceso, la accesibilidad, la pérdida de control de los propios datos, compatibilidad con otras herramientas y la falta de un estándar bien definido. Se deben sopesar los beneficios y los riesgos y optar por la solución que más convenga al tipo de organización en cuestión.

Otra opción de uso de BI en Internet es implementar un sistema propio de BI y que éste sea accesible a través de Internet. En este tipo de solución se debe preparar los componentes del sistema para que las herramientas BI puedan extraer la información del Data Warehouse, independientemente de dónde se encuentre situado el puesto de trabajo.

El Data Warehouse del sistema de BI en Internet debe prepararse para su acceso remoto, las interfaces de usuario deben estar preparadas para conectarse a través del navegador y la información a acceder puede ser interna o externa a la empresa, se accedería a través de la Intranet o Internet. También se puede habilitar una Extranet con permisos restringidos para permitir que personal externo a la empresa pueda conectarse, como clientes, socios,...

La web 2.0 y las redes sociales han dado lugar a la aparición de BI 2.0. Actualmente las redes sociales como Facebook, Twitter, LinkedIn, ... absorben la mayor parte del tráfico en Internet, por lo tanto la forma de hacer negocios en Internet también está cambiando. Por ese motivo el BI 2.0 analiza el uso de dichas redes sociales y propone estrategias de negocio para poder sacarles el máximo provecho. BI 2.0 detecta nuevas oportunidades de negocio de la organización en las redes sociales [21].

También se puede utilizar Internet para obtener información valiosa para la organización a través del Web Mining. Web Mining es el proceso global de descubrir información o conocimiento potencialmente útil a partir de datos existentes en Internet [15].

7. BI de Open Source

Uno de los grandes obstáculos a los que se enfrenta un proceso de implantación de BI es el alto coste que va asociado al mismo, este coste llega en muchos casos a superar el coste de la inversión del hardware de la organización. Este es el principal motivo por el cual cada vez más empresas optan por una solución Open Source frente a las costosas soluciones de las empresas tradicionales. Hay que tener en cuenta que las organizaciones que se decantan por un producto BI Open Source no quedan abandonadas a su suerte, ya que ofrecen un soporte profesional a través de pago de licencias anuales [9] [10].

Las soluciones BI Open Source reducen considerablemente los costes de implantación sin comprometer el nivel funcional o la calidad del servicio. Al no pagar una licencia por el software (se paga por mantenimiento y soporte) los clientes evalúan el producto y lo aceptan sin sentirse atados a una tecnología específica [3].

El Open Source actualmente ha alcanzado un gran grado de madurez y calidad, un ejemplo de ello es el sistema operativo Linux se utiliza cada vez más en empresas de hosting y en supercomputadoras¹.

En resumen las ventajas del BI Open Source frente a otras soluciones de carácter privativo son [1]: la reducción de costes de implantación; independencia de los proveedores; empezar con una solución de bajo coste e ir incrementado progresivamente sus funcionalidades; resultados en estándares abiertos y compatibilidad con otras herramientas.

¹<http://www.top500.org/charts/list/36/osfam>

8. Suite Pentaho

Pentaho es una Suite de BI Open Source basada en Java. En su suite Pentaho integra diferentes proyectos open source y los hace funcionar de forma integrada en su plataforma siguiendo el flujo de trabajo. Los proyectos open source que ha integrado en su suite permite que se cubran las siguientes áreas: análisis de información, reportes, cuadros de mando, flujos de trabajo y minería de datos [8][10][12]. En la siguiente figura se representa la arquitectura de la Suite Pentaho.

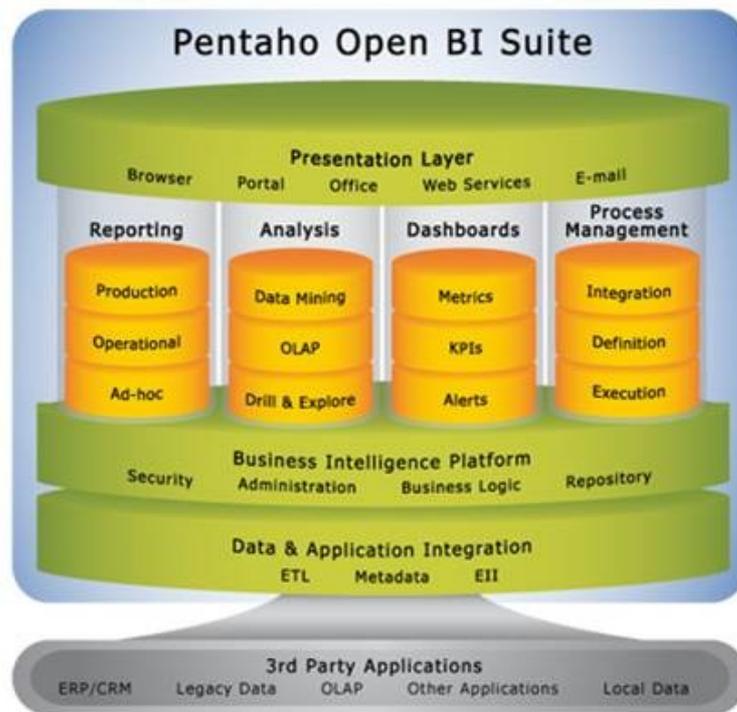


Figura 6: Plataforma BI Pentaho

Pentaho Data Integration: Es el proceso ETL (extracción, transformación y carga) de la suite Pentaho. Utiliza la herramienta Kettle ETL para realizar dicho trabajo. Utiliza una interface gráfica “Spoon” para diseñar los trabajos de transformación de datos que pasan a ser procesados por Kettle [21] tal como se puede visualizar en la siguiente figura.

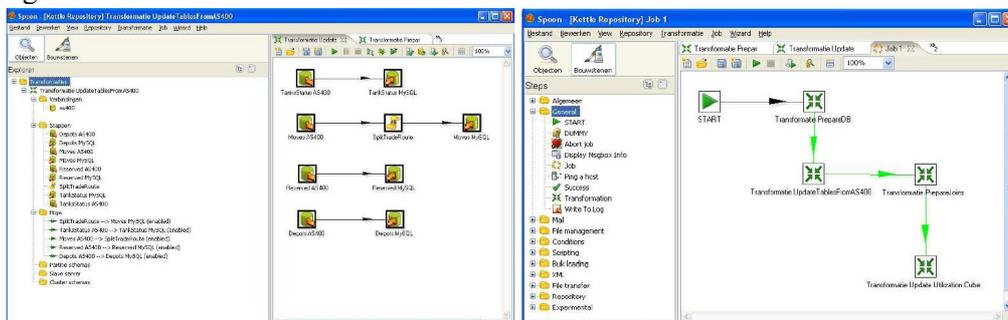


Figura 7: Proceso ETL

Pentaho Analysis: Forma parte del motor OLAP integrado en la Suite BI de Pentaho. Pentaho utiliza por defecto el cliente Jpivot (ejemplo de visualización en la figura 9) para trabajar con el servidor OLAP Mondrian. Pentaho permite al usuario seleccionar el esquema Mondrian a utilizar y muestra una visualización para un análisis rápido. Un ejemplo del flujo de datos sería [12][22]:

1. El cliente Jpivot envía una solicitud
2. El servidor OLAP Mondrian recibe la solicitud comprueba si tiene la petición en caché, y si es así responde al cliente.
3. Genera nuevas consultas si los datos no están en caché.
4. Se almacenan los nuevos datos en caché para posteriores consultas.
5. Se envían los resultados al cliente.

Este proceso queda representado en la siguiente figura.

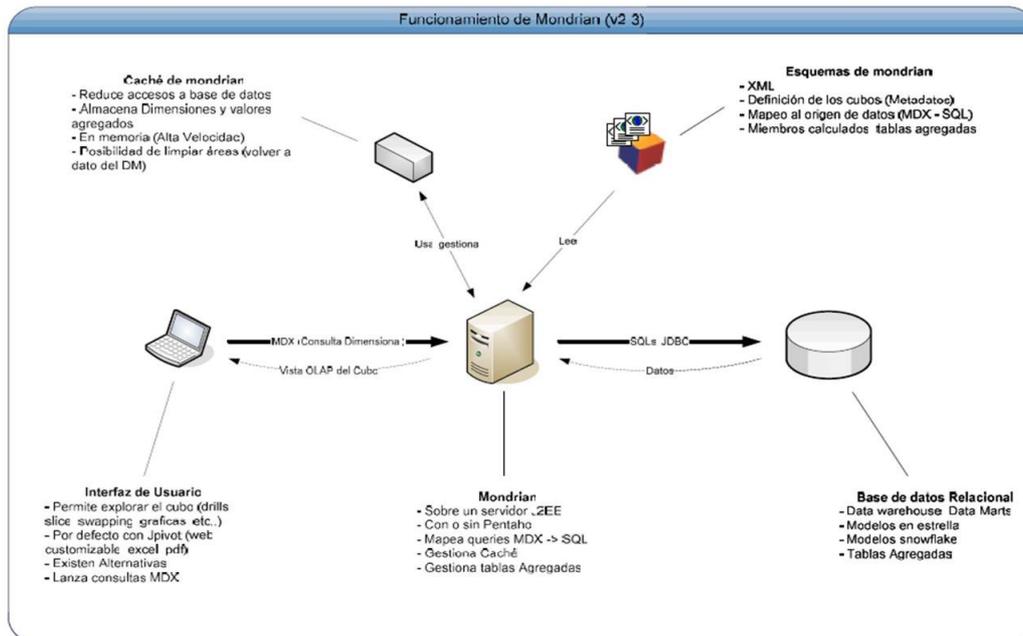


Figura 8: Funcionamiento Mondrian

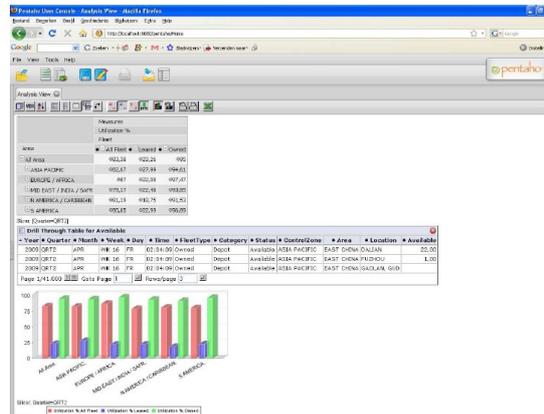


Figura 9: Visualización Jpivot

Pentaho Reporting: Basada en el proyecto JfreeReport, permite generar informes de forma ágil y de gran capacidad. Soporta la distribución de los resultados del análisis en múltiples formatos, todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto. Los reportes Pentaho se pueden programar e incluso que sean generados periódicamente [2][22]. En la siguiente figura se pueden visualizar diferentes tipos de informes soportados por Pentaho Reporting.



Figura 10: Pentaho Reporting

Pentaho Dashboards: Son los cuadros de mando de Pentaho. Integrados completamente con Pentaho Analysis y Pentaho Reporting. En los cuadros de mando se definen los KPI del sistema BI estos deben reflejar factores genuinos de éxito y ser cuantificables [12].

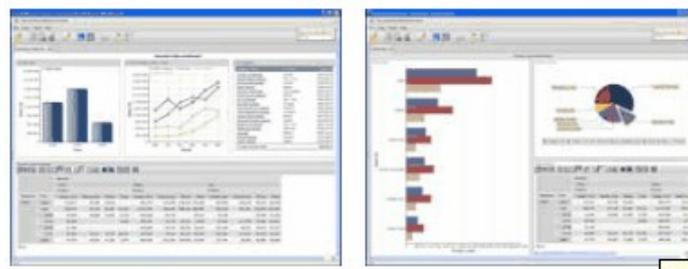


Figura 11: Pentaho Dashboards

Pentaho Datamining: Utiliza la herramienta Weka. Provee un completo conjunto de algoritmos que automatizan los procesos de transformación de datos a la forma en que la minería de datos puede explotarlos. Los resultados pueden ser visualizados en modo gráfico ya sea agrupado, segmentado, de árbol de decisión, bosque aleatorio, redes neuronales y componentes de análisis [22].

Pentaho es una suite que integra diferentes herramientas BI que en conjunto ofrecen una solución completa a las necesidades BI de una organización. Pero en contrapartida a las suites comerciales para poder hacer uso de Pentaho, es necesario tener conocimientos técnicos de TI para la construcción de cuadros de mando, esquemas y conocer lenguajes software como SQL, PHP, Java, XML,... También es necesaria una documentación más extensa para poder hacer uso de esta herramienta, existe mucha información esparcida en Internet pero ésta no está unificada. En el artículo "Implementing BI concepts with Pentaho, an evaluation" se indica que Pentaho es una buena herramienta que cubre todo el espectro de una solución BI, pero para poder utilizarla, es necesario poseer mayores conocimientos técnicos que con otras soluciones comerciales [12].

9. Tendencias en BI

Las tendencias en el uso del BI son varias, pero básicamente todas se centran en alinear los objetivos de la organización con la solución de BI a implantar. Para ello se deben tener en cuenta diversos factores como: que para satisfacer al usuario los datos a analizar deben estar lo más actualizados posible, acceso a la información independientemente de la ubicación, predicciones, compartir información entre diferentes departamentos, análisis de datos no estructurados, aprovechar la información disponible en Internet, tener en cuenta Internet y las redes sociales como una más de las líneas de negocio,...

Algunas de metodologías que definen las tendencias anteriormente nombradas son: Análisis Predictivo, Análisis en Tiempo Real, Análisis en Memoria, Integración con CPM, Cloud BI, BI 2.0 y Agile BI Governance.

El Análisis Predictivo de los datos está relacionado con el Data Mining, gracias a la minería de datos y a la evolución de los mismos a lo largo de un determinado periodo se pueden realizar previsiones de demanda de ciertos productos, análisis de riesgos,... [15]

El Análisis en Tiempo Real de los datos es ahora una de las principales tendencias, tradicionalmente los datos se volcaban al Data Warehouse cada cierto tiempo, lo que conllevaba que el BI fuese concebido como plan estratégico a medio y largo plazo. Actualmente la demanda de la actualización de los datos es mayor para poder tomar las decisiones con mayor rapidez y mayor acierto [15][16][18]. Un ejemplo podría ser las decisiones a tomar en el cambio de vuelos de pasajeros y equipajes de una compañía aérea en caso de retrasos. Para conseguir este objetivo es necesario que los datos del Data Warehouse se actualicen en tiempo real.

El Análisis en Memoria se basa en guardar en memoria datos pre-calculados para obtener una mayor rapidez en futuras consultas[15]. Esto es posible gracias a los avances tecnológicos y el abaratamiento de los costes del hardware.

La Integración de BI con CPM (Corporate Performance Management) es la unión de la información proporcionada por las herramientas BI con las estrategias corporativas de la organización, llevando esto a tomar acciones concretas [15].

Como ya se ha comentado anteriormente la unión de BI con Internet [15][20][21] es una de las principales tendencias, tanto en su uso para obtener información (Web Mining) y detectar oportunidades de negocio (BI 2.0) como en la externalización de BI a través del Cloud Computing (Cloud BI).

Agile BI Governance es otra forma de concebir el BI, más enfocado al usuario y a la organización que a las TI. Empleando metodologías Ágiles (entregas frecuentes, diálogo, simplicidad,...) con las que el usuario se siente más cómodo y adaptando la solución BI a las necesidades de la organización, se consiguen mejores resultados y una mayor satisfacción del usuario[14].

Otra de las tendencias a tener en cuenta es la unificación de los diferentes Data Marts dentro de una organización en un único Data Warehouse [15][16]. Con esto se consigue evitar los silos de información existentes entre los diferentes departamentos dentro de la organización. Otro efecto de esta unión es que la información ya no se envía a los usuarios, sino que son estos los que acceden a través de diferentes herramientas a la información deseada, produciéndose así una democratización de la información dentro de la organización.

10. Conclusiones

En este artículo se ha expuesto el estado actual de los sistemas de información de los sistemas de BI y las herramientas que se pueden utilizar para realizar una implantación de dicho sistema.

Los sistemas BI no son la panacea, existen organizaciones en las que la implantación de un sistema BI no ha resultado exitosa y ello se debe principalmente a la mala o no definición de los factores críticos de éxito (CSF) que se pretenden alcanzar con la implantación del sistema. Un sistema BI bien definido aporta grandes beneficios, pero a veces estos no son fácilmente cuantificables, por ello se deben definir objetivos

realistas y sobre todo promocionar el uso del BI y sus ventajas entre los usuarios del mismo.

La implantación de los sistemas tradicionales de BI (normalmente suites comerciales) conllevan un alto coste tanto económico, como humano y temporal, lo que conlleva que las pequeñas y medianas empresas no se decidan por implantar este tipo de soluciones. Actualmente han aparecido en el mercado suites BI completas open source que requieren un menor coste de licencias, esto ha animado a las pymes a implantar soluciones BI open source en su empresa, si bien hay que remarcar que generalmente para la implantación de estas soluciones open source es necesario poseer mayores conocimientos técnicos que con las suites BI privativas y que el coste total de la implantación de la solución dependerá de cada caso concreto.

Las pymes también están empezando a utilizar el Cloud BI, a través del cual no hace falta realizar una gran inversión en infraestructura TI, ya que se ofrece el BI como un servicio. Este tipo de soluciones no es factible para todas las empresas, ya que el volumen de los datos a tratar, la sensibilidad de los mismos, los requisitos en los tiempos de accesos y la seguridad ofrecida en el Cloud BI pueden no satisfacer a todos los tipos de organización.

Como conclusión del presente artículo se puede apreciar que existen multitud de herramientas y tipos de soluciones BI para implantar en una empresa, pero que el tipo de herramienta seleccionada no es el principal factor determinante del éxito de la implantación, sino que el éxito dependerá principalmente de un buen conocimiento de la organización y la involucración de los usuarios en todo el proceso.

En futuras líneas de trabajo se puede investigar más acerca del Cloud BI y como van evolucionando los servicios ofrecidos a través de éste, para poder solventar los inconvenientes existentes tanto a nivel tecnológico como legal, debido a la sensibilidad de los tipos de datos. La explotación de las líneas de negocio planteadas a través del BI 2.0 en las redes sociales también sería una futura línea de investigación.

Referencias

- [1] Eduardo Nicolás Campazzo; Virginia Inés Santos. *Business Intelligence: Negocios Inteligentes para Empresas Inteligentes*. Universidad Nacional de la Rioja. Rene Favaloro esq. Laprida, CP5300, La Rioja, Argentina. http://www.imamu.edu.sa/DContent/IT_Topics/negocios_inteligentes_para_empresas_inteligentes.doc. 2006
- [2] Josep Curto Díaz. *Comparativa herramientas reporting open source*. *Stratebi* - http://www.telefonica.net/web2/todobi/Oct07/Reporting_OS.pdf . 2007
- [3] Chip Nickolett. *La Informática Decisional en Open Source (Business Intelligence)* .Ingres Corporation. <http://downloads.ingres.com/online/collaterals/wp/SP/DeliveringBusinessIntelligence-SP.pdf> .2008
- [4] Zack Jourdan; R. Kelly Rainer; Thomas E. Marshall. *Business Intelligence: An Analysis of the Literature*. Department of Management, College of Business, Auburn University, Auburn, Alabama, USA. *Engineering Management Review*, IEEE - http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4648888 . 2008
- [5] José Luis Regidor. *Reflexiones sobre el estado del arte del Business Intelligences*. Revista BIT - <http://dialnet.unirioja.es/servlet/dcart?info=link&codigo=1263196&orden=49819> Septiembre 2005
- [6] Christian Thomsen; Torben Bach Pedersen. *A Survey of Open Source Tools for Business Intelligence*. Department of Computer Science, Aalborg University. Lecture

- Notes in Computer Science, 2005, Volume 3589/2005 -
<http://www.springerlink.com/content/0djlyx6ywtzewch2/fulltext.pdf> 2005
- [7] William Yeoh; Andy Koronios. *Critical Success Factors for Business Intelligence Systems*. University of South Australia. SA 5095 Australia. Journal of computer information systems, 2010.
http://www.iacis.org/jcis/pdf/Yeoh_Koronios_2010_50_3.pdf . 2010
- [8] Stratebi (Los autores no aparecen en el documento). *Comparativa B.I. Open Source*. 2010
- [9] Roland Bouman, Jos van Dongen. *Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL*. ISBN: 978-0-470-48432-6. 2009
- [10] Jordi Conesa i Caralt (coord.); Josep Curto Diaz. *Introducción al Business Intelligence*. Universitat Oberta de Catalunya. ISBN: 9788497888868. 2010
- [11] Sixto Ortix Jr. *Taking Business Intelligence to the Masses*.
<Http://ComputingNow.computer.org>. 2010
- [12] Orhan Tuncer, Jan van den Berg. *Implementing BI concepts with Pentaho, an evaluation*. Delft University of Technology. Netherlands. 2010
- [13] Lyndsay Wise. *Cinco pasos para lograr un proyecto de BI exitoso*. TEC – Technology Evaluation Center. Junio 2007
- [14] J. Fernández, E. Mayol, J.A. Pastor. *Agile Business Intelligence Governance: Sus justificación y presentación*. Universitat Politècnica de Catalunya. Universitat Oberta de Catalunya.
- [15] Carmen Cámara Núñez. *Análisis de los Sistemas de Business Intelligent y su Aplicación Práctica en Proyectos Software*. Universidad Carlos III de Madrid. 2010
- [16] Josep Lluís Cano. *Business Intelligence: Competir con Información*. Depósito Legal: M-41185-2007. 2007
- [17] Jayanthi Ranjan. *Business Intelligence: Concepts, Components, Techniques and Benefits*. University of Management Technology, Ghaziabad, India. 2009
- [18] Hugh J. Watson. Barbara H Wixom. *The Current State of Business Intelligence*. University of Georgia, University of Virginia. Septiembre 2007
- [19] Estibaliz Rotaecche Cortés. *Business Intelligence. El Conocimiento Compartido* Ibermática. 2007
- [20] Marinela Mircea, Bogdan Ghilic, Marian Stoica. *Combining Business Intelligence with Cloud Computing to delivery agility in actual economy*. Department of Economic Informatics. The Bucharest Academy of Economic Studies. 2011
- [21] Gregory S. Nelson. *Business Intelligence 2.0: Are we there yet?*. ThotWawe Technologies, Chapel Hill North Carolina. 2010
- [22] Alexandro Salinas. *Introducción a Pentaho*.
<http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/> . 2008