

Herramientas de Explotación de Información de Bases de Datos

Gabriela Torres Rodríguez¹, Oscar Hilario Salinas Aviles¹, Ángel Estrada Arteaga¹, y Joquebed Guerrero González¹

^{um} Centro de Desarrollo de Software de la Universidad Tecnológica Emiliano Zapata del Estado de Morelos (CDS-UTEZ),
Emiliano Zapata, Morelos.
{gabrielatorres, oscarsalinas, angeleestrada, joquebedguerrero}@utez.edu.mx

RESUMEN

Se realizó la prueba de factibilidad para la implementación de la herramienta SpagoBI en el Centro de Desarrollo de Software de la Universidad Tecnológica Emiliano Zapata (CDS-UTEZ). Los resultados indican que es posible introducir en el CDS-UTEZ la nueva tendencia de análisis y manejo de datos llamado inteligencia de negocios (BI, por sus siglas en inglés *Business Intelligence*). SpagoBI fue seleccionado debido a que es el que mejor se adapta a la dinámica de trabajo del CDS-UTEZ. De las características de esta aplicación la que se probó, en este trabajo inicial, es la conectividad a diferentes sistemas gestores de bases de datos y la generación de reportes. En nuestro caso el BI se complementó siguiendo las mejores prácticas del PMBOK para la administración de proyectos.

Palabras clave: Inteligencia de negocios, bases de datos, información, administración.

1 Introducción

En la actualidad se genera información de manera considerable, ya sea en papel o de forma electrónica dentro de pequeñas, medianas y grandes empresas. Con el transcurso del tiempo se obtienen grandes cantidades de datos que adquieren un valor importante en cualquier industria para la toma de decisiones, convirtiendo la información como un activo más de las organizaciones.

Por tal razón surge el área de la inteligencia de negocios, donde se tiene como principal función optimizar la toma de decisiones utilizando diferentes tecnologías que entre sus actividades permitan reunir, almacenar y analizar información que se encuentra guardada por largos periodos dentro de una empresa [1].

Cada año en el Centro de Desarrollo de Software de la Universidad Tecnológica Emiliano Zapata (CDS-UTEZ), se realizan proyectos externos e internos donde se generan soluciones a la medida de las necesidades de los clientes. Es muy frecuente que después de un tiempo el mismo cliente vuelva con la necesidad de extraer

información más detallada sobre su organización, la cual está almacenada en las bases de datos de la solución inicialmente propuesta, dejando de lado con esto los reportes solicitados en un principio. Esto provoca la necesidad que los usuarios finales manipulen directamente las bases de datos, generándoles más costos o incluso que desistan y eliminen la solución inicial.

En este artículo se presentan los resultados de una investigación realizada a diferentes herramientas de explotación de bases de datos orientadas a la BI y los resultados obtenidos de una de ellas. En la sección 2 la situación actual del CDS-UTEZ. En la sección 3a qué se refiere la inteligencia de negocios. En la sección 4 las diferentes herramientas de explotación de información de bases de datos. En la sección 5 la implementación de SpagoBI. En la sección 6 casos de prueba y resultados. En la sección 7 las restricciones presentadas para el proyecto. Y, finalmente en la sección 8 los trabajos futuros a realizar en SpagoBI dentro del CDS-UTEZ.

2 Situación actual

El CDS-UTEZ realiza proyectos de desarrollo y mantenimiento de software tanto de proyectos internos como externos. Con el transcurso del tiempo cada uno de los clientes que solicitó una solución ha generado grandes cantidades de datos, por lo cual surgen nuevas necesidades basadas en el histórico de la información donde se requiere responder a preguntas concretas a través de reportes con características cuantitativas y cualitativas, como por ejemplo la producción de un producto y su éxito en el mercado durante un cierto periodo, llevando esto a proporcionar soluciones más versátiles en la recuperación de la información y así hacer la consideración a la implementación de herramientas de explotación de información de bases de datos con un enfoque en la BI.

3 Inteligencia de Negocios (BI)

La BI se puede definir concretamente como la habilidad de transformar los datos en información, y la información en conocimiento, de tal forma que se pueda optimizar el proceso de toma de decisiones en los negocios. Para ello deben responderse los siguientes cuestionamientos cuando se inicia el involucramiento con el tema:

1. ¿Cuántas personas disponen de más información y de menos tiempo para sus análisis?
2. ¿Los sistemas generados proporcionan una solución factible a la hora de tomar decisiones?
3. ¿Los responsables de generar información directiva están desbordados por las peticiones de información urgente, continua y no coordinada?

La respuesta a la primera pregunta es unánime: cada vez se tiene más información y menos tiempo para analizarla; La generación de la información proviene de diferentes

fuentes tanto internas como externas, por lo cual es obvio que los cambios que se realicen en los mercados sea vertiginosa. En la segunda pregunta variará la respuesta de acuerdo a la madurez de los sistemas de información de las organizaciones, algunos cubrirán cada una de las necesidades y otros no, pero cuando se llega a aunar en cada uno de ellos se puede observar que no se están planteando las preguntas adecuadas ni las tecnologías idóneas. La respuesta de la tercera pregunta es: están desbordados, no todos los responsables de las áreas definen sus necesidades de información de la misma forma, en algunos de los casos la definición de los conceptos de gestión son distintos [2].

Las nuevas condiciones que operan sobre las organizaciones y la gestión (en particular aquellas derivadas del intercambio con el entorno, de la complejidad de los negocios y de la inmediatez de las actividades comerciales), han generado la necesidad de nuevos desarrollos, capaces de producir información en un contexto multivariado, dinámico, incierto e instantáneo. Como respuesta a esta creciente necesidad, ha surgido una nueva área de desarrollo, BI, que ha permitido modelar sistemas automáticos de monitoreo, seguimiento y control. Sistemas diseñados para administrar y relacionar bases de datos que permitan retroalimentar en forma instantánea la toma de decisiones, es decir, sistemas automáticos, en línea e instantáneos, que informan sobre las variables del negocio y que anticipan tendencias. Sistemas que en forma continúa están monitoreando diversas variables de acuerdo a parámetros de control y modelando respuestas por medio de la simulación de escenarios que pueden ser informados automáticamente [3].

4 Herramientas de explotación de información de bases de datos

En esta sección se presentan las diferentes herramientas de explotación de bases de datos investigadas dentro del área de BI.

Ariacom: Es una base de datos de información y una herramienta multidimensional de análisis de generación de consultas SQL (Structure Query Language) dinámicas, donde usuarios sin conocimientos técnicos pueden ejecutar y realizar reportes complejos desde cualquier base de datos relacional [4]. Actuate BIRT: Es una oferta comercial basada en el código abierto de BIRT (Business Intelligence and Report Tools) y los beneficios de las capacidades de la plataforma Eclipse y una muy activa comunidad de usuarios. Actuate BIRT proporciona la conveniencia de instalación con un solo botón, y soporta los despliegues de producción a través de Actuate-backed [5]. JasperSoft Business Intelligence Suite: Es la denominación del conjunto de herramientas de JasperSoft que permiten a una organización generar información basada en sus propios datos de gestión para la evaluación y toma diaria de decisiones, en forma dinámica y en línea. El marco de trabajo de JasperSoft permite integrar fácilmente las diversas fuentes de datos disponibles en la empresa, y por medio de técnicas de análisis multidimensional obtener indicadores que, presentados en tableros de control y reportes dinámicos, proveen de esta sensible información a la alta gerencia [6]. PENTAHO: Es una plataforma de BI orientada a la solución y centrada en procesos. Pentaho consiste en una Suite Completa de Inteligencia de negocio que

incluye todos los principales componentes requeridos para implementar soluciones de BI (reportes, cuadros de mando, integración de datos, etc.) [7]. SpagoBI: Es la OSBI suite (Open Source BI) que marca la diferencia en el campo de código abierto. Cubriendo todas las áreas analíticas de los proyectos de BI, apropiado para cualquier requerimiento y es la solución adecuada para ambientes empresariales, proporcionado con un completo rango de soporte de servicios [8,9]. SQL Power BI: Ofrece una completa suite de herramientas de productividad innovadores que permiten simplificar la minería de datos, inteligencia de negocios y medición de desempeño. Dichas herramientas ayudan a los equipos de desarrollo a facilitar el diseño de bases de datos, desarrollo ETL, limpieza de datos y reportes robustos de BI [10]. A continuación se muestra un cuadro comparativo resumiendo las diferentes características de las herramientas de BI investigadas:

Tabla 1. Cuadro comparativo de herramientas de explotación de información de bases de datos, BI.

Módulo	Ariacom * Producto comercial	ActuateBirt *Opciones completas en su versión comercial	JasperSoft *Opciones completas en su versión comercial	Pentaho	SpagoBI	SQLPower
Servidor de Aplicación	Aplicación Conexión Web a través del uso de un complemento	Integrado *iServer	JBoss	JBoss, Tomcat	JBoss, Tomcat.	Integrado
Colaboración BI	Integrado	LDAP	-	-	Dossier	Integrado * Versión Enterprise
Cuadro de mando	-	-	JFreeChart	JFreeChart	Openlaszlo	Integrado
Minería de datos	Integrado	Integrado	-	Weka	Weka	-
Sistema Manejador de Base de Datos	MS Access, MS SQL Server, Oracle y MySQL *Conexión con otras bases de datos, requiere la instalación del controlador	Informix, Oracle, PostgreSQL, DB2.	MySQL, Oracle, MSSQL Server, PostgreSQL, etc.	MySQL, Oracle, MSSQL Server, PostgreSQL, etc.	MySQL, Oracle, MSSQL Server, PostgreSQL, etc. *Conexión con otras bases a través de Drivers (archivos .jar).	Oracle, MySQL, MS SQL Server, SQL Stream, PostgreSQL, HSQLDB

Módulo	Ariacom * Producto comercial	ActuateBirt *Opciones completas en su versión comercial	JasperSoft *Opciones completas en su versión comercial	Pentaho	SpagoBI	SQLPower
	ODBC.					
ETL	Integrado	-	JasperETL	Pentaho Data Integration	Talend Open Studio	-
Geo- referencia	-	-	Google Maps	Google Maps	GEO	-
Cronograma de trabajo	Integrado	Integrado	Quartz	Quartz	Quartz	Integrado * Versión Enterprise
OLAP ¹	Integrado	Integrado como motor interno en tablas cruzadas	Mondrian y JPivot	Mondrian y JPivot	Mondrian y JPivot	Mondrian
Reportes	Integrado	ActuateBirtRe portDesigner	JasperReport	Pentaho Report Designer, JasptReport, Birt	JasperReport, Birt	Integrado
Single SingOn (SSO) ²	Integrado	-	Acegi	CAS	CAS	-

De acuerdo al Tabla 1, JasperSoft, Pentaho y SpagoBI son las más completas a nivel de funcionalidad, la diferencia es que con SpagoBI realizar la implementación de todas las características es completamente libre y abierto a personalización del usuario. Permitiendo con esto incluir un sentido de identidad y pertenencia de la adaptación de la aplicación hacia la organización y/o usuarios.

¹OLAP (Procesamiento Analítico En Línea): categoría de herramientas de software que proporcionan análisis de datos almacenados en bases de datos de forma sofisticada y con la capacidad de realizar cálculos complejos. Las herramientas OLAP permiten a los usuarios analizar diferentes dimensiones o multidimensiones de datos. Por ejemplo, proporcionan series en tiempo y tendencia de análisis de vistas.

²Single sign-on (SSO) es un procedimiento de autenticación que habilita al usuario para acceder a varios sistemas con una sola instancia de identificación.

5 Implementación de SpagoBI

SpagoBI es una herramienta de BI que cubre varias funcionalidades que normalmente sólo corresponderían a versiones comerciales de otra aplicación, caracterizándolo su enfoque 100% de código abierto, multiplataforma, soporte a diferentes Sistemas Manejadores de Bases de Datos y personalización de la aplicación para una organización en particular.

5.1 Instalación

Para su instalación es necesario que el servidor cumpliera con las siguientes especificaciones:

1. JDK 1.5 o superior.
2. Sistemas operativos soportados: cualquiera que soporte JVM 1.5 o superior, ejemplo: Windows o Linux.
3. Sistema Gestor de Bases de Datos para Metadatos: MySQL, Postgres, Oracle, HSQL, MSSQL Server.
4. Servidor de aplicación: servidores que soporten J2EE, como por ejemplo Tomcat, JBoss, WebSphere, etc.

Entre las configuraciones que se realizaron, se puede destacar la parametrización del archivo `server.xml`, donde especifica la conexión a la base de datos del servidor para la instalación de SpagoBI, Figura 1, la generación de variables de entorno, administración de hilos para el pool de conexiones, entre otras.

```
<Resource name="jdbc/spagobi" auth="Container" type="javax.sql.DataSource"
maxActive="100" maxIdle="10" maxWait="-1"
username="root" password="admin" driverClassName="com.mysql.jdbc.Driver"
url="jdbc:mysql://localhost:3306/spagobi"/>
```

Fig. 1. Configuración de *DataSource* en el archivo `server.xml`.

5.2 Personalización

Como ya se mencionó, Spago BI permite personalizar de manera institucional y de usuario el ambiente de trabajo. El desarrollador puede generar temas que identifiquen al usuario con la institución para la cual está laborando, creando sentido de identidad y pertinencia. En la Figura 2, se observa la ventana de trabajo con la identidad institucional del CDS – UTEZ.

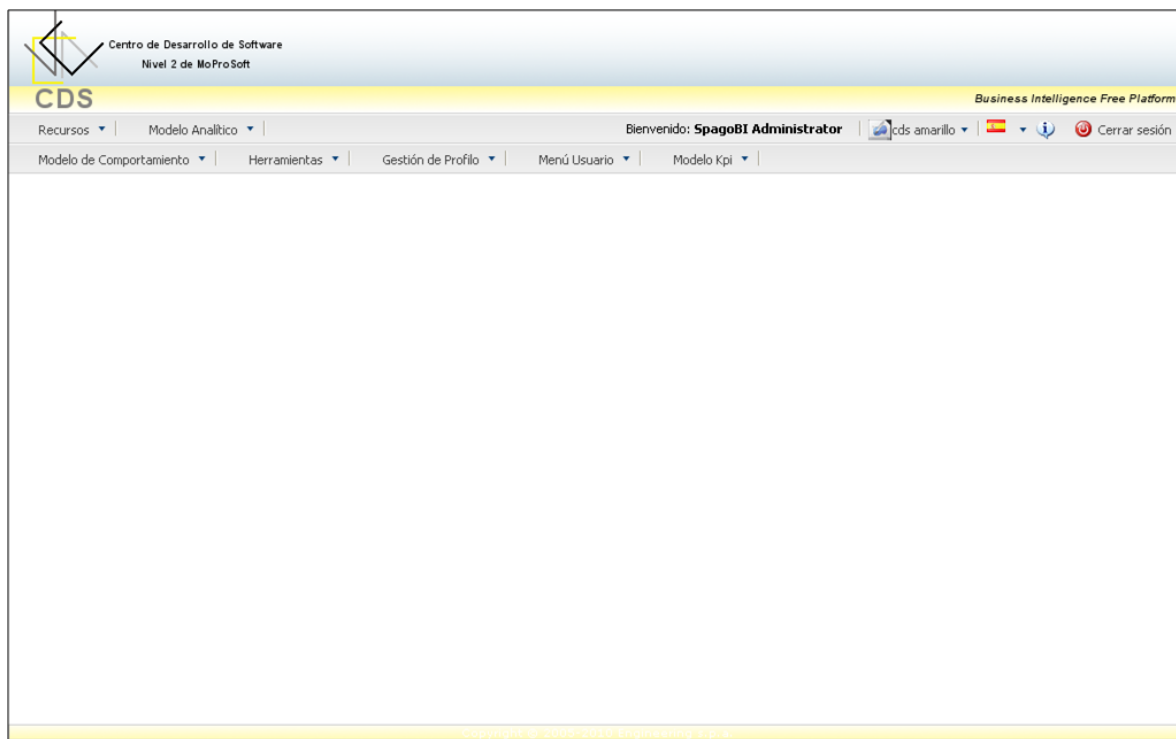


Fig. 2. SpagoBI, personalizado para el CDS-UTEZ.

A continuación se muestran algunas líneas de código las cuales fueron configuradas para la implementación del tema en la aplicación.

```
<THEMES>
  <THEME name="cds_theme" default="false" view_name="cds" ext_theme="ext-all.css" />
  <THEME name="sbi_default" default="false" view_name="default" ext_theme="xtheme-gray.css" />
  <THEME name="cds_yellow_theme" default="true" view_name="cds amarillo" ext_theme="ext-ux-slidezone.css" />
  <!-- <THEME name="other_theme" default="false" view_name="other" /> -->
</THEMES>
```

Fig. 3. Líneas de configuración para implementación de temas personalizados en SpagoBI.

6 Casos de prueba y resultados

Los casos de prueba se elaboraron basándose en la generación de reportes donde se explota la recuperación de información de diferentes de bases de datos reales de manera analítica. Para la generación de reportes, se desarrollaron tres diferentes reportes en los cuales se presentaban resultados de diferentes proyectos desarrollados en CDS-UTEZ. Por ejemplo se tomó como caso de prueba la base de datos de una

institución educativa, donde se requería que se mostraran los alumnos asignados por materia, donde el reporte debería especificar el número de alumnos que se encontraban cursando determinada materia, así mismo se determinaba el sexo, masculino o femenino, con el fin de definir si eran mayor el número de mujeres u hombres que cursarán determinada materia, los resultados obtenidos fueron los siguientes:

1. Se implementó como IDE de generación de reportes BIRT, integrado como complemento dentro del IDE de Eclipse. En este punto se generó una plantilla con los datos que se requerían consultar.
2. Al contar con la plantilla se llevaron a cabo las configuraciones necesarias dentro de SpagoBI para la carga de datos que fueran necesarias de la plantilla.
 - (a) Configuración de parámetros.
 - (b) Asignación de permisos de visualización a usuarios.
3. Se instaló la plantilla a la aplicación de SpagoBI.
4. Se realizó la prueba de visualización de un usuario que no fuera el administrador. Se obtuvo un reporte de alumnos asignados a una materia en el cual se muestra la materia y la cantidad de alumnos asignados a la misma. El reporte se puede obtener en forma de tabla, Figura 4 y/o tipo histograma, Figura 5.



REPORTE **ALUMNOS ASIGNADOS POR MATERIA**

Materia	Número de alumnos
Administración	1
Computación	1
Francés	2
Negociació E.	1

Fig. 4. Vista del reporte generado desde SpagoBI.

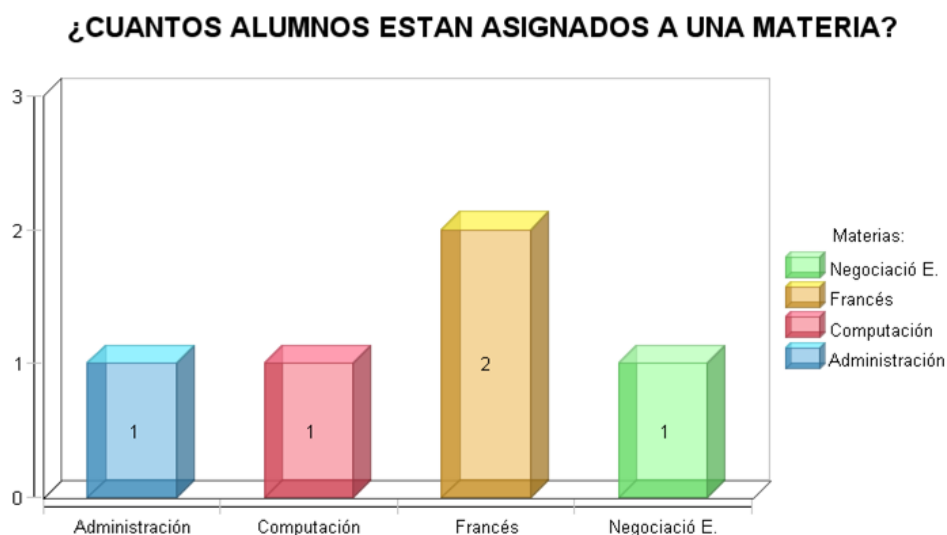


Fig. 5. Gráfica generada con los mismos datos y reutilizando la configuración del reporte anterior desde SpagoBI.

En la Figura 6 se pueden observar los pasos secuenciales que se realizaron para los casos de prueba a nivel de generación de reportes. Se ingresa a la aplicación con el nombre de usuario y contraseña, se realiza la conexión a la base de datos que se desea manipular, posteriormente se carga la plantilla correspondiente a la información que se desea recuperar, y finalmente se configuran los parámetros para la plantilla.

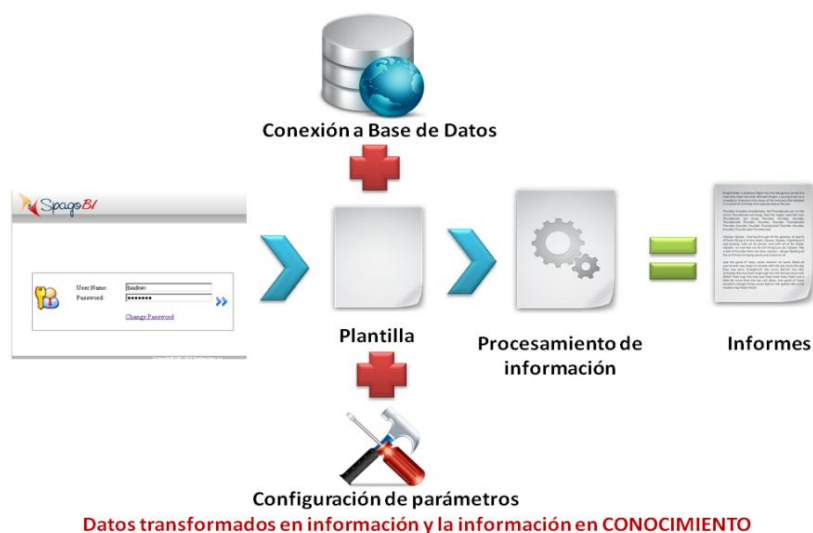


Fig. 6. Arquitectura de implementación de reportes.

Los resultados con pruebas tales como la anterior fueron exitosos, cabe mencionar que cada una de ellas fueron desarrolladas de manera local en un equipo de cómputo de escritorio, para que las pruebas sean completas se recomienda que se realicen en un servidor de producción.

7 Restricciones

La investigación de herramientas de explotación de información de bases de datos no pretende cubrir con todas las herramientas existentes en el mercado, considerando que día con día hay avances tecnológicos que pueden superar a las actuales. Asimismo la herramienta de BI a implementar dependerá de la organización ya sea que esté definida por el costo de la aplicación, el tamaño de la empresa, funcionalidades, etc. Para el caso del CDS-UTEZ se realizó la investigación de herramientas de BI que cubrieran principalmente la generación de reportes analíticos de manera completa y que su implementación inicial fuera probando con reportes comunes solicitados por los clientes.

Conclusiones

Es posible la implementación de una herramienta como SpagoBI para la explotación de las bases de datos del CDS-UTEZ, tanto de los proyectos actuales así como de los que ya concluyeron. La implementación de una herramienta como la propuesta permite a empresas como el CDS-UTEZ, explotar información diversa contenida en bases de datos de cualquier tipo, ajustándola a las necesidades y requerimientos de los clientes o de la misma empresa; pudiendo hacer extracciones o consultas en más de una ocasión, siempre y cuando la información permanezca en la base de datos. SpagoBI tiene otras muchas funcionalidades, sin embargo en esta etapa del proyecto, se eligió la generación de reportes como la prueba de factibilidad para la implementación.

Trabajo futuro

Continuar con la implementación de la funcionalidades de SpagoBI en una primer etapa como prueba y finalmente concluyendo con su puesta en marcha para toda la organización. Para ellos se iniciará con la realización de reportes analíticos más complejos y posteriormente la combinación de estos con el resto de funcionalidades.

Agradecimientos

Agradezco al CDS-UTEZ por proporcionarme una beca durante el desarrollo del proyecto y al Cuerpo Académico de Desarrollo de Software por su orientación para la implementación de la herramienta.

Referencias

- [1] López, G.: Inteligencia de Negocios. *Negocios Internacionales* , 1-2. (2011)
- [2] Lluís Cano, J.: *Bussiness Intelligence: Competir con Información*. ESADE (2007)
- [3] Chalico, C., LI, CISA, CISSP, & CISM.: Valor, Riesgos y Crisis. *Puntos Finos* , 20-27. (2009)
- [4] Business Reports, <http://www.ariacom.com/brBusinessReports.asp>
- [5] Birt Exchange, <http://www.birt-exchange.com/be/home/>
- [6] JasperSoft, <http://www.jaspersoft.com/js4-webinar-series>
- [7] pentahoCommunity, <http://wiki.pentaho.com/>
- [8] SpagoWorld,
<http://fisheye.ow2.org/browse/~raw,r=2783/SpagoBI/trunk/SpagoBIIns taller /spagobi-docs/QuickStart-0.9.3.pdf>
- [9] SpagoBI WIKI, http://wiki.spagobi.org/xwiki/bin/view/spagobi_server/
- [10] DATAPRIX, <http://www.dataprix.com/empresa/productos/sql-power-software>
- [11] talend open data solutions, <http://es.talend.com/solutions-data-integration/etl-for-analytics.php>