

Envío: 16-10-2013

Aceptación: 05-11-2013

Publicación: 27-12-2013

SISTEMA DE INTELIGENCIA DE NEGOCIOS PARA ACUEDUCTO Y ALCANTARILLADO

BUSINESS INTELLIGENCE SYSTEM FOR WATER AND SEWERAGE

Eriberto Vanegas Lago¹

1. Ingeniero especialista en desarrollo de sistemas de inteligencia de negocios y de información geográfica. Facultad de Ciencias Informáticas de la Universidad de Granma. Centro de desarrollo, departamento de soluciones de gestión. Cuba. E-mail: evanegas@grm.uci.cu

RESUMEN

Gran parte del éxito de las empresas modernas depende del aprovechamiento de sus recursos intangibles. El correcto flujo de los datos e información es vital para un acertado proceso de toma de decisiones. Debido al aumento del volumen de los datos almacenados, los directivos se enfrentan a un ambiente de incertidumbre y complejidad crecientes. Muchas veces no se cuenta con las herramientas necesarias para manipular eficientemente estos datos y convertirlos en información que genere conocimiento. Este trabajo tiene como objetivo el desarrollo de un Sistema de Inteligencia de Negocios que permita almacenar, procesar, analizar y mostrar de manera eficiente, los datos que se generan en los departamentos de las empresas de acueducto y alcantarillado, mejorando de esta forma el apoyo al proceso de toma de decisiones. Se propone utilizar la metodología HEFESTO para el desarrollo de almacenes de datos y la suite del Pentaho Community Edition en su última versión estable (4.8.0). El objeto de estudio práctico es la Empresa de Acueducto y Alcantarillado Granma.

ABSTRACT

Much of the success of modern enterprises depends on the use of intangible resources. The flow of data and information is vital to the successful decision making process. Due to the increased volume of data stored, the managers face an environment of uncertainty and increasing complexity. Many times we do not have the tools to handle this data efficiently into information and to generate the necessary knowledge. The main goal of this work is to develop a Business Intelligence System that allows to store, process, analyze and display efficiently the data generated in the departments of a random company, and thereby, to improve the decision making process. It is proposed to use the HEFESTOS methodology for developing data warehouses and the Pentaho Suite Community Edition (version 4.8.0 or a higher stable version). The practical study object is the Enterprise of Water and Sewerage Granma.

PALABRAS CLAVE

Decisiones, información, inteligencia de negocios.

KEYWORDS

Decision, information, bussines intelligence.

1. INTRODUCCIÓN

En la Empresa de Acueducto y Alcantarillado de la provincia Granma (EAALG), Cuba, se desarrolló un sistema de gestión que responde a la necesidad de una herramienta informática capaz de gestionar de manera rápida y eficiente la enorme cantidad de datos que allí se manejan. Hasta ese momento, el control de esos datos, así como la información generada, se almacenaba en hojas de cálculo.

El sistema, desarrollado completamente con tecnologías y herramientas libres, permite gestionar los datos de los procesos departamentales, relacionados con los equipos de transporte (tanto de construcción como los complementarios), los grupos electrógenos, la energía, el combustible y las estaciones de tratamiento, bombeo y saneamiento de agua, así como las plantas potabilizadoras. A partir de la información que se genera en cada una de estas áreas se obtiene una serie de reportes que permiten verificar el funcionamiento general de la empresa.

Aunque fueron resueltos los problemas que condujeron al desarrollo del sistema de gestión, como parte de su perfeccionamiento, mejora y evolución, se han identificado limitaciones y necesidades que no pueden ser solucionadas.

Los Sistema de Procesamiento de Transacciones (OLTP, por sus siglas en inglés), dentro de los que se incluyen los sistemas de gestión, son la solución típica cuando es necesario gestionar datos; pueden ser desarrollados y manejados fácilmente. Sin embargo no son capaces de resolver muchas de las necesidades actuales de las empresas. Estas limitaciones son inherentes a estos sistemas. Entre ellas se pueden mencionar:

- no permiten análisis detallados de los datos almacenados;
- deficiente gestión de los reportes;
- no responden a las necesidades específicas de información para los directivos;
- no apoyan eficientemente al planeamiento ni el proceso de toma de decisiones;
- no generan ni evalúan de forma sistemática las alternativas y escenarios previstos;
- no están optimizados para trabajar con grandes volúmenes de datos. A medida que estos aumentan también lo hace el consumo de recursos de hardware y el tiempo de respuesta de las consultas.

Debido a las limitaciones identificadas se plantea como problema científico de esta investigación *la contradicción existente entre la disponibilidad de datos en el sistema de gestión de Información de la EAALG y la ausencia de reportes que muestren la información necesaria para apoyar el proceso de toma de decisiones.*

El análisis anterior no indica que haya sido un error desarrollar dicho sistema de gestión. Este resolvió todos los problemas que estaban dentro de su alcance. La presente investigación se enfoca solamente en la necesidad de utilizar otro tipo de tecnología para el procesamiento de los datos y su conversión en información útil. Esto será posible si se desarrolla un sistema que

resuelva las limitaciones antes mencionadas y además las nuevas necesidades de la empresa, entre las que se encuentran:

- herramienta de fácil uso para directivos y demás trabajadores;
- reportes dinámicos e interactivos;
- información con alto grado de detalle;
- varias perspectivas de análisis de la información;
- tiempos mínimos de respuestas del sistema.

En correspondencia con lo planteado se propone como hipótesis investigativa que *el desarrollo e implantación de un Sistema de Inteligencia de Negocios (BIS, por sus siglas en inglés) para las condiciones específicas de la EAALG permitirá mejorar la disponibilidad de información para el apoyo al Proceso de Toma de Decisiones (PTD).*

Se define como objeto de estudio teórico de esta investigación la *Inteligencia de Negocios*. Tomando como objeto de estudio práctico a la EAALG.

Para dar solución al problema planteado se propone como objetivo general *desarrollar un BIS bajo las condiciones actuales de la EAALG, que soporte toda la información necesaria para el apoyo al PTD*. Como campo de acción se identifica *los BIS*.

Entre los objetivos específicos trazados para la investigación se encuentran:

- Definir la metodología y herramientas para el desarrollo del BIS
- Definir y desarrollar los Mercados de Datos (DM, por sus siglas en inglés) que conformarán del Almacén de Datos (DWH, por sus siglas en inglés) final.
- Diseñar y desarrollar los reportes que faciliten la visualización y análisis de los datos contenidos en el almacén.

La investigación propuesta está encaminada a desarrollar e implantar un BIS que soporte la información necesaria para el PTD en la organización objeto de estudio.

2. GESTION DE LA INFORMACION

A nivel mundial los directivos de las organizaciones modernas están convencidos de que la información es el activo estratégico que los diferencia de sus competidores e impulsa su éxito. Una encuesta realizada en el año 2010 por la revista Forbes demostró que el 85% de los encuestados coinciden en que actualmente la información es considerada como un activo estratégico, y la gran mayoría (95%) considera que la gestión de información es esencial para el éxito del negocio [1].

Una adecuada gestión de información permite monitorizar constantemente el comportamiento de las principales variables que intervienen en el negocio. En dependencia de los cambios, será posible tomar las medidas necesarias en el momento preciso, que permitan adaptarse a las nuevas condiciones y sacar provecho de las mismas. De esto se infiere que uno de los procesos más beneficiados es la toma de decisiones.

En la actualidad el PTD recibe un gran apoyo por parte de las Tecnologías de la Información y las Comunicaciones (TICs). Han surgido muchas herramientas con el objetivo de mejorar y hacer más competitivos los ambientes empresariales. La mayoría de las etapas del proceso estudiado pueden ser soportadas por dichas herramientas, permitiendo que la eficacia de las decisiones dependa en mayor medida de la cantidad y calidad de los datos disponibles. El término que engloba esta tendencia actual surge a mediados de los años 90 y se conoce como Inteligencia de Negocio (BI, por sus siglas en inglés).

3. INTELIGENCIA DE NEGOCIOS

En el mundo de los negocios, los tomadores de decisiones necesitan tener acceso a información precisa y oportuna para lograr sus objetivos. Históricamente la BI había sido utilizada por los analistas para procesar los datos haciendo uso de herramientas complejas y hojas de cálculo. En la actualidad la toma de decisiones implica una amplia gama de roles de negocio. Los mayores vendedores de BI se centran en proporcionar suites completas, lo que permite a los tomadores de decisiones acceder a los datos de origen en casi cualquier ambiente [2].

Mediante el uso de la BI se logra unir el mundo de los datos y el de los negocios. Esta permite a las empresas analizar grandes cantidades de datos de forma rápida y sencilla, que pueden ser procesados utilizando diversas reglas o criterios del negocio en cuestión, de forma inteligente. La BI integra los datos y puede cruzarlos entre varios procesos, evitando de esta forma los análisis aislados e incompletos, a la vez que facilita y apoya el PTD.

Para Negash la BI es un conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información no estructurada (interna y externa a la compañía) en información estructurada, para su explotación directa o para su análisis y conversión en conocimiento [3]. Carlos Gameiro plantea que la BI agrupa un conjunto de sistemas que combinan la recopilación, adquisición y almacenamiento de datos de diferentes fuentes con herramientas analíticas, presentándolos en el orden y la forma lógica de los tomadores de decisiones, generando una rápida vista de la situación de negocios en el pasado, presente y futuro [4].

A partir de los distintos puntos de vistas de estos y otros autores consultados, se puede resumir que la BI es el resultado de la estrecha relación entre las metodologías y herramientas que permiten el procesamiento completo de los datos (captura, almacenamiento, tratamiento y visualización) hasta su conversión en información; todo esto con el objetivo de dotar a los tomadores de decisiones de los análisis necesarios que permitan conocer y estudiar el pasado, controlar el presente y prever el futuro de las organizaciones.

4. ARQUITECTURA DE INTELIGENCIA DE NEGOCIOS

De igual forma que el concepto de BI, varios autores han propuesto arquitecturas básicas que presentar los BIS. A continuación se analizan los considerados por el autor como más significativos.

Vercellis define que la arquitectura de los BIS está compuesta por los siguientes elementos (figura 1): las fuentes de datos, los DWH, DM y las metodologías [5].

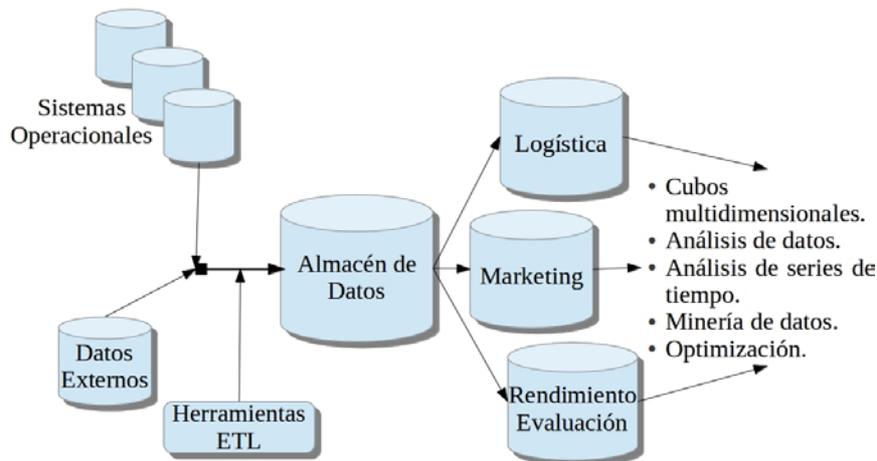


Figura 2. Arquitectura típica de inteligencia de negocio [5].

Teniendo en cuenta los entornos heterogéneos de hoy en día, donde existen sistemas diferentes y los dominios tienen diferentes partes de los datos necesarios, Liya, Gilad y Claudio definen la arquitectura de BI como se muestra en la figura 2 [6].



Figura 2. Arquitectura de BI [12].

Luego de comparar las arquitecturas presentadas anteriormente y otras estudiadas, se propone que los elementos coincidentes que deben estar presentes en las arquitecturas de los BIS, son los siguientes: fuentes de datos, proceso de ETL, los DM y DWH, las herramientas de OLAP, la presentación de informes, la minería de datos y los cuadros de mando, señalando que los dos últimos no son obligatorios.

Luego de definir las pautas principales que deben tenerse en cuenta para el desarrollo del BIS, es necesario definir el conjunto de buenas prácticas que guiarán el proceso de desarrollo.

5. METODOLOGIA DE DESARROLLO DE SOFTWARE

Diferentes tipos de sistemas requieren diferentes procesos de desarrollo. En consecuencia, estas actividades genéricas deben ser organizadas de diferentes formas y descritas a diferentes niveles de detalle. De cualquier manera, el uso de un proceso inapropiado puede reducir la calidad y usabilidad del software producido, lo que se traduce en aumento de los costos [7], [8].

Desde el surgimiento de los DWH se han propuesto varias metodologías para su desarrollo. Luego del estudio de varios de los principales enfoques, el autor de esta investigación decidió guiarse por los principales exponentes de este tipo de tecnología: William H. Inmon y Ralph Kimball. Luego de comparar los enfoques propuestos por cada autor se optó por utilizar el de Kimball. Este plantea la definición y construcción de cada uno de los DM para luego integrarlos en el DWH hasta que el sistema esté completo. No se necesita tener previo conocimiento del negocio. La metodología se caracteriza por estar basada en experimentos y prototipos. Este es un método flexible que permite disminuir los costos [9], [10].

Durante la investigación se identificaron otras metodologías que no siguen estrictamente uno de los enfoques antes mencionados, sino que toman las mejores prácticas de alguno de ellos. Entre ellas se destaca la metodología HEFESTO de la cual se utilizaron las mejores prácticas. La metodología final está compuesta por seis fases: 1) inicio, 2) análisis de los requerimientos, 3) análisis de las fuentes de datos, 4) modelado del DWH, 5) integración de datos y 6) representación de la información.

Luego de definida la metodología se procede a definir las herramientas que soportaran el desarrollo del BIS.

6. HERRAMIENTAS PARA EL DESARROLLO

En el caso específico de esta investigación, que se desarrolla en el marco de las empresas cubanas, es necesario incluir un criterio importante y hasta cierto punto obligatorio a la hora de definir las herramientas a utilizar: la condición de ser software libre o, al menos, de código abierto. Esto responde a la necesidad de lograr la independencia y soberanía tecnológica, ante la imposibilidad de acceder a las herramientas privativas y de pagar los altos precios de las licencias.

Varios autores coinciden en que los principales BIS de código abierto son: *SpagoBI* y *Pentaho*. Existen varias comparaciones entre estas herramientas teniendo en cuenta aspectos como la representación de gráficos, reportes, cuadros de mando, ETL, OLAP, minería de datos y otros [11], [12]. Finalmente, y tomando como base las comparaciones mencionadas, el autor del presente trabajo utilizó la suite comunitaria de Pentaho en su versión estable más reciente (4.8.0) para el desarrollo del BIS bajo las condiciones actuales de la EAALG. El DWH estará almacenado en PostgreSQL en su versión 9, este será administrado desde el PgAdmin III.

7. RESULTADOS

Una de las ventajas del uso de la metodología HEFESTO es que puede ser utilizada tanto para DM como para DWH. Como en este caso se optó por el enfoque de Kimball, uno de los primeros pasos para el desarrollo del BIS fue definir cada uno de los DM que conformarían el DWH final, así como delimitar claramente el alcance de cada uno.

De dos departamentos que existen en la EAALG se identificaron ocho MD, cuatro por cada departamento:

Departamento de energía y transporte:

- Grupos electrógenos.
- Transporte.
- Combustible.
- Energía.

Departamento de operaciones:

- Estaciones de saneamiento.
- Plantas potabilizadoras.
- Estaciones de bombeo.
- Estaciones de tratamiento.

Las principales actividades realizadas en cada fase para cada DM fueron las siguientes:

1. identificar necesidades de información de cada área, así como perspectivas e indicadores;
2. determinar estado de los sistemas fuentes;
3. conformar los indicadores (utilizando el procedimiento de normalización lineal);
4. crear modelos para pronósticos utilizando series de tiempo;
5. identificar dimensiones y hechos;
6. diseñar el modelo físico;
7. mapear y cargar los datos;
8. diseñar reportes y análisis de datos en las herramientas de análisis.

Luego de realizar cada una de las actividades mencionadas anteriormente se completó el BIS. Este presenta las siguientes funcionalidades:

- gestión de usuarios y roles;
- gestión de reportes;

- análisis de la información;
- gestión de fuentes heterogéneas;
- creación de cuadros de mando integrales;
- configuración y guardado de consultas.

Es necesario destacar que este sistema se caracteriza además por ser de fácil instalación, configuración y explotación. Presenta una arquitectura cliente-servidor flexible y de alta calidad. La tecnología que lo soporta lleva bastante tiempo de desarrollo. Puede integrarse de forma rápida a la infraestructura existente debido a que es multiplataforma y de código abierto. Puede mencionarse el alto desempeño, estabilidad y escalabilidad del sistema, además de que facilita el camino para conseguir una completa solución de BI y es una de las más utilizadas a nivel mundial. A continuación imágenes del DM de grupos electrógenos.

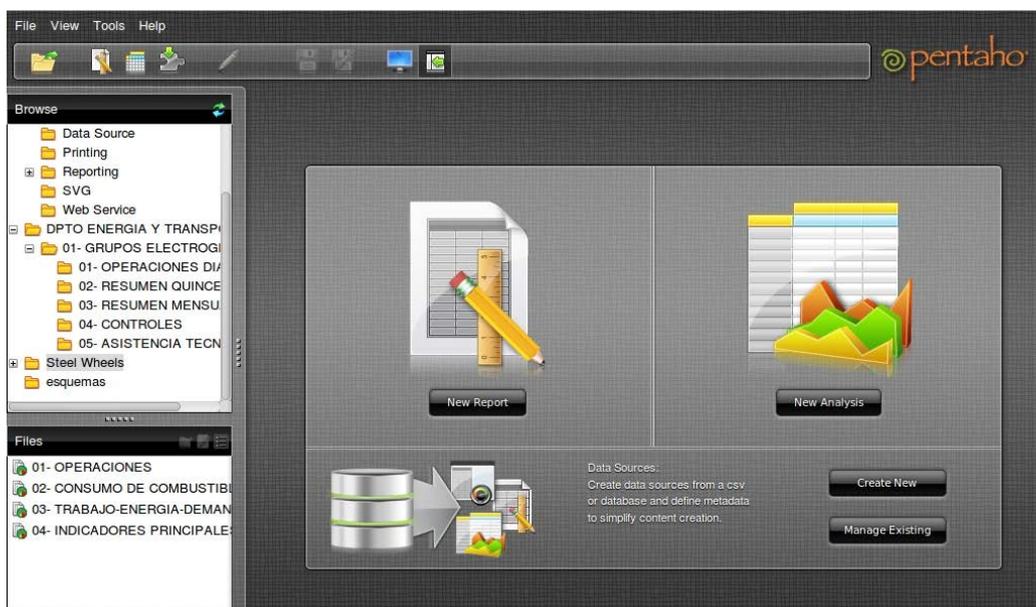


Figura 3. Página de inicio.

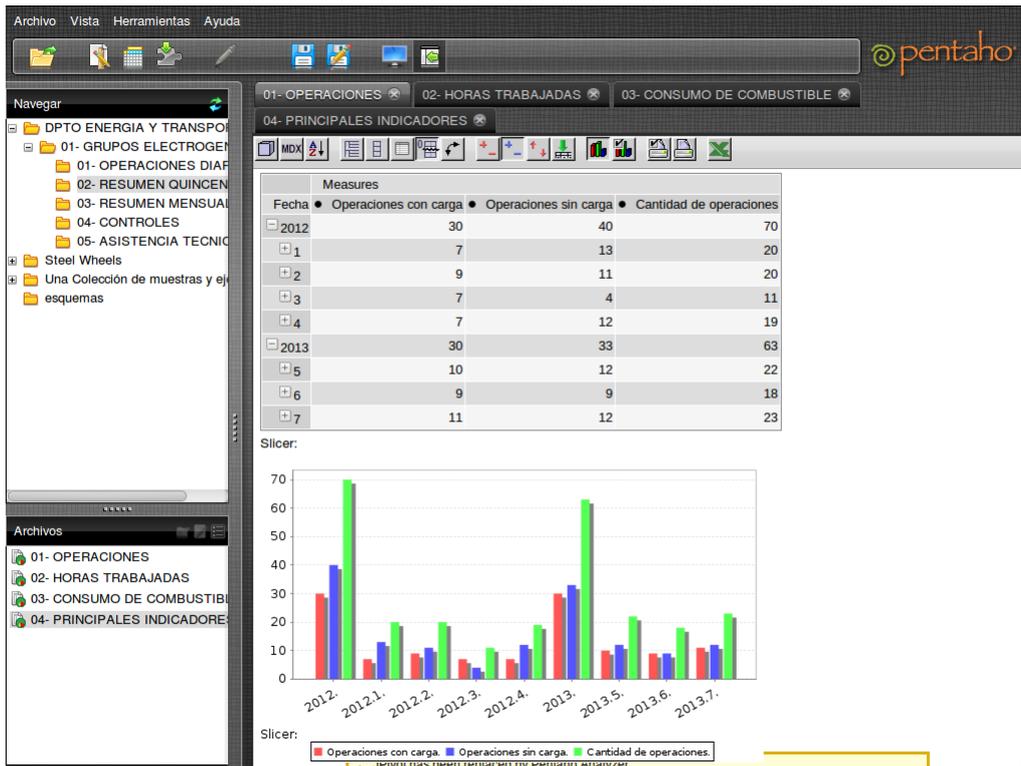


Figura 4. Ejemplo de reporte.

8. CONCLUSIONES

La caracterización epistemológica del proceso de gestión de datos en la EAALG confirmó la existencia de un problema científico que exigía la utilización de las TICs. Para darle solución a dicho problema se definió la metodología que sirvió como guía para el desarrollo del sistema propuesto. Fueron definidas además las herramientas necesarias, teniendo en cuenta en todo momento la necesidad de que fueran libres o al menos de código abierto. Se definieron y desarrollaron ocho mercados de datos que responden a cada una de las áreas departamentales existentes en la EAALG. Para cada DM se definieron y desarrollaron los reportes que facilitan la visualización y análisis de los datos contenidos en el almacén.

Se desarrolló un BIS que gestiona de forma eficiente los datos necesarios para apoyar el PTD en la organización objeto de estudio. Específicamente el sistema brinda información actualizada, relacionada con la planificación y control del consumo de los portadores energéticos y del transporte, así como del funcionamiento de los grupos electrógenos. Además permite controlar el funcionamiento de la infraestructura hidráulica, instalaciones de desinfección del agua, plantas potabilizadoras y lagunas de estabilización. Los reportes y análisis de la información son realizados de forma dinámica, lo que posibilita la extracción de información de forma rápida y desde varias perspectivas. El modelo matemático utilizado permite la realización de pronósticos a partir de los datos temporales almacenados en el DWH.

Con la información generada por el sistema, el trabajo de los especialistas y directivos de la EAALG, y de cada municipio de la provincia, se facilita considerablemente, sin embargo el mayor beneficio lo obtuvo la sociedad. Mejoró además la atención a las necesidades de la población en materia de acueducto y alcantarillado, así como los servicios que se prestan en las UEB. Otro de los aportes de esta investigación es que los resultados obtenidos podrán generalizarse en las demás provincias del país debido a que el funcionamiento de la EAALG es similar a las demás de este tipo. Se dio un importante paso en cuanto a la evolución y mejoramiento de los mecanismos apoyo al PTD en las Empresas de Acueducto y Alcantarillado.

9. REFERENCIAS BIBLIOGRÁFICAS

- [1] **Forbes.** *Managing Information in the Enterprise: Perspectives for Business Leaders* [online]. Available from: www.forbes.com/forbesinsights. 2010.
- [2] **Ballard Chuck, Abdel-Hamid Ahmed, Frankus Robert, Hasegawa Fabio, Larrechart Julio, Leo Pietro and Ramos, Jo.** *Improving Business Performance Insight . . . with Business Intelligence and Business Process Management* [online]. First Edit. S.l.: IBM Corp. Available from: <http://www.ibm.com/redbooks>. 2006.
- [3] **Negash Solomon.** Business intelligence. In: *Communications of the Association for Information Systems*. Vol. 13, no. July, pp. 450. 2004.
- [4] **Gameiro Carlos.** *Implementation of Business Intelligence tools using Open Source Approach*. In: ACM. 2011.
- [5] **Vercellis Carlo.** *Business Intelligence: Data Mining and Optimization for Decision Making*. S.l.: A John Wiley and Sons, Ltd. 2009. ISBN 9780470511381.
- [6] **Wu Liya, Barash Gilad and Bartolini Claudio.** *A Service-oriented Architecture for Business Intelligence*. 2007.
- [7] **Pressman Roger S.** *Software Engineering. A practitioners approach*. Seventh. 2010. ISBN 2008048802.
- [8] **Sommerville Ian.** *Software Engineering*. 2011. Ninth. S.l.: Wiley Publishing, Inc. ISBN 9780137035151.
- [9] **Kimball Ralph, Reeves Laura, Ross Margy and Thornthwaite Warren.** *The Data Warehouse Lifecycle Toolkit*. S.l.: Wiley Publishing, Inc. 2007.
- [10] **Kimball Ralph and Ross Margy.** *The Data Warehouse Toolkit*. Second. S.l.: Wiley Computer Publishing. 2002. ISBN 0471200247.
- [11] **Golfarelli Matteo.** *Open Source BI Platforms: a Functional and Architectural Comparison*. 2009.
- [12] **Bernardino Jorge.** *Open Source Business Intelligence Platforms for Engineering Education*. 2011. Portugal.