

# DATAMART PARA EL ANÁLISIS DE INFORMACIÓN DEL SISTEMA ACADÉMICO DE LA UNIVERSIDAD TÉCNICA DEL NORTE CON HERRAMIENTAS DE SOFTWARE LIBRE

Tana Paspuel Gloria Estefanía  
Universidad Técnica del Norte  
Av. Fray Vacas Galindo 4- 27 y Mariano Acosta  
tana\_estefania@yahoo.es

**Resumen.** La inteligencia de negocios es un proceso que permite la integración de datos para convertirlos en información valiosa, con la finalidad de que el usuario sea capaz de determinar oportunidades y amenazas para la empresa a través de los reportes generados; para que la implementación tenga éxito se requiere de datos históricos o datos que provengan de distintas fuentes de datos. Actualmente existen muchas soluciones de software que permiten realizar inteligencia de negocios, pueden ser de software libre o software propietario, la decisión por optar alguna de éstas soluciones la tiene el personal encargado del departamento de tecnologías de información junto con su equipo de desarrollo, ellos deben de considerar algunos factores como son: los recursos con los que cuenta la empresa, la predisposición que tienen los altos mandos en el apoyo del proyecto, la cooperación de todo el personal que integra la organización y la disponibilidad de información requerida para ser analizada.

## Palabras claves

Inteligencia de negocios, software libre, software propietario

## 1. Descripción

### 1.1 Descripción del problema

Existe poco conocimiento sobre herramientas de software libre para la creación de datamarts por lo que se hace necesario investigar alternativas más económicas y eficientes para realizar inteligencia de negocios, conocer sus ventajas y limitaciones para solucionar problemas de tipo real.

En la actualidad las empresas están emigrando sus sistemas a herramientas libres para no tener que comprar licenciamiento, lo que se pretende al realizar este estudio de varias herramientas de software libre es brindar opciones que permitan la creación de datamarts para que las empresas o instituciones interesadas en migrar su información a herramientas software libre sepan cuál es su mejor opción así como las ventajas y desventajas de cada una de estas herramientas.

Cada día la tecnología evoluciona a gran escala, es por ello que las pequeñas y medianas empresas migran su información a grandes almacenes de datos para tomar decisiones gerenciales, las herramientas más conocidas con licenciamiento de las casas productoras de Microsoft y Oracle no son accesibles para las pequeñas y medianas empresas que están en auge de desarrollo, es por ello que se hace necesaria la investigación de herramientas de software libre para contrastar con las herramientas de software propietario para así poder determinar ventajas y desventajas.

## 1.2 Objetivos

### 1.2.1 Objetivo general

Estudiar e implementar un Datamart para el análisis de información del Sistema Académico de la Universidad Técnica del Norte con herramientas de software libre.

### 1.2.2 Objetivos específicos

- Investigar las herramientas libres Birt, Attivio y Pentaho que permiten la creación de datamarts.
- Seleccionar entre las herramientas libres propuestas con la finalidad de determinar la herramienta más adecuada.
- Aplicar la investigación en la realización de un prototipo de datamart con la información del sistema Académico de la FICA.
- Presentar informes y reportes generados a través de los cubos de información.

## 1.3 Definición de términos asociados

**1.3.1 Inteligencia de negocios.-** Es la habilidad para transformar los datos en información entendible, capaz de permitir a los usuarios sacar conclusiones sobre sus movimientos económicos, financieros, transaccionales, etc., éste proceso juega un papel estratégico muy importante para una empresa en su desarrollo, generando una potencial ventaja competitiva y

aportando en la eliminación de información redundante e innecesaria.

**1.3.2 Datawarehouse.-** Podremos definirlo como una colección de datos integrado, orientado a un determinado campo de acción que puede ser una empresa, una organización, una institución educativa, etc., su objetivo es ayudar a la toma de decisiones en el lugar donde esté implementado.

**1.3.3 Datamart.-** Son almacenes de datos más pequeños que por lo general son utilizados por las pequeñas y medianas empresas a diferencia de los cubos de información que nos ayudan a visualizar la información clasificada.

**1.3.4 Birt.-** Es un proyecto de Eclipse de código abierto, ofrece un gran número de tareas, sobre todo para aplicaciones web, con capacidades de inteligencia de negocios para clientes pesados. Birt recibe soporte de una activa comunidad de usuarios y desarrolladores de eclipse.org.

**1.3.5 Attivio.-** Desde el año 2008 la empresa Attivio Inc. ha sido reconocida por como una de las empresas innovadoras y con mayor potencial para brindar soluciones de inteligencia de negocios. Es una empresa privada de software empresarial que produce y vende información de acceso unificado, que permite a los usuarios encontrar todo tipo de información con una sola consulta.

**1.3.6 Pentaho.-** La Suite de Pentaho es un paquete de software libre que nos permite generar inteligencia empresarial, este paquete incluye otras herramientas de software libre integradas para generar informes, minería de datos, ETL, etc., fue creado por 5 fundadores en el año 2004, el objetivo es ofrecer un paquete de código abierto de inteligencia de negocios.

#### 1.4 Comparativa de las herramientas

Al realizar la comparativa de 3 herramientas de software como son Attivio, Birt, Pentaho, se lo hizo mediante un estudio comparativo de las características, ventajas y desventajas de cada una de ellas, enlistando las fortalezas más importantes, con la finalidad de identificar la herramienta más ventajosa en cuanto a funcionalidades en el momento de su implementación.

Para la selección de la herramienta se ha tomado en cuenta la facilidad en el uso y manejo, las

tecnologías soportadas, base de datos compatibles, tiempo de implementación, sistemas operativos sobre los cuales corre la arquitectura, seguridad, licenciamiento, la integración con herramientas externas de gestión, el soporte de las herramientas y las plataformas compatibles, de acuerdo a la investigación realizada de las 3 herramientas, se ha logrado determinar ventajas y desventajas de cada una de ellas.

#### 1.4.1 Cuadro comparativo de herramientas de software libre

|  | ATTIVIO (AIE)       | BIRT                   | PENTAHO                      |
|--|---------------------|------------------------|------------------------------|
| SEGURIDAD                              | X                   | X                      | ✓                            |
| LICENCIA OPEN SOURCE                   | X                   | ✓                      | ✓                            |
| SSOO COMPATIBLES                       | Linux, Mac          | Linux, Windows         | Linux, Windows               |
| ETL                                    | X                   | X                      | ✓                            |
| CUBOS OLAP REPORTING                   | X                   | ✓                      | ✓                            |
| VERSIONES COMMUNITY Y EMPRESARIAL      | X                   | X                      | ✓                            |
| ESPACIO EN MEMORIA                     | Ocupa mucho espacio | Ocupa bastante espacio | Ocupa poco espacio           |
| WIZARDS E INTERFACES GRAFICAS          | X                   | X                      | ✓                            |
| PERMITE DEFINIR DIMENSIONES            | X                   | X                      | ✓                            |
| DATAMINING                             | X                   | ✓                      | ✓                            |
| BASE DE DATOS COMPATIBLES              | MySQL               | MySQL                  | MySQL, POSTGRES, ORACLE, SQL |
| SERVIDOR DE APLICACIONES               | X                   | X                      | ✓                            |
| DASHBOARDS                             | X                   | ✓                      | ✓                            |
| PLANIFICACIÓN DE TAREAS EN TIEMPO REAL | X                   | X                      | ✓                            |
| TOTAL                                  | 5                   | 5                      | 12                           |

Tabla 1: Comparativa de Herramientas

#### 1.4.2 Selección de la herramienta adecuada

El proceso de selección de una herramienta para el desarrollo de inteligencia de negocios puede resultar complicado y tedioso, debido a que existen múltiples alternativas en el mercado actual, para el desarrollo del prototipo del presente trabajo de grado se escoge a la suite de Pentaho para el desarrollo de cubos y generación del datamart, PostgreSQL como motor de base de datos y DBDesigner4 para el diseño de la base de datos.

#### 2. Diseño de la solución

Para el diseño del prototipo se tomará en cuenta el modelado físico y el modelado lógico, la base

de datos creada para ser utilizada por el datamart se denominó: TESISDW.

### 2.1 Herramientas de Software Libre que se utilizarán

| Software                        | Descripción   |
|---------------------------------|---|
| <b>Pentaho BI server</b>        | Es un servidor OLAP.  |
| <b>Pentaho Report Designer</b>  | Es una herramienta para sacar reportes.   |
| <b>Pentaho Data Integration</b> | Herramienta que nos ayuda en la integración de datos, nos permite realizar proceso ETL. |
| <b>Pentaho Schema WorkBench</b> | Herramienta que nos sirve para diseñar cubos dimensionales.                             |
| <b>Base de datos PostgreSQL</b> | Repositorio del datawarehouse.  |
| <b>DBDesigner 4</b>             | Herramienta que ayuda en el diseño de la base de datos.                                 |

**Tabla 2.** Herramientas de software libre para desarrollar un datawarehouse

#### 2.1.1 Pentaho Report Designer

Esta herramienta se enfoca principalmente en la creación, generación y distribución de reportes de manera organizada de todas las fuentes de datos, permite distribuir los resultados del análisis de información en varios formatos como pdf, txt, html, etc., e inclusive tiene la opción para imprimir.

#### 2.1.2 Pentaho Data Integration

Es una muy buena herramienta gráfica que permite realizar procesos de ETL con la información de la base de datos transaccional para luego obtener tablas dimensionales.

#### 2.1.3 Pentaho Schema WorkBench

Es una interfaz gráfica de la suite Pentaho que permite elaborar, modificar y probar esquemas de cubos OLAP de manera gráfica, para luego ser publicados en el servidor Mondrian, ésta herramienta genera archivos de esquema XML, permite diseñar cubos de información con las respectivas dimensiones y medidas del cubo dimensional.

#### 2.1.4 Servidor Mondrian

El servidor OLAP open source Mondrian contiene como núcleo un archivo Jar, este archivo trabaja como un conector JDBC que ayuda hacer conexiones con las distintas bases de datos y permite ejecutar consultas SQL de la base de datos transaccional, Mondrian está empaquetado de dos formas: como un paquete WAR que está conformado de Jpivot y el otro paquete conformado de Jpivot con una base de datos Derby.

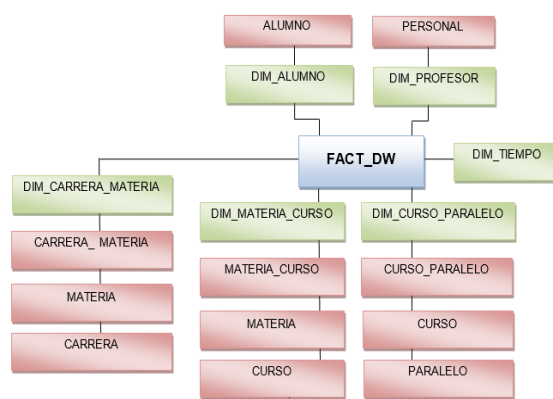
### 3. Construcción de la solución

#### 3.1 Modelado dimensional

El modelamiento dimensional surge del modelo relacional, así la información podrá ser almacenada en grandes bases de datos, de manera que los datos puedan ser extraídos y analizados mediante consultas OLAP, el modelo dimensional consiste en una sola tabla grande de datos que son descritos usando dimensiones y medidas.

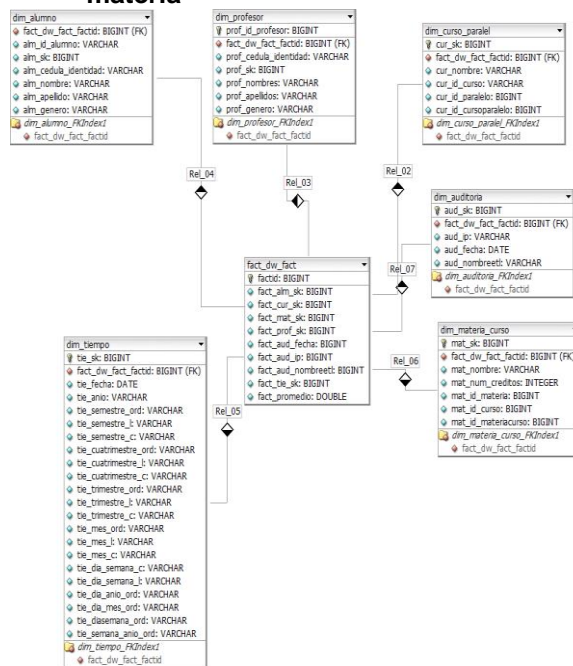
##### 3.1.1 Diagrama de origen de datos

A continuación se describirá mediante un diagrama relacional las dimensiones y la tabla de hechos que contiene el proyecto con la base de datos de origen.



**Figura 1.** Diagrama de origen de datos

##### 3.1.2 Diseño de la base de datos del modelo dimensional para el cubo alumno\_materia



**Figura 2.** Base de Datos dimensional alumno\_materia





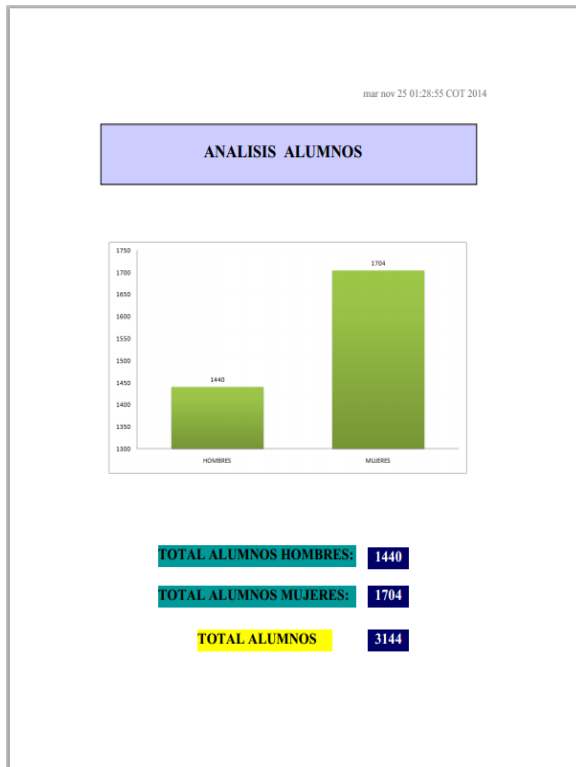


Figura 7. Vista del informe generado Total alumnos

REPORTE DE LISTADO DE CARRERAS

| id_carrera | nombre_carrera   |
|------------|--|
| 1          | Carrera de Ingeniería en Electrónica y Redes de Comunicación |
| 2          | Carrera de Ingeniería en Mecatrónica                         |
| 3          | Carrera de Ingeniería en Sistemas Computacionales            |
| 4          | Carrera de Ingeniería Industrial                             |
| 5          | Carrera de Ingeniería Textil                                 |

TOTAL DE CARRERAS: 5

Figura 8. Informe generado del listado de Carreras

### 3.5 Puesta en marcha

En la mayoría de empresas el personal encargado de la gestión de proyectos debe hacer uso de las herramientas y establecer procedimientos estandarizados con la finalidad de optimizar tiempo y lograr el éxito de los proyectos; es importante tomar en cuenta que el proyecto no ha finalizado mientras no se haya realizado la puesta en marcha, se hayan realizado las respectivas pruebas y se hayan obtenido los resultados finales del proyecto.

### 3.6 Pruebas

La finalidad de las pruebas es poder verificar el correcto funcionamiento de las tareas por las cuales se diseñó la aplicación y también para validar las expectativas del usuario, estas pruebas son realizadas por el personal encargado de supervisar los resultados obtenidos, la ejecución de las pruebas debe cumplir con los alcances y limitaciones del proyecto planteado, para esto es necesario cumplir los siguientes requerimientos:

Verificación del datamart

- Pruebas de ETL
- Procedimientos ETL probados
- Diseño del cubo aprobado
- Pruebas de reportes
- Pruebas de funcionamiento

### 3.7 Resultados de las pruebas obtenidas

Luego de haber analizado y aplicado el proceso de pruebas se obtienen los respectivos resultados, es muy importante que el sistema o prototipo cumpla con los requerimientos que exigen al hacer las pruebas, dependiendo del tipo de evaluación que haga el encargado de monitorear el sistema deberá de emitir su informe con el porcentaje respectivo de cumplimiento de los parámetros propuestos, como resultado deberá emitir un informe con porcentajes de evaluación y recomendaciones o correcciones a ejecutar por el personal encargado de desarrollo.

## 4. Conclusiones y recomendaciones

### 4.1 Conclusiones

- A lo largo del desarrollo del presente trabajo de grado se ha podido observar que, la implementación de datamart con herramientas de software libre es muy confiable y seguro, obteniendo estabilidad al momento de integrar los datos, logrando buenos resultados.
- La creación de datamarts es un paso previo a la implementación de datawarehouse, conforme la organización o empresa va creciendo los datamarts van aumentando, permitiendo toma de decisiones gerenciales de acuerdo al área de negocio o departamento que necesita mayor énfasis en ser administrado.
- La implementación de business intelligence en una organización, pequeña o mediana hace que la información permanezca ordenada y controlada, registrando las transacciones diarias que realiza la empresa hasta convertirse en información histórica valiosa para ser analizada.
- Para tener una buena implementación de inteligencia de negocios es muy importante conocer bien los requerimientos académicos del área de negocio y hacer un análisis completo de la información que va a ser analizada.
- La implementación de datawarehouse en grandes volúmenes de información resulta un proceso largo pero esencial cuando se necesita hacer análisis de información, ya que permite agrupar y clasificar en grupos significativos la información más importante.
- Los beneficios que se obtienen al realizar la comparación de tres herramientas de software libre, resulta muy provechoso a la hora de implementar este tipo de proyectos,

con la demostración de la herramienta seleccionada, el correcto funcionamiento, la fácil utilización y depuración de la misma.

- De los resultados obtenidos se destaca la valiosa experiencia obtenida en el área de Inteligencia de Negocios en el transcurso del desarrollo del presente prototipo.
- Se investigó las herramientas de software libre Birt, Attivio y Pentaho lo que permitió determinar a lo largo del desarrollo del presente trabajo de grado que la Suite de Pentaho, tiene muchas más funcionalidades y es una herramienta confiable para la elaboración de datamarts.
- La suite de Pentaho permitió la elaboración de datamart con datos reales con el objeto de demostrar la el tema planteado en el presente trabajo de grado.
- La herramienta Pentaho Reporting permitió la elaboración de reportes luego de haber realizado el cubo de información.

#### 4.2 Recomendaciones

- Para una exitosa implementación y para futuras adiciones es necesario trabajar en equipo tanto las áreas de negocio involucradas como el departamento de tecnología de Información.
- Es muy importante tomar en cuenta la arquitectura y los componentes de software a utilizar, su compatibilidad y unificación con otras plataformas porque pueden existir futuras adiciones o migración a una nueva plataforma.
- Al momento de haber terminado la implementación se recomienda dar seguimiento a la información que se carga en el datamart con la finalidad de mantener información actualizada y confiable que sea de beneficio para los usuarios que está encargados en la toma de decisiones.
- Las asignaciones de roles para las distintas áreas del negocio deben ser definidas por el personal encargado de monitorear la información.
- Es recomendable utilizar herramientas de software libre en las pequeñas y medianas empresas ya que permiten reducir costos, ahorrar recursos económicos, la suite Pentaho es una solución que nos permite la gestión y análisis de información que puede ser utilizada por cualquier organización que desee implementar una solución de Inteligencia de Negocios.
- La correcta identificación de los requerimientos y necesidades se lo debe hacer directamente con los usuarios y gerentes de las organizaciones a implementar Inteligencia de Negocios con el objetivo de que la información sea segura y

confiable para evitar futuros problemas en el desarrollo del sistema.

- Se recomienda a los docentes incentivar el uso de herramientas de software libre que faciliten el desarrollo de proyectos de software.
- Se recomienda realizar estudios comparativos con herramientas de software libre que permitan la elaboración de minería de datos.

#### 5. Referencias

- Kimball Ralph. (2008). The Data warehouse Life Cycle Toolkit. Indiana: Wiley.
- Wrembel Robert. (2007). Data warehouses and OLAP: concepts, architectures and solutions. Estados Unidos: IRM Press.
- Antonio Rivas. (23 de Febrero de 2011). Análisis del cuadrante de Gartner –Business Intelligence 2011. Parte 3. Obtenido de <http://www.bi.dev42.es/2011/02/23/olap-molap-rolap/>.
- Attivio. (16 de Octubre de 2013). Company Overview. Obtenido de <http://www.crunchbase.com/organization/attivio>.
- Attivio. Inc., (20 de Junio de 2014). Complete Unified Information Access for Business Innovation. Obtenido de <http://www.attivio.com/how-we-do-it/the-platform/active-intelligence-engine>.
- Christopher Nuzum. (17 de Noviembre de 2010). Attivio Search System Requirements. Obtenido de <https://support.tractionsoftware.com/traction/#/single&proj=AttivioSearch&rec=4&brief=n>.
- Deinsa. (25 de Abril de 2014). Inteligencia de Negocios. Obtenido de <http://www.idensa.com/>
- EOS. (25 de Febrero de 2014). Business Intelligence QlickView. Obtenido de <http://www.eopensolutions.com/productos-y-servicios/business-intelligence-qlikview-ecuador>
- fabFORCE.net. (25 de Enero de 2014). DBDesigner4. Obtenido de <http://www.fabforce.net/dbdesigner4/>
- Hide Julian. (23 de Agosto de 2014). Mondrian Documentation. Obtenido de [http://mondrian.pentaho.com/documentation/installation\\_es.php](http://mondrian.pentaho.com/documentation/installation_es.php)
- Honorable Consejo Universitario. (12 de Noviembre de 2005). REGLAMENTO GENERAL - UTN. Obtenido de <http://www.utn.edu.ec/web/portal/images/doc-utn/reglamento-general-utn.pdf>
- Kimball y Wrembel. (2 de Abril de 2011). Los datos de transacciones crecen. Obtenido de <http://www.informatica.com/es/vision/harnes-sing-big-data/oltp-and-analytics/>.

- Kimball, R. (2008). The Data warehouse Life Cycle Toolkit. Indiana: Wiley.
- Mariana Isabel Rojas. (20 de Junio de 2009). Data warehouse. Obtenido de <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonoAdsDiseno.pd>.
- Marysol Tamayo, Javier Moreno. (20 de Junio de 2010). Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP. Obtenido de [http://www.scielo.org.co/scielo.php?pid=S0120-56092006000300016&script=sci\\_arttext](http://www.scielo.org.co/scielo.php?pid=S0120-56092006000300016&script=sci_arttext).
- Meteorite Consulting. (16 de Enero de 2014). Saiku Enterprise Analytics. Obtenido de <http://www.meteorite.bi/>
- Microsoft. (02 de Enero de 2013). Procesamiento y modos de almacenamiento de particiones. Obtenido de <http://msdn.microsoft.com/es-es/library/ms174915.aspx>.
- Roberto Espinosa. (19 abril de 2010). Kimball vs Inmon. Ampliación de conceptos del Modelado Dimensional. Obtenido de <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimENSIONAL/>.
- Sinnexus. (20 de Noviembre de 2012). Datawarehouse. Obtenido de [http://www.sinnexus.com/business\\_intelligence/datawarehouse.aspx](http://www.sinnexus.com/business_intelligence/datawarehouse.aspx).
- Sinnexus. (21 de Enero de 2014). ¿Qué es Business Intelligence? Obtenido de [http://www.sinnexus.com/business\\_intelligence/](http://www.sinnexus.com/business_intelligence/)
- Thomas. (03 de Julio de 2014). Reporting Tales. Obtenido de <https://www.on-reporting.com/what-is-pentaho-reporting/>
- Wikipedia. (26 de 04 de 2013). HOLAP. Obtenido de <http://es.wikipedia.org/wiki/HOLAP>
- Wikipedia. (17 de 10 de 2013). OLTP. Obtenido de <http://es.wikipedia.org/wiki/OLTP>
- Wikipedia. (05 de Noviembre de 2013). ROLAP. Obtenido de <http://es.wikipedia.org/wiki/ROLAP>
- Wikipedia. (20 de Marzo de 2014). Almacén de Datos. Obtenido de [http://es.wikipedia.org/wiki/Almac%C3%A9n\\_de\\_datos](http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos)
- Wikipedia. (13 de Noviembre de 2014). Attvio. Obtenido de <http://en.wikipedia.org/wiki/Attvio>
- Wikipedia. (26 de Marzo de 2014). Business Intelligence and Reporting Tools. Obtenido de [http://es.wikipedia.org/wiki/Business\\_Intelligence\\_and\\_Reporting\\_Tools](http://es.wikipedia.org/wiki/Business_Intelligence_and_Reporting_Tools)
- Wikipedia. (15 de Noviembre de 2014). Datamining. Obtenido de [http://en.wikipedia.org/wiki/Data\\_mining](http://en.wikipedia.org/wiki/Data_mining)
- Wikipedia. (17 de Junio de 2014). Granularidad. Obtenido de <http://es.wikipedia.org/wiki/Granularidad>
- Wikipedia. (21 de 06 de 2014). MOLAP. Obtenido de <http://es.wikipedia.org/wiki/MOLAP>
- Wikipedia. (05 de 06 de 2014). OLAP. Obtenido de <http://es.wikipedia.org/wiki/OLAP>
- Wikipedia. (16 de Julio de 2014). Pentaho. Obtenido de <http://es.wikipedia.org/wiki/Pentaho>
- Wikipedia. (18 de Junio de 2014). Sistema de planificación de recursos empresariales. Obtenido de [http://es.wikipedia.org/wiki/Sistema\\_de\\_planificaci%C3%B3n\\_de\\_recursos\\_empresariales](http://es.wikipedia.org/wiki/Sistema_de_planificaci%C3%B3n_de_recursos_empresariales).
- Wood Sherman. (27 de Abril de 2014). Documentación Mondrian. Obtenido de <http://mondrian.pentaho.com/documentation/workbench.php>