DATAMART ANALYSIS OF ACADEMIC INFORMATION SYSTEM TECHNICAL UNIVERSITY OF NORTH WITH FREE SOFTWARE TOOLS

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

Abstract. Business intelligence is a process that enables the integration of data into valuable information with the aim for the user to be able to identify opportunities and threats for the company through reports generated; for successful implementation requires historical data or data coming from various data sources. Currently there are many software solutions that enable business intelligence, can be free software or proprietary software, the decision for choosing one of these solutions lies with the personnel department of information technology along with its development team, they must consider some factors including the resources available to the company, the predisposition of high command in supporting the project, cooperation of all staff that integrates the organization and availability of information required for analysis.

Keywords
Business intelligence, free software, proprietary software.

1. Description

1.1 Description of the problem
There is little knowledge about free software tools for creating data marts so it is necessary to investigate more economical and efficient for business intelligence alternatives, know their advantages and limitations to troubleshoot real dude.

Today companies are migrating their systems free tools to not having to buy licensing, what is intended to conduct this study several free software tools is to provide options to create data marts for companies or institutions interested in migrating your data to free software tools know what your best option and the advantages and disadvantages of each of these tools.

Every day technology evolves a large scale, is why small and medium enterprises migrate your information to big data warehouses to make management decisions, the best known with licensing production houses Microsoft and Oracle tools are not accessible to small and medium businesses are booming development, which is why research free software tools to contrast with proprietary software tools in order to determine the advantages and disadvantages is necessary.

1.2 Objectives

1.2.1 General Objective
Study and implement a Datamart for analysis Academic Information System Technical University North with free software tools.

1.2.2 Specific Objectives
- Investigate free tools Birt, Attivio and Pentaho that allow the creation of datamarts.
- Select between free tools proposed in order to determine the most appropriate tool.
- Apply research in the realization of a prototype datamart with Academic information system FICA.
- Reporting and reports generated through the hubs of information.

1.3 Definition of terms associated

1.3.1 Business Intelligence.- The ability to transform data into understandable information, able to allow users to make conclusions about their economic movements, financial, transactional, etc., this process plays an important strategic role in a company in its development, generating a competitive advantage and providing potential for eliminating redundant and unnecessary information.

1.3.2 Datawarehouse.- We can define it as an integrated data collection, oriented to a particular field of action that may be a company, an organization, educational institution, etc., its objective is to help decision making in the place where it is implemented.

1.3.3 Datamart.- They are smaller data warehouses that usually are used by small and medium enterprises unlike cubes information to help us view classified information.

1.3.4 Birt.- It is an Eclipse project open source, provides a large number of tasks, especially for web applications, with business intelligence capabilities to thick clients. Birt Supported by an active community of users and developers eclipse.org.

1.3.5 Attivio.- Since 2008 the company Attivio Inc. has been recognized as one of the innovative companies with the greatest potential to provide business intelligence solutions. It is a private enterprise software company that produces and sells unified information access, which allows users to find all kinds of information with a single query.

1.3.6 Pentaho.- Pentaho Suite is a free software package that allows us to generate business intelligence, this package includes other open source tools integrated reporting, data mining, ETL, etc., was created by five founding in 2004, the aim is to offer a package of open source business intelligence.

1.4 Comparative of the tools

To make the comparison 3 software tools like Ativio, Birt, Pentaho, it is done through a comparative study of the characteristics, advantages and disadvantages of each, listing the most important strengths, in order to identify the tool more advantageous in terms of functions at the time of implementation.

For the selection of the tool is taken into account the facility in the use and management, supported technologies, supported database, implementation time, operating systems upon which runs the architecture, security, licensing, integration with external tools management, support tools and supported platforms, according to research carried out of the 3 tools, has been determined advantages and disadvantages of each.

1.4.1 Comparison chart free software tools

	ATTIVIO (AIE)	BIRT	PENTAHO
SAFETY	X	X	✓
OPEN SOURCE LICENSE	X	✓	✓
SSOO COMPATIBLES	Linux, Mac	Linux, Windows	Linux, Windows
ETL	X	X	✓
OLAP CUBES	X	✓	✓
REPORTING	X	✓	✓
COMMUNITY AND BUSINESS VERSIONS	X	X	✓
SPACE IN MEMORY	Occupy much space	Occupy much space	Occupies litle space
WIZARDS AND GRAPHICS INTERFACE	X	X	✓
PERMITTED TO DEFINE DIMENSION	X	X	✓
DATAMINING	X	✓	✓
BATABASE COMPATIBLE	MySQL	MySQL	MySQL, POSTGRES ORACLE, SQL
APLICATION SERVER	X	X	✓
DASHBOARDS	X	✓	✓
PLANNING TASKS TIME REAL	X	X	✓
TOTAL	5	5	12

Table 1: Tools comparative

1.4.2 Selecting the appropriate tool

The process of selecting a tool for developing business intelligence can be complicated and tedious, because there are many alternatives on the market today, to develop the prototype of the present paper grade is chosen to suite Pentaho for development and creation of datamart cubes, PostgreSQL as database engine and DBDesigner4 to design the database.

2. Design the Solution

For the design of the prototype will be considered physical modeling and logical modeling, database created for use by the datamart was called: TESISDW.

2.1 Free Software Tools to be used

Software	Description
Pentaho BI server	It is a server OLAP.
Pentaho Report Designer	It is a tool for getting reports.
Pentaho Data Integration	Tool that helps in the integration of data, allows us to perform ETL process.
Pentaho Schema WorkBench	Tool that helps us design dimensional cubes.
Database PostgreSQL	Repository datawarehouse.
DBDesigner 4	Tool that helps in the design of the database.

Table 2. Free software tools to develop a datawarehouse

2.1.1   Pentaho Report Designer
This tool focuses primarily on the creation, production and distribution of reports organized of all data sources so, can distribute the results of the analysis of information in various formats like pdf, txt, html, etc., and even has the option for printing.

2.1.2   Pentaho Data Integration
It is a very good graphical tool that allows ETL processes with information from the transactional database and then obtain dimensional tables.

2.1.3   Pentaho Schema WorkBench
It is a graphical interface Pentaho suite that allows to develop, edit and test schemes OLAP cubes graphically, soon to be published in the Mondrian server, this tool generates XML schema files, allows designing cubes of information with the respective dimensions and measures dimensional cube.

2.1.4   Mondrian Server
The open source Mondrian OLAP server contains as kernel a Jar file, this file works as a JDBC connector which helps make connections to different databases and allows you to run SQL queries from the transactional database, Mondrian is packaged in two ways: as packet WAR Jpivot is shaped and other package forming Jpivot a Derby database.

3.   Construction of the solution

3.1 Dimensional modeling
Dimensional modeling arise of relational model, so the information may be stored in big database, so that data can be extracted and analyzed using queries OLAP, the dimensional model consists of a single large table of data which are described using dimensions and measures.
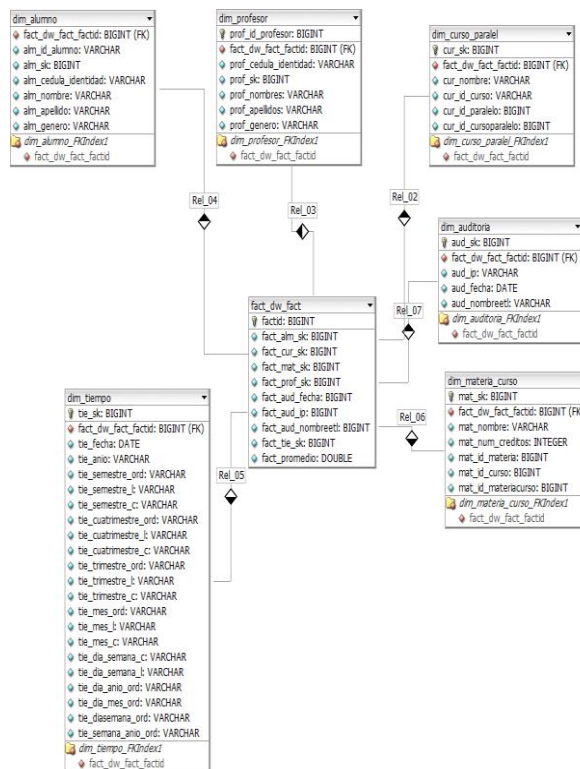
3.1.1   Chart data source
Below is described by a relational diagram dimensions and fact table that contains the project with the database of origin.
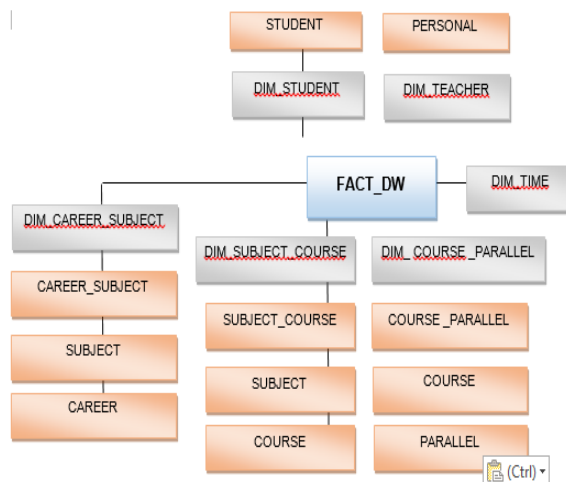
Figure 1. Diagram of Data Source

3.1.2   Design dimensional database model for the cube student_subject

Figure 2. Dimensional database student_subject

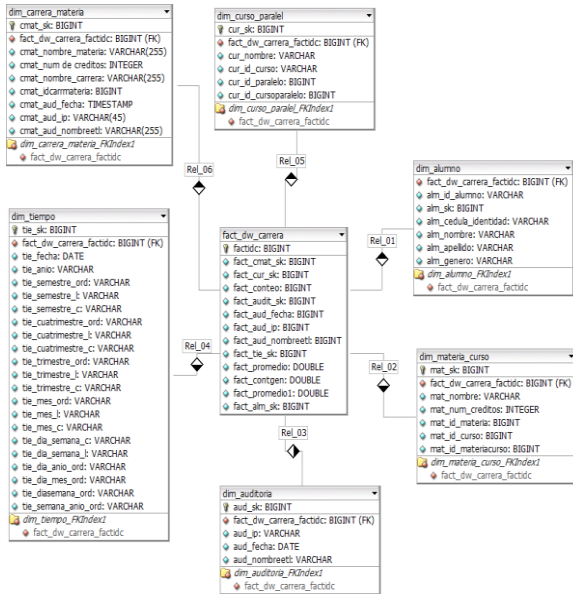### 3.1.3 Design dimensional database model for the cube career_subject

Figure 3. Dimensional database career_subject

3.2 Construction of the prototype

After selecting the tools to build the prototype, we will proceed to develop the prototype which works with the power of information, thus causing an interaction between the user and equipment, with the aim of validate important requirements.

3.3 Publishing and viewing information hub

For publication of the hubs of information is necessary to use the biserver tool, this tool allows creating and assigning users and roles. To access the management console, you must open a window of a web browser and type: ip_machine: 8099, it will enter a page which allows you to perform activities such as creating users, assigning permissions and roles to each user created , connection to the database, group management, etc.
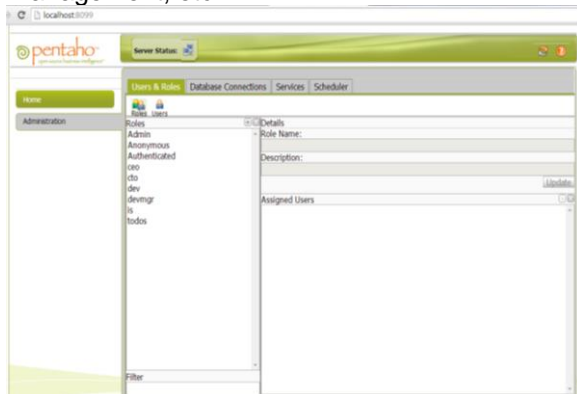

Figure 4. Main page of the administration console

To access the navigation console, you must open a window of a web browser and type: ip_machine 8080, will enter a page where you will allow you to interact directly with cubes desk can make analysis and generate reports.
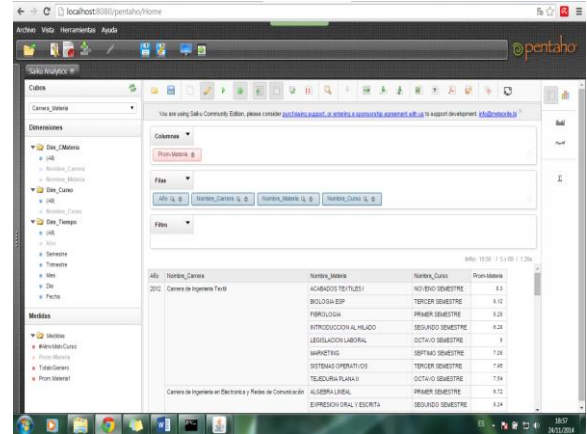

Figure 5. View of the cube information

3.4 Publication of reports

Pentaho Reporting is a set of reporting tools for creating open source relational reporting and analysis of a wide gamma of data sources, can create PDF, Excel, HTML, text, rich text, XML and CSV files. The reporting engine is a Java library of light weight with minimal global footprint, which can be embedded in other desktop applications or Java server for publishing reports and reports makes it through internet.
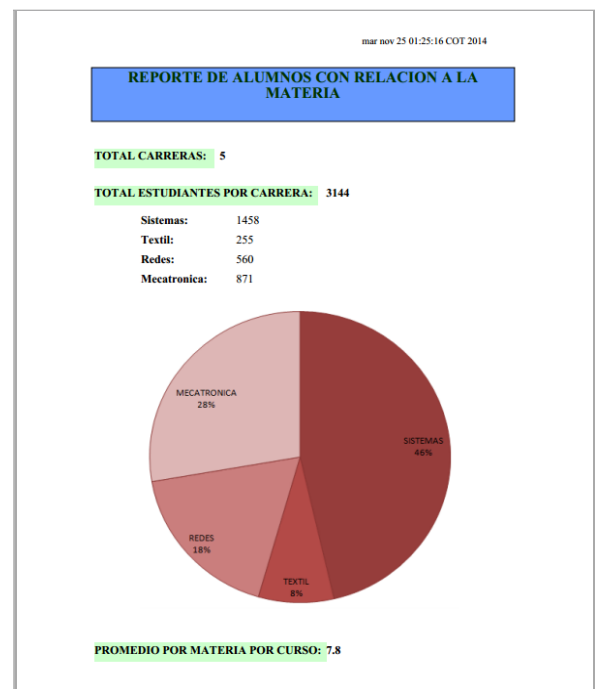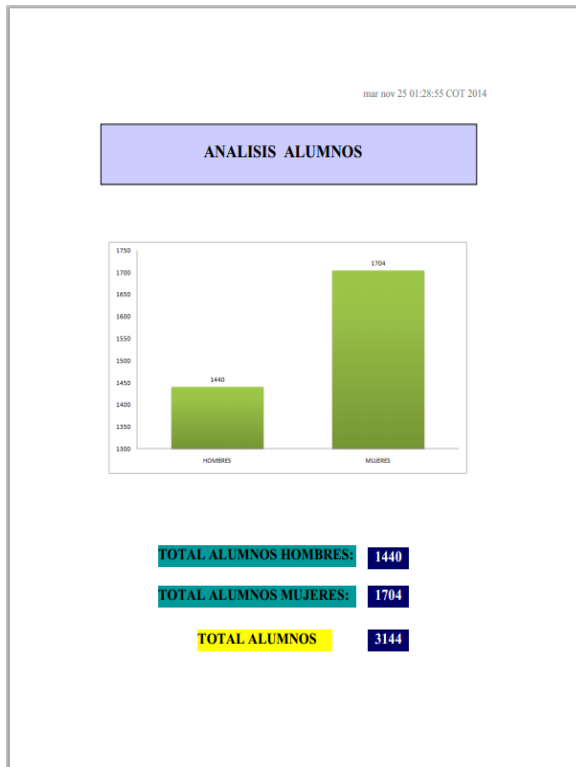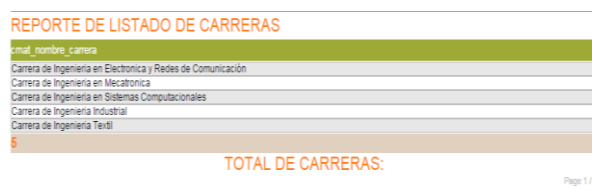

Figure 6. View the report generated students in relation to the subject

Figure 7. View the report generated total students

Figure 8. Report generated list career

3.5 Commissioning

In most companies the staff responsible for project management should make use of the tools and establish standardized with the aim of optimize time and success of projects procedures; is important to note that the project is not finished until it has completed the implementation, have made the respective tests and have obtained the final results of the project.

3.6 Testing

The purpose of testing is to verify proper operation of the tasks for which the application was designed and also to validate user expectations, these tests are performed by the staff responsible for overseeing the results, the test execution should comply with the scope and limitations of the proposed project, this is necessary to meet the following requirements:

Datamart checking
ETL test

ETL tested procedure s
Approved cube design
Reports tests
Performance testing

3.7 Test results obtained

After having analyzed and explained the process of testing the respective results are obtained, it is very important that the system or prototype complies with the requirements demanded by making tests, depending on the type of assessment by the responsible monitoring system should be issue its report to the respective percentage of compliance with the proposed parameters, as a result it must issue a report with percentages of assessment and recommendations or corrections to run by staff development.

4. Conclusions and recommendations

4.1 Conclusions

- Throughout the development of the present work grade has been observed that the implementation of datamart with free software tools is very reliable and secure, obtaining stability when integrating data, achieving good results.
- The creation of datamart is a prerequisite to the implementation of datawarehouse step, according to the organization or company grows datamarts are increasing, allowing management decisions according to business area or department that needs greater emphasis to be given
- The implementation of business intelligence in an organization, small or medium makes the information remains orderly and controlled, recording daily transactions made by the company to become valuable historical information for analysis.
- To have a good implementation of business intelligence is very important to know the academic requirements of the business area and do a full analysis of the information is going to be analyzed.
- The implementation of datawarehouse large volume of information is a long but essential process when you need to do data analysis because it allows grouping and classifying them into meaningful groups the most important information.
- The benefits obtained by performing the comparison of three free software tools, it is very helpful when implementing such projects, demonstrating the selected tool, smooth operation, easy use and cleaning of the same .

- From the results obtained valuable experience in the area of Business Intelligence during the development of this prototype.
- Free software tools Birt, Attivio and Pentaho determine which allowed throughout the development of this work to degree than Pentaho Suite has many more features and is a reliable tool for the development of datamarts is investigated.
- The Pentaho suite allows the development of datamart with real data in order to demonstrate the issue raised in this work grade.
- The Pentaho Reporting tool allows the preparation of reports after making the hub of information.

4.2 Recommendations

- For successful implementation and future additions necessary teamwork both business areas involved as the information technology department.
- It is very important to consider the architecture and software components to use, compatibility and unification with other platforms because there may be future additions or migrating to a new platform.
- When you have finished implementing it is recommended to monitor the information that is loaded into the data mart with the aim of maintain current and reliable information that is of benefit to users who are responsible in making decisions.
- Role assignments for different business areas should be defined by the personnel responsible for monitoring information.
- We recommend using free software tools for small and medium enterprises and that reduce costs, save financial resources, Pentaho suite is a solution that allows the management and analysis of information that can be used by any organization wishing to implement a Business Intelligence solution.
- Correctly identifying the requirements and needs it to do directly with users and managers of organizations implementing Business Intelligence in order to make information secure and reliable to avoid future problems in developing the system.
- It is recommended to teachers encourage the use of free software tools that facilitate the development of software projects.
- It is recommended that comparative studies with free software tools that enable the development of data mining.

5. References

- Kimball Ralph. (2008). The Datawarehouse Life Cicle Toolkit. Indiana: Wiley.
- Wrembel Robert. (2007). Datawarehouses and OLAP: concepts, architectures and solutions.Estados Unidos: IRM Press.
- Antonio Rivas. (February 23, 2011). Quadrant of Gartner Analysis -Business Intelligence 2011. Part 3. Retrieved from http://www.bi.dev42.es/2011/02/23/olap-molap-rolap/.
- Attivio. (16 de Octubre de 2013). Company Overview. Retrieved from http://www.crunchbase.com/organization/attivio.
- Attivio. Inc., (20 de Junio de 2014). Complete Unified Information Access for Business Innovation. Retrieved from http://www.attivio.com/how-we-do-it/the-platform/active-intelligence-engine.
- Christopher Nuzum. (17 de Noviembre de 2010). Attivio Search System Requirements. Retrieved from https://support.tractionsoftware.com/traction#/single&proj= AttivioSearch&rec=4&brief=n.
- Deinsa. (25 de Abril de 2014). Business Intelligence. Retrieved from http://www.idensa.com/
- EOS. (25 de Febrero de 2014). Business Intelligence QlickView. Retrieved from http://www.eopensolutions.com/productos-y-servicios/business-intelligence-qlikview-ecuador
- fabFORCE.net. (25 de Enero de 2014). DBDesigner4. Retrieved from http://www.fabforce.net/dbdesigner4/
- Hide Julian. (23 de Agosto de 2014). Mondrian Documentation. Retrieved from http://mondrian.pentaho.com/documentation/installation_es.php
- Honorable Consejo Universitario. (12 de Noviembre de 2005). GENERAL REGULATIONS - UTN. Retrieved from http://www.utn.edu.ec/web/portal/images/doc-utn/reglamento-general-utn.pdf
- Kimball y Wrembel. (2 de Abril de 2011). Transaction data grow. Retrieved from http://www.informatica.com/es/vision/harnessing-big-data/oltp-and-analytics/.
- Mariana Isabel Rojas. (20 de Junio de 2009). Datawarehouse. Retrieved from http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonoAdsDiseno.pd.
- Marysol Tamayo, Javier Moreno. (20 de Junio de 2010). Analysis MOLAP storage model versus ROLAP storage model. Retrieved from

http://www.scielo.org.co/scielo.php?pid=S01
20-56092006000300016&script=sci_arttext.

- Meteorite Consulting. (16 de Enero de 2014). Saiku Enterprise Analytics. Retrieved from http://www.meteorite.bi/
- Microsoft. (02 de Enero de 2013). Processing and storage models partition. Retrieved from http://msdn.microsoft.com/es-es/library/ms174915.aspx.
- Roberto Espinosa. (19 abril de 2010). Kimball vs Inmon. Large dimensional modeling concepts. Retrieved from http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimensional/.
- Sinnexus. (20 de Noviembre de 2012). Datawarehouse. Retrieved from http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.
- Sinnexus. (21 de Enero de 2014). ¿Whys is Business Intelligence? Retrieved from http://www.sinnexus.com/business_intelligence/
- Thomas. (03 de Julio de 2014). Reporting Tales. Retrieved from https://www.on-reporting.com/what-is-pentaho-reporting/
- Wikipedia. (26 de 04 de 2013). HOLAP. Retrieved from http://es.wikipedia.org/wiki/HOLAP
- Wikipedia. (17 de 10 de 2013). OLTP. Retrieved from http://es.wikipedia.org/wiki/OLTP
- Wikipedia. (05 de Noviembre de 2013). ROLAP. Retrieved from http://es.wikipedia.org/wiki/ROLAP
- Wikipedia. (20 de Marzo de 2014). Datawarehouse. Retrieved from http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos
- Wikipedia. (13 de Noviembre de 2014). Attivio. Retrieved from http://en.wikipedia.org/wiki/Attivio
- Wikipedia. (26 de Marzo de 2014). Business Intelligence and Reporting Tools. Retrieved from http://es.wikipedia.org/wiki/Business_Intelligence_and_Reporting_Tools
- Wikipedia. (15 de Noviembre de 2014). Datamining. Retrieved from http://en.wikipedia.org/wiki/Data_mining
- Wikipedia. (17 de Junio de 2014). Granularity. Retrieved from http://es.wikipedia.org/wiki/Granularidad
- Wikipedia. (21 de 06 de 2014). MOLAP. Retrieved from http://es.wikipedia.org/wiki/MOLAP
- Wikipedia. (05 de 06 de 2014). OLAP. Retrieved from http://es.wikipedia.org/wiki/OLAP

- Wikipedia. (16 de Julio de 2014). Pentaho. Retrieved from http://es.wikipedia.org/wiki/Pentaho
- Wikipedia. (18 de Junio de 2014). Enterprise resource planning. Retrieved from http://es.wikipedia.org/wiki/Sistema_de_planificaci%C3%B3n_de_recursos_empresariales.
- Wood Sherman. (27 de Abril de 2014). Documentación Mondrian. Retrieved from http://mondrian.pentaho.com/documentation/workbench.php