

Diseño de un Datamart para el apoyo en la toma de decisiones del Departamento de Admisión de la Universidad Peruana Unión, Filial Tarapoto

Guadaña Quiroz, Iván¹

Recibido 13 de setiembre de 2016, Aceptado 10 de octubre de 2016
Received: September 13, 2016 Accepted: October 10, 2016

RESUMEN

La presente investigación tiene como objetivo diseñar un Datamart (DM) para el apoyo en la toma de decisiones del Departamento de Admisión de la Universidad Peruana Unión, Filial Tarapoto, que sigue el proceso normal de una solución de Inteligencia de Negocios (IN); reunir información de las distintas fuentes de datos, diseñar un prototipo de extracción, transformación y carga de data (ETL), para después mostrar la información a través de gráficos estadísticos. Este proceso se basó en la metodología flexible HEFESTO 2.0 que sirvió para bosquejar todas las tareas a realizar, las que se organizaron en las siguientes fases: levantamiento de información, identificación de requerimientos, análisis de los requerimientos, identificación de los orígenes de datos, modelado multidimensional, integración de los datos, procesamiento y explotación de la información. El tipo del estudio es tecnológico y descriptivo, debido al énfasis de obtener conocimiento e interactuar con la información recolectada. Se obtuvieron los siguientes resultados: conocer la cantidad de ingresantes según lugar de procedencia y carrera profesional, cantidad de ingresantes según modalidad y campaña de admisión, cantidad de ingresantes según religión y carrera profesional, cantidad de ingresantes según campaña y sector de admisión. Concluimos que la Suite de Pentaho es de gran utilidad por su sencillez y su tecnología necesaria para realizar soluciones de IN, destacamos la herramienta Spoon que nos permitió realizar con éxito el proceso de ETL; por último, nuestro cubo y generación de reportes tuvo éxito por la enorme utilidad de Schema Workbench y Analysis Report.

Palabras clave: Datamart, ETL, inteligencia de negocios, HEFESTO, data warehouse.

¹ Ingeniero de Sistemas. Universidad Peruana Unión. email: ivanguadania@gmail.com

ABSTRACT

This research aims to design a Datamart (DM) to support the decision making process of the admission department of the Peruvian Union University, subsidiary Tarapoto which follows the normal process of a Business Intelligence (BI): gather information from different data sources, design a prototype of extraction, transformation and load of data (ETL), then display the information through statistical graphs. This process was based on the flexible methodology HEFESTO 2.0 that served to sketch out all the tasks to be performed which were organized in the next phases: gathering of information, identification of requirements, analysis of requirements, identification of data sources, multidimensional modeling, integration of data, processing and use of data. This study is technological and descriptive, due to the emphasis on obtaining knowledge and interacting with the information collected. The following results were obtained: know the exact number of entrants according to their place of origin and professional career, number of entrants according to modality and admission campaign, number of entrants according to religion and professional career, number of entrants according to campaign and admission sector. We conclude that the Pentaho Suite is very useful because of its simplicity and its technology is necessary for IN solutions. We highlight the Spoon tool that allowed us to complete successfully the ETL process. Finally, our bucket and reports generation was successful because of the enormous utility of Schema Workbench and Analysis Report.

Keywords: Datamart, ETL, business intelligence, HEFESTO, data warehouse.

INTRODUCCIÓN

El Departamento de Admisión, de la UPeU FT, cumple con dos roles importantes dentro de la institución: primero marketing y publicidad, y segundo admisión propiamente dicho; por consiguiente tiene un objetivo fusionado de ambos roles: cubrir cada una de las vacantes asignadas para cada carrera. (Tarrillo, comunicación personal, 11 de mayo, 2015).

Cada campaña de admisión lleva consigo la distribución de los recursos para lograr una mayor cantidad de alumnos postulantes sostenedores de obra. Las campañas de admisión siguen ciertos parámetros para poder lograr las metas de postulantes establecidas. Es necesario, para poder realizar una buena planificación de campaña, conocer información estadística detallada de campañas anteriores. Al final de cada campaña se obtiene información de suma importancia que sirve como input para próximas campañas. (Tarrillo, comunicación personal, 11 de mayo, 2015)

En tal sentido, la dificultad para tener información explícita de ayuda para la toma de decisiones se da por las siguientes causas: Inadecuada definición, y estructura de los reportes, complejos procesos para su generación, largos periodos para la

generación y/o actualización de información estadística y problemas derivados de la deficiente confianza en la información generada; que como consecuencia genera: incertidumbre en la distribución de los recursos en las campañas de admisión, se desconoce los lugares que tienen la mayor cantidad de postulantes sostenedores de obra y dificultad en el uso de las Tecnologías de Información y Comunicación (TIC) en las campañas de admisión. Los principales afectados son: el área administrativa de la universidad, el jefe del departamento de admisión y el personal que labora en el área. Esto ocurre al planificar la distribución de los recursos, con una frecuencia trimestral o semestral, de acuerdo a las campañas programadas. Trayendo una magnitud de impacto alta.

Instituciones y empresas están utilizando IN como recurso principal para la toma de decisiones, tal es el caso de la Universidad de Tarapacá, Arica, Chile que incorpora este recurso para crear una plataforma permitiendo la visualización y el análisis de información relacionados con el proceso de admisión y matrícula (Fuentes & Valdivia, 2010).

Zambrano, Rojas, Carvajal, & Acuña, (2011) presentan el análisis de rendimiento académico estudiantil, usando DM para analizar información de sistemas que producen varias operaciones las cuales son almacenadas en bases de datos transaccionales, con el fin de predecir, en base a la información histórica, el desempeño académico de los estudiantes y las asignaturas aprobadas. Gracias a la arquitectura descrita se obtuvo la salida deseada, al obtener gráficos estadísticos de asignaturas aprobadas, por alumno, en un semestre determinado. Es preciso agregar que el beneficio de contar con un DM radicó en la facilidad de análisis exploratorio de los datos para la creación de reportes, en tanto esto mejoró la asertividad en la predicción del rendimiento académico.

Al realizar este trabajo, se comprobó la importancia del DM, en cual no ha sido muy aplicado en instituciones universitarias. De esta manera, el área de admisión pudo obtener información más detallada, siendo de gran beneficio para la toma de decisiones. Como guía para alcanzar los objetivos de estudio, se fusionó la metodología HEFESTO 2.0 y el ciclo normal de una solución de IN, dando prioridad a la metodología HEFESTO para tener un panorama claro al identificar y cubrir los requerimientos. Como resultado de esta investigación, los beneficiarios son la administración de la universidad, ya que les permite tomar decisiones basadas en hechos (representados por gráficos); la dirección de admisión, definiendo el recurso humano según la ubicación geográfica, tipo de instituciones educativas y zonas de alto impacto, basadas en una data histórica. Además, los usuarios directos serán el personal que labora en el área.

El propósito planteado es diseñar un modelo de inteligencia de negocios, útil y necesario en la toma de decisiones para el departamento de Admisión de la Universidad Peruana Unión, Filial Tarapoto. Para lo cual fue necesario explorar las fuentes de datos, proponer un modelo estándar de extracción, transformación y carga, explotar la información para ser presentados a través de gráficos estadísticos.

MATERIALES Y MÉTODOS

Las fases desarrolladas en la investigación fueron:

Levantamiento de información

El levantamiento de información se hizo a través de reuniones pactadas con el coordinador del departamento de Admisión, dando lugar a: entrevistas y lluvia de ideas. Esto permitió detectar el problema al que nos enfrentábamos e iniciar con esta investigación.

Identificación de requerimientos

En esta fase se procedió a identificar los requerimientos del departamento y transformarlos en objetivos los cuales permitieron seleccionar las principales ideas y necesidades del cliente. También se aplicó el método log files (análisis de archivos) para identificar las métricas e indicadores y tener lo necesario para justificar la investigación.

Algunos de los requerimientos obtenidos fueron: conocer las zonas o lugares de los que provenían más ingresantes, saber qué sector de admisión es el que más registra ingresantes y tener un comparativo entre las campañas de admisión.

Análisis de los requerimientos

Esta fase permitió identificar las medidas e indicadores con la información que se extrajo de la fase anterior (Ingresantes según religión, modalidad, etc.). Asimismo, fue de gran ayuda para definir el nivel de granularidad del modelo de IN (por campañas); por otro lado, también se definió el modelo conceptual a utilizar. Esto se evidencia en la figura 5.

Identificación de los orígenes de datos

En esta fase se estudió y analizó el origen de los datos (base de datos de Admisión) que fue provista por el Área de Desarrollo de la Dirección General de Tecnologías de Información (DIGETI). También se normalizó y desnormalizó toda la data necesaria para el diseño del modelo de IN. Al final se estableció el modelo lógico ampliado apto para cumplir los propósitos. Ver figura 1.

Modelado multidimensional

En esta fase, teniendo ya el modelo lógico bien establecido, se procedió al modelado y/o armado del modelo multidimensional tipo estrella, creando la tabla principal (hecho_admision) y las siguientes tablas de dimensiones: adm_carrera, adm_modalidad, adm_postulante, adm_procedencia (procedencia del postulante), adm_sectoradm (sector donde hay un promotor de admisión), adm_religion, adm_

tiempo cada cual con sus campos respectivos. Teniendo las dimensiones listas se unió hacia el hecho.

Integración de los datos

Lo primero que se realizó, en esta fase, es establecer correspondencias que entendemos como un tipo de vista o manera de generar una consulta de todos los datos que serán cargados al hecho_admision. En seguida pudimos cargar con normalidad los hechos al modelo dimensional establecido, utilizando una herramienta del Kit de Pentaho (Spoon, antes Kettle). Ver figura 3 y figura 4.

Procesamiento de la información

En esta fase se aplicó las herramientas de la Suite de Pentaho como (Schema Workbench) para la construcción del cubo. Cabe mencionar que hicimos un solo cubo para todas las dimensiones establecidas e importamos un tipo de archivo XML. Ver figura 6.

Explotar información

En esta fase se utilizó la Suite de Pentaho Business Analytics (versión 6.0.1 de prueba) para explotar toda la información y empezar a generar los reportes necesarios que muestren los resultados esperados de acuerdo a indicadores. Aquí se dio lugar a los reportes a través de gráficos estadísticos.

RESULTADOS Y DISCUSIÓN

Esta sección presenta todo el proceso de diseño de IN, de acuerdo a las necesidades (indicadores) manifestados por el cliente y su validación correspondiente.

Asimismo, se detalla la explotación del modelo propuesto a través de gráficos estadísticos, de acuerdo a cada indicador. Para todo este proceso se utilizó la Suite de Pentaho.

Resultados 1

Para obtener la información (data operacional) y manipularla, se hizo una consulta SQL (ver anexo 01) a la misma base de datos de Admisión y se exportó en formato xlsx, solo los datos que sean útiles para el diseño del modelo, ya que la estructura de su base de datos depende del modelo de base de datos del sistema académico. De esta manera no se toca para nada los datos fuente.

La figura 1 muestra la data exportada, solo con los datos necesarios para nuestro modelo de IN.

Respecto al proceso de ETL, realizado por Sánchez, (2009) no coincidimos en que realiza doble trabajo en pasar toda la data de su .xls a las tablas designadas sin relación alguna y luego las extrae y transforma, para luego pasarlas al modelo dimensional propuesto por el autor. En contraposición al modelo de ETL, propone seleccionar los datos necesarios de la fuente que es .xls y pasarlos al modelo dimensional estrella que propusimos, utilizando Select Values e Insert/Update que está dentro de Transformation en la herramienta Spoon del Pentaho.

Resultados 3

La figura 5 detalla el modelo lógico de datos, el cual ha sido creado en el gestor de base de datos Postgresql y trabajado con la herramienta Pentaho Report Designer con su extensión SQL Query Designer, a su vez la figura 14 muestra las dimensiones utilizadas a través de la herramienta de Pentaho Schema Workbench. El modelo es de tipo estrella.

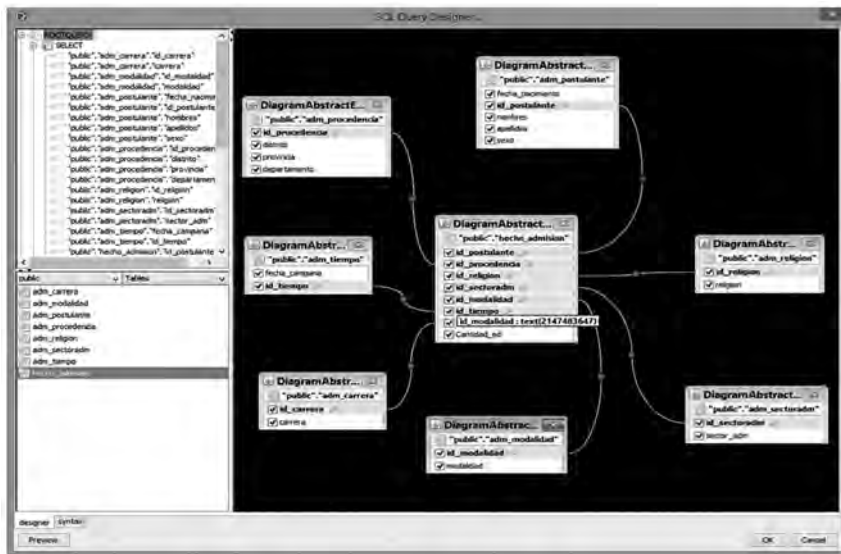


Figura 5. Modelo de datos estrella

Coincidimos con la idea del esquema dimensional de Tapia (2006), ya que cuando los indicadores están bien establecidos, las dimensiones se harán con más facilidad y si todos están relacionados, entonces no será necesario utilizar un modelo dimensional copo de nieve, sino un modelo de estrella, como el que utilizamos en el modelo. A su vez resaltamos que el hecho debe contener toda la información que se desea obtener para generar correctamente los reportes.

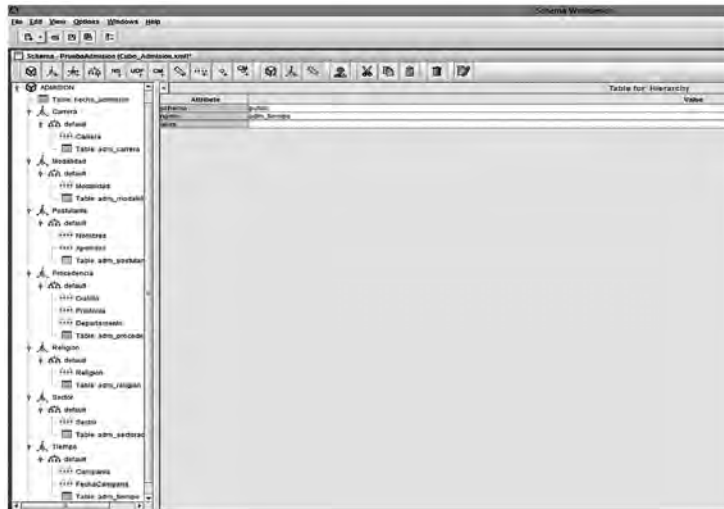


Figura 6. Dimensiones Cubo Admisión

El cubo Admisión se hizo en función a un solo hecho (hecho_admision), el cual determinará: cantidad de ingresantes por religión y procedencia, ingresantes por carreras y procedencia, cantidad de ingresantes por carrera y sectores de admisión (sector de promotores), cantidad de ingresantes por semestres de las distintas carreras y zonas de procedencia y cantidad de ingresantes por semestre, modalidad y procedencia.

Resultados 4

La validación de resultados se realizó en función a los reportes actuales de admisión, comparados con los reportes arrojados por el DM. Para lo cual se hizo un comparativo en el tiempo que demora en obtener tal o cual reporte.

Tabla 1
Comparativo entre reportes actuales y nuestro modelo de IN.

	Reportes actuales	Reportes Datamart Admisión
Tiempo por consulta a la base de datos	1 minuto	15 segundos
Tiempo en hacer un reporte	3 horas	1 minuto
Beneficios según tipos de sistemas		
Sistema Actual	Modelo de IN	
Reportes no aptos para la toma de decisiones	Reportes según la necesidad del Departamento	
Realizados en función de indicadores generales	Realizados en función a indicadores específicos	

Fuente: Creación propia.

Resultados 5

Después de considerar los requerimientos y analizar la base de datos operacional, se determinaron los siguientes indicadores:

El indicador se hizo con la finalidad de determinar la cantidad de ingresantes por religión y procedencia, el cual permite conocer, a detalle, de qué lugares provienen más estudiantes sostenedores de obra. Este análisis se hizo por provincia, ya que los distritos son más numerosos y sería de poco impacto.

El segundo indicador ayuda a conocer la preferencia de los postulantes a las distintas carreras profesionales según lugar de procedencia, lo cual sería muy útil para captar estudiantes para alguna carrera en particular.

El tercer indicador, permite saber la cantidad de ingresantes por carrera profesional o sector de admisión, este sector de admisión tiene que ver con las sedes establecidas para los promotores. Este indicador ayudará a tomar la decisión de la existencia de un promotor en los sectores establecidos.

El cuarto indicador está basado en la cantidad de ingresantes por semestre académico, carrera profesional o zona de procedencia. Este indicador es importante para saber qué campañas de admisión fueron más efectivas y en qué lugares tuvieron mayor impacto.

El quinto indicador, determina la cantidad de ingresantes por modalidad de ingreso, semestre o lugar de procedencia. El cual permitirá saber bajo qué modalidad prefieren ingresar los postulantes a la universidad.

En las siguientes figuras revelamos los gráficos según indicadores:

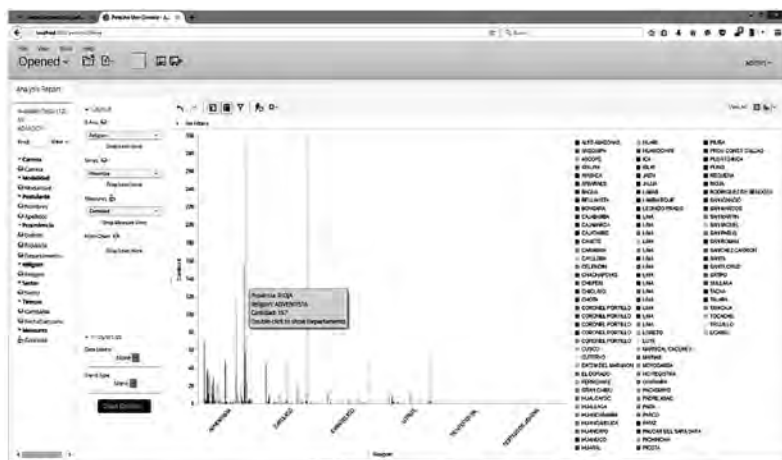


Figura 7. Cantidad de ingresantes por religión y lugar de procedencia

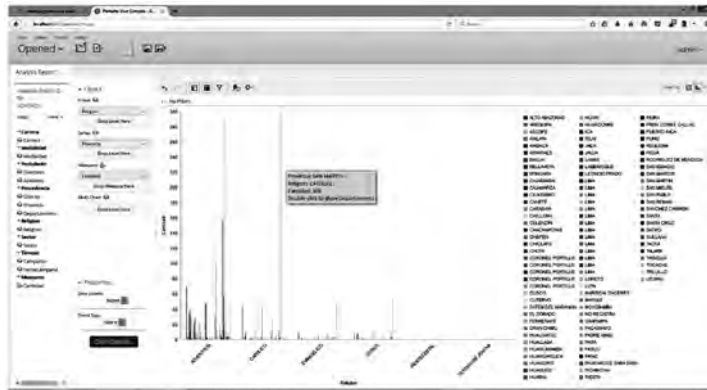


Figura 8. Cantidad de ingresantes por carreras y lugar de procedencia.

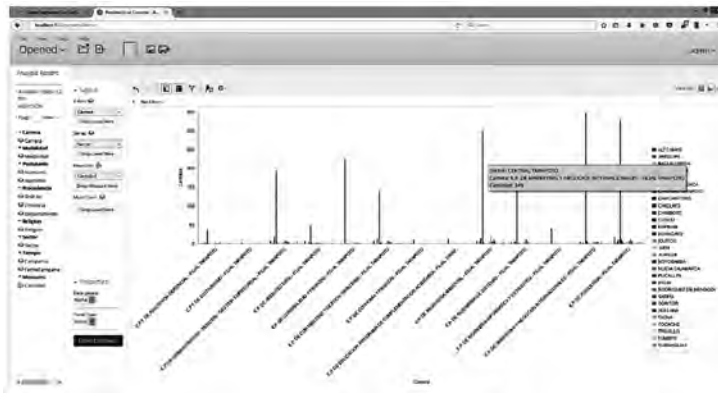


Figura 9. Ingresantes por carreras y sectores de admisión (sectores de promotores)

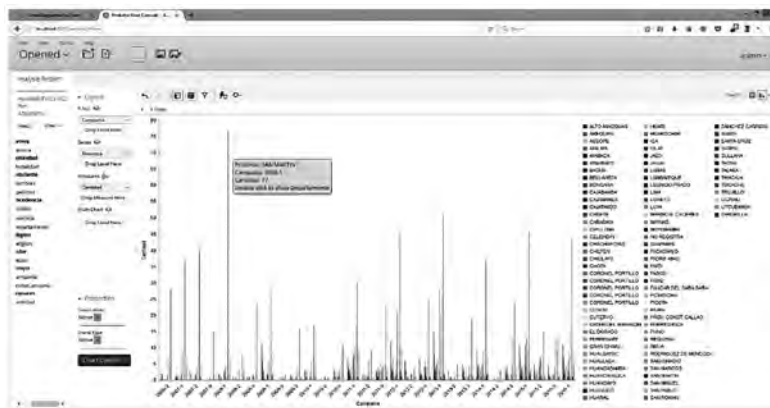


Figura 10. Cantidad de ingresantes según campaña y lugar de procedencia

También es importante comparar los resultados con otros autores, ya que permitirá ampliar el panorama de nuestra investigación, para ser mejorada o plantear nuevas soluciones.

RECOMENDACIONES

Recomendamos, al momento de manipular la data histórica de una institución o empresa, extraer un backup con extensión .xls o xlsx, debido a que este tipo de extensión es leída por cualquier sistema operativo que cuente con Microsoft Excel, y es de fácil manejo al momento de filtrar los datos.

El modelo comprende solo el Departamento de Admisión, y se recomienda ingresar al sistema de admisión la mayor cantidad de información de las campañas de admisión para que no existan datos nulos, lo que ocasiona perder algunos registros.

En relación a la recomendación anterior, es necesario estandarizar los datos de admisión para que se pueda obtener un mejor resultado a futuro.

Finalmente, recomendamos usar la metodología desarrollada para resolver importantes soluciones a las diferentes áreas de la universidad como: financiera, académica y contable; o para soluciones de BI a instituciones educativas u otras que no tengan que ver con giros comerciales.

Referencias

- Bello, F. (2006). REFLEXION:La investigación tecnológica: o cuando la solución es el problema. *Revista FACES*, 13, 14. Retrieved from <http://servicio.bc.uc.edu.ve/faces/revista/a6n13/6-13-3.pdf>
- Bernabeu, R. (2010). Hefesto. Córdoba, Argentina. Retrieved from <http://www.dataprix.com/data-warehouse-y-metodologia-hefesto/hefesto-metodologia-propia-para-la-construccion-un-data-warehousehttp://www.dataprix.com/data-warehouse-y-metodologia-hefesto/ii-hefesto-metodologia-propia-para-la-construccion-un-data-wa>
- Bustamante, A., Galvis, E., Gonzáles, M., García, A., & Benavides, L. (2011). *Soluciones de inteligencia de negocios en la práctica: apoyo a la toma de decisiones en proyectos educativos para población infantil vulnerable en el caribe Colombiano*. (Spanish). *UIS Ingenierías*, 10(2), 123–135. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=zbh&AN=86166562&lang=es&site=ehost-live>
- Bustos, S., & Mosquera, V. (2013). *Análisis, diseño e implementación de una solución BUSINESS INTELLIGENCE para la generación de indicadores y control de desempeño, en la empresa OTECEL*. Retrieved from <http://repositorio.espe.edu.ec/handle/21000/6305>
- Chavez, J. D. (2013). *Análisis, diseño e implementación una solución de inteligencia de negocios orientada a controlar los procesos de generación y emisión del DNI (Documento Nacional De Identidad)*

- en el Reniec (*Registro Nacional de Identificación Y Estado Civil*). Universidad Católica del Perú. Retrieved from <http://tesis.pucp.edu.pe/repositorio/handle/123456789/5314>
- Cruz, G. R. (2003). *Utilización de la tecnología Data Warehouse en instituciones educativas*. Universidad Autónoma del Estado de Hidalgo. Retrieved from <http://www.uaeh.edu.mx/docencia/Tesis/icbi/maestria/documentos/Utilizacion de la tecnologia.pdf>
- Dall'Orto, L., & Wu, R. (2006). Construcción y pruebas de una herramienta de desarrollo de soluciones para inteligencia de Negocios Módulo de Extracción. *Test*. Pontificia Universidad Católica del Perú. <http://doi.org/10.1017/CB09781107415324.004>
- Fuentes, L., & Valdivia, R. (2010). Incorporation of business intelligence Elements in the Admission. *Revista Chilena de Ingeniería*, 18, 383–394.
- Guillén, F. (2012). *Desarrollo de un Datamart para mejorar la toma de decisiones en el Área de Tesorería de la Municipalidad Provincial de Cajamarca*. Universidad Privada del Norte. <http://doi.org/10.1017/CB09781107415324.004>
- Sánchez, J. A. (2009). Tesis. Universidad Peruana Unión.
- Tapia, I. (2006). *Una metodología para sectorizar pacientes en el consumo de medicamentos aplicando Datamart y Datamining en un hospital nacional*, 111. Retrieved from <http://cybertesis.unmsm.edu.pe/xmlui/handle/cybertesis/273>
- Zambrano, C., Rojas, D., Carvajal, K., & Acuña, G. (2011). Análisis de rendimiento académico estudiantil usando data warehouse y redes neuronales. Analysis of students ' academic performance using data warehouse and neural networks. *Revista Chilena de Ingeniería*, 19, 369–381.